Convergence of Dynamical Features Under ICA Processing With Application to Fetal ECG

by

Michael James Andrew Potter

A Thesis
Submitted to the Faculty of Graduate Studies
in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Department of Electrical and Computer Engineering
University of Manitoba
Winnipeg, Manitoba

Thesis Advisor: W. Kinsner, Ph.D. P.Eng.

xxxiv + 397 + A63 + B9 + C29 +D9 = 541 pages

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION

Convergence of Dynamical Features Under ICA Processing with Application to Fetal ECG

BY

Michael James Andrew Potter

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirement of the degree

Of

Doctor of Philosophy

THE UNIVERSITY OF MANITOBA
FACULTY OF GRADUATE STUDIES
FINAL ORAL EXAMINATION OF THE PHD THESIS

The undersigned certify that they have read, and recommended to the Faculty of Graduate Studies for acceptance, a PhD thesis entitled:

CONVERGENCE OF DYNAMICAL FEATURES UNDER ICA PROCESSING WITH APPLICATION TO FETAL ECG

by

MICHAEL JAMES ANDREW POTTER

in partial fulfillment of the requirements for the PhD Degree

Dr. W. Kinsner, Advisor
Electrical and Computer Engineering

Dr. E. Shwedyk, Internal Examiner
Electrical and Computer Engineering

Dr. S. Balakrishnan, Internal Examiner
Mechanical & Manufacturing Eng.

Dr. J. Brewster, Internal Examiner
Statistics

Dr. B. Dony, External Examiner
University of Guelph
School of Engineering
Guelph, ON, Canada

Date of Oral Examination: **September 8, 2008**
The student has satisfactorily completed and passed the PhD Oral Examination.

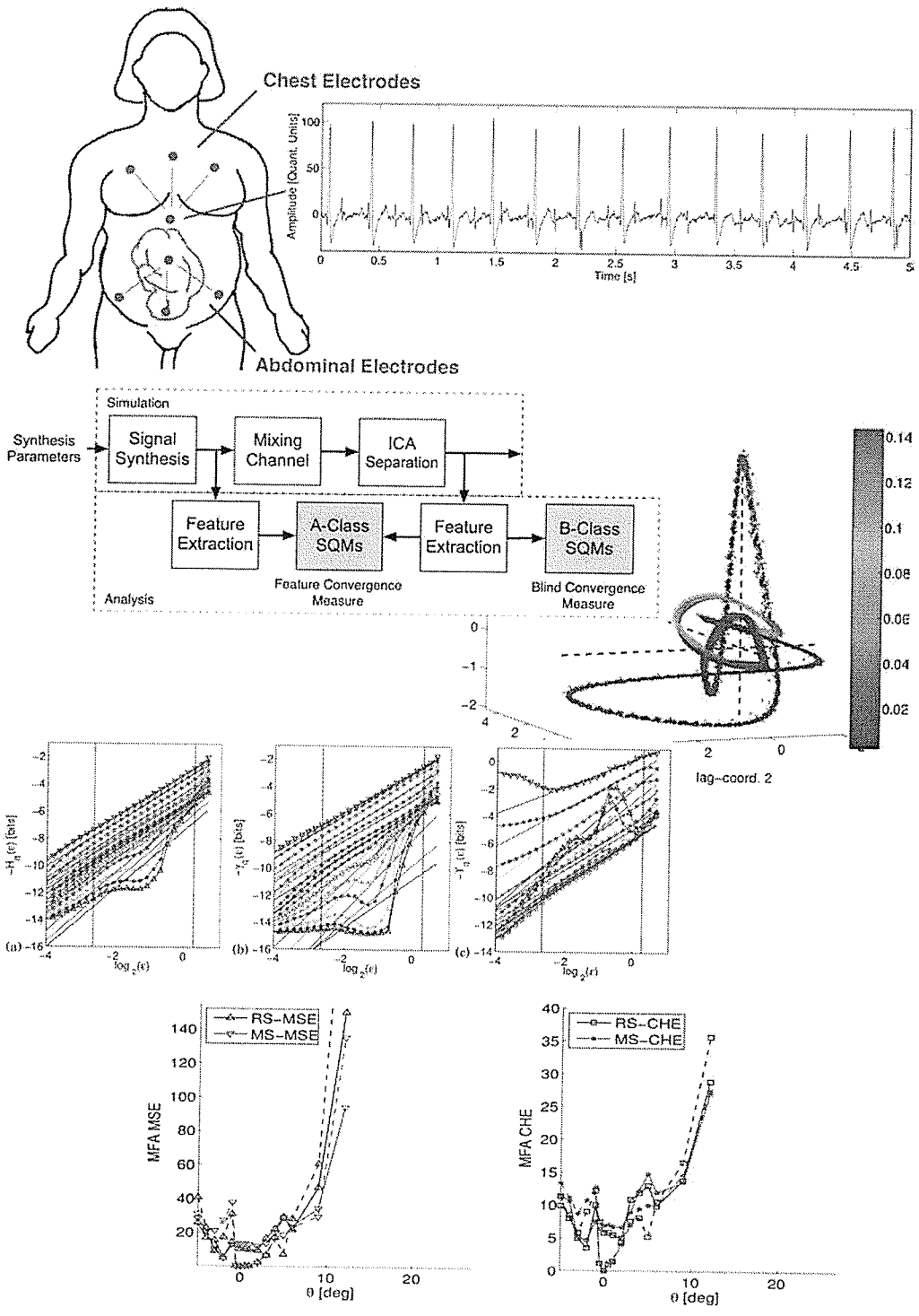
Dr. W. Kinsner, Advisor

Dr. E. Shwedyk, Internal Examiner

Dr. S. Balakrishnan, Internal Examiner

Dr. J. Brewster, Internal Examiner

Dr. B. Dony, External Examiner
University of Guelph
School of Engineering
Guelph, ON, Canada


Chest Electrodes
Abdominal Electrodes
Amplitude [Quant. Units]
Time [s]
Simulation
Synthesis Parameters
Signal Synthesis
Mixing Channel
ICA Separation
Feature Extraction
A-Class SQMs
Feature Extraction
B-Class SQMs
Feature Convergence Measure
Blind Convergence Measure
Analysis
lag-coord. 2
MFA MSE
RS-MSE
MS-MSE
θ [deg]
MFA CHE
RS-CHE
MS-CHE
θ [deg]

To my wife, who, at the prime of her life,

married not only a man of many weaknesses,

but a thesis as well.

Abstract

Independent component analysis (ICA) is a revolutionary class of algorithms for the blind separation of independent sources from an instantaneous mixture. In theory, ICA is stronger than classical approaches to signal demixing because ICA is higher-order, non-parametric, and data-driven. The application of ICA to electrophysiological systems is of particular interest, and the fetal *electrocardiogram* (ECG) separation problem is a specific example that has been longstanding in the literature. Research in ICA, however, has not provided a definition of performance measurement that is consistent with both ICA theory and practice. In particular, the practical performance measures of classical signal processing are limited by being second-order and model-driven. Since traditional measures do not apply, how can a signal recovered blindly through ICA processing be measured for accuracy?

This thesis proposes that feature convergence be used as a practical measure of ICA separation performance. Arguments in principle are made regarding the significance of this approach to signal processing. Using a novel simulation of the fetal ECG separation problem, an experimental study of feature convergence for the measurement of separation quality is also presented. In particular, this experiment is designed to investigate the relationships of nonlinear and higher-order convergence measures to the fundamental converging elements of ICA optimization (*i.e.*, cost functions and weight matrices). Results demonstrate that (i) two categories of convergence apply to ICA cost functions (sub- and superlinear), (ii) two similar categories of convergence apply higher-order features (entropy and kurtosis), and (iii) the convergence of nonlinear dynamical features is similar to entropy but is sensitive to nonlinear

effects through the embedding dimension. In particular, these results demonstrate that demixing levels equivalent to 20 dB *signal-to-noise ratio* (SNR) can be identified with a statistical match of signal kurtosis, but is insufficient for the estimation of entropy or nonlinear features (*i.e.*, multifractal spectra). Furthermore, preliminary results indicate that random phase correlation in the source signals can introduce bias in most feature estimation.

This thesis presents several novel contributions, including: (i) a discussion of ICA performance using convergence profiles; (ii) an analysis of multifractal feature convergence under ICA; (iii) an improved model for the synthesis of ECG from beat annotations; and (iv) a technique for the direct calculation of the multifractal spectrum of scaling indices from a correlation partition.

Acknowledgements

My heart goes out in thanks to all the family and friends who supported me through the long days of this work. Specific thanks go to my colleagues Stephen Dueck, Aram Faghfouri, Sharjeel Siddiqui, and Neil Gadhok in the Signal and Data Compression Lab for all the stimulating discussions on technical topics. You made the geek life fun.

Thanks so much to the friends at the Winnipeg Church of Christ who offered much prayer and support in the final year. I am especially grateful to David Jung for his caring guidance, and Robert Borgersen for engaging the support of "Team Michael" and providing encouragement daily. Special thanks also go to Erica Jung and Dr. Frank Bovell for volunteering to proof-read portions of my manuscript. Your devotion will not be forgotten.

Thanks also to the examiners who had to undergo the monumental task of making this work official. May you be given rest in twice the abundance of the work you invested here.

To my advisor, Dr. Kinsner, I owe so much. Thanks for taking in an interdepartmental stray and giving him the opportunities of a lifetime. You have prepared me for an academic career like no other.

To my parents, Bev and Don — who have sponsored me continually along this wandering path and loved their crazy son even when they did not understand how it all works — words cannot express how grateful I am for your influence in my life. A lifetime is not long enough to repay all the debt of love I owe.

Last, to my wife and my Lord, who have walked every day with me, shown endless grace, and never given up or lost belief in me: I would have never completed this without you. You alone have made it worth it.

Table of Contents

List of Figures

List of Tables

List of Abbreviations and Acronyms

ADE	Absolute deviation of entropy
ADK	Absolute deviation of kurtosis
ANN	Artificial neural network
API	Amari performance index
ASI	Amari-squared index
bpm	Beats per minute
BSP	Body surface potential
BSS	Blind source separation
cdf	Cumulative distribution function
CP	Correlation partitioning
CSS	Cycle-shuffle surrogate
dB	Decibel
DCPMS	Direct correlation partition Mandelbrot spectrum
DOG	Derivative of Gaussian
ECG	Electrocardiogram
EEG	Electroencephalogram
EMG	Electromyogram
fECG	Fetal electrocardiogram
FFT	Fast Fourier transform
FM	Frequency modulation
FMPI	Frequency-modulated phase-interpolation
FNN	False nearest neighbour

HOS	Higher-order statistics
HP	Histogram partitioning
HRV	Heart rate variability
IAAFT	Iterated amplitude-adjusted Fourier transform (surrogate)
ICA	Independent component analysis
id	Identically distributed (random variables)
IHR	Instantaneous heart rate
iid	Independent and identically distributed (random variables)
IPFM	Integral-pulse frequency modulation
ISP	Intelligent signal processing
MFA	Multifractal analysis
MFE	Matrix Frobenius error
MI	Mutual information
MRI	Magnetic resonance imaging
MS	Mandelbrot fractal dimension spectrum
MVUB	Minimum-variance unbiased estimator
NSRD	(Physiobank) Normal Sinus Rhythm Database
PCA	Principal component analysis
pdf	Probability distribution function
PPS	Pseudo-periodic surrogate
RND	Real noisy (fECG) data
RS	Rényi fractal dimension spectrum
SAD	Simulated analytical data
SFD	Simulated noise-free (fECG) data

SND Simulated noisy (fECG) data

SNR Signal to noise ratio

sps samples per second

SQM Separation-quality metric

T21 Type-2.1 (surrogate)

TR Tachogram resampling

VNE Vector norm error

WT Wavelet transform (continuous)

WTMM Wavelet-transform modulus-maxima

Xinfomax Extended-infomax algorithm (by Lee *et al.* [125])

xMI Cross mutual information

List of Mathematical Symbols

$\boldsymbol{u}$	Vectors are bold-faced
$\mathbf{U}$	Matrices are bold-faced caps
$\underline{u}$	Random variables are underlined
$\underline{\boldsymbol{u}}$	Arbitrary random vector
$\underline{\overline{u}}$	Centered random variable
$\underline{\overline{\boldsymbol{u}}}$	Centered random vector
$\underline{\widehat{u}}$	Estimate of a variable
$\stackrel{\text{def}}{=}$	"Defined as"
$\backsim$	"Distributed as" relationship (*i.e.*, between a random variable and its pdf)
$\sim$	"Scales as" relationship between two functions
$\approx$	"Box-counting (log-log) ratio" for scaling estimation
$\leftarrow$	"Updated to"
$*$	Convolution operator
$\langle\cdot\rangle$	Arithmetic mean
$\lvert\cdot\rvert$	Absolute value
$\lVert\cdot\rVert$	Norm
$\lVert\cdot\rVert_F$	Frobenius norm
$\lfloor\cdot\rfloor$	Floor function (greatest lower integer bound)
$\lceil\cdot\rceil$	Ceiling function (least upper integer bound)
α	Hölder exponent (of a measure)
$\mathcal{A}$	Attractor in $\mathbb{R}^{N_e}$
A_ϵ	Amari performance index (matrix-based convergence metric)

$\aleph$	Separation quality metric (SQM)
b	Number system base
C	Arbitrary constant
C	Cantor set
$C^n(\mathcal{X})$	Family of n-times continuously differentiable functions over $\mathcal{X}$
CCI	Cross-correlation index
δ_1	Kronecker delta function
δ_∞	Dirac delta "function"
δ_P, δ_T	Event intervals (PR- and RT-intervals, respectively)
D_q	Rényi fractal dimension spectrum
D_0	Hausdorff dimension
D_2	Correlation dimension
Δ_R	RR interval
$\frac{\overline{\Delta}}{n}$	Scalar relative difference operator for multifractal analysis
$\frac{\Delta}{n}$	Functional relative difference operator for multifractal analysis
$\mathbf{e}_i$	Canonical basis vector
ϵ	Coarse-graining scale parameter
$\mathcal{E}\{\cdot\}$	Expectation operator
f	Probability distribution function
$\mathcal{F}$	Fourier transform
F	Cumulative distribution function
$\mathfrak{g}$	Gaussian random variable
$\mathbf{G}_{N_c}$	Givens matrix (in N_c)
h	Lipschitz-Hölder exponent (of a function)

H	Entropy
H_q	Rényi generalized entropy
H_g	Negentropy (difference from Gaussian entropy)
Ξ	Heaviside function
$\mathbf{I}_{N_c}$	Identity matrix ($N_c \times N_c$)
$\mathcal{I}$	ICA cost function
K	"K" = 1024 (base 2 unit)
k	Normalized kurtosis
k'	Unnormalized kurtosis (fourth-order auto-cumulant)
κ	Cumulant
λ_n	Autocorrelation time
λ_0	First zero of the autocorrelation function
λ_{0^2}	Second zero of the autocorrelation function
$\widetilde{\lambda}$	Theiler window λ_n stretchfactor
$\mathcal{L}$	Space of all linear transformations
$\mathcal{L}_O$	Stiefel manifold
L^p	Classical L^p spaces (Lebesgue p-integrable functions, $\left\{ g \,\middle|\, \int \lvert g\rvert^p < \infty \right\}$)
l^p	Classical l^p spaces (p-summable sequences, $\left\{ \{\xi_i\}_{i=0}^{\infty} \,\middle|\, \sum_i^{\infty} \lvert\xi_i\rvert^p < \infty \right\}$)
$\mathbf{M}$	Simulation mixing matrix
$\mathcal{M}$	Mutual information
μ	Measure
μ_1	Mean
μ_2	Variance
μ_H	Hausdorff measure

η	Arbitrary nonlinear function
$\mathbb{N}$	Set of natural numbers (non-negative integers)
N	Number of points in time series
N_c	Number of channels in time series
N_θ	Number of demixing angles parameterizing $\mathcal{L}_O$
N_e	Embedding dimension
N_ϵ	Number of scales in an MFA
N_g	Number of elements in an MFA histogram partitioning
N_o	Number of observations in an attractor
N_q	Number of elements in an MFA q-vector
Δn	Embedding (sample) lag
O	Big-"O" algorithm complexity (infinte asymptotic upper bound)
Ω	Domain of a function
Ω_1	Support of a function or measure
$\mathcal{P}$	Correlation partitioning
$\pi_m[x]$	Type-m surrogate of the time series x
Pj	Projection operator
Pm	Permutation operator
P_n	Polynomial of order n
Φ	Fractal generator operator
φ	Generalized Cantor measure generating vector
Pr	Probability of an event
θ	Angle or phase
$\mathbb{R}$	Set of real numbers

$\mathbb{R}^+$	Set of positive real numbers
R	Autocorrelation function
rem	Remainder
$\mathfrak{R}_\alpha$	Relative Hölder exponent of two measures
$\mathfrak{R}_{D_0}$	Relative Hausdorff dimension
$\mathfrak{R}_{D_q}$	Relative Rényi fractal dimension spectrum
$\mathfrak{R}_{\mathsf{KL}}\left(\underline{\boldsymbol{u}}_1 \middle\| \underline{\boldsymbol{u}}_2\right)$	Kullback-Leibler divergence
$\mathfrak{R}_{\mathsf{KL}_q}\left(\underline{\boldsymbol{u}}_1 \middle\| \underline{\boldsymbol{u}}_2\right)$	Generalized Kullback-Leibler divergence
$\mathfrak{R}_Z$	Relative multifractal partition function
$\mathcal{R}(\mathcal{X})$	Set of all random variables over the space $\mathcal{X}$
$\rho_\heartsuit$	Heart rate
$\boldsymbol{s}$	Source time series
$\mathcal{S}$	Fourier spectrum
σ	Standard deviation
σ^2	Variance
σ_{PPS}	PPS noise radius
$\boldsymbol{\Sigma}_{\underline{\boldsymbol{u}}}$	Covariance matrix of $\underline{\boldsymbol{u}}$
SNR^*	Standardized SNR
sort	Sort operator
τ_r	Learning rate
$\mathcal{T}$	Wavelet-transform modulus-maxima tree in time-scale space
$\tau(q)$	Partition scaling exponent function
$\tau_{\mathfrak{R}}$	Relative partition scaling exponent function

ν	Spectral analog frequency (*i.e.*, in $\mathbb{R}$)
ν_s	Sampling frequency
$\mathcal{V}$	Open cover
V_r	Member of an open cover
$\boldsymbol{\upsilon}$	State vector of phase space
$\mathbf{W}$	ICA weight matrix
$\mathbf{W_0}$	PCA sphering matrix
$\mathring{\psi}$	Digamma function
ω	Spectral cyclic frequency (*i.e.*, in Nyquist interval)
W_T	Theiler window
$\mathcal{W}$	Wavelet transform (continuous)
ψ	Mother wavelet function
$\boldsymbol{x}$	Observed time series
χ_Ω	Characteristic function of the set Ω
$\mathcal{X}$	Arbitrary vector space
γ_q	Hölder entropy
Υ_q	Mandelbrot entropy
ζ	Realization index for random processes
$\mathbb{Z}$	Set of integers
Z	Multifractal partition function
Z_B	Boltzmann partition function
Z_C	Correlation sum
$\overline{z}^{[m]}$	m-spacing statistic

List of Copyrighted Material for which Permission was Obtained

Chapter I

INTRODUCTION

In theory, there is no difference between theory and practice.

But, in practice, there is.

—Jan L. A. van de Snepscheut

1.1 Motivation

The difference between theory and application is often a challenging problem in the design of signal acquisition and processing systems. Specifically, practical systems that deal with the classic problem of noninvasively acquiring, processing, and interpreting *in vivo* biological signals are subject to that challenge, and must undergo theoretical and practical scrutiny. The class of *body surface potentials* (BSP) is a broad class of practical biomedical signals used in current medical practice. These signals are recordings of electrical potentials at the skin surface, such as: the *electromyogram* (EMG) concentrating on muscle tissue; the *electroencephalogram* (EEG) concentrating on the skull; and the *electrocardiogram* (ECG) concentrating on the the thorax. The BSP methodology is a convenient method for clinical study of internal systems since the non-invasive nature of these recordings is beneficial to the health and comfort of the patient. The BSPs, however, also share the drawback of measuring the electrophysiological activity of their respective internal sources (the muscles, the brain, and the heart) only indirectly, and are subject to the *signal-mixing problem*. Though the biochemical reactions producing electric currents are independent and separately localized, they are all

transported to the skin surface and recorded at the electrodes in a mixture. This signal-mixing problem is a common issue in applications involving indirect measurement techniques. The acquired many-sensor signals are composed of multiple source components and affected by an indirect mixing channel. To analyze the properties of the original sources, therefore, one must counteract the mixing channel by performing *signal separation.*

There is a long history of signal separation approaches in signal processing. Classical signal separation theory is focused primarily on the conceptualization of signals in Fourier-space: *i.e.*, power distributions over frequency bands. If the sources dominate different regions of frequency (spectral regions), then separation can be accomplished by *filtering* the acquired signal to the appropriate bandwidth. For this reason, the application of electrophysiology analysis has matured in spite of the signal-mixing problem. The ECG is primarily a recording of the heart because, in practice, the geometrical proximity of the sensors to the heart, the frequency-dependent conductance of the body, and the absolute amplitudes of the heart signals make them *relatively dominant* at the sensors. Likewise for the EEG and EMG signals. Over time, the utility of the classical theory of "power-distribution" (*i.e.*, Fourier decomposition) has been enhanced with modern extensions, such as *time-frequency decompositions.* These extensions have resulted with practical signal-separation techniques such as decorrelation, radar, and denoising. Notwithstanding these advances, the foundation of classical signal processing remains rooted in distributions of *power*, and the signal-mixing problem is overcome by power separability.

New solutions to the signal-mixing dilemma, however, have been proposed recently that relax the assumptions made in classical signal processing. These new solutions are based on a theory of *statistics* and not power, and have the potential to radically change the practice of signal separation. As a new theory, however, there must also a potential for new discrepancies

Statistical Signal Processing	
Classical Signal Processing	**Intelligent Signal Processing**
Gaussian Distributions	Non-Gaussian Distributions
Power / Variance	HOS
PCA Separation	ICA Separation

Fig. 1.1 The development of "intelligent" signal processing.

between theory and practice.

1.2 Problem Definition

This new theory for signal separation revolutionizes the statistical assumptions commonly used in engineering: it allows for *non-Gaussian* distributions and statistics. In its broadest form, this development in signal processing has been coined *"intelligent signal processing"* (ISP) [80, Haykin (2001)], since the methodologies are flexible, and thus effectively "smart" by adapting to the arbitrary distribution of the data. That is, the Gaussian-based *model-driven* approach of a century of signal processing is replaced by an adaptive, *data-driven* approach. This extension beyond the physical, analytical, historical, and computational elegance of the Gaussian model has been a very difficult challenge, only recently empowered by pioneering minds, soft computing paradigms, and vigorous computer technology.

The power-based solutions of classical signal processing can be considered optimal in a Gaussian sense, since power is essentially a second-order statistic (*i.e.*, variance). The tools of ISP are, consequently, the *higher-order statistics* (HOS), Fig. 1.1. It is this new theory of HOS

that requires evaluation in practical application. Since the underlying assumptions of classical theory have been questioned, and new approaches developed, the research community must also address the new challenges of applying these approaches in practice, since the classical theory and common experience no longer apply.

To address the signal-mixing problem in particular, a class of ISP techniques structured on HOS-based transformations have been proposed called *independent component analysis* (ICA). ICA is a theoretical extension to a foundation in traditional signal processing: namely, *principal component analysis* (PCA). PCA is a linear transformation which takes the data to orthogonal output components of maximum variance (*i.e.*, second-order statistics, or power). In the context of the signal-mixing problem when localized and uncorrelated sources are observed as a mixture through a linear instantaneous mixing channel, PCA is optimum in the sense of signal-to-noise ratio. ICA, however, is based on new and different assumptions. Here, the channel model is equivalent, but ICA takes the data into maximally *independent* output components, Fig. 1.2, in the presence of higher-order statistics. Specifically, while PCA is blind to HOS, ICA uses the HOS present in the observed distributions to adapt its linear transformation to account for "higher-order correlations" in the mixture. Considerable interest has been shown in ICA, since it is adaptive to the HOS in the unknown sources non-parametrically, and therefore it offers a means of achieving *blind source separation* (BSS). That is, the processing of a linear mixture of unknown sources (either time signals or images) into separate sources can be done *without* first modelling the unknown sources parametrically.

While the ISP paradigm breakthrough has spawned a rapid development of BSS signal-demixing algorithms, the study of the performance of these algorithms has remained relatively unchanged from the classical approach. Primarily, the evaluation of an ICA algorithm proceeds in two parts: (i) by theoretically or empirically evaluating the algorithm applied to a

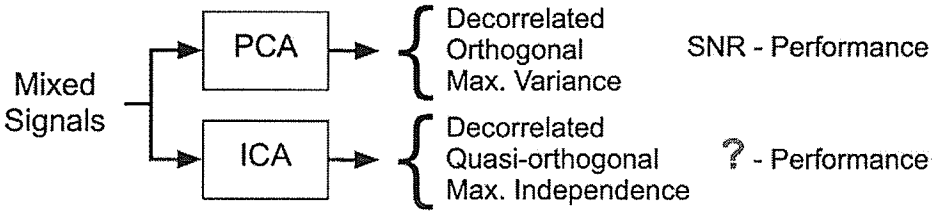


Fig. 1.2 PCA and ICA separation achieve different output components.

simple (non-blind) signal model; and (ii) by applying the algorithm to actual (blind) data for subjective evaluation. In the second part, numerical measures are very difficult to apply, given that the sources are not available for analysis. In the first part, the primary tools for measuring ICA accuracy is either: (i) *signal-to-noise ratio* (SNR), or (ii) *Amari's performance index* (API), which is a recent confusion-matrix-based separation measure that matches the ICA conditions. Apparently, it has not concerned the researchers in the field that

(a) SNR is power-based and inconsistent with the ISP paradigm;

(b) Amari's performance index, while theoretically consistent with ISP, provides little practical insight into the accuracy of feature extraction from the ICA output; and

(c) The relationship between the theoretical analysis and the practical application in the two-part evaluation scheme is often very weak.

These three limitations are persistent in the literature, and demonstrates a compelling need to re-evaluate the concept of *meaningful performance measures* in the context of ISP in general, and specifically, ICA. Otherwise, beyond any power-based characterization, the accuracy of solutions to signal-mixing problems will be in question. Thus, it is critical that researchers make headway in the use of signal processing quality measures that are meaningful in the application of ISP. In particular, the distinction between the relevance of the separation quality

measure *in theory*, its relevance *in practice*, and its relationship to other *practical* quality measures is significant.

As described more fully in the next section, this work attempts to evaluate the connection between the traditional ICA metrics (*e.g.*, SNR, API) that are applied in non-blind theoretical settings to the features which would be extracted in practical cases (*i.e.*, BSS). In doing this, an evaluation of the theoretical and practical measures of separation quality can be provoked, and the analysis of an advancing technology can, itself, advance.

1.3 Research Questions

The effect of ICA in practice has been demonstrated in a multiple areas of image and signal processing, including BSP processing [104, (2000)][51, (2000)][174, (2001)]. As described in the last section, however, the literature exhibits several different approaches to performance quantification, but none consider the connection between the two disjoint parts of ICA papers in a practical way. In effect, the evaluation of ICA in practice is disjoint from the evaluation of ICA in theory. This work is interested in exploring this unresolved issue in ICA performance analysis. What connections can be made between the traditional metrics of ICA performance, which are limited to known trials, and the accuracy of ICA processing in a BSS application?

Some mitigating factors are known to affect the direct connection between theory and practice. The HOS on which ICA is based are known to be susceptible to outliers in the data [93][63]. What influence can be expected, then, from observations contaminated with observation noise? Furthermore, what effect can be expected on the features that a signal processing system extracts from the output signal? While a subjective evaluation of output features is common, can this effect of ICA be measured quantitatively?

To evaluate the possibilities for *feature-based* separation quality metrics (SQM), *feature*

convergence under ICA optimization is analyzed here in an empirical experiment. This experiment is designed to reduce the disparity between theory and practice in the context of the chosen test case of fetal ECG separation from an abdominal BSP [51]. This signal separation test case, demonstrated in Fig. 1.3, has been a problem of significant interest for many years in the signal processing community [234, (1975)][228, (1987)][201, (1996)][51, (2000)]. It is a particularly suitable benchmark test case for the desired ICA analysis since:

(a) The desired source signals are nonstationary;

(b) The desired source signals occupy similar bandwidth;

(c) The desired source signals are from the same class of sources (same type and function);

(d) Models are available for synthesizing the desired source signals [140, (2003)];

(e) Source separation has never been successful to the point of developing a clinical practice of fetal ECG analysis from (electrical) abdominal signals;

(f) Traditional metrics are limited in significance, and the utility of nonlinear dynamical metrics has been established [207, (2000)][211, (2002)][90, (2004)].

ICA involves a theory of mathematical approximations, and, in practice, can also be sensitive to data error. Can ICA, therefore, be considered robust enough to be an effective and reliable processor for electrophysiological recordings? Will SNR-alone imply convergence of non-power-based features? Is it possible to successfully identify a "separation outlier" (*i.e.*, an unsuccessful feature representation)? Within the context of fetal ECG separation, this work addresses these questions in a scoped sense.

Of course, some features of practical interest are specific to each application (*e.g.*, ECG analysis). Generic signal processing features with universal applicability, however, are used

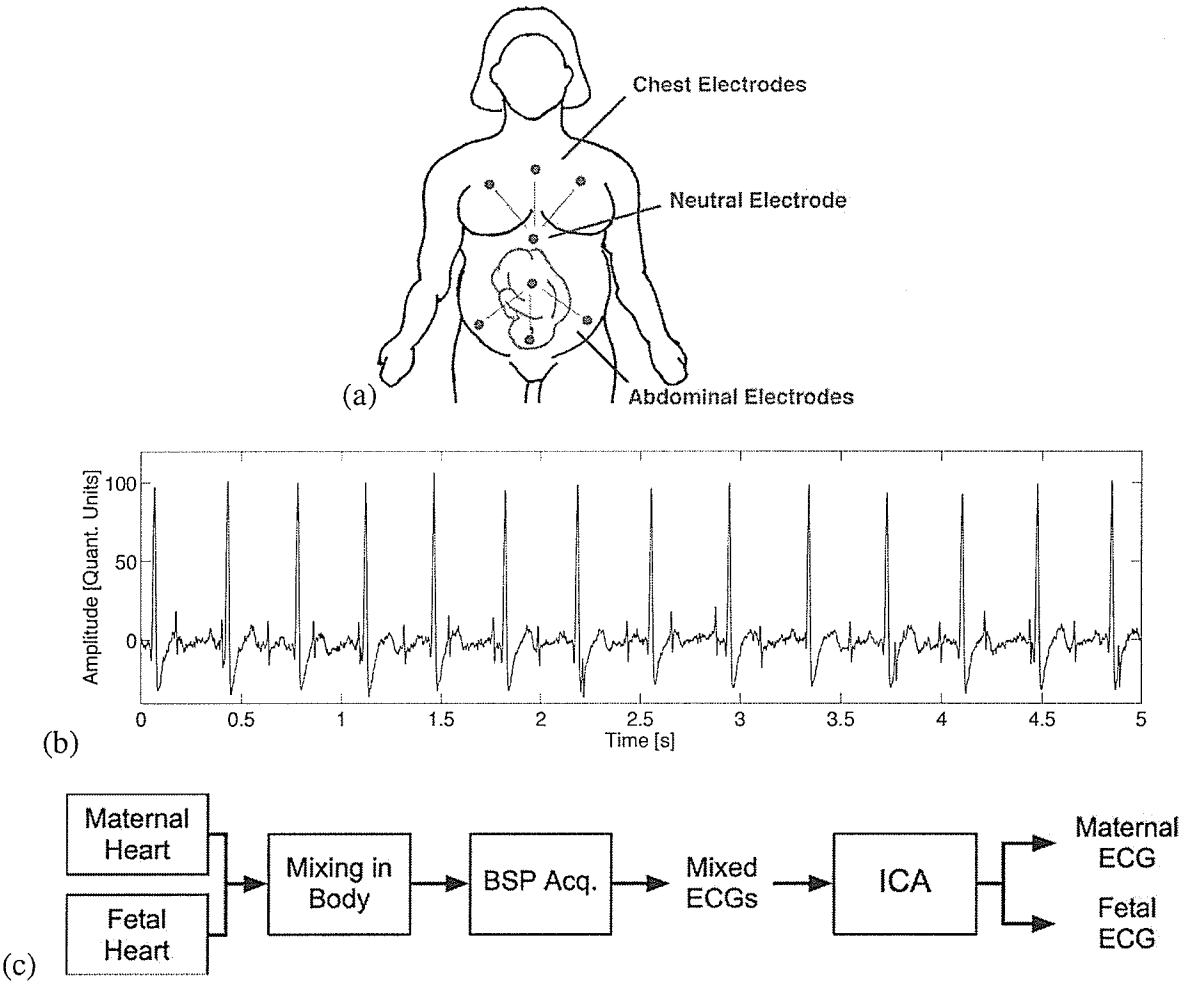


Fig. 1.3 The selected test case application of ICA: fetal ECG extraction. (a) BSP acquisition (after [234]); (b) Acquired BSP time series with maternal dominance over fetal signal, (Data from [51, (2000)]); (c) separation problem schematic.

often in engineering (*e.g.*, SNR). To maintain a feasible complexity of research, the focus of feature convergence under ICA processing in this work is restricted to generic signal processing features. In particular, a specific class of nonlinear features that are known to characterize nonlinear dynamics, *multifractal features*, will be analyzed for convergence. Since this class of features is nonlinear, it is not intuitively clear how they will converge under ICA separation. Thus, the thesis question for this work is formalized as

> *What measure of ICA performance can uniformly assure safe multifractal feature characterization of a fetal ECG?*

1.4 Thesis Statement

This thesis demonstrates that (i) not all ICA cost functions converge at the same rate, (ii) not all higher-order statistics converge at the same rate, (iii) nonlinear dynamical features extracted from ICA estimates may not converge to the features of the original independent sources because of nonlinear sensitivities; and (iv) traditional bounds using second-order performance metrics fail to account for sensitivities in some higher-order and dynamical feature extraction.

This is accomplished in an ICA simulation of Fig. 1.3(c) by measuring the convergence of statistical and multifractal scaling features extracted from an ICA separated signal to those features extracted from the known independent source.

Furthermore, by demonstrating that the scaling features do not converge uniformly with the other metrics, it is concluded that ICA convergence alone is not sufficient to validate any scaling feature extraction from signals processed by ICA.

1.5 Thesis Objectives

1.5.1 Grand Objectives

As explained, this thesis is interested in validating predefined ICA processes in a context that bridges engineering theory and practice. From the signal processing point of view, the most important elements in practice are the *features* extracted from the signals. Simply put, it is the practical feature quality of the output signals that matters, and compromised feature quality is the most concerning issue. Thus the objective here is to evaluate signal *feature convergence* as a measure of ICA performance.

Moreover, it was mentioned that features in practice can very often be application specific, (*e.g.*, clinical features in the cardiological analysis of ECG do not apply to colour images). In general theory, signal processing is limited by the restricted meaning of "feature quality" found in mathematical measures. Here, the convergence of *generic* features that have shown to be useful in the study of natural signals are analyzed; these include statistical, information-theoretic, and multifractal dynamical scaling features. This choice of features is consistent with the history of research on the characterization of natural signal "feature quality" at the University of Manitoba Signal and Data Compression Laboratory.

Considering the nature of these grand objectives, this work is, therefore, not a true "biomedical engineering" thesis. This work is not intended to examine deeply the ECG signals themselves, or try to satisfy a physician directly. A biomedical thesis would directly focus on the development of "*important* component analysis", Fig. 1.4. Instead of attempting to design yet another "new ICA algorithm for fetal ECG separation", the objective here is to consider how such an algorithm should be evaluated. It is an objective of this work to identify the engineering limitations of ICA in signal processing applications, and the limitations of methods that evaluate ICA. This work will introduce a novel approach to the evaluation of ICA algorithms,

fECG Separation	
Biomedical Engineering	**Signal Processing**
Clinical Features Physician Approved Develop ECG "Important Component Analysis"	Generic Complexity Features Engineer-based Study Contemporary ICA
☒	☑

Fig. 1.4 The emphasis of this thesis is from the perspective of signal processing, and not biomedical engineering.

and demonstrate the complexities involved in "bridging theory and practice".

Notwithstanding the caveats above, this research also establishes a necessary groundwork for the maturation of ICA separation from an exploratory research initiative into a signal processing module available for use in the processing routines of intelligent clinical devices. In expectation of increasing social emphasis on medical screening, this analysis of ICA in an ECG context can contribute to proposed designs of intelligent and autonomic [95] ambulatory diagnostic assistants. Considering these objectives, it is then necessary to place *some* significance on proper ECG dynamics for both the analysis and simulation of ECG.

1.5.2 Method Objectives

The objective of the chosen methodology is to apply a scientific approach to the analysis of the feature convergence of ICA separated signals. For this analysis, independent noise-free uncontaminated signals modelling ECG behaviours are simulated and then mixed linearly for ICA processing, as shown in Fig. 1.5. Although it is not the predominant goal of this work to develop new dynamical models for the ECG, a requirement for these simulated ECG signals

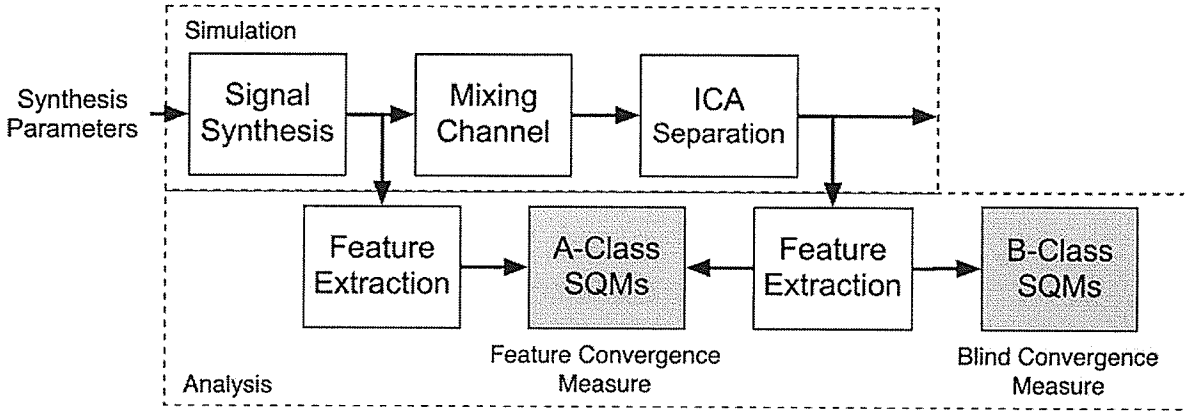


Fig. 1.5 The experimental method to assess ICA performance.

is that they exhibit dynamical fidelity to natural ECG (*e.g.*, regular morphology, statistical variability, attractors, and scaling). A simulation of a simple fetal ECG system is designed to represent the mixing of abdominal ECG. The complexity of this mixing model is kept quite low in order to concentrate on the signals themselves and the features to be analyzed. In fact, the demixing introduced by multiple ICA algorithms (as measured by the convergence of their cost functions), will be compared by parameterizing the ICA demixing matrix.

To study the effect of ICA, the separation performance is measured by several simultaneous *separation-quality metrics* (SQM): (i) *A-Class* SQMs based on *a priori* knowledge of the independent signals, and (ii) *B-Class* SQMs based only on the observed (BSP) data at hand: namely, the ICA cost function itself. The mutual relationship of the SQMs define an ICA "error-space" from different statistical, fractal, and signal features characteristic of the BSP data. Patterns can determine the relationship between A-Class errors and B-Class errors and identify non-convergence. Specifically, the multifractal (dynamical scaling) characterization of the ECG attractor reconstruction is used as one feature-based SQM. The objectives for the analysis of the ICA error-space specifically include:

(a) To determine if ICA can recover the HOS of the original source from its BSP (A-Class

convergence);

(b) To determine if ICA can recover the multifractal characterization of the original source from its BSP (A-Class convergence).

1.5.3 Novelty and Contributions

This work represents the first feature-based analysis of ICA performance (*i.e.*, where signal *quality* is measured generically in a sense beyond simple matrix or power measures). This novelty is tandem with the introduction of the new ICA error-space approach that bridges "theory" and "practice". This work also clearly demonstrates the complexity involved in maintaining an experimental paradigm suitable to connect the two "ends" of engineering. Furthermore, this work presents the analysis of multifractal (dynamical scaling) feature convergence as a new paradigm for ICA performance measurement.

Four contributions to science and technology are made with this work. First, the convergence profile methodology for ICA analysis is proposed and demonstrated, provoking a new discussion of what separation quality means in theory and practice. Second, the comparative analysis of ICA SQM-convergence establishes the first relationship of multifractal feature convergence to other ICA convergence measures. Third, the direct $f(\alpha)$ correlation-integral technique is defined, providing an extension to the canonical theory of multifractal analysis. Last, the surrogate ECG method for ECG synthesis improves the dynamical fidelity of the state of the art Oxford ECG dynamical model.

This work also contributes several resources to the community by (i) publishing code for the generation of surrogate ECG, (ii) publishing code for the calculation of an attractor multifractal analysis from a correlation partition, (iii) synthesizing a historical and geometrical background on ICA, (iv) synthesizing a survey on the fetal ECG separation problem, and (v)

synthesizing a background on multifractal analysis from a measure-theoretic perspective (that applies to attractor or wavelet methodologies), and the significance of surrogate data for the validation of multifractal analysis.

1.6 Thesis Organization

The remainder of this thesis is organized into three groups of material: (i) four chapters on background theory; (ii) three chapters on experimental design, results, and conclusions, and (iii) appendices with details on the experimental techniques used in the thesis.

The first chapter on background, Ch. 2, scopes the fundamentals of signal processing and biomedical engineering into the experimental requirements for the objectives outlined in the last section. In particular, the limitations of signal-to-noise ratio as a quality measure is addressed, the biological facts and features of fetal ECG are reviewed, and the approaches to the mathematical characterization and synthesis of ECG signals are considered. This chapter provides the greater context of the mathematical studies of features that follow.

The second background chapter, Ch. 3, is a mathematical study of higher-order statistics for the definition of SQMs and the theory of ICA signal separation. Rényi generalized entropies are presented in Sec. 3.4.7 as an extension to the traditional Shannon formula. The significance of non-Gaussianity and the details of ICA as an extension of PCA is presented in Sec. 3.5. Furthermore, the fundamental construction of ICA algorithms is described, with specific attention to the significance of ICA cost functions and the geometry of matrix space under the ICA degeneracies.

Then Ch. 4 presents a measure-theoretic approach to multifractal scaling analysis. In particular, procedures for (i) reconstructing fractal attractors, and (ii) characterizing them using the Rényi fractal dimension spectrum and Mandelbrot singularity spectrum are described. The

novel extension of multifractal theory, the direct $f(\alpha)$ correlation-integral technique, is defined in Sec. 4.7.3.

The final background chapter, Ch. 5, presents the theory of surrogate data for the validation of multifractal analysis. Surrogate data provides an important context in the analysis of fractal attractors, and, by extension, a context for the novel ECG synthesis algorithm. The difficulties that can arise in the interpretation of multifractal analysis are discussed.

The first of the experimental chapters is Ch. 6, in which the design of experiments and implementation is presented. The experiments are focused on the characterization of fractal attractors from a simulated ECG time series. Here, the techniques and procedures for the synthesis and analysis of the fetal ECG separation simulation are described. Specifically, this includes a discussion of how the statistical and fractal features identified in the background chapters are used to implement the SQM error-space. The details of the ECG synthesis method are left for a more complete discussion in App. A.

Results from the experiments that demonstrate a non-uniform convergence among HOS and multifractal features are discussed in Ch. 7.

Conclusions are finalized in Ch. 8.

Several appendices are also included to expand on some of the ancillary theory or implementation details of the thesis. Details of the newly developed surrogate ECG model are presented in App. A. A survey of the literature examining the fetal ECG separation problem is presented in tabular form in App. B. Details of the implementation for the calculation of the correlation partition of Sec. 4.6.1 is presented in App. C. Moreover, details on a novel unified algorithm for the direct calculation of all multifractal entropies is included in Sec. C.4. Last, an introduction to the wavelet-based approach of multifractal analysis which was considered for (but ultimately dropped from) the experimental method is preserved in App. D, since it

represents a notable and harmonized extension to Ch. 4. To save paper, only select code for the most significant routines in the implementation of experiments is included as text in the appendices. Complete code does appear as an appendix in electronic form on CD versions of the thesis.

Chapter II

Background on the Processing, Analysis, and Modelling of ECG

2.1 Overview

As stated in the previous chapter, this thesis studies the performance of ICA at blind source separation using generic feature-based metrics. Therefore, it is necessary (Fig. 1.5) to (i) simulate the chosen biomedical test case in controlled experiments, and (ii) extract features on which to design ICA performance metrics, the SQMs. Specifications of these simulations and features are required, and should be "as simple as possible, but no simpler"[1]. This chapter reviews the fundamentals in signal processing and biology for sufficient specifications for the experimental design. In particular,

(a) The deficiencies of signal-to-noise ratio as an SQM will be identified;

(b) The biological origins of the ECG will be surveyed to identify appropriate ECG features for (i) SQM feature extraction, and (ii) the modelling of noise-free uncontaminated ECG;

(c) The biomedical engineering background on the chosen test case, *fetal ECG* (fECG) separation is presented; and

[1] Albert Einstein

(d) Some necessary background in general signal processing and data acquisition is finalized.

Note that the "SQM-friendly" ECG features that are identified in this chapter will be given a more rigorous mathematical treatment in the subsequent chapters. Furthermore, only the basic principles of the chosen ECG model will be presented here. The details of the ECG synthesis algorithm will be relegated to App. A. Beyond this chapter, there will be no further discussion of ECG biology or fetal ECG signal acquisition.

This chapter will proceed in the following order:

(a) Discussions of signal-to-noise ratio as the prototypical generic SQM and its deficiencies;

(b) The physiology behind heart function, heart rate, and ECG electrophysiology;

(c) Fetal electrocardiography (including the acquisition process);

(d) Paradigms for analytical ECG characterization; and

(e) The modelling and simulation of the fetal ECG problem.

The final two items, in particular, involve the discussion of how to analyze and model ECG signals for feature extraction and the problem of abdominal BSP simulation . It will be demonstrated that, for this purpose, dynamical features are generic, yet complex enough to be physiologically relevant and still within the scope of this thesis. Also, specific properties of the fetal ECG separation problem will be identified that will validate the assumptions required in the ICA approach (subsequently discussed in Ch. 3).

2.2 Significance and Deficiencies of Signal-to-Noise Ratio

Signal-to-noise ratio (SNR) is the prototypical generic measure of performance in signal processing. It is used in image processing, communications, audio processing, and instrumentation. (It is especially significant in audio processing where average power on a logarithmic scale is perceived as loudness.) Its universal importance is a consequence of its simplicity and the traditional noise models that are assumed.

As it is based on a quadratic measure of error, SNR is related to the natural Euclidean L^2 and l^2 metrics. As a ratio, SNR measures the relative contribution of average energy (power) from the unwanted noise. So why not simply use SNR to measure the quality of ICA signal separation? The truth about SNR is that, due to its simplicity, it is universally applicable, but not universally significant. There are two compelling reasons to consider SQMs beyond SNR alone: (i) the significance of higher-order statistics in intelligent signal processing; and (ii) the limitations of global metrics.

It is easy to show that sum-of-squared-error (and thus SNR) is the optimal maximum-likelihood solution in the case of zero-mean Gaussian noise [180, Ch. 15][198, Ch. 9]. Furthermore, the spread of the distribution, and often the distribution itself, can be completely characterized by the SNR or noise power among a *fixed class* of zero-mean noise models. Thus, SNR is a useful tool to compare noises from the same model (*e.g.*, Gaussian to Gaussian), but not as significant when comparing noise from different models (*e.g.*, Gaussian to shot, white to correlated). The fundamental concept, here, is that for SNR to be uniquely significant, the class of noise models must be chosen *a priori*. However, as was discussed in the previous chapter, the goal of intelligent signal processing is to shun *a priori* models and let the data itself influence the technique. Why should SNR be considered a sufficient metric,

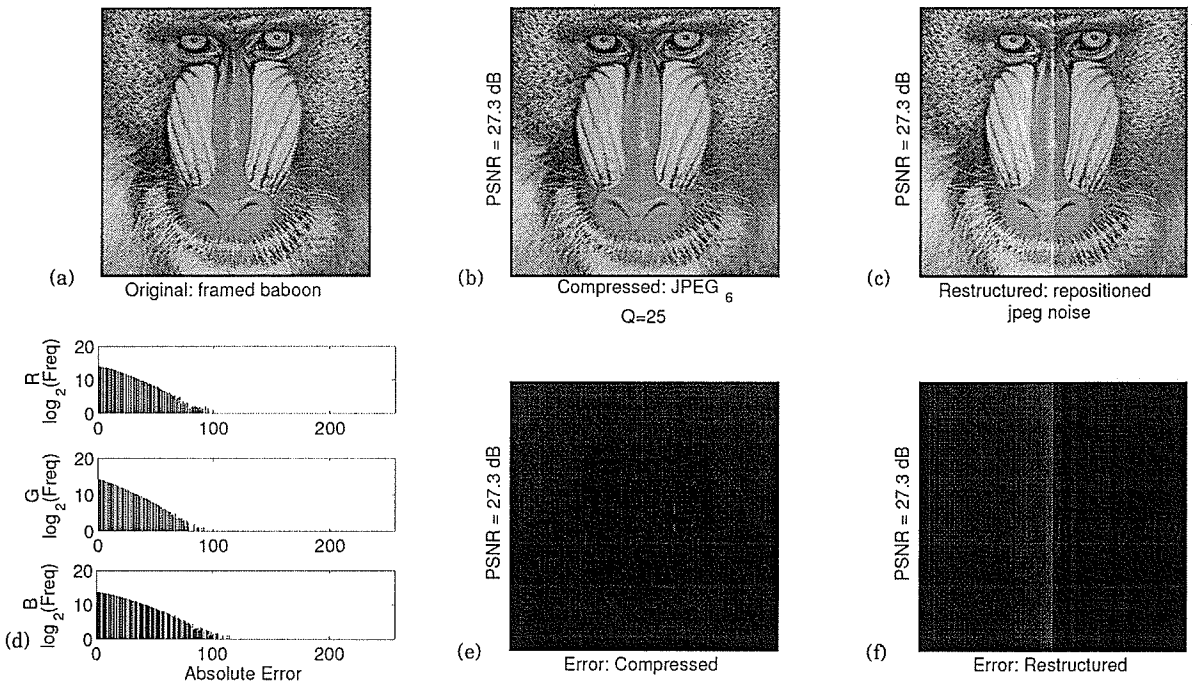


Fig. 2.1 SNR global invariance: (a) Original raster image (classic test image); (b) JPEG compressed form of panel (a); (c) An image constructed by locally resorting the JPEG compression error of panel (b); (d) Global error histogram common to panels (e) and (f); (e) Absolute error of panel (b); (f) Absolute error of panel (c); Images in (b) and (c) have identical signal-to-noise ratio with respect to (a) and identical peak-signal-to-noise ratio since the errors are globally equivalent. Perceptual image quality is distinguishable, however, because of the significant difference in local error behaviour.

then, when the noise is not restricted to a Gaussian model? If higher-order statistics are considered relevant for signal separation, they must also be considered relevant for signal quality. Notwithstanding this observation, SNR or its equivalent is still the most predominant quality metric used in the ICA literature.

The second reason to consider SQMs beyond SNR, is that SNR is a global metric, which means it is insensitive to some non-statistical properties of noise, such as "local structure". As an example, consider Fig. 2.1 which demonstrates the impact of the local structure in noise on image perception. The two noisy images in panels (b) and (c) have the exact same global

noise statistics (as displayed in panel (d)). Even though they have the same SNR, panel (c) is perceptively of lower quality. Only the organization in the local structure of the noise is different, since the error of panel (c) has been sorted into a more uniform gradient (as shown in panel (f)), instead of the scattered distribution of the original compression error (shown in panel (e)). Thus it is observed that "signal quality" (here, image perception quality) has some properties in practice that are not well represented by SNR.

Since there is a motivation to consider "signal quality" for measuring ICA separation performance, SQMs should consider the signal's other features that are useful in the signal processing paradigm. In the remainder of this chapter, the properties of ECG signals are surveyed so that generic data-driven features can be determined for SQMs. Some awkward ECG-specific metrics for the quality measurement of practical fetal ECG separation have been developed in several papers [136][196], but these are not applicable to general ICA SQM analysis. These will not be pursued, because, as mentioned in the previous chapter, it is the goal of this work to use the fetal ECG benchmark only as a specific test case for general methods.

2.3 Heart Physiology and Function

Now the biomedical context for the required elements of this work will be reviewed. In order for this work to be considered data-driven, it is important to consider the systematic origins of the electrical signals under study. This will proceed in two sections. First, in Sec. 2.3.1, the function and coupling of the heart at the system level will be described and the equivalent behaviour of the fetal and adult heart will be demonstrated. This justifies the use of a single model for the synthesis of both the fetal and maternal uncontaminated ECG. Second, in Sec. 2.3.2, an overview of the origins of cardiac bioelectricity at the cell and fibre level will be presented. This is a source of important features for both the feature extraction

and synthesis modules of the chosen experimental method.

2.3.1 Adult and Fetal Heart Function

The heart serves as the hydrodynamic pump of the body and is a vital function for both adult and fetal humans. The contraction of the blood chambers in the heart pushes blood through the vascular network carrying oxygen, nutrients, waste, and cells throughout the entire body. The heart is both a mechanical and electrical oscillator. It contracts mechanically to create vascular flow, then restores itself to a preparation state, before pushing again. This mechanical behaviour of the heart muscle is controlled by oscillating electrical signals among the heart cells, to be discussed in the next subsection. The oscillation of this joint-system is not independent of other systems. The electrical behaviour is specifically influenced by the autonomic nervous system as it coordinates the mechanical cycle (via the *vagal* and *sympathetic nerves*) with the rest of the autonomic functions [183]. Consequently, the vascular and electrical cycles of the heart, dynamically coupled to the oxygenation systems (*i.e.*, pulmonary or umbilical), are *not perfectly periodic*. Specifically, the coordination of the heart cycle rate with oxygen level is clinically known as the *sinus arrhythmia*. In a post-natal system, this is the observation that the heart cycles will speed up on inhalation (*i.e.*, high levels of oxygen in the lungs) and slow down on exhalation (low levels, respectively). In a fetal system (pre-natal), oxygen is acquired from a hand-off with the mother's blood in the placenta. Here, it is observed that the fetal heart beats slower during a constriction of the umbilical cord [142], thus reducing its connection to the oxygen supply in the placenta. Even though it indirectly uses the maternal circulation as the introductory system of oxygen (and nutrients, and also the depository of waste), the fetus maintains its own independent circulatory system and its heart coordinates with its perceived oxygen flow.

The heart is the first functioning organ of the fetus to develop, even though it is in a

primitive state. After 21-22 days of gestation (post-ovulation), the embryo has two primitive heart tubes that merge into a single tubal heart. At this stage, the cardiomyocytes become capable of beating [227]. Once the vascular circuit is completed, the heart begins beating and pumping blood. This involves the activation of myosin and actin in the presence of Ca^{2+} [227]. During this stage of development the fetal heart is beating and pushing blood, but the heart cells are still immature and have a less-organized structure than a mature heart. With development, the cells become larger in size, and come under increasing levels of tension [227]. As they grow, the cellular system of Ca^{2+} pumps and myosin myofibrils become more complex, until these structures eventually become the dominant feature in the mature cardiac cell.

By the fifth week of gestational development, Fig. 2.2, the fetal cardiovascular system is fully functional, but its regulating mechanisms and vascular structure differ from an adult system. At this stage the fetal heart is only a tube and does not have four separate chambers. It is "looped" and asymmetrical, reminiscent of the vascular position of the mature heart. As the fetal heart continues growing through cell division, walls begin to extend into the heart tube to separate what will become the four heart chambers [227]. In the adult system, the left and right systems of the heart act in series with equivalent stroke volume. They differ, however, in the pressures at which they act (low-pressure into the lungs, high pressure into the vascular circulation system). In the fetal system, the subsections of the fetal heart and the fetal vasculature operate in parallel. An adult has separate deoxygenated venal and oxygenated arterial flows, whereas the fetal vascular system has oxygenated and deoxygenated blood mixed throughout. A series of shunts in the fetal vascular pathways direct the deoxygenated blood to both the umbilical and body system, and reduce the blood flow to the lungs. Similarly oxygenated blood from the umbilical system is shunted to all vascular systems, including back to

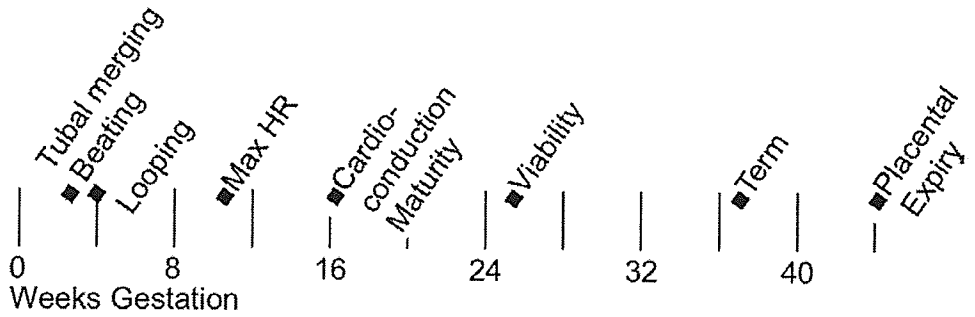


Fig. 2.2 Timeline of fetal development: Specifically, the cardiac conduction system is fully developed by week 16.

the placenta [227]. About 40% of the fetal blood flow from the heart goes through the umbilical system [189]. These vascular differences, however, affect only the flow of the blood, and not the electrical oscillation of the heart that is described in Sec. 2.3.2. By the 16th week of gestation, the conduction system of the heart is functionally mature [240, Wood and Huhta (1999)]. Significantly, the mother's heart is not responsible for fetal circulation and the two hearts conduct, contract, and beat independently. As a result of this analysis of heart function, the systematic and electrical function of the fetal and post-natal systems are sufficiently comparable that the same model for ECG synthesis and the same ECG features for SQM analysis can be used for both systems.

As above, the fetal ECG and maternal ECG can be considered the results of equivalent processes and use the same ECG synthesis model for the simulations, though parameters of the models must be different. One such parameter to consider is heart rate. As mentioned, the fetal heart rate is much higher than that of an adult. The average baseline (resting) heart rate of an adult male is near 70 *beats per minute* (bpm), and near 75 bpm for a female [237]. Baseline heart rates lower than 60 bpm or higher than 100 bpm are considered exceptional and are clinically termed *bradycardia* and *tachycardia* respectively [69]. Of course, heart rate is *time-varying*; that is, the cardiac system characteristically changes over time. It is therefore

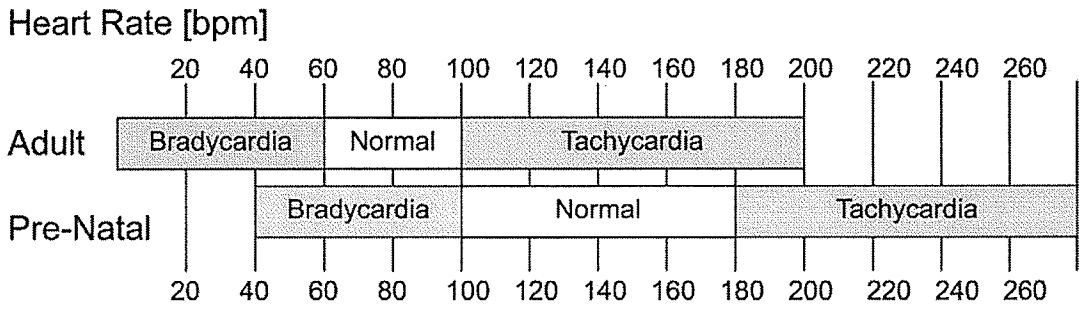


Fig. 2.3 Normal heart rates for adult and fetal function.

natural (and not unhealthy) for tachycardia to occur during exercise, and bradycardia to occur during sleep.

In contrast, fetal baseline heart rates are much faster, and change over the course of fetal development as the nervous system's control on the heart's pacemaker increases. From the fifth week of gestation, the heart rate increases until a maximum average of 180 bpm in weeks 8 to 10. [240] Fetal heart rate begins to decrease in the later stages as the nervous systems establishes control over the sinoatrial node [227]. (Fetal heart rate is, thus, approximately 140 bpm in the 18th-24th weeks, and 130 bpm at term (37 weeks) [240].) From 16 weeks on, fetal baseline heart rate is considered healthy in the interval 100–180 bpm [199]. In clinical terms, then, fetal bradycardia and tachycardia are defined as a baseline heart rate less than, or greater than, respectively, that interval. Fetal heart rates in unhealthy situations can go higher than 300 bpm [64]. There is also a strong relationship between structural heart disease and fetal bradycardia [46].

For a comparison of the values of the heart rate between adult and fetus, and their intervals of health, see Fig. 2.3. Thus, in synthesizing ECG time series for abdominal BSP simulation, the mean heart rates are significant parameters for consideration.

The following section presents the electrophysiology of the ECG process that can be considered common to adult and fetal heart function (after 16 weeks of gestation). An analysis of

the origins of the ECG signals will serve to help derive features required for the ECG synthesis model as well as SQM definition.

2.3.2 ECG Electrophysiology

As mentioned earlier, the mechanical action of the heart is triggered through a spontaneous electrochemical process native to the heart and regulated by electrochemical influences from the vagal (accelerator) and sympathetic (brake) nervous system. First, the electro-cellular properties are surveyed. Then the coupling of the cells and the nervous regulation of the heart rate is discussed.

The cardiac cells are a unique class of modified muscle cells. As with other muscle and nerve cells, a cardiac cell fulfills its electrical function as a result of an *action potential.* A cardiac activation potential is distinguished from other muscle and nerve action potentials because of its long response (150–300 *milliseconds* (ms) compared to 1 ms for other fibres) [183]. The cardiac cell has a natural distribution of heavy ions throughout its structure that maintains a non-zero resting *transmembrane potential* across its cell wall. An action potential occurs when the transmembrane potential changes as the distribution of ions (up to 10 ions [67, pg. 61]) is disrupted by their flow across the cell membrane. The two most significant ions in cardiac function are Na^+ which is the first and fastest action, and Ca^{2+}, which, while the secondary and slower action, is that which initiates muscle contraction. The action potential consists of a *depolarization*, when the transmembrane potential changes from its non-zero resting state, and a *repolarization*, when the potential returns to its resting state. The action potential of any cell is dynamically native and self-consistent [67][66], but can be stimulated by the behaviour of neighbouring cells. In general, this is to mean neighbouring cardiac cells, but also applies to neighbouring nerve cells or pacemaking electrodes where they are established. Heart cells are rather inhomogeneous in the details of their different activation

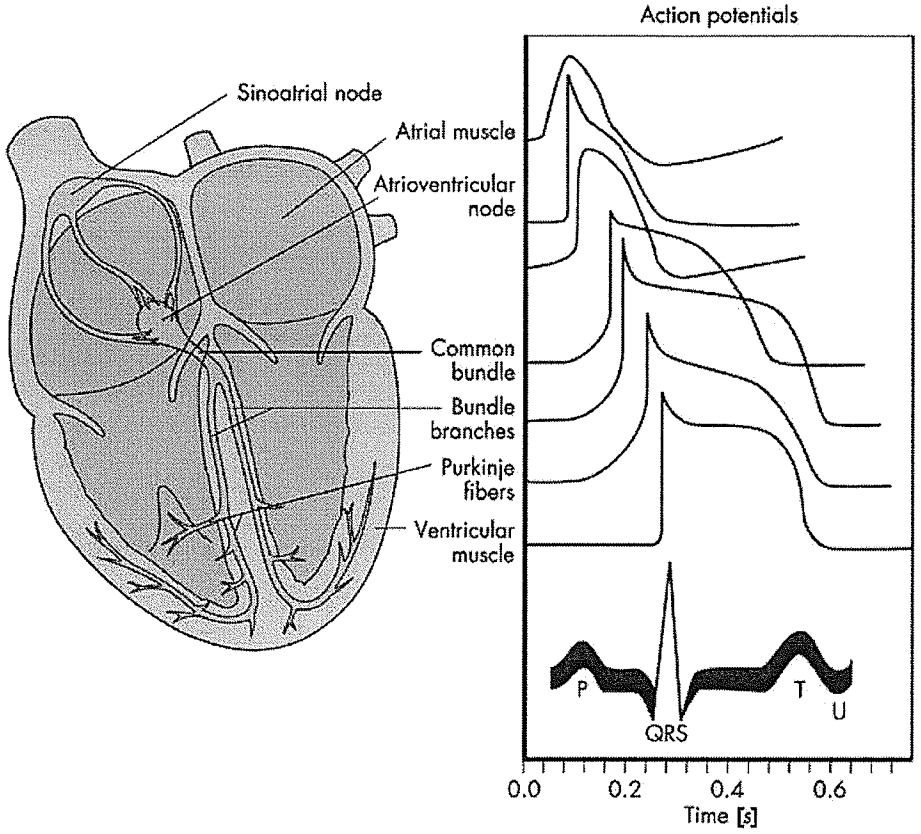


Fig. 2.4 The activation potentials of heart cells by fibre category. (from Morgan, Clinical Anesthesiology, 3rd ed., 2002, [151]. Copyright McGraw-Hill. Used by permission.)

potentials depending on their position in the electrocardiac structure, as shown in Fig. 2.4. Each type of cell is slightly different in its natural (autonomic) rhythm, refractory period, and pacemaking abilities. However, the coupling of the cells promotes a 1:1 phase lock with the fastest autonomic pacemaker in the network [67, pg. 135]. This resulting coupling is called *sinus rhythm*, with well defined *depolarization waves* and *repolarization waves* that move spatially over time through the heart's conduction network from a single pacemaking site. The collective effect of these directed wavefronts are macroscopic electrical current vectors through the cellular mass of the heart.

As mentioned before, the rate of cardiac oscillation is determined by the fastest pacemaker in the network, the *sinoatrial node*. As shown in the diagram of heart's conduction system in Fig. 2.5, this area is in the upper right atrium, ①. The depolarization wavefront then moves through the heart as depicted by the arrows in Fig. 2.5. The *atria* (upper chambers of the heart) are electrically connected and depolarize together, ②, with the wavefront moving at a slow rate (0.4 m/s [152]). The *ventricles* (lower chambers) are electrically isolated from the atria except at a special junction between the right atrium and ventrical, the *atrioventricular node*, ③. A collection of *internodal pathways* connect the sinoatrial node to the atrioventricular node so stimulation arrives before the general atrial depolarization wavefront. The special fibres of the atrioventricular node are designed to restrict ion flow and delay the conduction of the electrical impulse to the ventricles by about 9 ms. This allows for the depolarization and mechanical contraction of the atria before the ventricles are stimulated. Ventricular stimulation begins once a bundle of fibres, known as the *Bundle of His*, carry the potential impulse, ④, to the bottom of the heart in two branches. These fibres conduct very fast (2 m/s [152]), and then repetitivley subdivide into the so-called *Purkinje fibres*. These fibres deliver the stimulus to the ventricular muscle mass, ⑤, and the depolarization wave then starts moving through the ventricles from bottom to top. The mechanical contraction of the heart cells occurs after the depolarization wave has moved through the muscle mass, under the (depolarized) presence of the calcium ion.

After the mechanical contraction, repolarization of the heart proceeds in a similar fashion. However, atrial repolarization and ventricular depolarization occur concurrently. This tends to obscure any remote observation of atrial repolarization, since the ventricles have (i) more mass, and (ii) less activation delay, resulting in a stronger signal.

By this description, therefore, it is observed that each normal heart cycle is the result of

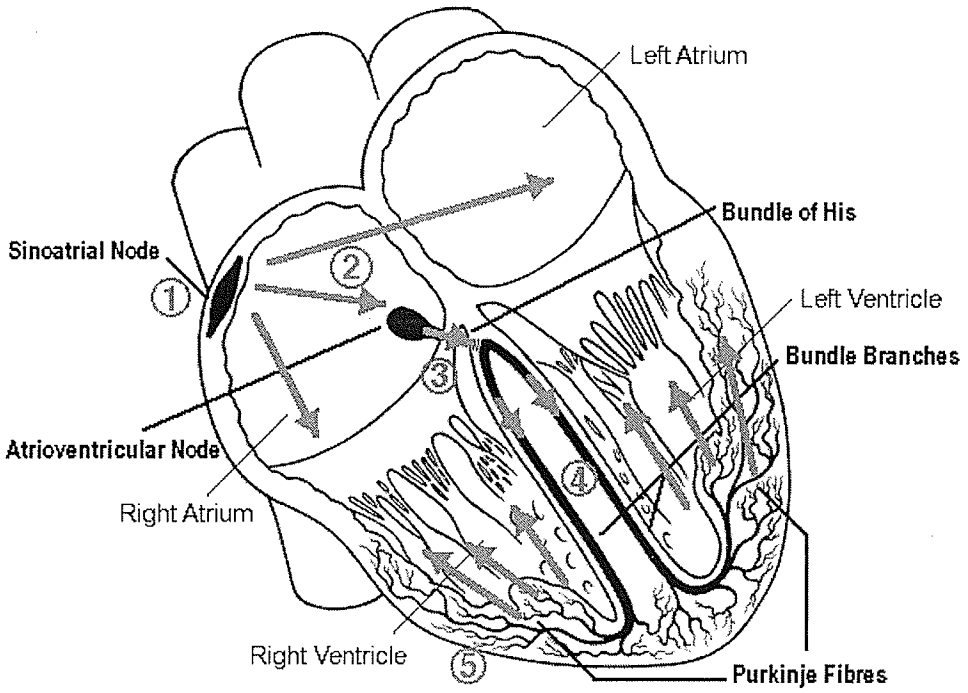


Fig. 2.5 The electrical anatomy of the heart. Modified from "Heart 3.tiff", P. Cull, ed., *The Sourcebook of Medical Illustration*: The Parthenon Publishing Group, 1989, Online. Available as of Oct. 2007 at `http://msjensen.education.umn.edu/Webanatomy/image_database/Cardiovascular/heart-conduction.tif` Copyright-free for educational use.

a massively coupled electrochemical cascade of cellular potential elements elaborately distributed through the anatomy of the heart. There is a regular structure common to the cycle, induced by the structural synchronization of the cardiac network. There is also a variability in the heart cycle as the sinoatrial node responds to the electrochemical influence of the sympathetic and vagal nervous system. The presence of nerve cells at the sinoatrial node is an important contributor to the changing heart rate of the developing fetus.

The global cyclic regularity and cyclic variability observed in the ECG is familiar to a class of nonlinear oscillators, as demonstrated in Fig. 2.6. The origins of the ECG signal is therefore analogous to snowflake growth or the dripping of water droplets: basic principles

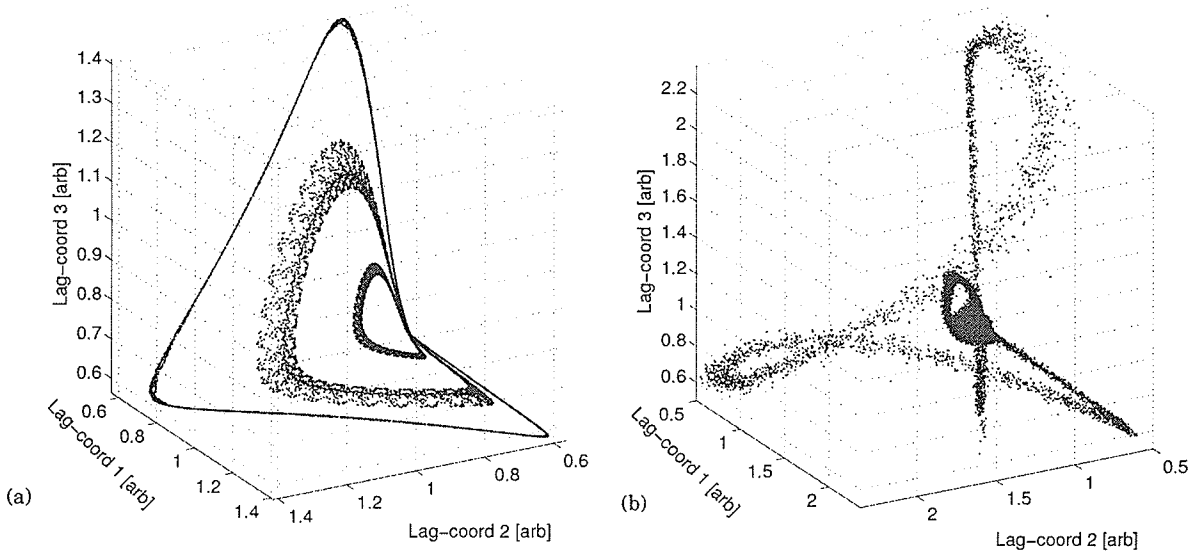


Fig. 2.6 Pseudoperiodic phase-space patterns of nonlinear oscillators: (a) Rössler oscillator (a=b=0.1,c=13); (b) Phase space of real ECG data.

of stimulation and coupling create similar but non-identical structures. Sinus rhythm is regular, and near periodic, but each cycle is a unique response to the current conditions of the cascade, as suggested by the resulting "thick curve" shown at the bottom right of Fig. 2.4. As mentioned, this description of the heart's electrical behaviour is suitable for both the fetus and adult.

In the next section the theory and practice of fetal electrocardiography is discussed. This includes:

(a) A historical review of fetal electrocardiography; and

(b) Principles of filtering and sampling for the acquisition of the ECG.

2.4 Fetal Electrocardiography

2.4.1 Signal Acquisition

The ECG is a collection of signals made from the body at large that are dominated by the electrical impulse originating from the activation potential wavefront in the heart, as described in the previous section. The fetal ECG can be taken from the skin of the fetus only by an invasion of the womb and the fetal sack. This is done, for example, during labor to monitor fetal distress [142]. Alternatively, a BSP can be taken over the mother's abdomen as shown in Fig. 2.7(a). This method leaves the fetal sack undisturbed and can be done throughout pregnancy and fetal development. This indirect method introduces an artifact from the mother's heart — effectively measuring her ECG as well, Fig. 2.7(b). Experimentally proposed as early as 1906 (by Cramer, [241]), fetal electrocardiology went through a period of research in the 1950s and '60s. Since the 1970s, it has been considered a benchmark separation problem in signal processing research [234, (1975)][228, (1987)][201, (1996)][51, (2000)]. The interest in the problem comes jointly from the significance of the source signals (and possible clinical benefit to pre-natal monitoring), as well as the difficulty of the problem due to the

(a) Unwanted signal sources (*i.e.*, noisy conditions);

(b) Reversal of power dominance (*i.e.*, the unwanted mother's ECG is dominant);

(c) Nonstationarity of the source signals; and

(d) Broad and similar bandwidth of the source signals.

Source separation attempts have never been successful to the point of developing a common clinical practice of fetal ECG analysis from abdominal signals (since the finer features of the fetal ECG are hidden). Instead, obstetric practice currently relies almost purely on

echocardiography in the form of (i) infrequent monitoring of the heartbeat by stethoscope, (ii) brief continuous or intermittent monitoring of the heartbeat by Doppler ultrasound, and (iii) infrequent Doppler ultrasound images and movies to analyze cardiac structure and function [188]. For these reasons, fetal ECG separation from abdominal signals remains an intriguing benchmark for signal processing methods, and an open field for developments in new screening techniques. As a field of research in fetal cardiology, Wood and Huhta (1999) [240, p. 817] have characterized fetal electrocardiography as follows:

> Electrocardiographic recordings of the fetal heart would be ideal to differentiate the various patterns of arrhythmias, but to date, transabdominal recordings are unable to reliably show fetal atrial P-wave morphology and the use of scalp electrode electrocardiography is only available at the time of delivery.

The BSP electrodes placed on the maternal abdomen are steel plates with low resistance which contact the skin through a sticky electrolyte paste. While electrode placement is standardized in adult electrocardiography, there is no standard electrode placement for abdominal fetal ECG. The raw analog electrical signal picked up at the electrodes is then guided by wires to the ECG recorder unit. At the recorder, there is an analog processor to define ECG *leads.* These leads are virtual wires that define the signals against standardized references and gain. Many lead configurations for adult cardiology have been developed. Two common types of leads are *bipolar leads* where the signal is taken as the difference between two electrodes (*e.g.*, Einthoven leads), or *unipolar leads* where the signal is taken from a reference ground system (*e.g.*, aVr lead) [69]. The signals at the ECG recorder unit are then scaled by amplifiers (*i.e.*, gain), low-pass filtered for discretization, and sampled using a sample-and-hold methodology and quantized [181]. This is a critical stage of the ECG acquisition. There is no universal standard for ECG digitization, and therefore care must be made in correctly cataloging for any

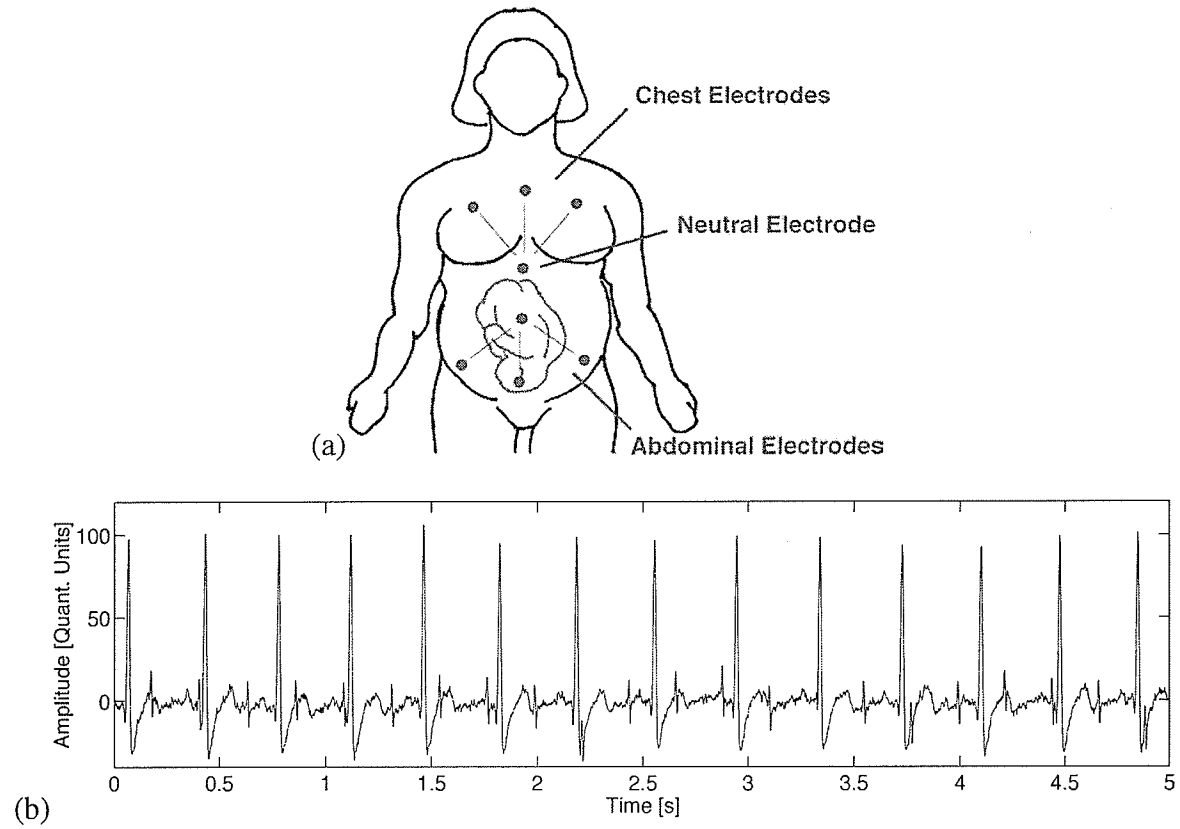


Fig. 2.7 The abdominal fetal ECG: (a) A rough schematic of electrodes that capture an fECG without disturbing the womb (After [234]); (b) the resulting waveform in a single lead with both the maternal and fetal QRS peaks with amplitude ratio ≈ 5.

record:

(a) The gain;

(b) The quantization depth and range;

(c) The cut-off frequency of the analog-to-digital low-pass filter; and

(d) The sampling rate.

The gain is important for correlating the quantization step to real units. Clipping occurs if the gain does not balance with the quantizer range. Amplitude resolution is based on the

quantization depth (*i.e.*, number of bits) as well as an optimized use of the quantizer range. The following section now discusses the principle that governs the cut-off frequency and sampling rate.

2.4.2 ECG Sampling and Non-Bandlimited Signals

2.4.2.1 Basic Sampling Theory

Now, some general principles of discretization are briefly discussed. First, the definition of an important logarithmic unit for spectral measurement is given.

Definition 2.1 (Decibel, dB [81]). A *decibel* (dB) is a logarithmic unit of quantity. Let u be a variable (on a typical linear scale). Then

$$u\,[\mathrm{dB}] = 10\log_{10} u. \tag{2.1}$$

The quantity 3 dB becomes important since it approximates a doubling; *i.e.*, $2\cdot u \to u+3$ [dB].

(A note about nomenclature: ν will be used in this subsection for an analog (linear) frequency on $\mathbb{R}$.)

The faithful conversion of analog signals into a time series is driven by the Shannon-Whittaker sampling theorem [163]. Underlying this fundamental theorem, however, lies the assumption that the analog signal at hand is *bandlimited*. That is, it is assumed that there exists an analog frequency ν_0 above which the power spectrum is zero. (Alternatively, it can be said that the signal has finite spectral support on $[-\nu_0, \nu_0]$.)

This description is impossible in practice, however, because only signals of *infinite time duration* can be bandlimited [163]. In practice, therefore, this condition is relaxed into signals with *separable* spectrum, as shown in Fig. 2.8. Here, the spectrum is separable into two distinct regions, a *passband* and a *stopband*. The spectral values in the stopband will, in

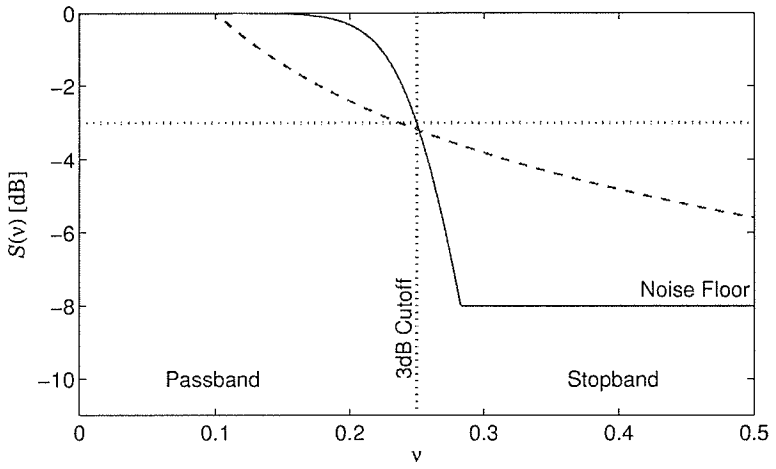


Fig. 2.8 Bandlimited signals in practice: Effectively bandlimited signals are separable into a passband and a stopband. A non-zero noise floor is present in the stopband, but the traditional concave-down skirt is well characterized by 3 dB cutoff frequency (solid). In contrast, a power-law spectral decay (dashed) is heavy-tailed and therefore is poorly represented by a 3 dB passband only.

fact, be non-zero (in opposition to the theory), but will act in all practicality as a *noise floor.* The noise floor, in most circumstances, is of sufficiently smaller amplitude to be considered negligible. The property of all classical signals (or linear filter response) is that some spectral transition, called a *skirt*, can be defined between the dominant passband (low frequencies) and the stopband (noise floor). Thus the class of "effectively bandlimited" signals are those with a well-defined spectral skirt that separates the analog spectrum. For convenience, the location of the spectral skirt's *cut-off frequency*, ν_0, is traditionally defined as the crossing point 3 dB lower than the maximum spectral power in the passband (*i.e.*, "max−3dB"), as shown in Fig. 2.8. Effectively, this considers high frequencies "negligible" if they are less than one half the maximum power. This provides the usual signal processing definition of signal *bandwidth* [163].

In the discretization of an effectively bandlimited signal, two important considerations

include: (i) the sampling frequency, and (ii) the cut-off frequency of the analog pre-filter. The sampling rate of the recorder, ν_s, determines the resolution in time of the time series, and the extent of its spectral content. The time series will demonstrate only those frequencies in the *Nyquist interval* of $[-\nu_s/2, \nu_s/2]$. Purity against anti-aliasing in the time series is provided by the recorder's analog filter and its cut-off frequency, ν_0. According to the sampling theorem, the time series correctly preserves the time/frequency structure of the original bandlimited analog signal *within the Nyquist interval*, or more accurately, below the cut-off frequency. Thus if the sampling frequency is chosen high enough such that the cut-off filter extends beyond the natural skirt of the signal's spectrum, no loss occurs in the conversion to a time series.

Note: the spectra of real-valued time series can be represented in Fourier space by conjugate-symmetric distributions of complex coefficients over the Nyquist interval, or equivalently, by (i) an even distribution of magnitude and (ii) an odd distribution of phase over the Nyquist interval [181]. All spectral images will use this symmetry and be simplified without loss of generality to show the positive Nyquist interval only.

2.4.2.2 ECG Sampling Theory

Next the general principles of Nyquist sampling are applied to the discretization of ECG and fetal ECG. A clear discussion on the non-bandlimited spectral nature of ECGs and its consequences is considered.

At issue with the sampling of ECGs is the Nyquist assumption of an effectively bandlimited analog signal. As discussed in the previous section, this is equivalent to a "spectrally separable" signal. Clearly, some signals, like sinusoidal signals, are easily separable. Others, like white noise, are not separable and, therefore, not effectively bandlimited. While white

noise and bandlimited signals may cover all traditional classes of signals, in reality other behaviours occur that must be considered. One such class is observed in Fig. 2.8, where a slow decay in the spectrum (dashed) complicates the usual theory. While the spectrum in the stopband is consistently below the max−3dB threshold, there is still a significant contribution to the overall distribution in the stopband. This would seem to violate its traditional classification as "negligible frequencies". To fill the theoretical gap, the definition below is offered.

Definition 2.2 (Broadspectral Signal)**.** A *broadspectral signal* is a signal $x(t)$ for which a 3 dB spectral cut-off changes severely the morphology, features, or structures of the signal. To clarify, let $\mathcal{S}(\nu)$ be the spectrum of $x(t)$. Furthermore, let ν_0 be the frequency at which the Fourier (power) spectrum last falls beyond 3 dB below the maximum amplitude; *i.e.*,

$$\nu_0 = \sup\left\{ \nu \in \mathbb{R}^+ \,\middle|\, \mathcal{S}(\nu) \geq \frac{1}{2} \max_{\nu'} \mathcal{S}(\nu') \right\} \tag{2.2}$$

Define a bandlimited version of the original signal $y(t) = H_{\nu_0} x(t)$, where H_{ν_0} is a low-pass filter operator with cut-off frequency ν_0. If $y(t)$ is a poor approximation to $x(t)$, then $x(t)$ is broadspectral.

In dealing with previously sampled signals, the test of (2.2) can be used, as long as only the Nyquist interval is required. That is, substituting $\omega \in [-\nu_s/2, \nu_s/2]$ for $\nu \in \mathbb{R}$ in (2.2) can be done as long as ν_0 is less than the cut-off frequency of the sampling filter.

According to this signal-processing definition, it can be observed from Fig. 2.9 on p. 38 that an ECG from the *Physionet Normal Sinus Rhythm Database* (NSRD) is broadspectral, since its bandlimited form becomes too smooth to accurately represent the edges of an ECG waveform.

What constitutes, then, an appropriate sampling rate ν_s for a broadspectral signal, if higher

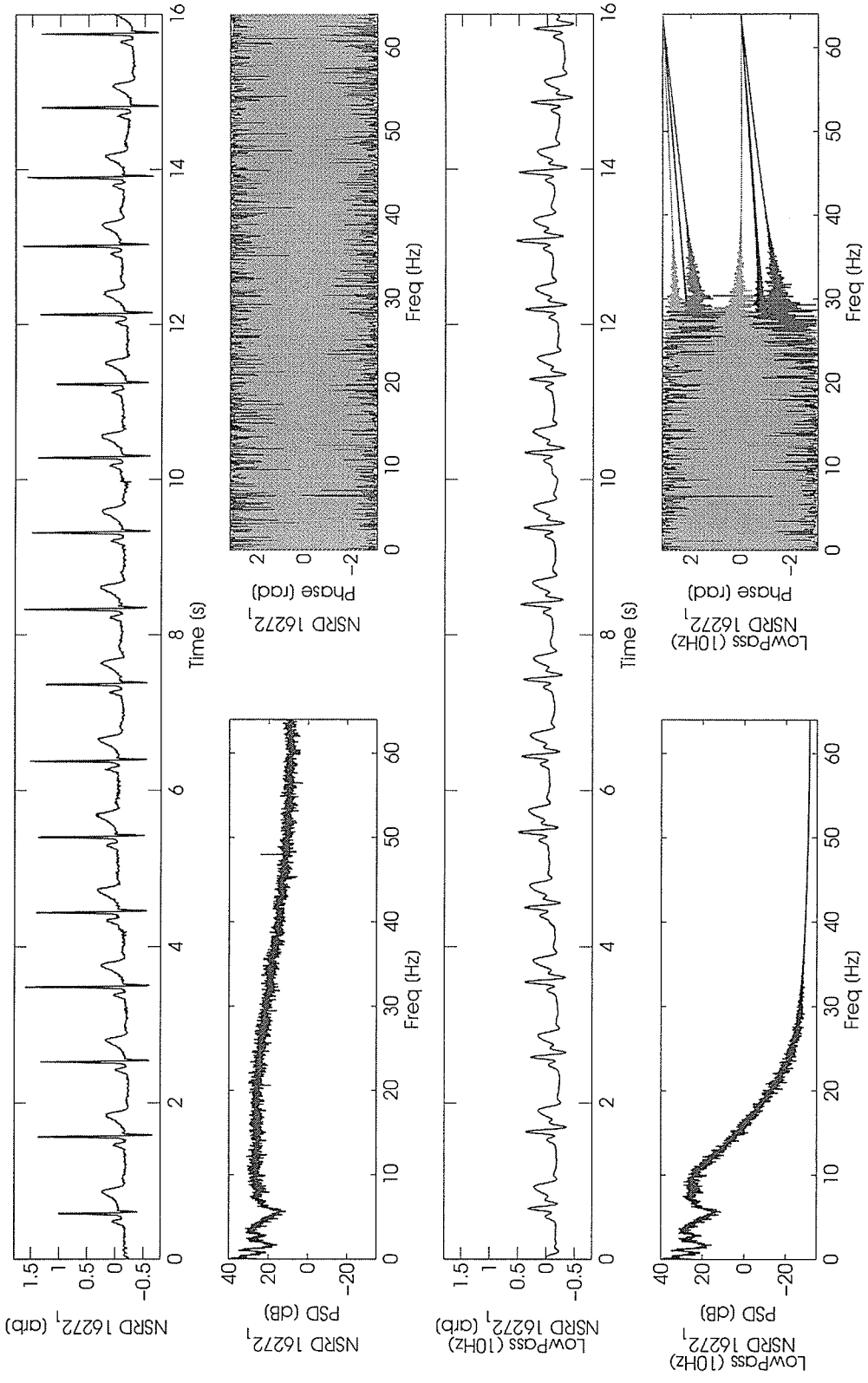


Fig. 2.9 ECG is broadspectral: an (a) NSRD time series with (b) its spectrum and (c) phase; its band-limited form as (d) a time series with (e) its spectrum and (f) phase. Note that the phase appears as a uniform distribution in the passband.

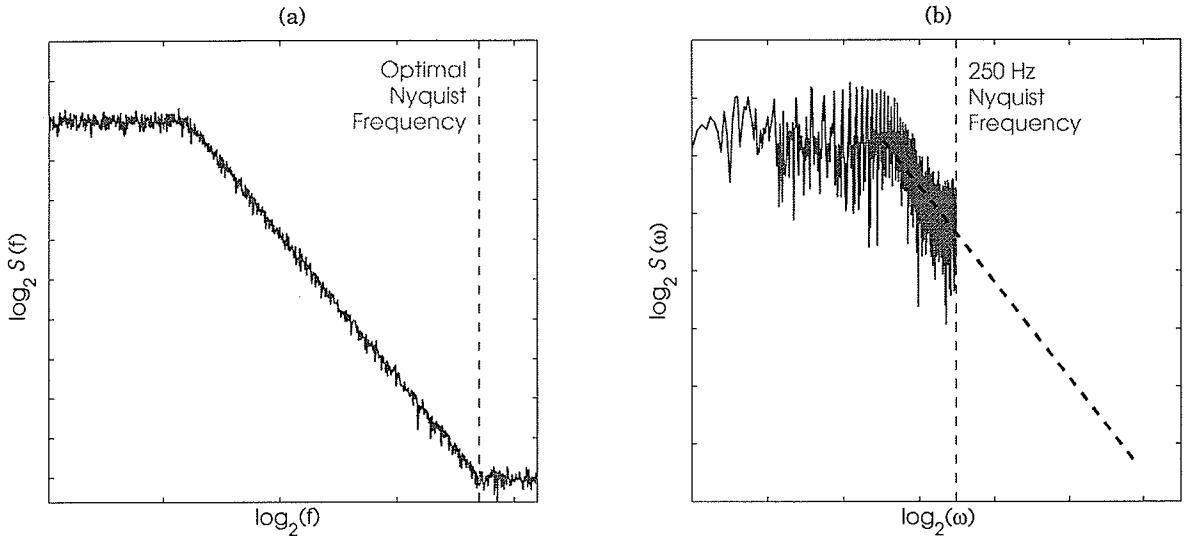


Fig. 2.10 Limitations of common ECG sampling: (a) idealized spectrum of a broadspectral signal with optimal sampling rate/Nyquist frequency in log-log scale; (b) spectrum of actual ECG in log-log scale.

frequency components are important? Broadspectral signals are conjectured to have an optimal sampling rate obtained where the analog spectrum $S(v)$ reaches the noise floor, as demonstrated in Fig. 2.10(a). At higher frequencies the signal is buried in noise, but all lower frequencies are significant. However, no recording is available in public databases such as the NSRD with high enough sampling frequency to observe the noise floor knee. The clinically used sampling rates have been determined through *qualitative* (visual) analysis of the ECG waveforms and the conditions of acquisition [186, Rijnbeek *et al.* (2001)]. As a result, no digitized ECG time series contains all of the spectral information available. This can be observed in the log-log spectrum of Fig. 2.10(b). Limitations of ECG feature analysis may occur because of this missing information [90].

2.4.2.3 ECG Sampling Practice

The practice of ECG sampling is determined mostly by the clinical context involved, rather than any information-theoretic or signal processing considerations as just discussed. Typically, the sampling rate is chosen as small as possible, within the constraints that the time series morphology is qualitatively still clinically useful. For long-term recordings (*e.g.*, 24-hour Holter monitoring), sampling at 125 *samples per second* (sps) will be considered acceptable since HRV and the presence of abnormal beats are usually studied. Vectorcardiographic studies, which provide a more detailed representation of cardiac conduction, are more likely to be sampled above 300 sps.

This work must also consider adaptations to the data acquisition practice in response to the special characteristics of fetal ECG. No clinical standards exist, yet it is recommended that a sampling frequency of 500-1000 sps be used for pediatric signal acquisition (again, measured qualitatively) [186, Rijnbeek *et al.* (2001)]. Since fetal heart rate typically exceeds that in pediatrics (*c.f.*, Sec. 2.3.1), at least that much should be considered for fetal ECG acquisition. Furthermore, since the maternal ECG is dominant and the fetal ECG is much weaker, it should be expected that the usual quantizer gain would suffice, but a greater quantizer dynamic range would be required to capture the effects of fetal ECG.

The *Physionet (Physiobank) ECG Archive* [1] has signals with sampling frequency ranging from 128 to 360 sps, and quantization from 8 bits to 12 bits, depending mostly on length and clinical use. The abdominal maternal ECG used by De Lathauwer *et al.* [51, (2000)] in Fig. 2.7 was sampled at 500 sps.

2.5 Paradigms for Analytical Feature Characterization of ECG

Body surface potentials like the ECG are very important electrophysiological recordings because, with enough *clinical experience*, they can provide information about the dysfunction of major internal systems without an invasive procedure. Clinical experience is defined here as the development of a "process of heuristic pattern recognition" [27, 1997] over a large high-dimensional space. It is the goal of the engineer to extend this qualitative process by designing a more quantitative pattern recognition process. This allows modern computers to participate in the theory and practice of ECG feature characterization, which is called "analytical feature characterization" in this work. Now, a discussion on several reasonable paradigms for developing analytical ECG characterizations will be presented.

2.5.1 ECG Morphology and Dynamics

2.5.1.1 Clinical Morphologies

As was analyzed through the previous section, the two important elements of the ECG are cyclic regularity and cyclic variability. The important features of clinical practice are regular morphological features of the ECG cycle, termed *waves*, because they can be identified with the electrophysiological behaviour of the excitation wavefront of the heart explained in Sec. 2.3.2. A typical example of such waves, labelled as (uppercase and/or lowercase [69]) P,Q,R,S, and T, are shown in Fig. 2.11. The deflection (*i.e.*, direction and amplitude) of these waves are different, however, depending on the ECG lead being analyzed. As such, the morphological features of an ECG lead are *relative* waveform features of the excitation wavefront. Each lead creates an orientation vector against which the wavefront is measured. If the wavefront moves parallel to the lead vector, than a positive wave is observed. If it moves antiparallel, a negative wave is observed, and under perpendicular motion, the wavefront produces no

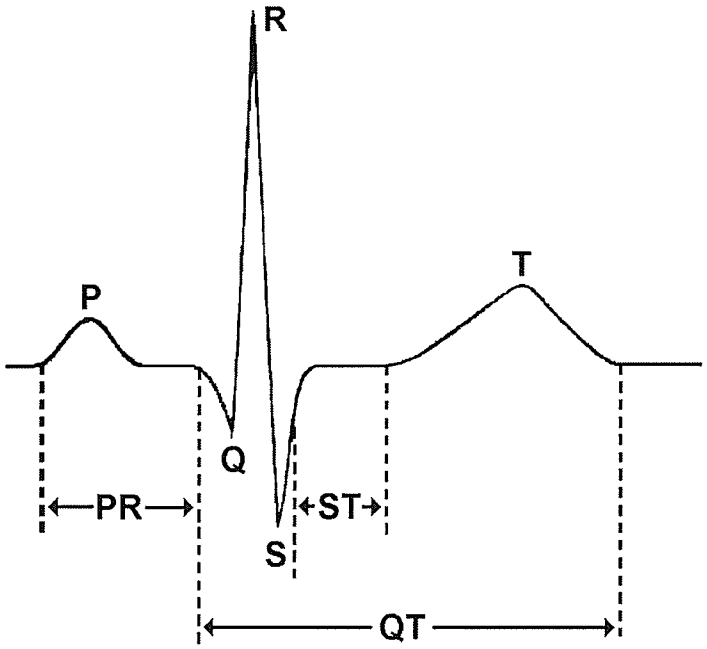


Fig. 2.11 The PQRST waveform of an ECG cycle.

wave in the lead at all. To observe the heart's electrophysiological function, a clinician notes the pattern of these waves (i) in time duration, (ii) in amplitude and form, and (iii) across the multiple leads [69, Goldberger (1999)]. As such, they perform a morphological analysis on a high-dimensional time series to discern system function using the power of clinical experience.

The first feature of the ECG cycle is the *P-wave*, which corresponds to the depolarization wavefront through the atria. It has a small amplitude and short width. The *QRS-complex* is the largest and most singular event in the ECG cycle. This corresponds to the large depolarization wavefront in the ventricular muscle (depicted as ⑤ in Fig. 2.5). Between the P and QRS-complex is a short *isoelectric* segment as the depolarization is constrained through the atrioventricular node and Bundle of His. After the QRS-complex, a normally isoelectric segment occurs as the ventricular cells stay at the plateau of the activation potential. Lastly, the

Table 2.1 Average ECG Waveform Properties

Wave	**Duration [ms]**	**Amplitude [mV]**
P-wave	60–80	0.1–0.2
PQ-segment	60–80	isoelectric
QRS-complex	80–100	1
ST-segment	100–120	isoelectric‡
T-wave	120–160	0.1–0.3

‡ Normal healthy function

T-wave is a small wave matched with the repolarization (relaxation) of the ventricular mass. (A "U-wave" is also sometimes discussed, but is rarely observed in standard ECGs.) These features identify a single cycle in the ECG (and consequently a single beat in the cardiovascular dynamics) and their properties are summarized in Table 2.1 [183, Rangayyan (2001)]. Each ECG cycle is very similar, but not identical. In particular the length and amplitude of any wave can vary, and the *baseline amplitude*, or average amplitude through the cycle, can drift. There are also many finer structures in the waveform that are not used in clinical analysis, but may contain information for automated feature classification.

2.5.1.2 ECG Cyclostationarity

Now, the cyclic variability of the ECG is analyzed, beginning with some more definitions.

Definition 2.3 (Cyclostationary [183]). A signal $x(t)$ is *cyclostationary* if it can be segmented into rhythmic cycles (or beats) of finite nominal mean length.

It follows from the definition that any periodic analog signal becomes cyclostationary when discretized. (Discrete periodic signals are a small subclass of discretized periodic analog signals [181].) Another example of a cyclostationary signal would be the frequency-modulation of a periodic carrier signal. Since the ECG has a well-defined beat of bounded length, normal ECGs are cyclostationary, as well.

Definition 2.4 (RR-interval [69]). The *RR-interval* is defined as the time duration between consecutive R-peaks (maximum positive deflection of QRS-complex) as shown in Fig. 2.12(a). (Note: some techniques use specific fiducial interpolation methods to reduce sampling jitter in RR-interval measurement.) RR-intervals are a measure of ECG cyclostationary variability.

The RR-interval, as a feature of cyclic variability, is very robust to (i) additive noise, and (ii) the low sampling frequency observed in clinical recordings. Some of the properties of natural RR-interval variability are shown in Fig. 2.12(b)-(d). Panel (b) shows how the RR-variability appears when comparing the morphology of consecutive beats of natural ECG. Panel (c) shows the trace of sequence of RR-intervals, or *tachogram*, which can demonstrate nonstationary behaviour. Quantization errors in the tachogram are inversely proportional to the ECG sampling frequency and autocorrelated.

The *mean heart rate* in bpm, $\rho_{\heartsuit}$, as used in clinical practice, is calculated from the mean RR-interval $\overline{\Delta_R}$ using

$$\rho_{\heartsuit} = \frac{60}{\overline{\Delta_R}} \tag{2.3}$$

where $\overline{\Delta_R}$ is measured over a 10 minute window [128]. (Note that the proportionality between heart rate and an RR-interval in (2.3) is only true for scalars (*i.e.*, mean rates, and not instantaneous rates) because of the nonlinear nature of the reciprocal. Further details about the interrelationship of heart rate and RR-intervals can be found in App. A.) That the RR-intervals are not constant implies a *heart rate variability* (HRV). Note that, in clinical research, there is a demonstrated relationship between HRV and adult health [98][97][172], while in obstetric practice, the patterns of fetal HRV are one of several clinical markers used to determine fetal distress [128]. HRV patterns in adult or fetal practice are typically measured from mean heart rates over small sliding windows (*e.g.*, 4-6 beats) [69][128]. However, the preferred form for studying HRV in this work will be tachograms of the form shown in Fig. 2.12(c); *i.e.*,

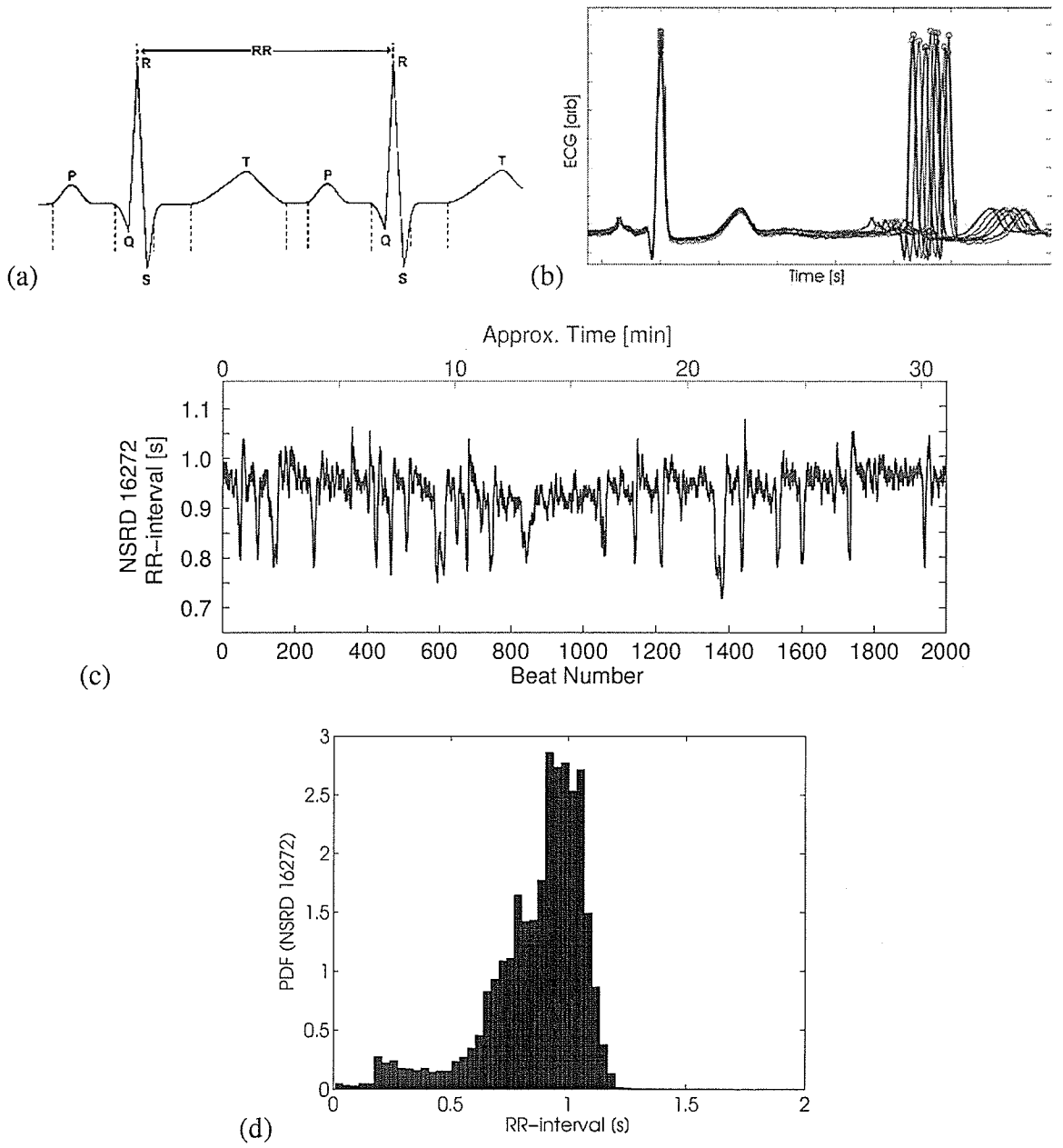


Fig. 2.12 RR-intervals measure cyclostationary variability; (a) the RR-interval definition; (b) RR variability in consecutive beats from an NSRD record; (c) extracted sequence of natural RR-intervals; (d) distribution of RR-intervals under 2 s.

RR-variability and not "rate variability" itself, though HRV will be used indiscriminantly for either quanitity.

2.5.1.3 Clinical and Analytical Characterization of Cyclostationarity

The cyclostationarity of the ECG is the dominant attribute that makes it suitable to clinical interpretation (*i.e.*, the patterns of regularity and variability in the morphological "waves"). In this way, the patterns of the waves can be connected to physiological interpretations of the internal processes of the ECG (*i.e.*, the excitation propagation of Sec. 2.3.2). Clinical experience amounts to the inference drawn from an analysis of the high-dimensional cyclostationary morphological patterns of multi-lead ECGs to the conceptual description of the internal cardiac system. The clinical analysis of HRV patterns has proven to be a more difficult characterization problem, with significant ambiguity and variation to interpretation [128, SOGC (2002)]. This is a considerable challenge to clinical practice, but is in keeping with the subtlety of the properties that HRV represents (*e.g.*, autonomic control, nervous influence).

Unlike the clinician, the engineer would prefer to use quantitative features that describe the physical processes of the system. What class of quantitative features should be considered? Figure 2.13 displays a cross-section of modelling classes with increasing complexity that can be considered: (i) morphological features, (ii) dynamical features, and (iii) physiological descriptive features.

The simplest model to consider would be morphological features (left side of Fig. 2.13). A limitation to a purely morphological characterization, however, is that no connection between the regularity and variability in the ECG process is characterized. The morphological behaviours are essentially considered in isolation of any physiological function or process it may represent. An added complication also arises because the morphological features of the ECG waves are dependent on the lead geometry, and therefore may not properly characterize

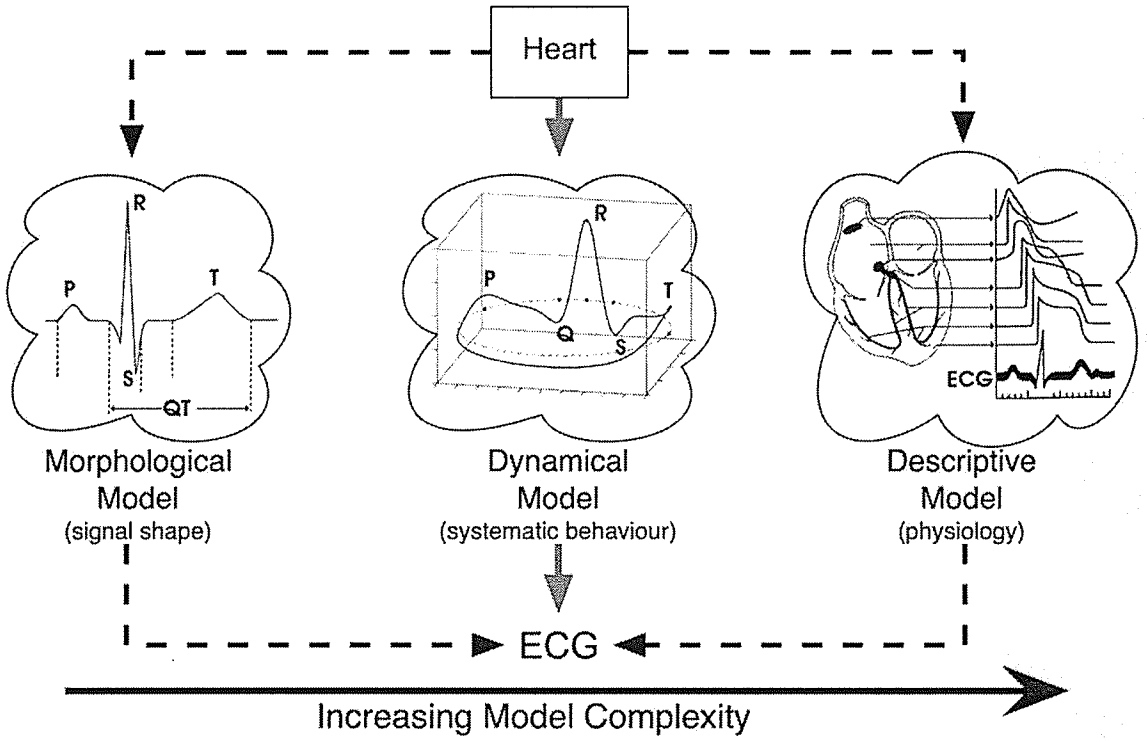


Fig. 2.13 The levels of ECG modelling. Dynamics-driven features will be used in this thesis.

the results from unstandardized abdominal ECG recordings. Furthermore, while it is true that the statistics and morphologies are the only significant elements for the ICA process, this work is interested in evaluating more than just statistical features for convergence.

However, to represent the physiological inference used in clinical experience *quantitatively* involves a difficult and exhaustive analysis to create a *descriptive model* of the heart (right side of Fig. 2.13). Here, the features could be physiologically significant (such as calculating a "complete heart-block coefficient"), but would also be highly specialized to ECG signals, and would not be applicable across signal processing in general. Essentially, this type of model is a biomedical engineering problem, and is well beyond the scope of this thesis and its goal to analyze ICA performance metrics.

As shown in Fig. 2.13, *dynamical* features are an important compromise because (i) they

support the signal morphology, yet capture some intrinsic relational structure; and (ii) do so without requiring a full descriptive model. With the origins of the signal in physiological cascades over many scales, and the connection to nonlinear oscillators demonstrated previously in Fig. 2.6, this suggests nonlinear models and complexity characterizations could be useful and sufficient. Thus, this work is justified to focus on using *generic dynamical features* from signal processing that can characterize, in part, the inner workings of a system that produces simultaneous regularity and variability.

In the next section, the generic signal processing models available for ECG feature extraction are presented. The models surveyed will provide the required dynamics, morphology, and inherent variability (at reasonable model complexity).

2.5.2 Mathematical Context of ECG Dynamics and Modelling

2.5.2.1 Signal Processing Basics

As was identified in the preceding section, the ECG is a cyclostationary signal with well-defined average morphology. This morphology, though, is dependent on the geometry of the sensors used to acquire the ECG signal. The cyclostationary variability in the ECG signal was also identified as a significant feature driven by the dynamics of the ECG. Now the analytical characterization of the identified morphology and dynamics are analyzed in the context of available mathematical models.

The most general description of an ECG signal is as a *random signal.*

Definition 2.5 (Random Signal)**.** A *random signal* is a random process for which each realization is in a family of Lebesgue square-integrable functions [22]; *i.e.*, a subset of L^2. A *random time series* has realizations in families of square-summable sequences, l^2.

Since the recording process inherently has some *a priori* uncertainty (otherwise taking a

measurement provides no information), the ECG acquisition process is essentially a random process.

The realization, x, of a random signal is the rawest form available to mathematically characterize a random signal in the signal processing paradigm. Typically, *feature extraction*, or the calculation of a quantitative descriptor of the realization, is an essential step for practical mathematical modelling.

Definition 2.6 (Signal Feature)**.** A *feature* of a signal is a characteristic quantitative variable or function onto which the family of random signals can be mapped. The set of all features describes the *feature space* for the random signals.

The purpose of feature extraction and the feature space is to provide a venue for analytical *characterization* of the random signal. Here, the patterns and structure of feature space are analyzed in order to determine distinguishable elements (*i.e.*, class identification). Depending on the context, the characterization of feature space may be a straightforward application of an *a priori* model (*i.e.*, Shannon communications), or a result from a data-driven process, such as supervised or unsupervised learning. A fully characterized feature space is then able to be used for the purposes of *classification*. Conceptually, classification is a segmentation of the feature space into different classes, but, practically, it appears as a mapping of a random signal realization onto a label. A well-designed feature extraction greatly reduces the dimensionality of the feature space compared to the raw realizations of the random signals without compromising the performance of the classification mapping. The author calls this property the *efficiency* of the feature extraction. For example, for a signal class in $l^2(\mathbb{N})$ (*e.g.*, a time series), a feature $g : l^2(\mathbb{N}) \mapsto \mathbb{R}^2$ is efficient. Good feature extraction enables proper classification with simpler mathematical structures.

2.5.2.2 *Statistical Models*

One large class of features is the statistical features, or *statistics*, of a random signal $\underline{x}$. These are the quantities (or functions) resulting from the nth-instance joint *probability distribution functions* (pdf), written f, of the random signal:

$$\begin{aligned} n = 1 \;&;\quad f(x,t) \\ n = 2 \;&;\quad f(x_1,t_1;x_2,t_2) \\ &\vdots \\ n \;&;\quad f(x_1,t_1;\ldots;x_n,t_n) \\ &\vdots \end{aligned} \tag{2.4}$$

where the realization $x(t)$ takes a possible amplitude x_i at time instance t_i.

A full *second-order characterization* of the random signal requires knowledge of these distributions up to $n = 2$. Common second-order statistical features of a random signal include [165, Ch. 9]:

(a) The mean function, $\mu_1(t) = \int_{\mathbb{R}} x f(x,t)\,dx$;

(b) The variance function, $\mu_2(t) = \int_{\mathbb{R}} (x-\mu_1(t))^2 f(x,t)\,dx$;

(c) The autocorrelation function, $R(t_1,t_2) = \iint_{\mathbb{R}^2} x_1 x_2 f(x_1,t_1;x_2,t_2)\,dx_1 dx_2$; and

(d) The autocovariance function, $C(t_1,t_2) = R(t_1,t_2) - \mu_1(t_1)\mu_1(t_2)$.

As can be observed from (2.4), a complete statistical characterization of a random signal is not efficient in the general case; *e.g.*,

$$\left\{x(t) \in L^2(\mathbb{R})\right\} \mapsto \left\{ f(x_1,t_1;\ldots;x_n,t_n) \in L^1(\mathbb{R}^{2n}) \mid n = 1,\ldots,\infty \right\} \tag{2.5}$$

Only when a symmetry exists can a statistical characterization be efficient.

Two important statistical feature symmetries that allow for efficiency are *stationarity* and *time-independence*.

Definition 2.7 (Stationarity)**.** A random signal is *stationary* if *all* its statistical properties are time-invariant (*i.e.*, invariant with respect to *translation* in time). That is,

$$\begin{aligned} n = 1 \quad ; \quad & f(x,t) = f(x,t+t_0) \quad \forall t_0 \\ n = 2 \quad ; \quad & f(x_1,t_1;x_2,t_2) = f(x_1,t_1+t_0;x_2,t_2+t_0) \quad \forall t_0 \\ \vdots \quad & \\ n \quad ; \quad & f(x_1,t_1;\ldots;x_n,t_n) = f(x_1,t_1+t_0;\ldots;x_n,t_n+t_0) \quad \forall t_0 \\ \vdots \quad & \end{aligned} \tag{2.6}$$

Stationarity implies, in particular, the second-order conditions [165, Ch. 9]

$$f(x,t) \equiv f(x) \tag{2.7}$$

$$R(t_1,t_2) \equiv R(t_2 - t_1)\,. \tag{2.8}$$

That is, the amplitudes are *identically distributed* (id), and the autocorrelation is only *lag-dependent*. If the two conditions (2.7) and (2.8) are met, the random signal is said to be *wide-sense stationary*, irrespective of the behaviour of *higher-order statistics*. Second-order assumptions like wide-sense stationarity are significant models in the theory of traditional signal processing.

Definition 2.8 (Time-independence)**.** A random signal is *time-independent* if the nth-instance joint pdfs factorize over time. That is,

$$f(x_1,t_1;\ldots;x_n,t_n) = \prod_{i=1}^{n} f(x_i,t_i) \quad \forall n\,. \tag{2.9}$$

This means that the amplitude at time t_1 has no bearing at all on the amplitude at time $t_2 > t_1$, and the random signal is completely characterized by the single-instance pdf $f(x,t)$.

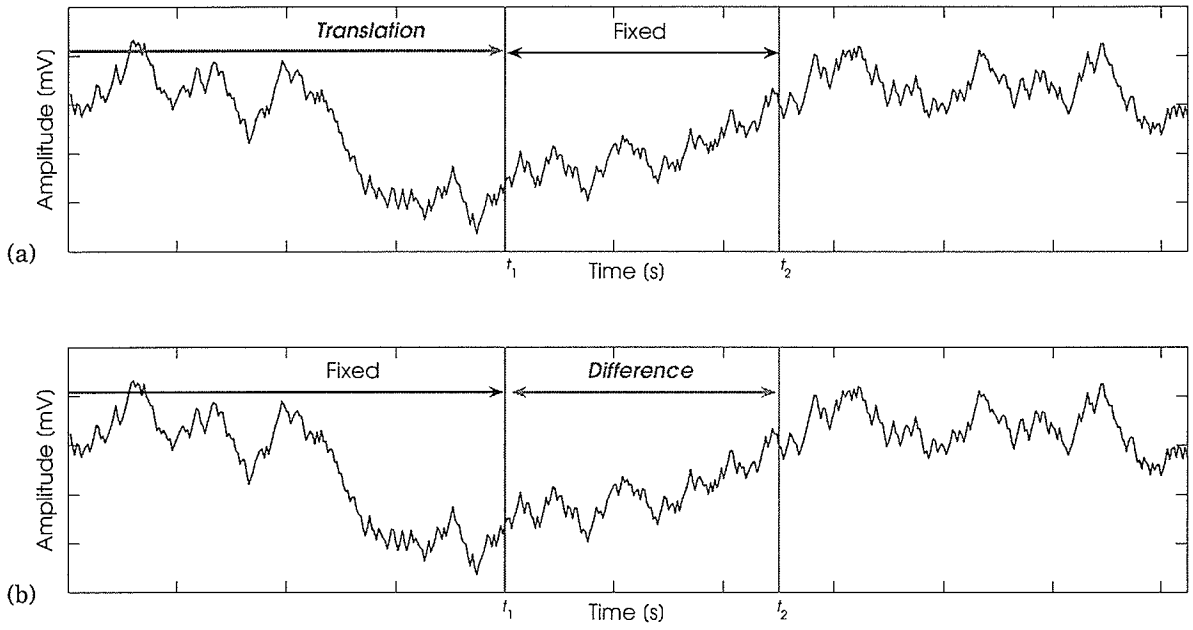


Fig. 2.14 The difference between (a) *stationary* statistics that are invariant over changing translations in time, and (b) *time-independent* statistics that are invariant over changing differences in time.

A random signal with time-independence up to second-instance is called an *uncorrelated* or *white* process. This occurs iff the autocovariance is zero except at identical times. (If the random signal is zero-mean, this is also equivalent to an impulse autocorrelation function.)

Stationarity and time-independence are complementary stochastic symmetries of random signals. Stationarity is an invariance of statistics with respect to changing *translations* in time, whereas time-independence is an invariance of statistics with respect to changing *differences* in time, as shown in Fig. 2.14.

A stationary *and* time-independent random signal is written *independent and identically distributed* (iid). This is one of the most tractable non-Gaussian models for a random signal. The class of iid signals are useful because their statistical representation reduces, essentially,

to $f(x)$; *i.e.*,

$$f(x_1, t_1; \ldots; x_n, t_n) = \prod_{i=1}^{n} f(x_i) \quad \forall n \; . \tag{2.10}$$

According to the definition used here, then, iid variables have an efficient statistical characterization.

2.5.2.3 *Deterministic Models*

Notwithstanding their efficient representation, iid random signals have no dynamical or cyclostationary structure, and are not good models of ECG signals (*c.f.*, Fig. 2.11). So a complementary analysis can be considered: random signals with *complete time-dependence.* These *deterministic* models of a random signal assume that a future amplitude is directly related to a present amplitude through some evolutionary mechanism of states. That is, for the realization $x(t)$ indexed by ζ, there exists a state vector $\boldsymbol{v}(t;\zeta)$ (of countably many components) and an evolution operator Φ which satisfies

$$\boldsymbol{v}(t';\zeta) = \Phi(\boldsymbol{v}(t;\zeta); t', t) \quad \text{where} \quad t' > t \tag{2.11}$$

and from which $x_2 = x(\boldsymbol{v}(t_2;\zeta))$ at time t_2 can be derived from $x_1 = x(\boldsymbol{v}(t_1;\zeta))$ at time t_1. From the stochastic perspective, a deterministic random signal has completely degenerate nth-instance pdfs, so that, using the Dirac delta "function" δ_∞,

$$f(x_1, t_1; \ldots; x_n, t_n) = f(x_1, t_1) \cdot \prod_{i=2}^{n} \delta_\infty(x_i - x(\boldsymbol{v}(t_i;\zeta))) \tag{2.12}$$

and all the stochastic nature of the signal is in the single-instance, or *initial conditions*, of the signal. (Contrast the deterministic symmetry defined in (2.12) with the time-independent symmetry defined in (2.9).)

An example of a deterministic random signal is $\underline{x}(t) = sin(\omega t + \underline{\theta})$. For a single realization indexed by ζ, the signal is a well-behaved deterministic function, with state vector

$\boldsymbol{\upsilon} = [t, \theta(\zeta)]^{\mathrm{T}}$. The intrinsic uncertainty in this random signal is the value of the phase, $\theta(\zeta)$. For this example, it is feasible that a regression on past points could estimate the intrinsic uncertainty of the random signal (*i.e.*, the phase) and the future amplitudes of the signal could then be predicted.

Thus, deterministic signals are characterized by the full set of (i) the evolution operator, and (ii) the initial state of the system. In some situations (*c.f.*, the example above), the evolution operator or equivalent can be "fitted" from predictable data. This approach to characterizing deterministic systems is very limited, however, and rather naive given the natural cyclostationarity variability that occurs in the ECG [107, Kantz and Schreiber (1998)]. *Predictability*, however, is not equivalent to determinism, and therefore there are other means to characterize deterministic dynamical systems.

It has become a fact of modern computer culture that *deterministic chaos* can result from simple nonlinear systems. Figure 2.15 shows the time series resulting from a deterministic system with state $\boldsymbol{\upsilon} = [r, x(0), n]^{\mathrm{T}}$ and evolution operator described by the equation

$$x(n) = r\ (1 - x(n-1))\ x(n-1) \tag{2.13}$$

with parameters $r = 3.997$, $x(0) = 0.5$, and $n = 0, \ldots, 255$. No simple regression on the time series (Fig. 2.15(a)) will uncover any state or evolution operator. However, the dependent nature of the time evolution can be seen from the constrained pattern in the next-amplitude map (Fig. 2.15(b)). The uncertainty, or lack of predictability, in deterministic chaos is a result of the incomplete specification of the *infinitely precise* elements in the state vector, and the "stretch-and-fold" dynamics of the nonlinear evolution operator [3, Addison (1997)]. This can be interpreted in information-theoretic terms as an information loss that occurs as the less significant bits of the current state vector become important bits in the next state. Future states, therefore, are dependent on information that is, essentially, beyond the current precision.

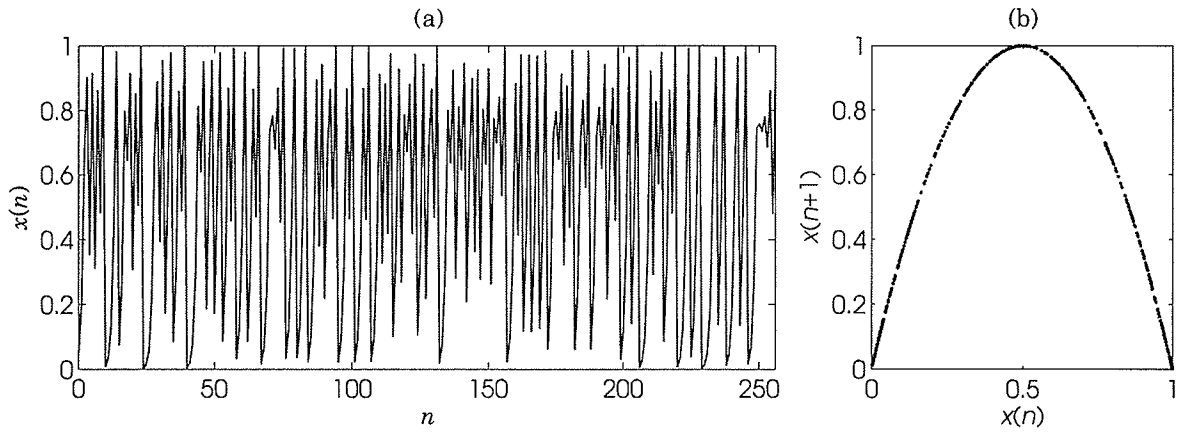


Fig. 2.15 The deterministic chaos of the logistic equation: (a) The time series of the logistic equation; and, (b) The next-amplitude map.

It is very difficult to characterize a deterministic nonlinear family of random signals exhibiting such a short *horizon of predictability* by estimating an evolution operator and the initial conditions from data. Instead, the *strange attractor* of the nonlinear dynamics is extracted as a feature and characterized.

The strange attractor is the highly complicated subset of states to which the evolution operator draws a large segment of initial conditions. This subset of initial states is called the *basin of attraction* for the attractor. The attractor is stable, in the sense that any deviation from the attractor returns to the attractor. The evolution operator, however, does not draw the system into a single state, but a complex *ergodic topology* of many states. This is in contrast with the attractors of linear deterministic systems which are topologically simple: either point attractors (*e.g.*, damped harmonic oscillator), or cyclic attractors (*e.g.*, damped-driven harmonic oscillator)[3]. Nonlinear dynamical systems, with these "strange" attractors, produce observations with the simultaneous, and seemingly contradictory, properties of equivalence and uniqueness:

Equivalence All initial conditions within the basin of attraction relax onto the attractor, and,

once there, the gross properties of their subsequent trajectory are necessarily equivalent and characteristic of the attractor topology (*e.g.*, the ergodic probability density);

Uniqueness Every observed trajectory on the attractor is unique at the smallest scale, similar to snowflake phenomena. Even close initial conditions in the basin of attraction will diverge at a fast rate, reducing the horizon of predictability.

Given these two properties of observed attractors, quantitative measures for attractor characterization focus on how the *natural measure* on the ergodic topology of the system attractor has a significant distribution (or geometry) over different topological scales [222, Theiler (1990)]. Simple attractors must lose *complexity* at smaller scales (*e.g.*, reducing to manifolds or points). The strange attractor, however, must remain complex — even at extremely small scales — in order to support unique trajectories within an invariant bounded distribution. Thus, the complexity of the strange attractor scaling, measured through *multifractal analysis* (MFA) [75, (1986)] [167, (1987)] is an important feature of the nonlinear dynamics of a random signal.

MFA measures the "scaling" (or power-law relationships) in highly singular objects with regularized complexity at every scale. An MFA of ECG for the purpose of signal compression has been studied by Huang and Kinsner in [90, (2004)]. They, as well as others [211, Small (2002)], have demonstrated the applicability of MFA of ECG signals by attractor reconstruction. This analysis of ECG feature characterization will be pursued in Ch. 4, since it is a generic dynamical signal processing analysis consistent with the cyclostationary form of ECG and its cyclic regularity and variability. As a nonlinear feature of the time series, MFA is a nontrivial feature of interest for an SQM convergence measure.

2.5.2.4 Chosen ECG Feature Model

The above has provided a reasonable justification for the restriction of random signal ECG models to the class of deterministic nonlinear families. This class of signals provides the joint properties of deterministic regularity and variability useful for ECG feature extraction (Fig. 2.16). This class of signals is also consistent with the treatment of biological dynamics as nonlinear oscillators (*c.f.*, [66, Glass (1983)][96, Glass (1992)]). It has been suggested by these laboratory experiments that the variable phase lock observed in biological systems may be consistent with deterministic dynamical models. Other experiments have demonstrated that the "quasiperiodic phase lock" (*i.e.*, aperiodicity) in ECG may be consistent with deterministic dynamical unpredictability [67, Glass (1988)][110, Kaplan (1994)][242, Small (1999)]. This chosen class of signals for analytical ECG characterization:

(a) Reduces the inefficient feature representation of the ECG as a dependent, non-stationary, and non-Gaussian random signal; and

(b) Avoids a detailed physiological approach to the ECG signal (as in Fig. 2.13), requiring (i) state vectors such as action potentials of individual cardiac cells, and (ii) dynamical relationships such as the highly complicated cardiac geometry. Such a nearly intractable model is outside the scope of this thesis.

With this restricted class of random signals identified, the next section examines the application of this signal class to the fetal ECG simulation problem.

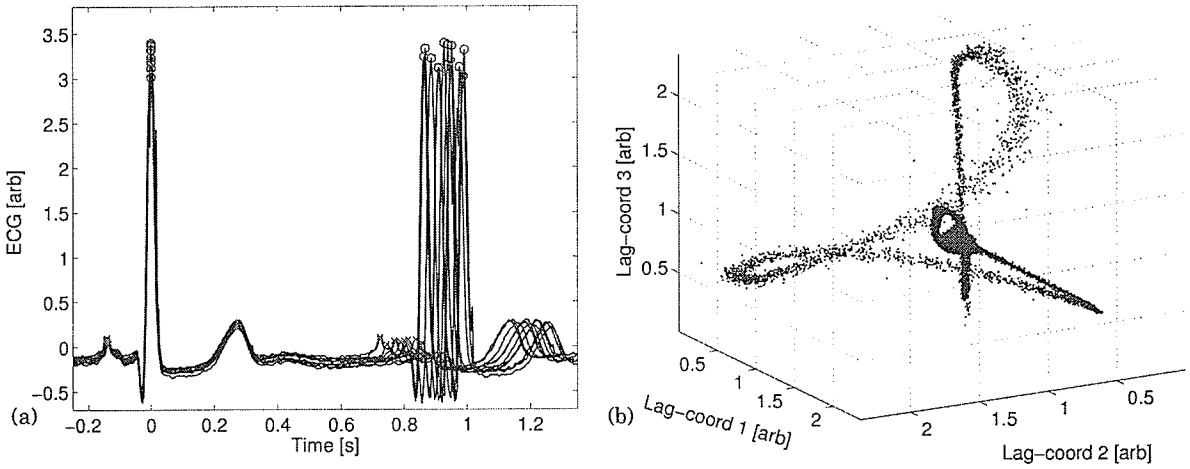


Fig. 2.16 Cyclic regularity and variability of natural ECG: (a) Morphological regularity and variability in consecutive beats from an NSRD times series; (b) Phase space representation of the ECG's regularity and variability.

2.6 BSP Mixing and Dynamical Modelling of ECG Signals

The simulation of the fetal ECG signal-mixing problem requires a discussion of (i) the characteristic mixing of an abdominal BSP, and (ii) the noise-free uncontaminated ECG modelling of the fetal and maternal ECG signals. These ideas are developed in the following subsections.

2.6.1 Instantaneous Linear Mixing Channel

Body tissue is composed mostly of water and has been proven experimentally to conduct electrical currents as a resistive volume conductor [26, Briller *et al.* (1966)]. This conductivity is nonuniform, especially considering the gas-filled areas (lower conductivity) of the lungs and blood-filled areas (higher conductivity) around the heart [170, Plonsey (1969)]. However, the electrical properties of conductive media generally supports the additivity of signals, and since the goal is not to try to recreate the internal electrical signals from the BSP (the *inverse problem of cardiology*), the inhomogeneous effects of the body tissue are irrelevant as long

as they are natural to the problem. Simply put, any time-varying artifacts in conductivity are considered part of the BSP signal, and is inconsequential to the signal-mixing problem.

More importantly, the additivity of the signals at the sensors are subject to delays introduced by conduction through the body. Again, frequency dependent delays at the electrodes due to the filtering of the body is considered native and a property of the BSP signals. However, geometry dependent delays are an important consideration in the signal-mixing problem. These, however, are of such small magnitude that all electrodes are effectively acquiring the fetal and maternal ECGs in simultaneous frames. Considering the upper bound represented by conductivity at the speed of light in water, delays would be less than $1.33/c = 4.4 \times 10^{-9}$ s/m, [238] resulting in delays no greater than 10 nanoseconds across the human body. Since kiloHertz sampling intervals are still orders of magnitude greater than these geometric delays, no effects are measurable.

According to these arguments, a BSP signal $u_i(t)$ recorded in the presence of two hearts independently and simultaneously generating action potentials in the mother's thorax can then be modelled with the simultaneous and additive (linear) equation

$$u_i(t) = \boldsymbol{m}_i{}^{\mathrm{T}} \boldsymbol{s}_{\mathrm{M}}(t) + \boldsymbol{m}'_i{}^{\mathrm{T}} \boldsymbol{s}_{\mathrm{F}}(t) \tag{2.14}$$

where the index i indicates the coefficients $\boldsymbol{m}_i$ change with the leads (*e.g.*, placement of the electrodes) and the vectors $\boldsymbol{s}_{\mathrm{M}}(t)$ and $\boldsymbol{s}_{\mathrm{F}}(t)$ represent isolated "uncontaminated" ECG signals from the maternal heart and the fetal heart respectively. These vectors are traditionally modelled as three-dimensional in the clinical literature, equivalent to the remote observation of a dipole representation of the cardiac wavefront [69, Goldberger (1999)][170, Plonsey (1969)].

Thus, the fundamental equation of the abdominal ECG $\boldsymbol{u}(t)$ can be written in matrix form

$$\boldsymbol{u}(t) = \mathbf{M}\begin{pmatrix} s_{\mathrm{M}}(t) \\ s_{\mathrm{F}}(t) \\ \text{artifacts} \end{pmatrix} \tag{2.15}$$

Here, the mixing matrix $\mathbf{M}$ is fundamentally due to the geometry of the electrodes and the positioning of the fetus within the womb. Therefore, as long as the fetus is not changing its orientation, $\mathbf{M}$ can be considered time-invariant during the signal acquisition [51, De Lathauwer (2000)]. It is again important to note that the signals $s_{\mathrm{M}}(t)$ and $s_{\mathrm{F}}(t)$ are not the potentials within the hearts' mass (or at their outer membrane), but rather the characteristic BSP of an independent system — an "uncontaminated ECG". Models for synthesizing a noise-free uncontaminated ECG are discussed next.

2.6.2 Uncontaminated ECG

The present study of separation performance in the fetal ECG mixing problem requires controlled simulations. As such, signals that represent "uncontaminated" ECG over many cycles are required. This is formalized as a definition.

Definition 2.9 (Uncontaminated ECG)**.** For the purposes of this thesis, an *uncontaminated ECG* is any noise-free *in vivo* BSP originating from heart electrophysiology alone.

2.6.2.1 Forward Problem Models

It is impossible to actually record an uncontaminated ECG (even if noise were allowed). This is because it is impossible to stop all other bioelectrical sources *in vivo* while the heart continues to function normally in isolation. Since there is no empirical solution to the uncontaminated ECG dilemma, mathematical solutions can be drawn from the *forward problem in electrocardiology* (which is a specific field of electrophysiological study) instead.

Here, source currents at the cell and tissue level are assumed in the mathematical model, and then minute electromagnetic currents are perpetuated through the body channel model where they assimilate into a measurable current at the body surface. The success of forward problem calculations rest on the accurate modelling of the source currents and body channel, not to mention rigorous and highly demanding computation. The ECG, for the purposes of this present study, reflects the properties of both the cardiac electricity and the body tissue (as a volume conductor).

In a previous study of forward problem solutions presented in the literature, Huang and Kinsner [90, (2004)] identified the work of Sachse *et al.* [190, (1998)] as a state of the art model to the forward problem. This group created a three-dimensional computer model of both heart and torso from the *magnetic resonance imaging* (MRI) dataset of the National Library of Medicine "Virtual Man" project [159]. From this high-resolution digital model (1/9 mm^3 per voxel), they applied a texture analysis to determine anisotropy in the conductive fibres. Lastly, they applied a cellular automaton model to generate the cellular currents, propagating the currents through the three-dimensional model to a BSP map.

This process correctly idealizes the acquisition of an uncontaminated ECG, having no interference from other sources, and the Sachse *et al.* result of a simulated bipolar lead is shown in Fig. 2.17. This model tries very hard to accurately represent the properties of the electric channel, and also tries to preserve a cellular model for the source currents. However, it only models the physiological origins of a single cycle, and not the dynamics of many cycles as is required for this work. Perhaps with future research, the model will account for more dynamical complexity. The current benefit of this model to this research is a verification of the ECG features in Fig. 2.11 without electrical signatures from other sources.

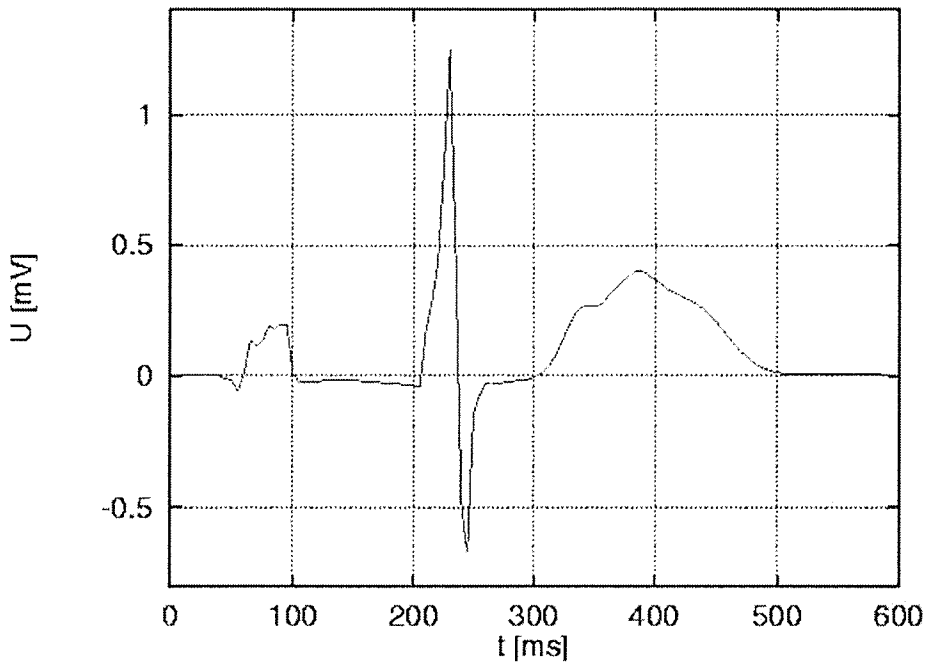


Fig. 2.17 A cycle of uncontaminated ECG from the Sachse *et al.* forward problem model. (From [233, (1998)]. Public Domain.)

2.6.2.2 ECGsyn Model and Extensions

Another recent model for the synthesis of uncontaminated ECG has been developed from a dynamical perspective. McSharry *et al.* [140, (2003)] have presented a nonlinear dynamical model for ECG simulation, called *ECGsyn*. This model is designed to recreate the morphological features of an ECG beat and also imitate the spectrum and statistics of the HRV. By first generating a random sequence of instantaneous heart rates with a specified spectrum, as shown in Fig. 2.18(a), a three-dimensional system of stochastic differential equations is integrated using a Runge-Kutta technique (Fig. 2.18(b)), and then projected down to generate a one-dimensional time series representing an ECG (Fig. 2.18(c)). The morphological and statistical features of the ECGsyn output can be influenced by the design parameters of the model. The considerable benefit of this model is that it provides a sequence of multiple beats

with a dynamical interrelationship, which imitates the real effect in the heart. In contrast with Fig. 2.17, however, there is no physiological motivation in this morphological design, (*i.e.*, no remote sensing of an activation potential wave).

For the purposes of this thesis, the work of McSharry *et al.* has been generalized in two ways to address limitations in the original model as discussed in App. A.

First, the stochastic RR-interval synthesizer has been eliminated, since there is some evidence that HRV exhibits nonlinear effects and may not be just simple stochasticity [107, Kantz and Schreiber (1998)][97, Ivanov *et al.* (1999)]. Instead, the algorithm has been modified to accept the RR-interval sequence as an argument, as shown in Fig. 2.19 and described in [175, Potter (2005)]. In this way, the natural variability properties of a real cyclostationary ECG can be expressed in the synthesized time series. RR-intervals are available directly from the Physionet databank or can be extracted from maintained ECG recordings, such as those in the NSRD. Since fetal RR-intervals are not available for use with the modified algorithm (*c.f.*, Sec. 2.4), an NSRD sequence of RR-intervals is rescaled into the range of fetal heart rate for fetal ECG synthesis.

Second, the module that transcribes the RR-interval sequence into an instantaneous angular velocity (Fig. 2.18(b), equivalent to *instantaneous heart rate* (IHR)), has been modified to keep RR-intervals invariant and preserve PR- and RT-intervals (Fig. 2.19(b)). As exhibited in the natural ECG of Fig. 2.16(a), the cyclic variability tends to be concentrated in the inter-beat (TP-) interval (*i.e.*, after the ventricular-repolarization and before the sinoatrial excitation) [107, Kantz and Schreiber (1998)]. With the modifications described in App. A, a *surrogate ECG* can be synthesized to reflect these natural variabilities. The cyclic regularity and variability of the surrogate ECG is shown in Fig. 2.20 (*c.f.*, Fig. 2.16 for comparison). The time series representation of a natural ECG and its surrogate ECG are shown in Fig. 2.21. From

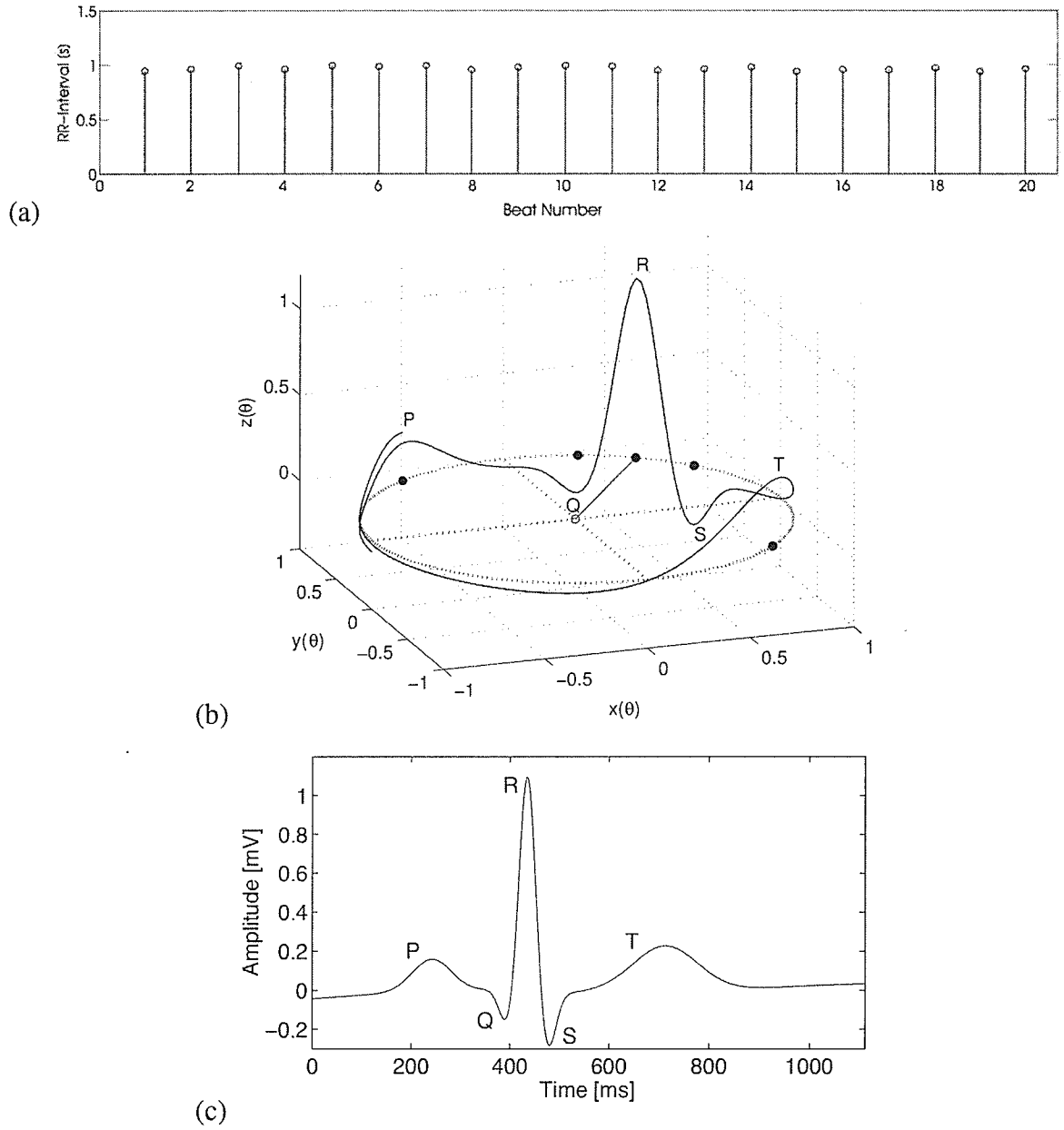


Fig. 2.18 The ECGsyn model: A random HRV sequence, panel (a), is generated and applied to a three-dimensional stochastic differential equation, panel (b) (after [140, (2003)]). A time series is created from the z-component of panel (b), which has a realistic beat morphology, panel (c).

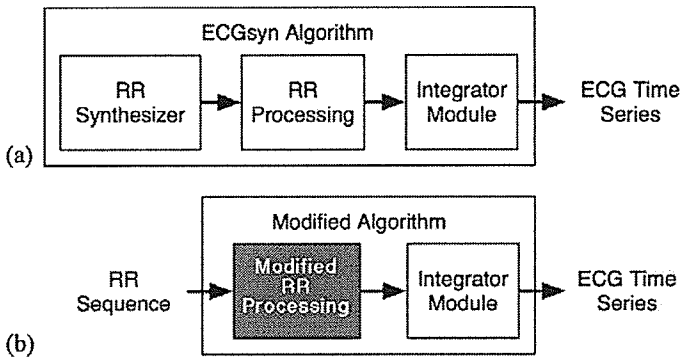


Fig. 2.19 Modifications to the ECGsyn model: (a) Original McSharry *et al.* model [140, (2003)]; (b) the new model uses natural RR-interval recordings and a modified RR-IHR algorithm.

this comparison, it is clear that the surrogate ECG represents an idealized recording from the same dynamical process as the natural ECG, but possibly from a different lead configuration.

The end result of the proposed synthesis model is an isolated noise-free simulation of a single ECG lead in the desired sampling frequency and quantized to IEEE double floating-point values. This model represents an important contribution to the state of the art in ECG modelling, but details of the surrogate ECG method will be left for App. A, since these details are not directly related to the analysis of ICA performance metrics. One further limitation that this thesis will not address at all, is the limitation of the model to produce a single lead ECG. As a single lead, many of the morphological parameters lack any physiological context. Other researchers have extended the model to the 3D-dipole context [192, Sameni *et al.* (2007)], but this is considered a biomedical engineering problem and beyond the scope of the thesis presented here.

2.7 Summary

This chapter has synthesized the relevant principles in (i) general signal processing and (ii) ECG biomedical engineering for the study of ICA at fetal ECG separation using generic

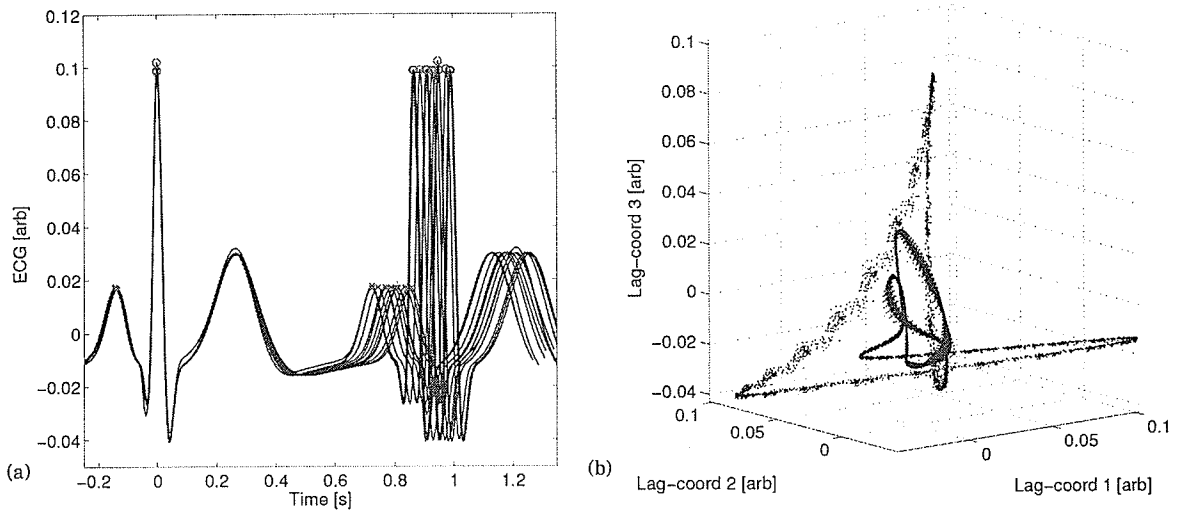


Fig. 2.20 Cyclic regularity and variability of dynamical ECG model: (a) Realistic morphology and variability in time series beats; (b) Phase space representation of cyclic regularity and variability.

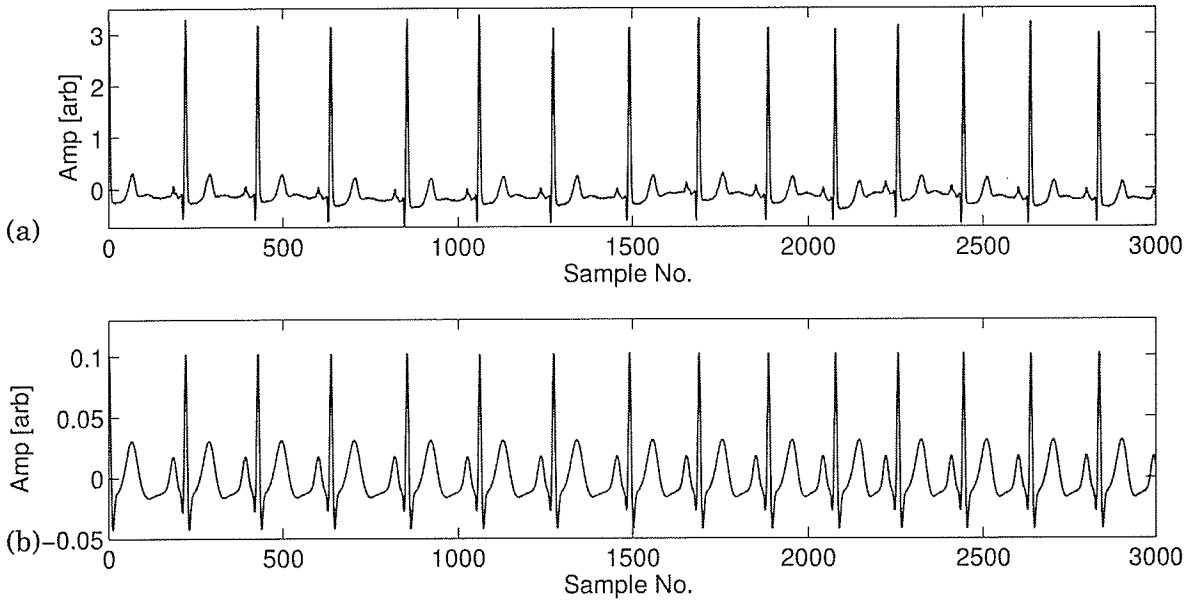


Fig. 2.21 Surrogate ECG results: (a) Dataset of actual NSRD ECG; (b) Synthesized surrogate ECG with the same event-interval dynamics. This surrogate ECG represents an idealized ECG from the same dynamical process as panel (a) but measured from a different lead (and thus with different morphologies and statistics).

feature-based performance metrics. The specifications for the experimental design identified here are intended to be "as simple as possible, but no simpler". That is, the background is sufficient to motivate and define the simulation and feature characterization to follow.

First, signal-to-noise ratio has been identified as an insufficient SQM because it is insensitive to (i) HOS, and (ii) local distribution effects. Thus other generic features should be considered for the SQMs.

Second, the biological origins of the ECG were surveyed to identify appropriate ECG features for SQM feature extraction and the modelling of noise-free uncontaminated ECG. The physiology behind heart function, and ECG electrophysiology identified that the same model for synthesis can be used in fetal and maternal cases, as long as heart rate can be set to the correct values.

Third, the mathematical context for ICA separation has been established. The linear mixing assumption for the fetal ECG signal-mixing problem has been validated. This is required specifically for the ICA assumptions to be discussed in Ch. 3.5. The clinical history and challenges of fetal electrocardiography has been described. A minimum sampling frequency of 500 sps has been identified, but not from a signal processing paradigm. Further experimental study would be required to validate the proposed means of identifying sampling frequency for broadspectral signals. These arguments complete the discussion of fetal ECG signal acquisition for the remainder of this work.

Fourth, it was demonstrated that ECGs have both a deterministic cyclic regularity and an intrinsic cyclic variability. The clinical features of ECGs were discussed, but the mathematical characterization of ECGs using both deterministic and stochastic random signal models were chosen for their universality. RR-intervals were presented as an important feature of ECG cyclostationary variability. It was demonstrated that a nonlinear dynamical paradigm would

acceptably represent both features. Therefore this paradigm will be applied in the analytical feature characterization of this work. Dynamical features are generic, yet complex enough to be physiologically relevant and still within the scope of this thesis. In particular, the statistics of the signal (stochastic) or the characterization of the attractor (deterministic) can be emphasized as appropriate for SQMs.

Last, the nonlinear dynamical paradigm has presented a means to synthesize a noise-free uncontaminated maternal and fetal ECG. A morphologically-inspired dynamical system is chosen to emulate the cyclic variability and regularity of an ECG time series over many cycles. Natural ECG-intervals extracted from recordings are used as an integral part of the simulation method in order to preserve natural HRV in the synthesized ECG. Although the basic principles of the chosen ECG model have been presented here, the complete details of the ECG synthesis algorithm can be found in App. A.

With the mathematical context for analytical feature characterization established, the "SQM-friendly" ECG features that have been identified in this chapter will now be given a more rigorous mathematical treatment. Statistical features for SQMs, as well as for the definition of ICA, will be presented in the next chapter. Attractor characterization by multifractal analysis will be discussed in Ch. 4. The final chapter of the background will be Ch. 5, which will present the theory and application of surrogate data for the validation of multifractal analysis.

Chapter III

Study of Statistics for ECG Features and Blind Source Separation

3.1 Overview

This chapter presents a study of

(a) the statistical features of ECGs for the definition of statistics-based SQMs, and

(b) the theory behind ICA separation.

It will begin with an introduction to the ECG distributions that will be characterized. In particular, their distinction from a random process is clarified, and their visualization and modelling is introduced. Thereafter, the definition and significance of statistical moments and cumulants as general features of uni- and multidimensional pdfs are reviewed. In particular, the higher-order statistics are distinguished from the more common second-order statistics. Next, the mathematical backbone behind ICA is reviewed in Sec. 3.4: namely the statistical independence of a joint distribution. Independence is then characterized in terms of pdf visualization, moments, and entropies. The most significant definition here will be the mutual information of a joint pdf. Also of interest, a generalization to (Shannon) entropies, called the Rényi generalized entropies are discussed, which will prove useful for Ch. 4. As a statistical feature, these entropies are introduced here.

Section 3.5 then presents the specific background on independent component analysis. First, a historical review of the ICA literature is presented to orient the reader with the contributors and developments in ICA's short history. (For a literature survey specific to the application of fetal ECG separation, including ICA techniques, see App. B.) Thereafter, in keeping with the signal processing paradigm of this work, the classical principal component analysis is reviewed (Sec. 3.5.2), from which ICA itself is formulated as a higher-order extension (Sec. 3.5.3). Theorem 3.22 in Sec. 3.5.3 reviews the seminal characterization of an ICA process and the degeneracies that must exist in an ICA solution. This is discussed further in Sec. 3.5.4, where the geometrical context of optimization over matrices is analyzed and some visualization of the ICA process in matrix space is presented. These geometrical features of ICA are significant in the state of the art ICA techniques. The application of the general ICA theory to the derivation of specific algorithms (Infomax and FastICA) is then exemplified in Sec. 3.5.5. Further considerations of how the general theory applies to practical data is discussed in Sec. 3.5.6.

The last element of ICA background is a section reviewing the performance metrics used in the literature to evaluate ICA separation (Sec. 3.5.7). Here it is shown how the indeterminacies of ICA must be considered in ICA separation quality metrics, and also that the majority of published metrics are second-order with only indirect consideration of higher-order features. From these, a methodology is proposed for this work to standardize the ICA indeterminacies, and a collection of interesting elements for convergence analysis is identified.

3.2 ECG Distributions

Statistics are tools used to characterize *distributions*. In the event that this is the distribution of a *stochastic* phenomenon, statistics are the complete characterization of the system. A

distribution, more generally, is just a multiplicity of a variable and occurs even for deterministic (non-stochastic) phenomena. As discussed in Ch. 2, deterministic systems simply require more for a complete characterization.

In this section, three questions about the application of statistics to ECG will be addressed. First, what are the ECG distributions being characterized, and how are they distinct from ensemble statistics? Second, how can these distributions be visualized? Third, what approaches are there for modelling these distributions?

3.2.1 Ensemble vs. Time-Integrated Distributions

In the last chapter, it was demonstrated that a random signal is the broadest general class for ECG characterization. That is, any ECG signal is a member of an *ensemble* of square-integrable functions. As such, the ECG process can be characterized by a distribution over an ensemble $\{x(t) \in L^2(\mathbb{R})\}$: that is, an *ensemble distribution* of time-functions. For the characterization of a *recorded ECG signal*, however, this is no longer the case, since one realization, $x(t)$, has already been instantiated (or extracted) from the ECG ensemble. A characterization of a single ECG recording by "statistics", therefore, is not in reference to the ensemble multiplicity.

Here, the *time-integrated distributions* of the signal is the source of the statistical characterization. Moreover, since a multichannel recording of a BSP is being considered, it is a multidimensional distribution from the observed signal, $\boldsymbol{x}(t)$. The dimension, or *number of channels*, of these signals will be noted by N_c. Thus the time-integrated distribution is from the ensemble of vectors $\left\{ \boldsymbol{x}(t) \in \mathbb{R}^{N_c} \mid t \in \mathbb{R} \right\}$ indexed by their time coordinate t.

This does not imply that ECG signals are stationary, time-independent, or stochastic as defined in the previous chapter. Neither is it saying that the time-integrated distribution of one ECG recording must be characteristic of an entire ensemble of ECGs. This demonstrates only

that the multiplicity is taken over time instances, and, at most, one need only treat distributions over N_c-dimensions. A statistical distribution f will be identified with a N_c-dimensional signal $\underline{\boldsymbol{x}}$ using the expression "distributed as", $\underline{\boldsymbol{x}} \sim f$. Furthermore, the distribution f is directly related to a unique random variable $\underline{\boldsymbol{u}} \in \mathcal{R}(\mathbb{R}^{N_c})$, where $\mathcal{R}(\mathbb{R}^{N_c})$ is the set of all random variables over $\mathbb{R}^{N_c}$. It follows from this formal connection that the statistical characterization of the signal $\underline{\boldsymbol{x}}$ and the random variable $\underline{\boldsymbol{u}}$ are identical, since they share the same distribution f. Therefore, one can characterize "random variables" in this chapter without signal processing consequence. These variables can be distinguished phenomenologically, however, since $\underline{\boldsymbol{u}}$ is random while $\underline{\boldsymbol{x}}$ need not be. (Further discussion of this association can be found in the chapter on surrogate data, Ch. 5.)

(Note: The definition of stationarity for a time series is therefore somewhat different from the corresponding definition for a random signal, since, again, there is no ensemble. A time series is considered stationary when the time-integrated distributions of a sliding interval of data are invariant. Apart from truly stochastic and iid signals, stationarity is uncommon.)

Using time-integrated distributions is perfectly consistent with the signal separation of instantaneously mixed sampled ECGs sampled at sampling frequency ν_s. Since

$$\boldsymbol{x}(t) = \mathbf{M}\boldsymbol{s}(t), \quad \forall\, t \in \left\{ n\nu_s^{-1} \mid n = 0, \ldots, N-1 \right\} \tag{3.1}$$

time integration results in two data sets, $\{\boldsymbol{x}\}$ and $\{\boldsymbol{s}\}$, related by a fixed matrix $\mathbf{M}$. The ordering of these data sets is *not important*, as long as the matching between the instantaneous vector-couples is maintained; *i.e.*, (3.1) holds true and $\mathbf{M}$ is constant. Although any time-shuffling of $\boldsymbol{x}(t)$ is no longer a meaningful ECG signal, the matrix $\mathbf{M}$ and the distributions of the components are all invariant, and ICA focuses on estimating $\mathbf{M}^{-1}$ from these distributions.

This study now proceeds by discussing the two approaches to the modelling of an ECG distribution.

3.2.2 Parametric vs. Non-parametric Modelling

The statistical characterization of an ECG is a feature mapping the samples from the ECG signal $\{\boldsymbol{x}(t_n)\}$ into a representation of the time-integrated distribution, f. This statistical modelling is organized into two broad classes: *parametric*, and *non-parametric* modelling.

Parametric modelling consists of mapping the statistical samples into some finite set of parameters that tune an *a priori* model to the data. In signal processing this approach is very common because the Gaussian model is chosen *a priori* to represent the statistics. The goal of parametric modelling is to estimate these parameters in an optimum way given the model. Optimality of these estimators is classically measured by considering bias and variance [198]. A complete parametric statistical characterization then amounts to a calculation of the *minimum-variance unbiased estimator* (MVUB) for the model from the samples.

The benefit of a parametric model is its simplicity and efficiency with finite data [198]. The risk with any parametric modelling, however, is that model mismatch introduces a *systematic bias* that is inconsistent with the actual data. If the underlying distribution of the data is represented poorly by the chosen model, the calculated parameters will not optimally characterize f. Moreover, while the parameters still represent *some* statistical feature of the data, the significance of their interpretation is lost.

Systematic bias can be avoided by using a non-parametric approach to distribution modelling. Instead of being model-driven, these techniques are data-driven, and more consistent with the objectives of ICA. The two most common non-parametric approaches to distribution modelling are (i) statistical moments, and (ii) kernel density estimation.

Moments are similar in spirit to the MVUB estimators in parametric modelling (and under certain conditions MVUB estimators can be moments). They are features calculated directly from the samples which represent some property of the underlying distribution. They are

universal, however, and not optimized to a particular distribution. Since moments, and specifically higher-order moments, are significant to ICA, they will be examined more deeply later on in this chapter.

Kernel density estimation assumes a small kernel of probability should be assigned to the region surrounding an observed sample [25]. The aggregation of these kernels provide an estimate of the complete distribution. This approach to distribution modelling requires a choice of mathematical kernel and its properties. The most important property of the kernel is the smoothing parameter, which assigns the spread of the kernel around the observed sample.

The price that generally comes with non-parametric modelling is an increase in *systematic variance* as compared to parametric modelling. That is, the consistency of modelling results from different instances of finite data is decreased.

This study now proceeds by discussing the visualization of these distributions, and then their characterization by moments and entropy. Then independence and both PCA and ICA will be developed.

3.2.3 ECG Distribution Visualization

Visualization is key to understanding the behaviour of distributions intuitively, and it will prove useful in the interpretation of ICA. The following defines four types of distribution visualization techniques.

Definition 3.1 (Scatterplot)**.** A *scatterplot* is the most raw form of distribution visualization. A finite dataset $\left\{ \boldsymbol{x}(t_n) \in \mathbb{R}^{N_c} \mid t_n \in \mathbb{R},\ n = 1, \ldots, N \right\}$ drawn from the given distribution is then presented by individually marking the data points in $\mathbb{R}^{N_c}$. As such, visualization can only occur unaided for $N_c = 1$ and 2. Using perspective-based projection, $N_c = 3$ can also be drawn on the page and occasionally proves useful.

Definition 3.2 (Intensityplot)**.** An *intensityplot* is a slightly more processed form of distribution realization. This expresses the behaviour of the distribution function f by approximating it with pixel colourings scaled to the amplitude of the distribution. This correctly represents the mapping nature of the distribution, and can be efficiently visualized for $N_c = 2$. Estimating an intensityplot from real data can be done using histogram, nearest-neighbour, or other non-parametric techniques [25]. In black and white printing, these plots can provide little detail, and will be used only infrequently in this thesis.

Definition 3.3 (Densityplot)**.** A *densityplot* is the formal visualization of a distribution's density function, $f : \mathbb{R}^{N_c} \mapsto \mathbb{R}$. This is best accomplished for one-dimensional variables $N_c = 1$, though for $N_c = 2$ some perspective-based projections are useful. Estimating the values of a distribution from real data can be accomplished by the same non-parametric estimation methods as for the intensityplot. The densityplot is the most natural form of visualization of an analytic distribution formula.

In particular, a technique to be used repeatedly will now be formalized in a definition.

Definition 3.4 (Histogram)**.** A *histogram* is a piecewise-constant non-parametric estimate of a distribution density from a finite data set in $\mathbb{R}^{N_c}$ calculated from non-overlapping volume elements. Here,

$$\widehat{f}(x) = \frac{N_i(x)}{N\,V_i(x)} \tag{3.2}$$

where $N_i(x)$ is the number of data points in the cell i containing x and having volume $V_i(x)$. For normalization, $N = \sum_i N_i$, which is the size of the data set. If the volumes of all cells are identical, $V_i = V$, then the histogram is called *regular*.

Figure 3.1 demonstrates the various visualizations of Gaussian noise in different dimensions. The (fetal ECG) abdominal BSP data from de Lathauwer *et al.* (2000) [51] is visualized

as time-integrated BSP distributions in Fig. 3.2.

3.3 Moments

Now, the characterization of distributions by moments is discussed. As described in Ch. 2, moment-based characterizations of a distribution are divided into two categories: (i) second-order moments, and (ii) higher-order moments. The second-order statistics capture the basic idea of a distribution's location and spread. Higher-order statistics measure different properties of a distribution's shape.

Before continuing, a few definitions are recalled. Here, for simplicity, distributions will be treated as random variables. Unidimensional random variables will be represented by an underscore $\underline{x}$, while (multidimensional) random vectors appear bold as well, $\underline{\boldsymbol{x}}$.

Definition 3.5 (Expectation)**.** Let $\underline{\boldsymbol{u}}$ be random vector in $\mathbb{R}^{N_c}$, written $\underline{\boldsymbol{u}} \in \mathcal{R}(\mathbb{R}^{N_c})$. The expectation of a function $g : \mathbb{R}^{N_c} \mapsto \mathbb{R}$ with respect to $\underline{\boldsymbol{u}}$ is

$$\mathcal{E}\left\{g(\underline{\boldsymbol{u}})\right\} = \int_{\mathbb{R}^{N_c}} g(\boldsymbol{u}) f(\boldsymbol{u})\, d\boldsymbol{u} \tag{3.3}$$

where $\underline{\boldsymbol{u}} \backsim f$. In general, one may view the expectation as an operator $\mathcal{E}\{\cdot\}$ on the argument with respect to the random variable. If g is a vector function, $g : \mathbb{R}^{N_c} \mapsto \mathbb{R}^{N_c'}$, then the expectation is taken component-wise to produce a unique vector $\mathcal{E}\left\{g(\underline{\boldsymbol{u}})\right\}$.

Definition 3.6 (Cumulative Distribution Function)**.** The *cumulative distribution function* (cdf) of a one-dimensional random variable $\underline{u}$ is the function

$$F(u) = \int_{-\infty}^{u} f(u')\, du' \tag{3.4}$$

where $\underline{u} \backsim f$.

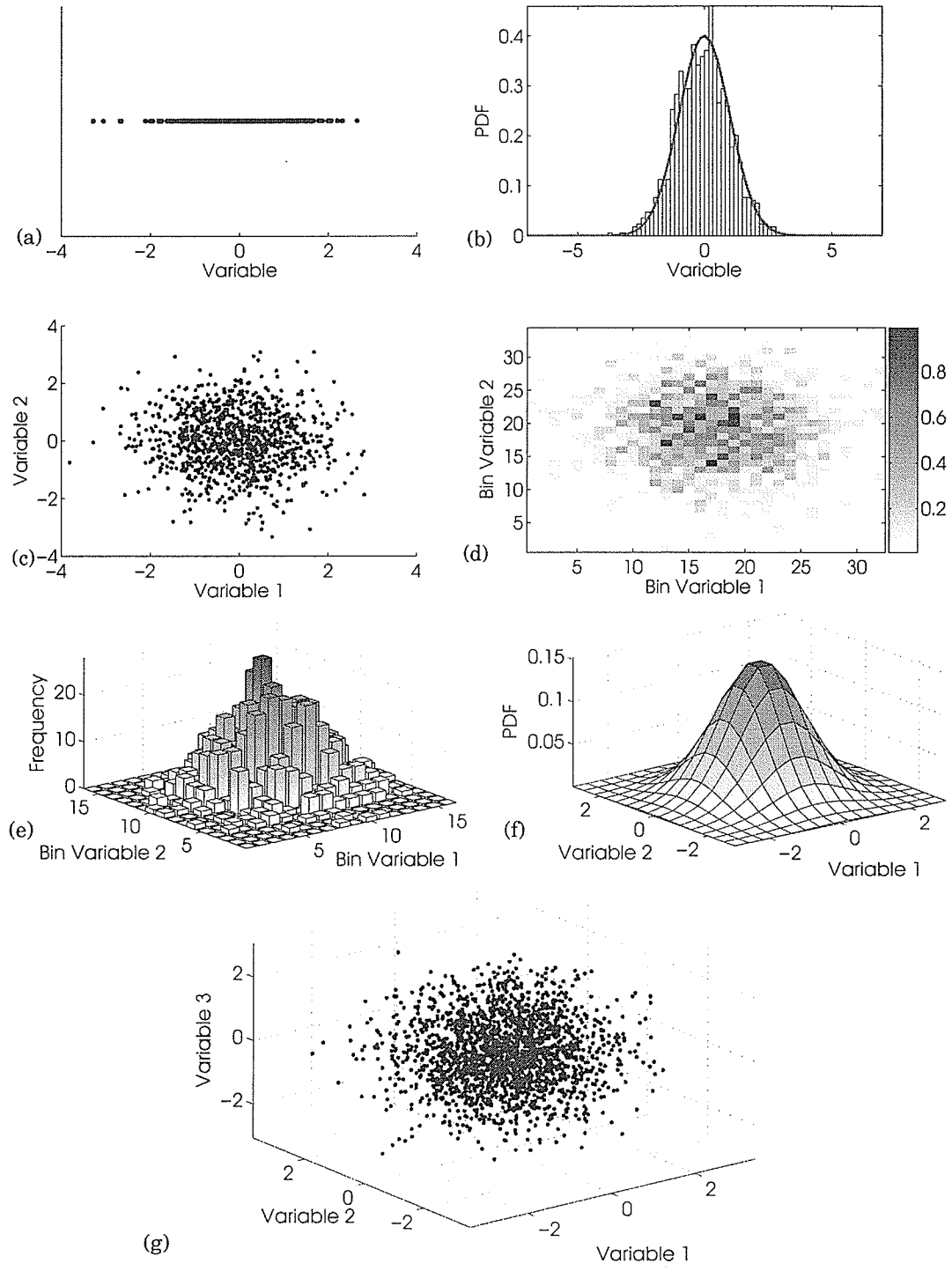


Fig. 3.1 The various visualizations of distributions: (a) 1-D Gaussian scatterplot; (b) 1-D Gaussian histogram and densityplot; (c) 2-D Gaussian scatterplot; (d) 2-D Gaussian intensity-plot; (e) 2-D Gaussian histogram; (f) 2-d Gaussian densityplot; (g) 3-D Gaussian scatterplot.

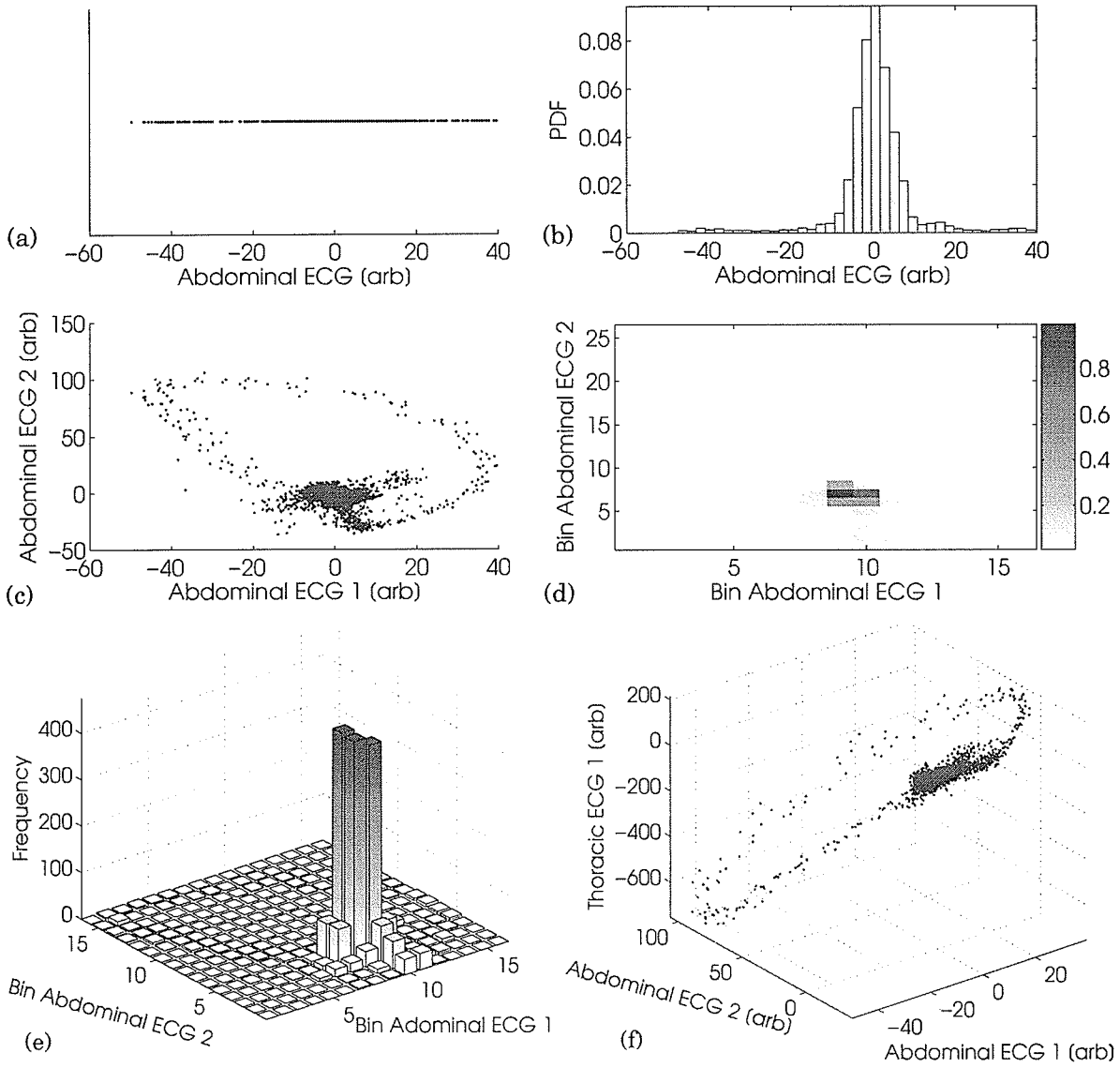


Fig. 3.2 Various visualizations of time-integrated BSP distributions of the de Lathauwer *et al.* data set [51] : (a) 1-D scatterplot (channel 2); (b) 1-D histogram (channel 2); (c) 2-D scatterplot (channels 2 and 3); (d) 2-D intensityplot (channels 2 and 3); (e) 2-D histogram (channels 2 and 3); (f) 3-D scatterplot (channels 2, 3 and 7).

3.3.1 Second-Order Moments

The second-order statistics capture the basic idea of a distribution's location and spread. Now, statistical moments are tensor representations of distributions. Most familiar to the reader will be the *central moments* of a one-dimensional distribution, given by

$$\mu_1 = \mathcal{E}\left\{\underline{u}\right\} \quad ; \quad \mu_n = \mathcal{E}\left\{(\underline{u} - \mu_1)^n\right\}, \quad \forall n > 1 \tag{3.5}$$

where $\mathcal{E}\{\cdot\}$ is the expectation operator. Then μ_1 and μ_2 are the mean and variance of the distribution. Also, being that they are achieved by polynomials of degree less than or equal to two, these statistics are called *second-order moments*. These two numbers roughly capture the main ideas of a distribution: (i) a (central) location, and (ii) a spread of values around the centre. (In effect, a second-order characterization is akin to representing all distributions by the interval $[\mu_1 - \sqrt{\mu_2}, \mu_1 + \sqrt{\mu_2}]$.) Alternatively, *raw moments* about the origin can be calculated from simple polynomial functions, $\mathcal{E}\left\{\underline{u}^n\right\}$. These, however, make for a more difficult interpretation, except for the raw second-order moment (which represents the energy of the signal).

Moments of multidimensional random vectors become more complicated, being not scalars, but *tensors* of order n equal to the order of the polynomial. It suffices, therefore, that the second-order moments of random vectors can be described by vectors (tensors of order 1) and matrices (tensors of order 2). For the random vector $\underline{\boldsymbol{u}} = [\underline{u}_1, \ldots, \underline{u}_{N_c}]^{\mathrm{T}}$, these are the mean

vector $\boldsymbol{\mu}$, and covariance matrix $\boldsymbol{\Sigma}$, defined as

$$\boldsymbol{\mu} = \mathcal{E}\left\{\underline{\boldsymbol{u}}\right\} = \int_{\mathbb{R}^{N_c}} \boldsymbol{u} f(\boldsymbol{u})\, d\boldsymbol{u}, \tag{3.6}$$

$$\boldsymbol{\Sigma} = \mathcal{E}\left\{(\underline{\boldsymbol{u}} - \boldsymbol{\mu})(\underline{\boldsymbol{u}} - \boldsymbol{\mu})^{\mathrm{T}}\right\} \tag{3.7}$$

$$= \int_{\mathbb{R}^{N_c}} \begin{pmatrix} (u_1 - \mathbf{e}_1^{\mathrm{T}}\boldsymbol{\mu})^2 & \dots & (u_1 - \mathbf{e}_1^{\mathrm{T}}\boldsymbol{\mu})(u_{N_c} - \mathbf{e}_{N_c}^{\mathrm{T}}\boldsymbol{\mu}) \\ (u_2 - \mathbf{e}_2^{\mathrm{T}}\boldsymbol{\mu})(u_1 - \mathbf{e}_1^{\mathrm{T}}\boldsymbol{\mu}) & \dots & (u_2 - \mathbf{e}_2^{\mathrm{T}}\boldsymbol{\mu})(u_{N_c} - \mathbf{e}_{N_c}^{\mathrm{T}}\boldsymbol{\mu}) \\ \vdots & \ddots & \\ (u_{N_c} - \mathbf{e}_{N_c}^{\mathrm{T}}\boldsymbol{\mu})(u_1 - \mathbf{e}_1^{\mathrm{T}}\boldsymbol{\mu}) & & (u_{N_c} - \mathbf{e}_{N_c}^{\mathrm{T}}\boldsymbol{\mu})^2 \end{pmatrix} f(\boldsymbol{u})\, d\boldsymbol{u}.$$

From the above, it is clear that $\boldsymbol{\Sigma}$ is symmetric (*i.e.*, $\boldsymbol{\Sigma}^{\mathrm{T}} = \boldsymbol{\Sigma}$) and positive definite [40].

In a scatterplot, the nonzero mean vector of a unimodal distribution can usually be identified as an offset from the origin, as is shown in Fig. 3.3(a) for the case of a 2-dimensional Gaussian. The diagonal elements of the covariance matrix give an idea of the spread of the distribution in space. In particular, the larger the element is, the farther the spread along that axis. For off-diagonal elements, the sign denotes a linear correlation in the spread of the distribution along those axes, as shown in Fig. 3.3(b).

It will be very common in the remainder of this thesis to assume that the distributions are zero-mean. This is true without any loss in generality, since any distributed vector $\underline{\boldsymbol{u}}$ can be *centralized* to

$$\underline{\overline{\boldsymbol{u}}} = \underline{\boldsymbol{u}} - \mathcal{E}\left\{\underline{\boldsymbol{u}}\right\} \tag{3.8}$$

which has zero mean. This thesis will continue to use the notation $\underline{\overline{\boldsymbol{u}}}$ for centralized variables.

Acquired BSPs are very likely to have non-zero mean. It would appear appropriate for the baseline as defined in Ch. 2 to be the mean. However, it is not. The baseline is the most frequent value attained in the BSP (the mode), and therefore strongly influences the mean, but the large values that occur during the waves also influence the result, as shown in Fig. 3.4.

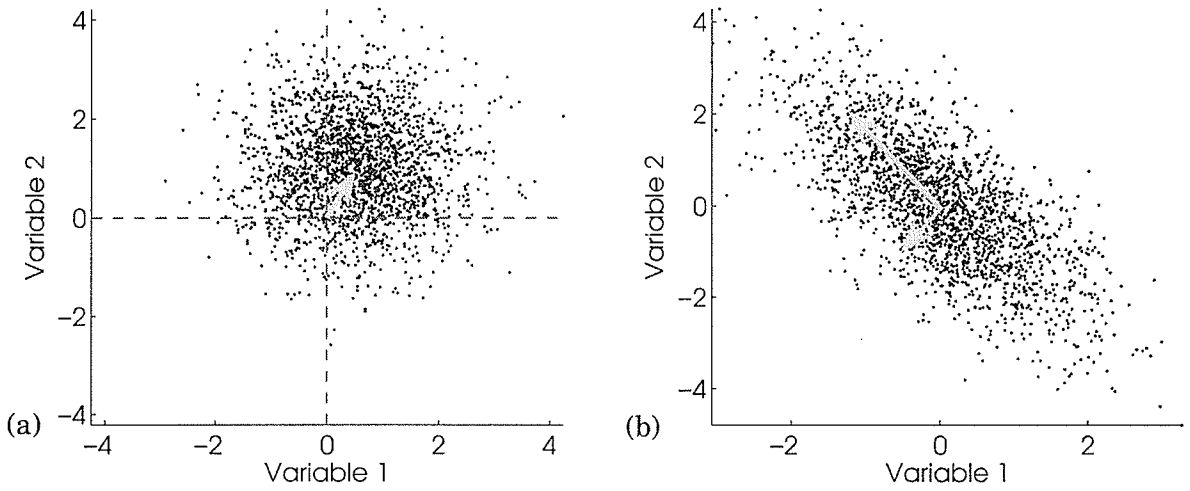


Fig. 3.3 Second-order moments: (a) the mean of a 2-D Gaussian distribution ($\mu = [0.5, 1]$); and (b) the spread of a 2-D Gaussian with $\Sigma = [1, -1; -1, 2]$.

As such an appropriately centralized BSP will often have a non-zero baseline. It is important that, in the processing of the BSPs into fetal and maternal ECG estimates, the mean of the joint system be removed before processing. It is a simple process to reintroduce after the ICA processing is complete. Without loss of generality, this thesis will use centralized data in the performance analysis of ICA.

3.3.2 Higher-Order Moments

3.3.2.1 One-Dimensional Case

The higher-order moments (μ_n, $n > 2$) characterize quantitatively the properties of $\underline{\boldsymbol{u}} \sim f$ other than location and spread: in effect, the "shape" of f. The central moments are known, however, to be inefficient scalar characterizers of even the simplest distributions. For example, it can be shown (as done in Stuart and Ord [218, Sec. 3.4 (1994)]) that the Gaussian distribution $\mathrm{g}(\mu_1, \mu_2)$, has diverging even moments, *i.e.*, $\mu_{2k} \to \infty$ as k increases. (Since the Gaussian is symmetric about zero, all odd moments vanish, $\mu_{2k-1} = 0$.) Given the special distinction

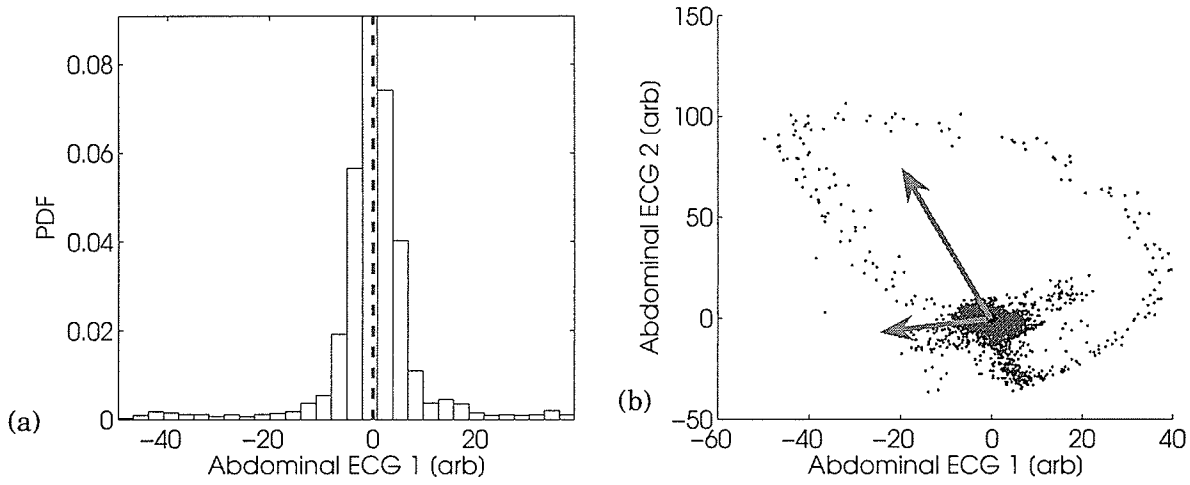


Fig. 3.4 Means of the time-integrated BSP distributions of the de Lathauwer *et al.* data set [51] : (a) 1-D histogram (channel 2) has estimated mean $\mu = 0.026$; (b) 2-D scatterplot (channels 2 and 3) has estimated mean $\boldsymbol{\mu} = [0.026, -0.199]$ and covariance $\boldsymbol{\Sigma} = [88.8, -68.6; -68.6, 329.4]$.

of the Gaussian distribution in statistical theory and engineering applications, therefore, preference is given to other (more complicated) moment parameters that represent the Gaussian compactly. A set of parameters called the *cumulants* of the distribution, κ_i, in effect, are the standardized higher-order moments [218]. These are rather cumbersomely defined as the ith Taylor coefficient of $\log \mathcal{F}(f)$, where $\mathcal{F}(f)$ is the Fourier transform of $f(\boldsymbol{u})$ (called the *characteristic function* of the distribution). Using these parameters the Gaussian distribution is compactly represented with only two non-zero moments: $\kappa_1 = \mu_1$ and $\kappa_2 = \mu_2 = \sigma^2$, which are the usual mean and variance.

Thus, following the tradition of ICA theory, *higher-order moments* implicitly assumes a cumulant system, though for any finite moment order, HOSs can be expressed combinatorially and calculated in terms of the central moments. (Note, therefore, that as the moment order increases, the cumulant expression in terms of central moments becomes more complicated.) One similar property of cumulants and central moments is that all odd moments are zero if the

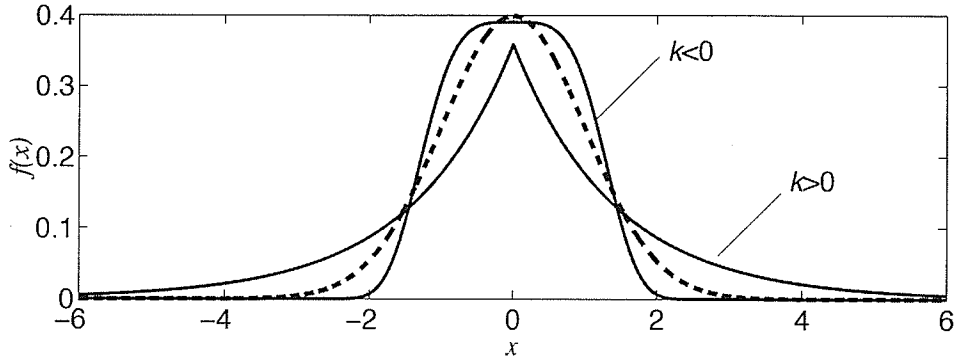


Fig. 3.5 Kurtosis: three distributions of equivalent variance. The Gaussian pdf is overlayed by a super-Gaussian ($k > 0$) and sub-Gaussian ($k < 0$) pdf.

distribution f is symmetric, [218] *i.e.*,

$$f(x) = f(-x)\ \forall x \quad \Rightarrow \quad \mu_n = \kappa_n = 0,\ \forall n \text{ odd} \tag{3.9}$$

Therefore, the first *symmetric measure* of non-Gaussianity by HOSs is κ_4. Termed the *unnormalized kurtosis*, $k' = \kappa_4$, it is expressed in central moments by the formula [218]

$$k' = \mu_4 - 3(\mu_2)^2 \tag{3.10}$$

The kurtosis is related to the sharpness in a unimodal distribution, as shown in the densityplots of Fig. 3.5. Distributions with positive kurtosis tend to be sharply peaked unimodal curves with heavy tails, and are called super-Gaussian. Negative kurtosis means a flatter (or often multimodal) distribution with thin tails. Distributions of this type are called sub-Gaussian. The kurtosis can also be expressed in a unitless *normalized form*,

$$k(\underline{u}) = \frac{\mathcal{E}\left\{\underline{\overline{u}}^4\right\}}{\mathcal{E}\left\{\underline{\overline{u}}^2\right\}^2} - 3 \tag{3.11}$$

3.3.2.2 Multidimensional Case

Multivariate extensions to these HOS definitions and principals are essential for ICA. These are quite standard and straightforward, guarding, as noted before, that higher-order

moments require tensor analysis in the multidimensional case. Of equivalent distinction to the kurtosis as the first symmetric measure of non-Gaussianity, the centralized random vector $\underline{\overline{u}} = [\underline{\overline{u}}_1, \ldots, \underline{\overline{u}}_{N_c}]^{\mathrm{T}}$ now has a 4-th order (cross) cumulant tensor [218]

$$\kappa_{wxyz}(\underline{\overline{u}}) = \mathcal{E}\left\{\underline{\overline{u}_w}\,\underline{\overline{u}_x}\,\underline{\overline{u}_y}\,\underline{\overline{u}_z}\right\} - \sum_{wxyz} \mathcal{E}\left\{\underline{\overline{u}_w}\,\underline{\overline{u}_x}\right\}\mathcal{E}\left\{\underline{\overline{u}_y}\,\underline{\overline{u}_z}\right\} \tag{3.12}$$

where the sum is over all distinct combinations. For $\underline{\boldsymbol{u}} \in N_c$, this makes an $N_c \times N_c \times N_c \times N_c$ array. From this, one can observe that the unnormalized kurtosis k' (3.10) is the special case of the fourth-order *auto-cumulant*, *i.e.*, the diagonal tensor elements where $w = x = y = z$. As such, the N_c diagonal elements of the 4-th order cumulant tensor (3.12) are the

$$\kappa_{iiii}(\underline{\overline{u}}) = \mathcal{E}\left\{\underline{\overline{u}_i}^4\right\} - 3\mathcal{E}\left\{\underline{\overline{u}_i}^2\right\}^2 = \mu_4(\underline{\overline{u}_i}) - 3\left(\mu_2(\underline{\overline{u}_i})\right)^2 \tag{3.13}$$

$$= \kappa_4(\underline{\overline{u}_i}) \tag{3.14}$$

$$= k'(\underline{\overline{u}_i}) \tag{3.15}$$

Therefore, the diagonal elements of this tensor contain information on the symmetric non-Gaussianity of the marginal distributions. The non-diagonal elements, by analogy, have some measure of joint (or coupled) non-Gaussianity, which is difficult to interpret, and nearly impossible to visualize. The diagonal elements become emphasized in ICA and these are more open to conceptualization. For this purpose, the definition of (3.11) is extended to a *normalized kurtosis vector* for the multidimensional case in an element-wise way,

$$k(\underline{\overline{u}}) = k\left(\sum_{i=1}^{N_c} \mathbf{e}_i \underline{\overline{u}}_i\right) \stackrel{\text{def}}{=} \sum_{i=1}^{N_c} \mathbf{e}_i k(\underline{\overline{u}}_i) \tag{3.16}$$

$$= \sum_{i=1}^{N_c} \mathbf{e}_i \left(\frac{\mathcal{E}\left\{\underline{\overline{u}_i}^4\right\}}{\mathcal{E}\left\{\underline{\overline{u}_i}^2\right\}^2} - 3\right) \tag{3.17}$$

where $\underline{\overline{\boldsymbol{u}}} = \underline{\boldsymbol{u}} - \mathcal{E}\left\{\underline{\boldsymbol{u}}\right\}$ is the centralized random variable.

The ECG signal in general, and all the acquired BSPs of de Lathauwer *et al.* (2000) [51] have non-Gaussian time-integrated distributions, as shown in Fig. 3.6.

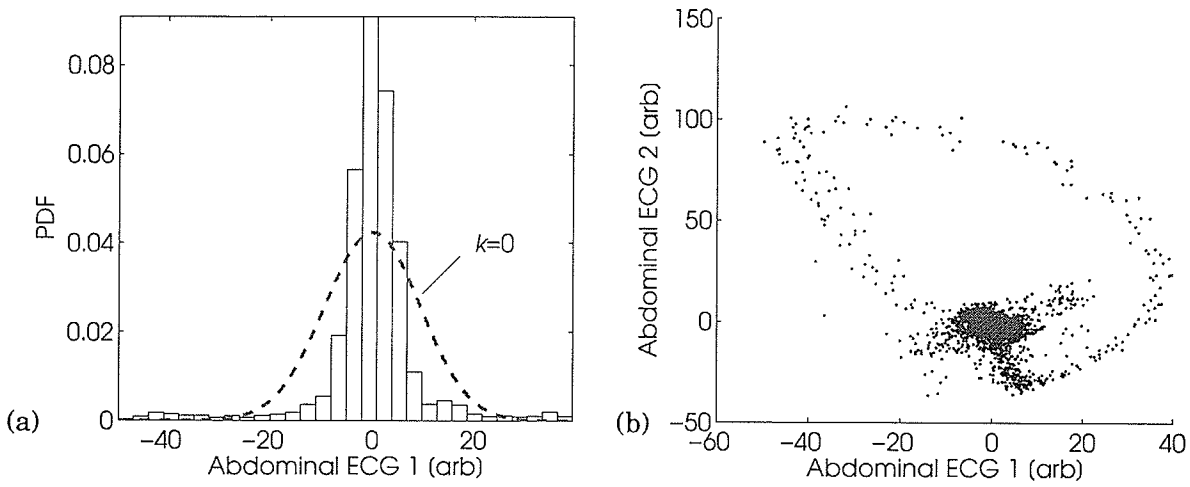


Fig. 3.6 HOS of the time-integrated BSP distributions of the de Lathauwer *et al.* data set [51]: (a) 1-D histogram (channel 2) has estimated kurtosis $k = 8.3$; (b) 2-D scatterplot (channels 2 and 3) has estimated kurtosis vector $k = [8.3, 13.8]$.

3.4 Statistical Independence

ICA is concerned about transforming a distribution of $\underline{\boldsymbol{x}} \in \mathcal{R}(\mathbb{R}^{N_c})$ into a linearly related independent distribution, its *independent components*, $\widehat{\underline{\boldsymbol{s}}} \in \mathcal{R}(\mathbb{R}^{N_c})$. The statistical background required for a mathematical treatment of *independence* is presented here.

3.4.1 Definition

The following recalls some well-known properties [165].

Definition 3.7 (Marginal PDF). Let a vector $\underline{\boldsymbol{u}} = [\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_{N_c}]^{\mathrm{T}}$ have a distribution f over $\mathbb{R}^{N_c}$. Then the *ith marginal* of $\underline{\boldsymbol{u}}$ is the distribution obtained by integrating out all other random variables $j \neq i$. That is,

$$f_i(\underline{u}_i) = \int_{\mathbb{R}^{N_c-1}} f(\underline{\boldsymbol{u}}) \prod_{j \neq i} du_j \tag{3.18}$$

Definition 3.8 (Statistical Independence)**.** Let a vector $\underline{\boldsymbol{u}} = [\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_{N_c}]^{\mathrm{T}}$ have a distribution f over $\mathbb{R}^{N_c}$. Then $\underline{\boldsymbol{u}}$ has *independent components* if the distribution *factorizes*; *i.e.*,

$$f(\underline{\boldsymbol{u}}) = \prod_{i=1}^{N_c} f_i(\underline{u}_i) \tag{3.19}$$

where $f_i(\underline{u}_i)$ is the ith marginal pdf.

3.4.2 Visualizing Independence

The statistical independence of a 2-dimensional distribution is visualized by its Cartesian symmetry. Figure 3.7 shows the scatterplot visualizations of some Gaussian and non-Gaussian independent distributions as well as their marginals. Edges, where they can be perceived, are mutually orthogonal and aligned with the coordinate axes.

3.4.3 Independence in Moments

It follows from the independence of $\underline{\boldsymbol{u}}$ that the expectation operator with respect to $\underline{\boldsymbol{u}}$ becomes separable on the off-diagonal; *i.e.*, $\mathcal{E}\left\{\underline{\overline{u}}_i \underline{\overline{u}}_j\right\} = \mathcal{E}\left\{\underline{\overline{u}_i}\right\} \mathcal{E}\left\{\underline{\overline{u}_j}\right\} \; \forall \, i \neq j$. Since the independent pdf $f(\underline{\boldsymbol{u}})$ can also be described by tensor statistical moments, these inherit diagonizability,

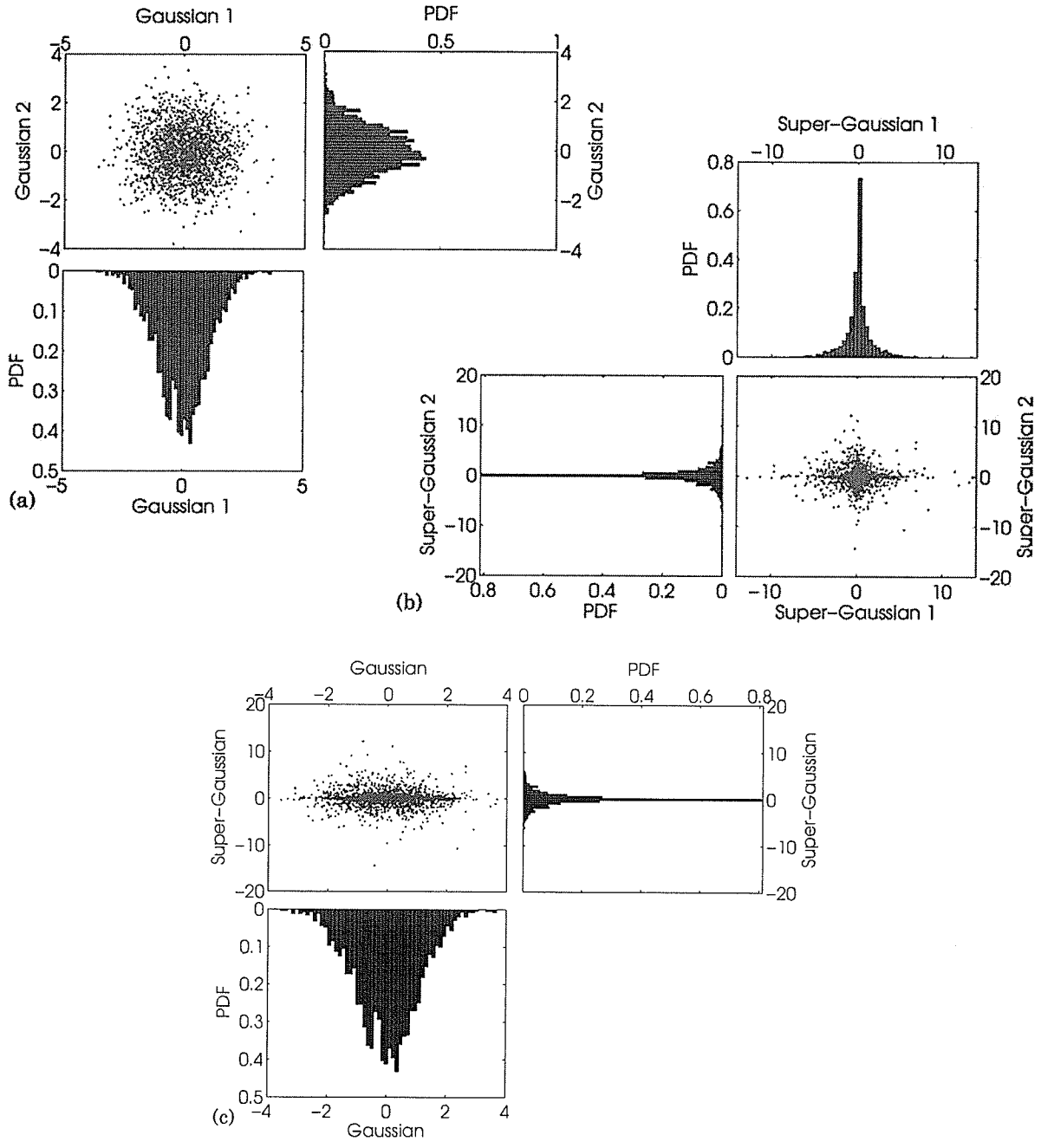


Fig. 3.7 Independent distributions and their marginals: (a) Gaussian; (b) Super-Gaussian; (c) Gaussian and Super-Gaussian

as shown below:

$$\boldsymbol{\Sigma}(\underline{\overline{u}}) = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & \sigma_{N_c}^2 \end{pmatrix} = \delta_1(i,j)\, \sigma_i^2 \mathbf{e}_j \mathbf{e}_i^{\mathrm{T}} \tag{3.20}$$

$$\kappa_{wxyz}(\underline{\overline{u}}) = \mathcal{E}\left\{\underline{\overline{u}}_w \underline{\overline{u}}_x \underline{\overline{u}}_y \underline{\overline{u}}_z\right\} - \sum_{wxyz} \mathcal{E}\left\{\underline{\overline{u}}_w \underline{\overline{u}}_x\right\} \mathcal{E}\left\{\underline{\overline{u}}_y \underline{\overline{u}}_z\right\} \tag{3.21}$$

$$= \delta_1(w,x)\delta_1(w,y)\delta_1(w,z)\, \sigma_w^2 k(\underline{\overline{u}}_w) \tag{3.22}$$

$$= \begin{cases} \sigma_i^2 k(\underline{\overline{u}}_i) & : \; w = x = y = z = i \;\; \forall i \in \{1, \dots, N_c\} \\ 0 & : \; \text{otherwise} \end{cases} \tag{3.23}$$

where δ_1 is the Kronecker delta function.

In fact, using the general form of the cumulant derived from the Taylor expansion and Fourier transform, the factorization of f implies the factorization of its Fourier transform, which further implies its logarithm is separable as sums, *i.e.*,

$$\log \mathcal{F}(f) = \log \mathcal{F}(\prod_i f_i) = \log \prod_i \mathcal{F}(f_i) = \sum_i \log \mathcal{F}(f_i) \tag{3.24}$$

From (3.24), *any* order of mixed derivative, and thus the mixed Taylor coefficients, will be zero. That is, *any off-diagonal* element of *any cumulant tensor* from an independent distribution will vanish. Thus the cumulant characterization of an *independent* distribution can be achieved without tensors at all. Only the infinitely many vectors consisting of all the auto-cumulants are required. The kurtosis vector of (3.16) is, therefore, important as the *first symmetric higher-order characterization* of *independent* distributions.

Similar to our discussion of random signals in Ch. 2, the diagonalization at second-order and higher-orders are often distinguished. Of special historical consideration is *correlation*, or

second-order dependence. If a random variable (distribution) $\underline{\boldsymbol{u}}$ satisfies (3.20) without being independent, then it is *decorrelated*. In that case, all the relationship between its marginal components occur at higher order. It is possible, therefore, that 3.22 is not satisfied even though (3.20) holds.

3.4.4 Cumulant Additivity of Mixed Independent Sources

It was observed that the off-diagonal cumulants of a zero-mean factorable (independent) multivariate distribution are zero. What properties do cumulants have with respect to the linear mixing of independent sources? As shown in the following proposition, cumulants are additive (or "accumulative") elements, from which they derive their name [218].

Proposition 3.9 (Cumulant Algebra of Independent Sources). *Consider the statistically independent one-dimensional variables $\underline{s}_i$, for $i = 1, \ldots, N_c$. Then the cumulants for the one-dimensional random variable $\underline{x} = \sum_i c_i \underline{s}_i$, for any $c_i \in \mathbb{R}$, are additive.*

Proof. The pdf of $\underline{x}$ is defined from the joint pdf of the $\underline{s}_i$ as

$$f_{\underline{x}}(x) = \int_{\{s \mid x = c^{\mathsf{T}} s\}} f(s_1, \ldots, s_{N_c})\, d\boldsymbol{s} \tag{3.25}$$

$$= \int_{\{s \mid x = c^{\mathsf{T}} s\}} \prod_{i=1}^{N_c} f_i(s_i)\, d\boldsymbol{s} \tag{3.26}$$

Thus the cumulants of the distribution $f_{\underline{x}}$ are defined from the derivatives of

$$\log \mathcal{F} f_{\underline{x}}(\omega) = \log \int_{-\infty}^{\infty} e^{-j\omega x} f_{\underline{x}}\, dx \tag{3.27}$$

$$= \log \int_{-\infty}^{\infty} e^{-j\omega c^{\mathrm{T}} s} \int_{\{s|\ x=c^{\mathrm{T}}s\}} \prod_{i=1}^{N_c} f_i(s_i)\, ds dx \tag{3.28}$$

$$= \log \int_{\mathbb{R}^{N_c}} \prod_{i=1}^{N_c} e^{-j\omega c_i s_i} f_i(s_i)\, ds \tag{3.29}$$

$$= \sum_{i=1}^{N_c} \log \int_{\mathbb{R}} e^{-j\omega c_i s_i} f_i(s_i)\, ds_i \tag{3.30}$$

$$= \sum_{i=1}^{N_c} \log \mathcal{F} f_{\underline{s}_i}(\omega c_i) \tag{3.31}$$

From the above, one observes that all derivatives are additive with respect to the independent component sources. In fact, the general cumulant form for the mixture is

$$\kappa_n(\underline{x}) = \frac{\partial^n}{\partial^n \omega} \log \mathcal{F} f_{\underline{x}}(\omega)\Big|_{\omega=0} = \sum_{i=1}^{N_c} \frac{\partial^n}{\partial^n \omega} \log \mathcal{F} f_{\underline{s}_i}(\omega c_i)\Big|_{\omega=0} \tag{3.32}$$

$$= \sum_{i=1}^{N_c} c_i^n \frac{\partial^n}{\partial^n \omega'} \log \mathcal{F} f_{\underline{s}_i}(\omega')\Big|_{\omega'=0} \tag{3.33}$$

$$= \sum_{i=1}^{N_c} c_i^n \kappa_n(\underline{s}_i) \tag{3.34}$$

so that the nth-order cumulants are homogeneous of degree n on its components. This is a useful property not held by the central moments (except $\mu_1 = \kappa_1$ and $\mu_2 = \kappa_2$). Thus, in particular, it holds for the unnormalized kurtosis that

$$k'(x) = \sum_{i=1}^{N_c} c_i^4 k'(s_i) \tag{3.35}$$

□

3.4.5 Entropy

Entropy is a scalar measure of variability, or uncertainty, of a multidimensional distribution. It is another feature of f calculated from an expectation, just like statistical moments.

Unlike moments, however, it does not have a growing tensor structure with higher dimensionality. It is presented as a formal definition, but one important preliminary is required first:

Definition 3.10 (Support). Any distribution f over $\mathbb{R}^{N_c}$ has a *support*, Ω_1, defined as

$$\Omega_1 = \left\{ \boldsymbol{u} \in \mathbb{R}^{N_c} \mid f(\boldsymbol{u}) > 0 \right\} \subset \mathbb{R}^{N_c} \tag{3.36}$$

Definition 3.11 (Differential Shannon Entropy [47, Cover and Thomas (2001)]). Let $\underline{\boldsymbol{u}}$ be an N_c-dimensional random variable with continuous support $\Omega_1 \subset \mathbb{R}^{N_c}$ and distribution $f_{\underline{u}}$. Then the *differential Shannon entropy* H is

$$H(\underline{\boldsymbol{u}}) = \mathcal{E}\left\{ \log \frac{1}{f(\underline{\boldsymbol{u}})} \right\} = -\int_{\Omega_1} f_{\underline{u}}(u) \log f_{\underline{u}}(u)\, d\boldsymbol{u} \tag{3.37}$$

The units of entropy are determined by the base of the logarithm. Using $\log_2$, entropy is measured in *bits*.

Most well-known analytic distributions can be formulated as a maximum-entropy distribution of a specific support [165]. Entropy also has the following well-known properties [47].

Proposition 3.12 (Entropy). *Let $\underline{\boldsymbol{u}}$ be an N_c-dimensional random variable with continuous support $\Omega_1 \subset \mathbb{R}^{N_c}$ and distribution $f_{\underline{u}}$. Then the entropy $H(\underline{\boldsymbol{u}})$ satisifes the following*

(a) $H(\underline{\boldsymbol{u}}) \in [-\infty, \infty]$;

(b) $H(\underline{\boldsymbol{u}} + \boldsymbol{a}) = H(\underline{\boldsymbol{u}}) \quad \forall \boldsymbol{a} \in \mathbb{R}^{N_c}$;

(c) $H(a\underline{\boldsymbol{u}}) = H(\underline{\boldsymbol{u}}) + \log|a| \quad \forall a \in \mathbb{R}$;

(d) $H(\underline{\boldsymbol{u}}) \leq \sum_{i=1}^{N_c} H(\underline{u}_i)$ *and equal iff* $\underline{\boldsymbol{u}}$ *is independent.*

Note that (b) implies that entropy is invariant to centralization, $H(\overline{\underline{\boldsymbol{u}}}) = H(\underline{\boldsymbol{u}})$.

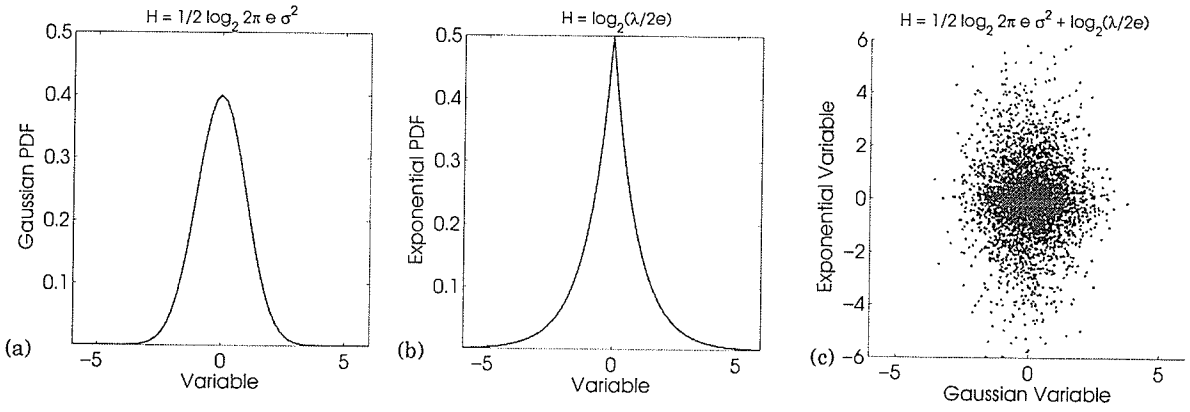


Fig. 3.8 Entropy: (a) Gaussian; (b) Exponential; (c) Exponential and Gaussian

Figure 3.8 shows the entropy of several analytic distributions and their independent joint-distribution. Calculating entropy from sampled data is much more complex than the analogous problem for moment or cumulant estimation, since the nonlinearity $\log f$ must be approximated. Details on implementation issues for entropy estimation will be described in Ch. 6.

Entropy is a HOS since it is sensitive to the non-Gaussian properties of a distribution. In fact, the following bound holds for continuous distributions f with support over $\mathbb{R}$:

Proposition 3.13 (Gaussian Entropy). *Let $\underline{u} \sim f$ be a zero-mean distribution with covariance matrix $\boldsymbol{\Sigma}$. Then*

$$H(\underline{u}) \leq H(\underline{\mathfrak{g}_u}) = \frac{1}{2}\log(2\pi e \det \boldsymbol{\Sigma}) \tag{3.38}$$

where $\underline{\mathfrak{g}_u}$ is the zero-mean Gaussian distribution of covariance $\boldsymbol{\Sigma}$.

Entropy is a very different HOS than the cumulants. First, entropy has a more complicated nonlinearity (*i.e.*, $\log f(u)$) than cumulant moments (which are essentially polynomials in u). Second, entropy is always a scalar, invariant with the dimensionality of the distribution, as shown in Fig. 3.9. Last, the entropy of a mixture $\underline{x} = \underline{s_1} + \underline{s_2}$ is not directly expressible by a simple algebra of the component entropies $H(\underline{s_i})$.

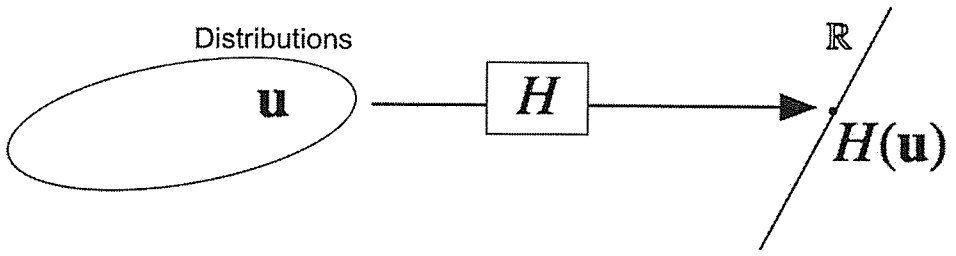


Fig. 3.9 Entropy maps all distributions to $\mathbb{R}$.

3.4.6 Mutual Information

Relative characterizations of distributions can also be defined from entropy. In particular, these are the *mutual information* (MI) and Kullback-Leibler divergence [47].

Definition 3.14 (Kullback-Leibler divergence)**.** Let $\underline{u}$ and $\underline{v}$ be two N_c-dimensional random variables with continuous supports $\Omega_1(\underline{u}), \Omega_1(\underline{v}) \subset \mathbb{R}^{N_c}$ and distributions $f_{\underline{u}}$, $f_{\underline{v}}$. Then the *Kullback-Leibler divergence* is calculated

$$\Re_{\mathrm{KL}}\left(\underline{u} \middle\| \underline{v}\right) = \mathcal{E}_{\underline{u}}\left\{\log \frac{f_{\underline{u}}}{f_{\underline{v}}}\right\} = \int_{\Omega_1(\underline{u})} f_{\underline{u}}(\boldsymbol{x}) \log \frac{f_{\underline{u}}(\boldsymbol{x})}{f_{\underline{v}}(\boldsymbol{x})}\, d\boldsymbol{x} \tag{3.39}$$

This scalar measure of two distributions is positive, is finite iff $\Omega_1(\underline{u}) \subset \Omega_1(\underline{v})$, vanishes iff $f_{\underline{u}} = f_{\underline{v}}$, and serves to define a scalar error between two distributions, as shown in Fig. 3.10. In particular, this scalar error can be used as a measure of dependence as follows.

Definition 3.15 (Mutual Information)**.** Let $\underline{u}$ be an N_c-dimensional random variable with continuous support $\Omega_1 \subset \mathbb{R}^{N_c}$ and distribution f and marginal pdfs f_i. Then the *mutual information* $\mathcal{M}$ is

$$\mathcal{M}(\underline{u}) = \Re_{\mathrm{KL}}\left(\underline{u} \middle\| \underline{u}' \sim \prod_{i=1}^{N_c} f_i\right) = \int_{\Omega_1} f(\boldsymbol{u}) \log \frac{f(\boldsymbol{u})}{\prod_{i=1}^{N_c} f_i(u_i)}\, d\boldsymbol{u} \tag{3.40}$$

This is a scalar measure of statistical independence, satisfying the very important property below straight from the definition.

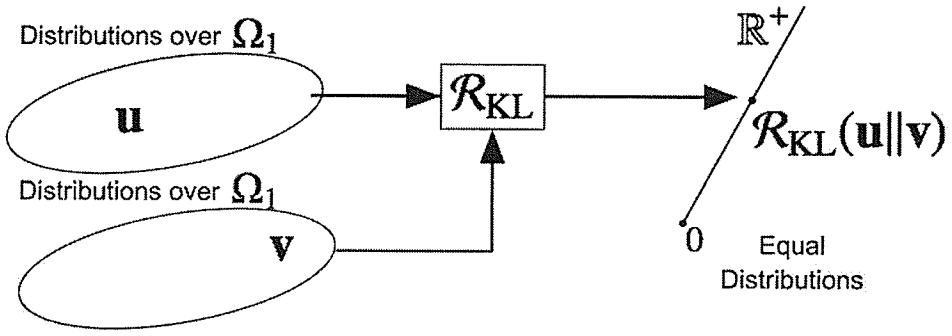


Fig. 3.10 The Kullback-Leibler divergence maps distribution pairs to $\mathbb{R}^+$.

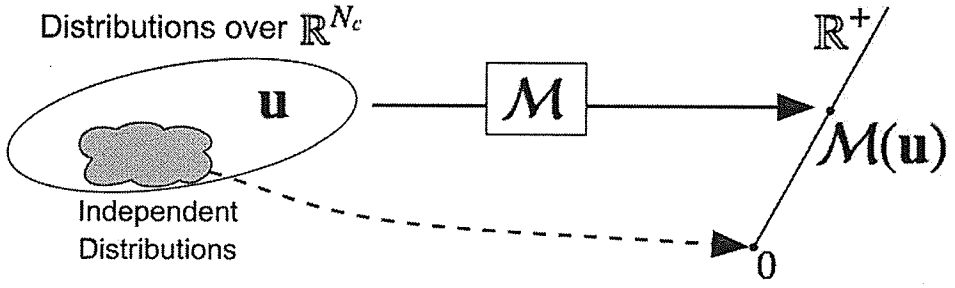


Fig. 3.11 The mutual information maps all higher-dimensional distributions to $\mathbb{R}^+$. Independent distributions are mapped to zero.

Theorem 3.16 (MI and Independence). *Let $\underline{u}$ be an N_c-dimensional random variable with continuous support $\Omega_1 \subset \mathbb{R}^{N_c}$ and distribution f and marginal pdfs f_i. Then $\underline{u}$ is independent iff $\mathcal{M}(\underline{u}) = 0$.*

The independence theorem above is a critical result in ICA theory, and is shown in Fig. 3.11.

Taking the Gaussian as the standard again, the Kullback-Leibler is again used to define a scalar measure called the negentropy.

Definition 3.17 (Negentropy). Let $\underline{u}$ be an N_c-dimensional random variable with continuous support $\Omega_1 \subset \mathbb{R}^{N_c}$ and distribution f with covariance Σ. Also let $\underline{g}_u$ be the Gaussian random

variable with the same mean and covariance. Then the *negentropy* of $\underline{u}$ is defined

$$H_{\mathrm{g}}(\underline{u}) = \Re_{\mathrm{KL}}\left(\underline{u} \,\middle\|\, \underline{\mathrm{g}}_u\right) \tag{3.41}$$

Negentropy will become an important measure of non-Gaussianity in the ICA paradigm. It explicitly quantifies the higher-order properties of entropy.

3.4.7 Rényi Generalized Entropies

Rényi's extension to Shannon's definition of entropy will be considered at this point. These alternative scalar measures of a distribution provide more quantitative statistical features for analysis. They are, as such, mostly unrelated to the theory of ICA, but will be used as higher-order ECG features in the analysis of ICA performance.

Definition 3.18 (Rényi Generalized Entropies [118, Kinsner (1994)])**.** Let $\underline{u}$ be an N_c-dimensional random variable with continuous support $\Omega_1 \subset \mathbb{R}^{N_c}$ and distribution f. Let q be an arbitrary real number. Then the *Rényi generalized entropy* of order q is

$$H_q(\underline{u}) = \frac{1}{1-q} \log \int_{\Omega_1} f_{\underline{u}}^q(\boldsymbol{u})\, d\boldsymbol{u} \tag{3.42}$$

$$= \frac{q}{1-q} \log \left\| f_{\underline{u}} \right\|_q \tag{3.43}$$

with the relationship to the classical L_q-norm as shown.

This is a true generalization of the Shannon differential entropy since $\lim_{q\to 1} H_q(\underline{u}) = H(\underline{u})$ [117, Kinsner (1994)]. What no longer holds in the generalization is the relationship between the entropy and the expectation operator, (3.37). Instead, $H_q(\underline{u}) = (1-q)^{-1} \log \mathcal{E}\left\{f^{q-1}\right\}$ in the generalized case.

The effect of the order q, apart from creating an ordered family of entropies which can be visualized as a bounded function $H_q : \mathbb{R} \mapsto \mathbb{R}^+$, Fig. 3.12, is to change the sensitivity of the

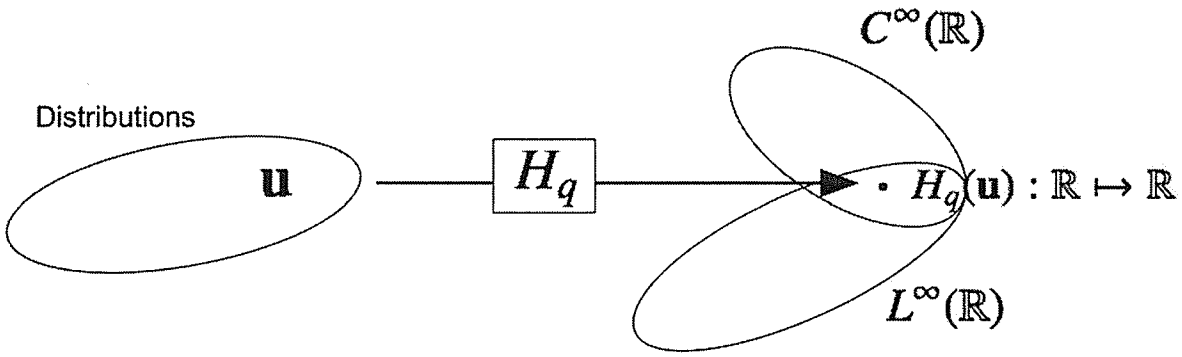


Fig. 3.12 Rényi entropy maps all distributions to $L^\infty(\mathbb{R}) \cap C^\infty(\mathbb{R})$.

entropy from small probabilities ($q << -1$) to large probabilities ($q >> 1$). In fact, infinite limits at $q = \pm\infty$ are well-defined as is shown in the following proposition.

Proposition 3.19 (Extrema of Rényi Entropy [117, Kinsner (1994)]). *The greatest lower bound of H_q is attained at the limit as $q \to -\infty$, given by $-\log \inf_{\Omega_1} f_{\underline{u}}$. $H_{-\infty}$ is therefore finite only for distributions that are bounded away from zero on their support. Similarly the least upper bound of H_q is attained at the Chebyshev limit, $q \to \infty$, and is given by $-\log \sup_{\Omega_1} f_{\underline{u}}$.*

Last, a generalized version of the Kullback-Leibler divergence is presented [48, Dansereau and Kinsner (2001)]. This can be used as a relative measure of statistical features.

Definition 3.20 (Generalized Kullback-Leibler Divergence). Let $\underline{u}$ and $\underline{v}$ be two N_c-dimensional random variables with continuous supports $\Omega_1(\underline{u}), \Omega_1(\underline{v}) \subset \mathbb{R}^{N_c}$ and distributions $f_{\underline{u}}$, $f_{\underline{v}}$. Let q be an arbitrary real number. Then the *generalized Kullback-Leibler divergence* of order q is calculated

$$\Re_{\mathrm{KL}q}\left(\underline{u} \middle\| \underline{v}\right) = \frac{1}{1-q} \log \int_{\Omega_1(\underline{u})} f_{\underline{u}}(x) \left(\frac{f_{\underline{u}}(x)}{f_{\underline{v}}(x)}\right)^{q-1} dx \,. \tag{3.44}$$

This is a true extension of the Kullback-Leibler divergence, for using Hôpital's rule to

evaluate (3.44) in the limit as $q \to 1$

$$\lim_{q\to 1} \Re_{\mathrm{KL}q}\left(\underline{u}\,\middle\|\,\underline{v}\right) = \lim_{q\to 1} \frac{\log \int_{\Omega_1(\underline{u})} f_{\underline{u}}(x)\left(\frac{f_{\underline{u}}(x)}{f_{\underline{v}}(x)}\right)^{q-1} dx}{-(q-1)} \tag{3.45}$$

$$= \lim_{q\to 1} \frac{\frac{\partial}{\partial(q-1)} \log \int_{\Omega_1(\underline{u})} f_{\underline{u}}(x)\,(e)^{(q-1)\log\frac{f_{\underline{u}}(x)}{f_{\underline{v}}(x)}} dx}{-1} \tag{3.46}$$

$$= \lim_{q\to 1} -\frac{\int_{\Omega_1(\underline{u})} f_{\underline{u}}(x)\frac{\partial}{\partial(q-1)}\,(e)^{(q-1)\log\frac{f_{\underline{u}}(x)}{f_{\underline{v}}(x)}} dx}{\int_{\Omega_1(\underline{u})} f_{\underline{u}}(x)\left(\frac{f_{\underline{u}}(x)}{f_{\underline{v}}(x)}\right)^{q-1} dx} \tag{3.47}$$

$$= -\frac{\int_{\Omega_1(\underline{u})} f_{\underline{u}}(x)\lim_{q\to 1}\,(e)^{(q-1)\log\frac{f_{\underline{u}}(x)}{f_{\underline{v}}(x)}} \log\frac{f_{\underline{u}}(x)}{f_{\underline{v}}(x)} dx}{\int_{\Omega_1(\underline{u})} f_{\underline{u}}(x)dx} \tag{3.48}$$

$$= -\int_{\Omega_1(\underline{u})} f_{\underline{u}}(x)\log\frac{f_{\underline{u}}(x)}{f_{\underline{v}}(x)}dx = \Re_{\mathrm{KL}}\left(\underline{u}\,\middle\|\,\underline{v}\right) \tag{3.49}$$

3.5 Independent Component Analysis

Now the theory of the signal separation method intended for analysis, ICA, will be presented. ICA has been a developing field for twenty years as a result of neural network experiments and research into unsupervised separation algorithms. ICA is a complicated topic because it involves (i) higher-order statistics (such as cumulant tensors) as features, and (ii) an optimization over matrices. The cumulants and HOS that are central to ICA have already been presented in this chapter, and so more focus can be made on the algebra and geometry implicit to the theory. In an attempt to give due emphasis to the significance and complexity of ICA theory and its dynamic evolution, some history of the topic's emergence is briefly presented first. Thereafter, the theory of ICA will be presented as an extension to principal component analysis (PCA). Though it has been argued that this is not the best pedagogy for ICA definition, [92, 1999], it is the most appropriate *signal processing* perspective. Also, given the

subject here, of analyzing SQMs for ICA above and beyond traditional SNR, the difference between the second-order approach of PCA and its higher-order relative, ICA, is of paramount importance.

3.5.1 ICA: A Historical Overview of the Literature

3.5.1.1 Prehistory

The development of ICA has been preceded by multiple signal separation techniques in array processing and noise cancellation. These include

(a) Blind deconvolution/Bussgang algorithms;

(b) Factor analysis;

(c) Projection pursuit; and

(d) Artificial neural networks (ANN).

The successful application of various principles in the ANN topic of unsupervised learning rules was the most prolific influence on the early development of ICA. These included information theoretic principles such as Barlow's principle of redundancy reduction to maximize neuron independence [17, (1961)], and the information maximization principle ("infomax") developed by Linsker [127, (1992)] and Nadal and Parga [155, (1994)]. The most significant contribution to ICA pre-history, however, came from Oja's modified Hebbian single-layer ANN learning rule for deflationary PCA, [161, 1982][162, 1989].

3.5.1.2 Discovery and Early Expansion

Conventional ICA history begins with French experiments by Jutten and Herault [86, (1986)][105, (1991)] on extensions to Oja's PCA neural network. These experiments demonstrated the ability for an ANN to isolate independent components from mixtures, but with

inconsistent results. The ICA terminology was first applied here, and with subsequent collaboration with Pierre Comon, the theoretical underpinnings of ICA were investigated. The 1994 seminal paper by Comon [45], clarified the general theory of independent component analysis as an extension to PCA (and also provides a good summary of the connections and independent contributions by many in field to that point). There, the identifiability criterion for noiseless ICA was formalized, which, in fact, represents (i) the necessary assumptions needed for ICA to apply, and (ii) the necessary and sufficient conditions for ICA cost functions. Comon also demonstrated that, by approximating the mutual information, acceptable cost functions can be defined which required much less complexity than the MI itself.

ICA research continued progressively in the French signal processing community through this period. Pham and Jutten collaborated on a maximum likelihood approach to ICA [168]. Cardoso analyzed and developed algebraic forms of ICA using cumulant eigenmatrices (joint approximate diagonalization of eigenmatrices, JADE) [38][37] as well as providing an efficient stochastic gradient algorithm, called the relative gradient (or EASI algorithm) [32].

American contributions began with Bell's investigations of Linsker's infomax principle in sigmoid ANNs under T. Sejnowski [21, (1995)]. Avoiding the complications required by cumulant analysis, this work recast the ICA problem into an information-theoretic framework, and consequently increased the popularity of ICA research in general. It also presented a stochastic gradient learning rule for the implementation of infomax. Two weaknesses of this method were (i) the requirement of super-Gaussian sources, and (ii) a matrix inversion in the algorithm. These were addressed later by Lee, Girolami, and Sejnowski in [125, (1999)] when they generalized the infomax learning rule to include sub-Gaussian signals, and included Amari's "natural gradient" to eliminate the matrix inversion.

Amari's work [7, (1996)][8][5] analyzed the non-Euclidean geometry of matrix space induced by the multiplication operation, and consequently characterized an optimal structure for matrix gradient updates. (This is discussed further in the later section on the geometry of matrix space.) When applied to ICA adaptive learning, the natural gradient update not only improves the general convergence of the gradient descent, but also conveniently eliminates the matrix inversions in infomax and related algorithms. Moreover, Amari also introduced one of the appropriate non-blind performance metrics for ICA, the Amari performance index [7, (1996)]. This matrix-based measure of ICA performance is a common tool in ICA performance analysis. The Riken research group of Amari and Cichocki have been a consistent influence on ICA [40].

Another significant contribution to ICA's early development came from Oja's school itself. There the one-unit PCA neuron was adapted to HOS and the collection of algorithms known as FastICA was developed [94, Hyvarinen and Oja (1997)] based on a principle of maximal non-Gaussianity. Since these algorithms were based on fixed-point optimization, and not gradient descent, the speed and simplicity of these algorithms distinguished themselves from the dominant methods in the literature. Through the turn of the century, and since, the Hyvarinen-Oja school has been both prolific and pragmatic at expanding and applying ICA.

3.5.1.3 Reconciliation

In the late 1990s, once the Finnish, Japanese, American, and French schools of ICA were established, the ICA literature expressed a movement of reconciliation, whereby the different techniques and algorithms were identified as equivalent under nominal conditions. Thus the varied principles for algorithm derivation, such as maximum-likelihood, information maximization, redundancy reduction, nonlinear PCA, and maximal non-Gaussianity all can be reduced to one another in noise-free situations if the pdfs of the blind sources are unimodal

[135, (1999)], smooth, and other wise well-behaved [156, (1997)][34, (1997)][160, (1998)][6, (1998)][92, (1999)]. Cardoso's relative gradient was also proven to be equivalent to Amari's natural gradient [36, (1998)][35, (1998)].

It is on this basis of equivalence that "toy problems" of mixing and ICA separation with nicely conditioned source distributions are limited in effect. Instead, the algorithms differ in how they react to real-world situations, specifically when the real conditions are far from the implicit assumptions or tunings of each algorithm. In one sense, therefore, it is the robustness *in practice* that distinguishes the original ICA algorithms, not their derivations in *theory*.

3.5.1.4 Applications

Applications of ICA to real-world data have been widespread, and so the full breadth of the literature cannot be reviewed here. As a non-exhaustive list with only representative papers or reviews to assist the interested reader, the effect of ICA has been observed in

(a) beamforming and array processing;

(b) speech and sound processing (*e.g.*, the "Cocktail Party problem") [79];

(c) EEG processing (*e.g.*, the "neural cocktail party") [101][229] [100];

(d) fMRI processing [29][144];

(e) image processing [111][126]; and

(f) fetal ECG processing.

As discussed in Ch. 2, the fetal ECG separation problem has been of interest to signal processing engineers for decades. Though EEG was the common application of ICA through the 1990's, it was the seminal work of De Lathauwer *et al.* [51, (2000)] who brought the fECG

issue to the ICA community in earnest. Table B.1 in App. B categorizes some interesting contributions in the application of ICA (or related signal processing methods) to fetal ECG. In particular, the literature summarized in Table B.1 is identified by the type of algorithms applied, the type of data used in experiments, and the type of performance metrics used in analysis. There, the type of experiments used in the fetal ECG literature are broken down into three subgroups:

(a) real fECG data, identified by *real noisy data* (RND);

(b) simulated but not noise-free data, identified by *simulated but noisy data* (SND) (*e.g.*, data is manipulated ECG recordings);

(c) simulated noise-free fECG data, identified by *simulated noise-free data* (SFD) (*e.g.*, data is generated from a mathematical ECG model); and

(d) simulated analytical data, identified by *simulated analytical data* (SAD) (*e.g.*, data is generated from a (non-ECG) mathematical model).

It is very common for papers to have a theoretical section using some sort of SFD or SAD experiment, as well as another section with an RND analysis. Moreover, these distinct sections within the references are typically only loosely connected. In contrast to this, the design for this work is to experiment somewhere between these "theoretical sections" and "practical sections". Here, performance is to be analyzed with metrics that are consistent with the theory of ICA *and* consistent with the practical conditions of blind source separation.

Also, considerable attention in Table B.1 is paid to the types of performance metrics, though more of this is discussed later in this chapter.

3.5.1.5 State of the Art

Since 2000, ICA has been maturing as a field in signal processing, and new alternative approaches to ICA have proliferated to achieve better robustness. Mihoko and Eguchi [145, (2002)][63] used an extension of the Kullback-Leibler divergence called beta-divergence to derive an ICA algorithm that can modify the contribution of each point to the weight-matrix update. In this way, their technique is more robust to outlying points. Principe's group at the U. of Florida introduced an ICA based on the Rényi generalized entropies defined previously in this chapter [55, (2002)]. This algorithm (MRMI) optimizes an efficient cost function of Rényi marginal entropies (estimated by Parzen windowing on the data) by a gradient descent on Givens rotations. This work is of theoretical interest for its unusual and efficient cost function, elegant derivations, and significant interpretations. It is also of practical interest since they claim the algorithm uses data more efficiently.

Several state of the art methods also go back to the fundamentals of ICA. The RADICAL algorithm by Miller and Fisher [123, (2003] is similar to the Principe marginal approach, but a modified Vasicek m-spacings entropy estimator is used for the cost function calculations of the marginals. It, too, uses data efficiently and is more robust to outliers. Similarly, Grassberger *et al.* [215, (2004)] approached ICA directly from an estimation of MI (MILCA), which is usually avoided in high-dimensional spaces because of the lack of good estimators. The use of MI, here, is enabled by a nearest-neighbour-based entropy estimate suitable for high-dimensional spaces [120, Grassberger *et al.* (2004)]. The difficulty with these methods, in general, is the cost function estimates are not smooth (*i.e.*, they are effectively noisy), and so the Givens rotations optimization of the algorithms needs some heuristic assistance: either via resampling averages (RADICAL) or Fourier filtering (MILCA). Note that Givens rotation parameterization is the common optimization space of JADE, MRMI, RADICAL, and MILCA

(though the optimization implementations are different). A fundamentally different, but seemingly straightforward, approach is also presented in [14, Jutten *et al.*, (2004)], whereby the differential of MI is presented in its stochastic form. The first order update to a mixing matrix can then be estimated from extremely crude non-parametric estimates of the joint pdf, which performs surprisingly well in comparison to similar algorithms, because the bias of the differential estimator is effectively independent of the coarseness of the pdf estimate. This approach is considered significant because it generalizes well to (non-instantaneous) convolutive mixtures.

Other methods have re-examined ICA and approached it from the proverbial "left field". Considering traditional ICA methods as an optimization of Kullback-Leibler nonlinear approximants, Bach and Jordan extended these approximants to include self-adapting *families* of nonlinear approximants using kernel spaces [15, (2002)]. This support vector machine approach to the problem, casts the ICA cost function in terms of the nonlinear correlation over a reproducing kernel space. Though the theoretical complications of this approach are many, in practical terms the algorithm involves numerous but rather efficient matrix computations, and is more robust to outliers and less confused by near-Gaussian data. Optimization of the kernel ICA objective function is done by conjugate gradient on the geodesics of orthogonal matrices. (Further discussion of this geometry appears later in this chapter.)

3.5.1.6 Books, Reviews, and Conferences

Several books have been published on the topic of ICA. Few, however, provide more insight than the original papers on which they are based. In particular, [124] by Lee is a one-sided account of the research pursued at the Salk Institute. Hyvarinen's book [93, (2001)] is nicely tutorial and more considerate to the contributions of other authors in the field. This book too, however, is just a collection and extension of previous work: namely Hyvarinen's

survey on ICA [92, 1999]. Alone, the survey paper is recommended as an introduction to the theory of ICA and its benefits and limitations, though the reprints in the book are helpful as well. Another recommended (and more recent) review is [194, Sanchez (2002)]. A very rigorous and algorithmic approach to the topic of blind source separation can be found in [40, Cichocki and Amari (2002)], however the reading is very difficult. Its appendices, however, make an excellent resource for matrix calculus identities that are beneficial in the derivation of BSS learning rules. Among books, the latest contribution by James Stone [216, (2004)] is by far the most approachable and affordable beginning for one's own personal collection.

For a broader review of ICA, paper collections such as the special issue of the *Proc. of the IEEE* on "Blind Identification and Estimation" (1998, v. 86 - invited papers by Amari [6] and Cardoso [35]), *Unsupervised Adaptive Filtering* [78, Haykin, ed. (2000)], and *Intelligent Signal Processing* [80, Haykin and Kosko, eds. (2001)] present superbly the many authors and ideas of significance. Other collections of interest include [65, Girolami, ed. (2000)] and [187, Roberts and Everson (2001)].

A regular *International Conference on Independent Component Analysis and Signal Separation* was begun in 1999, and has consistently grown. (The latest installment, number 7, was held in 2007.) Contributions in the field have also been growing in number and status at more established annual conferences such as the *International Conference on Acoustics, Speech, and Signal Processing* and the *International Conference on Neural Networks*, among others.

Now that a brief history of ICA is complete, the theory of ICA will be developed. This begins with the second-order equivalent to ICA, principal component analysis.

3.5.2 ICA Foundations: Principal Component Analysis

Principal component analysis (PCA) is a common preprocessing technique for signal processing [25] that is a special case of the ICA method. PCA is also referred to as "sphering" (or "spatial whitening") in the literature [93]. Recall that the covariance matrix of $\underline{x} = [\underline{x_1}, \underline{x_2}, \ldots, \underline{x_{N_c}}]^{\mathrm{T}}$ is non-diagonal if correlation exists between the observed signals $\underline{x_i}$. That is to say that the second-order crosswise moments satisfy

$$\mathcal{E}\left\{\overline{\underline{x_i}\underline{x_j}}\right\} \neq 0 \quad \text{for some } i \neq j \tag{3.50}$$

The signals are therefore dependent at the second order. However, these signals can be decorrelated in a straightforward manner. This common knowledge is presented here as a theorem in preparation for the higher-order analog to follow in the next subsection.

Theorem 3.21 (Sphering PCA Transformation). *Let the stochastic signal $\underline{x} \in \mathcal{R}(\mathbb{R}^{N_c})$ have finite covariance matrix $\boldsymbol{\Sigma}_{\underline{x}} = \mathcal{E}\left\{\overline{\underline{x}\underline{x}}^{\mathrm{T}}\right\} \neq \mathbf{I}_{N_c}$. Then there exists a transform matrix $\mathbf{W_0}$ unique up to a unitary transformation that decorrelates and spatially spheres $\overline{\underline{x}}$ (by rescaling). That is, $\mathcal{E}\left\{\overline{\mathbf{W_0}\underline{x}}(\overline{\mathbf{W_0}\underline{x}})^{\mathrm{T}}\right\} = \mathbf{I}_{N_c}$, where $\mathbf{I}_{N_c}$ is the N_c-dimensional identity matrix.*

Proof. The covariance matrix $\boldsymbol{\Sigma}_{\underline{x}}$ is positive definite, and therefore has an eigenvalue decomposition [40]

$$\boldsymbol{\Sigma}_{\underline{x}} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{-1} \tag{3.51}$$

where $\mathbf{Q}$ has orthonormal columns and $\boldsymbol{\Lambda}$ is positive diagonal. It follows, therefore, that the transformation matrix $\mathbf{W_0} = \boldsymbol{\Lambda}^{-1/2}\mathbf{Q}^{-1}$ can be applied to decorrelate and sphere $\underline{x}$. Last, since any unitary transformation (*e.g.*, rotation) leaves the covariance unchanged, any unitary transformation to $\mathbf{W_0}$ is also an acceptable transformation. □

Decorrelation is equivalent to second-order independence. For not only do the second-order moments factor for all distributions, as $\mathcal{E}\left\{\underline{\overline{x_i x_j}}\right\} = \mathcal{E}\left\{\underline{\overline{x_i}}\right\}\mathcal{E}\left\{\underline{\overline{x_j}}\right\} = 0$, but moreover, if higher-order moments are zero, then the pdf itself factors. Since it was observed that the Gaussian distribution is characterized by a mean vector, covariance matrix, and zero higher-order moments, then it is also true that the Gaussian distribution factors if and only if the covariance is diagonal, satisfying $\mathcal{M}(\underline{x}) = 0$.

The "white Gaussian" hypothesis of traditional signal processing enables and enforces the equivalence between decorrelation and independence. This has the benefit of turning the expectation operator $\mathcal{E}\{\cdot\}$ into an inner product, and thus a metric, on the second-order distributions, $\mathcal{R}_2(\mathbb{R}^{N_c})$. Orthogonal basis expansions of $\mathcal{R}_2(\mathbb{R}^{N_c})$ (like the Karhunen-Loeve basis) are possible because of the similarity between L^2 (functional) orthogonality and decorrelation orthogonality. The existence of higher-order elements, however, disrupts this coexistence and there is a much richer behaviour[1]. Thus, while sufficient for the Gaussian universe $\mathcal{R}_2(\mathbb{R}^{N_c})$, and undeniably useful, decorrelation and PCA fails as a method to factorize BSP distributions.

As mentioned, in the application of PCA on non-Gaussian distributions, the equivalence of decorrelation and (full) independence is lost, and, in particular, the invariance to rotation of independence is lost. In Fig. 3.13(a) one observes the distribution from independent (and decorrelated) non-Gaussian signals. However, by applying a rotation, Fig. 3.13(b), the system remains uncorrelated but the pdf is clearly not factorable and therefore dependent. Visually, the edges and symmetries of the joint-distribution make it very clear what rotation is required to return the system to independence. The challenge of ICA is in transferring this visual

[1] In an analogy to complex variables, this is similar to the introduction of complex elements to the real field: *e.g.*, the vectors $[1, -1]^T$ and $[1, 1]^T$ are *real* orthogonal, but introduce complex elements, such as $[1 + i, -1]^T$ and $[1 + i, 1]^T$, and orthogonality needs to be redefined.

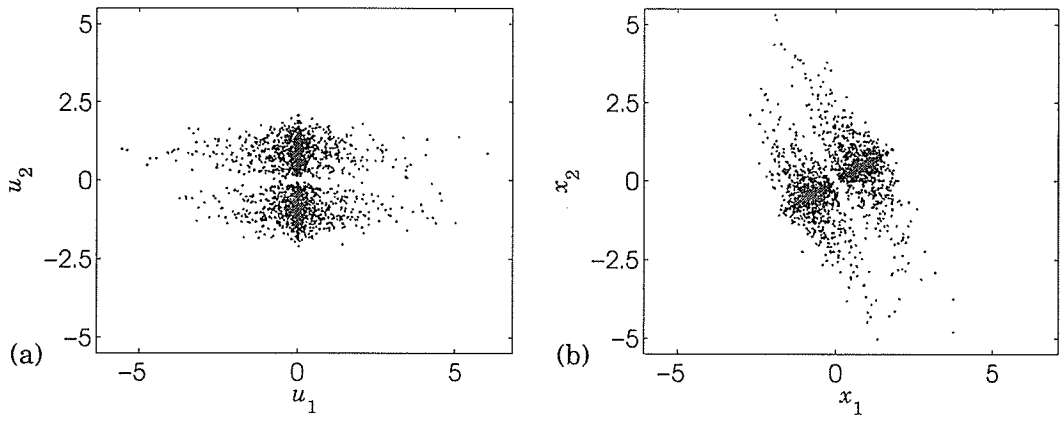


Fig. 3.13 (a) Independent non-Gaussian distribution, and (b) rotated into higher-order dependence.

property to something mathematically expressible, computable, and applicable in higher dimensions (where visualization fails completely). That is achieved in what follows.

3.5.3 General ICA Theory: Comon Theorem

The previous discussion of PCA and second-order dependence has made clear the importance of higher-order statistics to achieve independence. However, as was demonstrated in Theorem 3.21, the PCA transformation was a result of matrix theory being applied to the covariance matrix of the signals. The discussion of higher-order moments should make it clear that cumulant *tensors* must now be analyzed in order to find an equivalent ICA theorem. Comon completed this analysis [45, Comon (1994)], and provided a unified discussion on the many preceding experiments in BSS. He used the ICA acronym coined by Jutten and Herault [105, 1991]. The ICA theorem will be rephrased here without proof and its implications will be discussed.

Theorem 3.22 (ICA Transformation). *Let the stochastic signal* $\underline{x} \in \mathcal{R}(\mathbb{R}^{N_c})$ *be a full-rank mixture of independent sources; that is* $\underline{x} = \mathbf{M}\underline{s}$ *for some full-rank* $\mathbf{M}$ *and* $\underline{s} \in \mathcal{R}(\mathbb{R}^{N_c})$*, an independent stochastic signal. Let also that* $\underline{s}$ *has at most one Gaussian component and no*

point-like mass, and $\underline{\boldsymbol{x}}$ *has a finite covariance matrix* $\boldsymbol{\Sigma}_{\underline{\boldsymbol{x}}}$. *Then for any cost function* $\mathcal{I}$ *satisfying for arbitrary* $\underline{\boldsymbol{u}}$

$$(i)\ \mathcal{I}(f_{\mathbf{P}\underline{u}}) = \mathcal{I}(f_{\underline{u}}) \text{ for all permutations } \mathbf{P}; \tag{3.52}$$

$$(ii)\ \mathcal{I}(f_{\mathbf{D}\underline{u}}) = \mathcal{I}(f_{\underline{u}}) \text{ for all invertible diagonal } \mathbf{D} \tag{3.53}$$

and independent $\underline{\boldsymbol{u}}$

$$(iii)\ \mathcal{I}(f_{\mathbf{T}\underline{u}}) \geq \mathcal{I}(f_{\underline{u}}) \text{ for all invertible matrices } \mathbf{T} \tag{3.54}$$

$$(iv)\ \mathcal{I}(f_{\mathbf{T}\underline{u}}) = \mathcal{I}(f_{\mathbf{T}_2\underline{u}}) \text{ iff } \mathbf{T} = \mathbf{DP} \tag{3.55}$$

one has $\mathcal{I}(f_{\mathbf{W}\underline{x}}) = \mathcal{I}(f_{\underline{s}})$ *iff* $\mathbf{WM} = \mathbf{DP}$ *for some invertible scaling matrix* $\mathbf{D}$ *and permutation* $\mathbf{P}$. *Therefore, by minimizing* $\mathcal{I}(f_{\mathbf{W}\underline{x}})$ *over all matrices* $\mathbf{W}$, *a representation of the sources can be achieved.*[2]

The difficulty with this result is that it provides necessary and sufficient conditions for only the *existence* of the ICA transformation in terms of the existence of the cost function. It does not *construct* such a transformation like the PCA theorem. (Note that Comon also did present an example algorithm using this theorem in [45, 1994] as well.) On the other hand, this theorem presents the crucial practical benefit of eliminating the requirement of cumulants and mutual information from the construction of an ICA transformation. It can be understood in the following way. The MI of the signals *is* a scalar measure of independence satisfying the "contrast conditions" (3.52), (3.53), (3.54), and (3.55). One can therefore imagine that a cost function for linear transformations can be defined

$$C_{\underline{x}} : \mathbf{T} \in \mathcal{L} \mapsto \mathbb{R} \quad \text{such that} \quad C_{\underline{x}}(\mathbf{T}) = \mathcal{M}(\mathbf{T}\underline{\boldsymbol{x}}) \tag{3.56}$$

[2] *This thesis uses a minimization formulation of the theorem, since cost minimization is preferred in signal processing circles. Comon [45, 2004] in fact maximized* contrast functions, *which are the negatives of cost functions (without any loss of generality).*

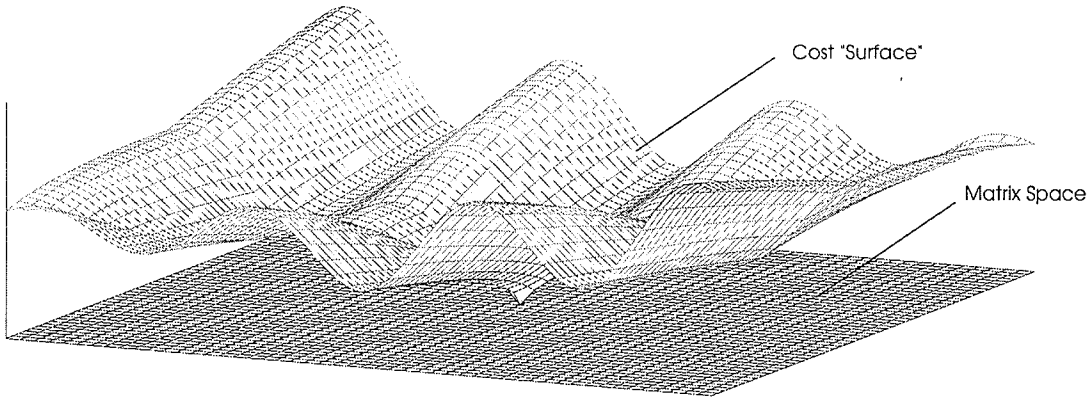


Fig. 3.14 A representation of the cost function "surface" over matrix-space.

and the ICA transformation is the solution $\mathbf{T}$ minimizing $C_{\underline{x}}$ to zero, Fig. 3.14. According to the degeneracy conditions, there will be multiple equivalent solutions (related by the degeneracy matrix group $\mathbf{P}$ and $\mathbf{D}$). Thus the problem breaks down to a nonlinear optimization over a matrix space. However, as is often done in nonlinear optimization problems, *substitute cost functions* can be used as long as the location of the minima are the same. By replacing the cost "surface" over the matrix space, one can perhaps hope to achieve faster convergence, better accuracy, or an easier implementation. The ICA theorem specifies the necessary and sufficient conditions to design a new ICA cost function $C_{\underline{x}}(\mathbf{T}) = \mathcal{I}(\mathbf{T}\underline{\boldsymbol{x}})$. What distinguishes different ICA algorithms is the cost functions used to derive the algorithm, or the tool used for nonlinear optimization over its "surface".

In spite of the implicit nature of the solution $\mathbf{W}\underline{\boldsymbol{x}}$, this design has the benefit of being non-parametric. The effect of ICA, overlooking the indirect means of solution, is an exact parallel to PCA. The statistics of the *data themselves* define a transformation $\mathbf{T}$ that sets a scalar measure of the data dependence to zero. Therefore, representative samples of the distribution to estimate the statistics are all that are required to estimate ICA — not an *a priori* model. The observation $\underline{\boldsymbol{x}}$ is all that is required to estimate ICA, and the sources can remain blind apart

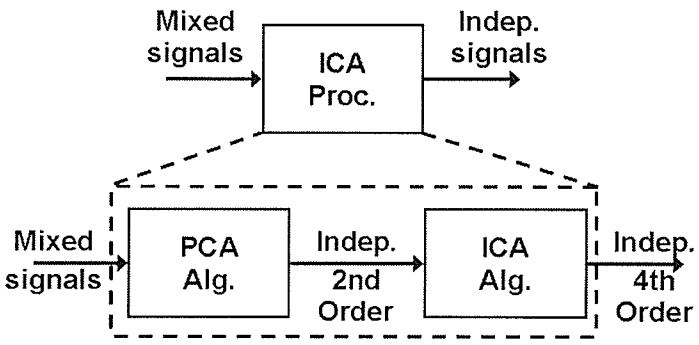


Fig. 3.15 The two-step implementation (second-order, higher-order) of an ICA process.

from the hypotheses of the theorem. With good data, good algorithms can separate sources.

The formulation of ICA has also explained the failure of PCA on many erratic signals having nonzero kurtosis. Removing the second-order correlation by rotating the space onto the orthogonal directions of maximum variance does not remove the mixing in the higher-order moments. Another linear transformation (*e.g.*, rotation) is required to isolate the independent components. Since higher-order statistics are the essence of the novelty of ICA, almost all ICA techniques focus on achieving an independence measurable in the simplest symmetric non-Gaussian higher-order moments: *i.e.*, of the fourth-order. Kurtosis becomes a very important feature in the design of cost functions. Most techniques also begin with the PCA sphering process and implement the ICA transformation in two steps: first, transforming to second-order independence; then, rotating to fourth-order independence, Fig. 3.15.

It is advantageous to use PCA *preprocessing* because the ICA demixing matrix $\mathbf{W}$ of sphered data can be constrained to be orthogonal. Also, since an ICA solution $\mathbf{W}$ only minimizes the independence cost function I (not reducing it to 0), there is no guarantee that the given solution will be decorrelated without explicit PCA.

3.5.4 ICA Theory: Optimization Geometry

3.5.4.1 Parameterization of $\mathbf{W}$

One of the significant advances of ICA in the 1990s (*i.e.*, the natural/relative gradient), and a common feature among the state of the art [55, (2002)][123, (2003)][215, (2004)][54, (2007)], is the explicit inclusion of the geometrical properties of ICA optimization. That is, the weight matrix $\mathbf{W}$ that is to be optimized is considered in the algebraic context of the space of linear transformations, $\mathbf{W} \in \mathcal{L}_{N_c}$, and not only in the context of its ${N_c}^2$ elements $W_{ij} = \mathbf{e}_i^{\mathrm{T}}\mathbf{W}\mathbf{e}_j$. Considering the $\mathbf{e}_i \in \mathbb{R}^{N_c}$ to be the canonical basis of $\mathbb{R}^{N_c}$ and the $\widetilde{\mathbf{e}}_i$ to be the canonical basis of $\mathbb{R}^{{N_c}^2}$, the canonical parameterization of $\mathcal{L}_{N_c}$ is by the embedding vector of simple matrix elements

$$\mathbf{W} \mapsto \sum_i \sum_j \left(\mathbf{e}_i^{\mathrm{T}}\mathbf{W}\mathbf{e}_j\right)\widetilde{\boldsymbol{e}}_{N_c(i-1)+j} \tag{3.57}$$

$$\mapsto \sum_i \sum_j W_{ij}\,\widetilde{\boldsymbol{e}}_{N_c(i-1)+j} \in \mathbb{R}^{{N_c}^2} \tag{3.58}$$

This is not, however, the only parameterization of $\mathbf{W}$ by ${N_c}^2$ independent values (*e.g.*, the matrix decomposition into symmetric and anti-symmetric parts, yielding $N_c(N_c+1)/2$ and $N_c(N_c-1)/2$ independent parameters respectively). Specifically, the additional algebraic (*e.g.*, multiplication) and pragmatic properties (*e.g.*, determinant and transpose) that exist in $\mathcal{L}$ must be expressed by awkward nonlinear relationships when using the $\mathbb{R}^{{N_c}^2}$ parameterization of (3.58). It is not surprising, therefore, that the simple Euclidean geometry of $\mathbb{R}^{{N_c}^2}$ alone does not characterize the full richness of $\mathcal{L}_{N_c}$.

3.5.4.2 Natural Gradient

The Amari natural gradient is a general principle of gradient optimization in Riemannian manifolds. For ICA, this applies because BSS optimization in $\mathcal{L}_{N_c}$ is restricted to the

submanifold of *nonsingular* matrices. Note that matrix addition is not a closed operator on this manifold (*e.g.*, $\mathbf{W} + (-\mathbf{W}) = 0$), while matrix multiplication necessarily is (*e.g.*, $\det(\mathbf{W}_1\mathbf{W}_2) = \det(\mathbf{W}_1)\det(\mathbf{W}_2) > 0$). Consequently, this manifold has a non-Euclidean geometry in $\mathcal{L}_{N_c}$, and even though the Euclidean gradient of a cost function

$$\nabla I(\mathbf{W}) = [\partial I/\partial W_{ij}]_{N_c \times N_c} \tag{3.59}$$

can be used to update the estimate of the weight matrix *additively*

$$\mathbf{W} \leftarrow \mathbf{W} + \nabla I(\mathbf{W}) \tag{3.60}$$

this does not take into account the natural "flow" of the Riemannian manifold of the solution space. (This is equivalent to the general observation that a local gradient of an arbitrary cost surface generally does not point directly away — *i.e.*, radially — from the surface minimum [25].) Instead, from arguments based on the Lie group invariance of the nonsingular matrices in $\mathcal{L}_{N_c}$, the natural step on the manifold (or natural gradient, $\widetilde{\nabla}$) involves a post-multiplication of the form [5][8]

$$\mathbf{W} \leftarrow \mathbf{W} + \widetilde{\nabla} I(\mathbf{W}) \tag{3.61}$$

$$\leftarrow \mathbf{W} + \nabla I(\mathbf{W})\mathbf{W}^{\mathrm{T}}\mathbf{W} \tag{3.62}$$

which actually forms the update *multiplicatively*

$$\mathbf{W} \leftarrow \left(\mathbf{I} + \nabla I(\mathbf{W})\mathbf{W}^{\mathrm{T}}\right)\mathbf{W} \tag{3.63}$$

and, consequently, the optimization becomes a more efficient process.

3.5.4.3 ICA Indeterminacies

The natural gradient takes into account the Riemannian-flow on the submanifold of $\mathcal{L}_{N_c}$ of the invertible matrices. Even more geometry exists in the BSS problem when the scaling indeterminacy and sphering are also considered.

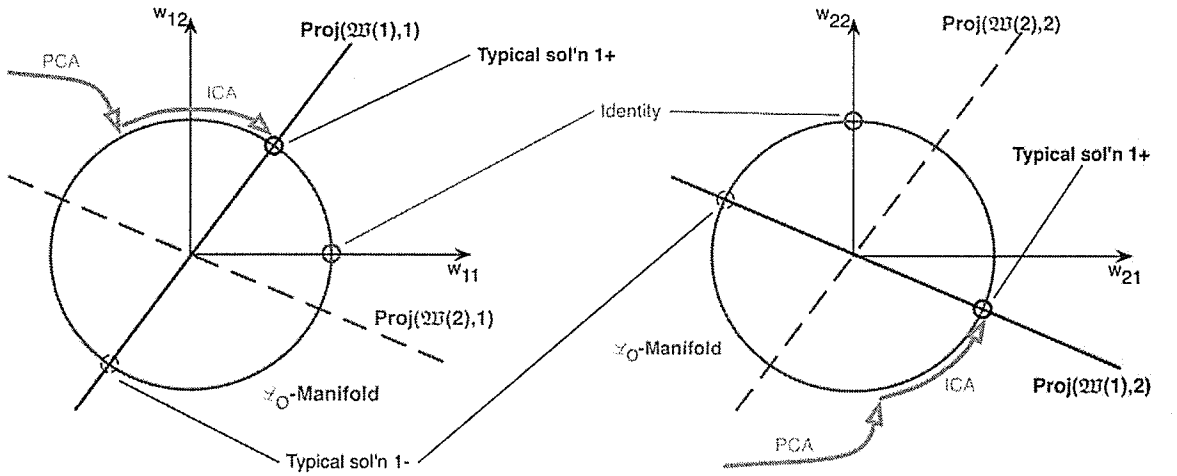


Fig. 3.16 Visualization of canonical ICA solution space for $\mathbf{W} = [W_{ij}] \in \mathcal{L}_2$: Each row of $\mathbf{W}$ appears as a 2D projection space (row 1 on the left, row 2 on the right), and $\mathbf{W}$ is their orthogonal product. The ICA scaling degeneracy makes an ICA solution $\mathfrak{W}_i$ jointly appear as lines through the origin (1D subspaces) in both left and right projections (solid). The ICA permutation degeneracy swaps these projective subspaces (dashed). Action of PCA (sphering) is to remove degeneracy by projecting the solution onto the $\mathcal{L}_O$ manifold. ICA higher-order optimization can then be considered as an optimization over $\mathcal{L}_O$ to the nearest intersection with the solution space $\mathfrak{W}_i$. Any Givens matrix optimization behaves equivalently.

Since an ICA solution $\mathbf{W}$ is invariant to scaling by a diagonal matrix, $\mathbf{W} \leftarrow \mathbf{DW}$, this means that under the canonical mapping of $\mathcal{L}_{N_c}$ into $\mathbb{R}^{N_c{}^2} = \bigotimes_{i=1}^{N_c} \mathbb{R}^{N_c}$, the ICA solution can be identified as a product of orthogonal subspaces. Let $\mathbf{e}_i$ be the canonical basis for $\mathbb{R}^{N_c}$, and let

$$\mathfrak{W}_i = \left\{ \boldsymbol{\xi} \in \mathbb{R}^{N_c} \,\middle|\, \boldsymbol{\xi} = d_i \left(\mathbf{We}_i\right)^{\mathrm{T}} \;\forall d_i \in \mathbb{R} \right\} \tag{3.64}$$

so that $\mathfrak{W}_i$ is identified as a line in $\mathbb{R}^{N_c}$ that is in fact the rowspace of the single row $\mathbf{We}_i$. It follows that an ICA solution $\mathfrak{W}(1)$ can be composed as the direct product of the lines

$$\mathfrak{W}(1) = \bigotimes_{i=1}^{N_c} \mathfrak{W}_i \tag{3.65}$$

This is visualized for the simplest $N_c = 2$ case in Fig. 3.16 as the two solid lines on the two

orthogonal projections (left/right) identified as row 1 and row 2 respectively.

Since ICA also has a permutation indeterminacy, the direct products of the $\mathfrak{W}_i$ in (3.65) can be arbitrarily reordered. In the simple case of Fig. 3.16, this is exhausted by $\mathfrak{W}(1) = \mathfrak{W}_1 \otimes \mathfrak{W}_2$ and $\mathfrak{W}(2) = \mathfrak{W}_2 \otimes \mathfrak{W}_1$ and the two possible forms are distinguished by the solid ($\mathfrak{W}(1)$) and dashed ($\mathfrak{W}(2)$) lines.

3.5.4.4 Orthogonal Constraints

As mentioned previously, ICA is often approached as two step-optimization, where second-order independence is attained first, and then is optimized to remove higher-order dependencies. When a sphering is applied (by $\mathbf{W}_0$, making the correlation matrix not only a diagonal matrix, but equal to the identity), this removes the scaling degeneracy from $\mathbf{W}$ and more geometrical interpretation can be considered.

Here, the ICA demixing matrix $\mathbf{W}$ now can be constrained to the set of *orthogonal* matrices, $\mathcal{L}_O$. This smaller search space appears as both a group under matrix multiplication and a submanifold of $\mathcal{L}_{N_c}$, and therefore can be interpreted geometrically. Here, $\mathcal{L}_O$ is a special case of a *Stiefel manifold* [59][171]. The Stiefel manifold in the simple $N_c = 2$ case appears as the two circles in Fig. 3.16, which are parameterized simultaneously by the single parameter θ (the right hand image is ahead in phase by 90 degrees, *c.f.*, $\mathbf{I}$ as marked). Note that, since the ICA solution $\mathfrak{W}$ is radial in nature, the Stiefel manifold is (tangentially) perpendicular to it everywhere. Theoretically, therefore, the sphering approach is effective since the natural gradient on $\mathcal{L}_O$ (3.63) is simply the Euclidean gradient ($\mathbf{W}^T\mathbf{W} = \mathbf{I}$). In practice, however, some difficulty is introduced by the curvilinear structure of $\mathcal{L}_O$. This nonlinearity means that even "perfect" updates along the Stiefel manifold's tangent introduce an error, lifting the matrix $\mathbf{W}$ from the manifold. True motion within the $\mathcal{L}_O$ manifold is by definition *geodesic motion*, which can be approached in two ways. First, the optimal gradient geodesic update can be

derived, but this requires, unfortunately, matrix logarithms and exponentials [171]. Second, a non-optimal geodesic update can be performed using *Givens matrices*. Givens matrices are real orthonormal skew-symmetric matrices of the type

$$\mathbf{G}_{N_c}(i,j,\theta) = \begin{pmatrix} 1 & & \cdots & & 0 \\ \vdots & \cos(\theta) & & \sin(\theta) & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & -\sin(\theta) & & \cos(\theta) & \vdots \\ 0 & & \cdots & & 1 \end{pmatrix} \tag{3.66}$$

which satisfies the identity

$$\mathbf{e}_m^{\mathrm{T}}\mathbf{G}_{N_c}(i,j,\theta)\mathbf{e}_n = \begin{cases} \cos(\theta) & : \ m = n \in \{i,j\} \\ \sin(\theta) & : \ m = i, n = j \\ -\sin(\theta) & : \ m = j, n = i \\ \delta_1(m,n) & : \ \text{otherwise} \end{cases} \tag{3.67}$$

where it is assumed by the notation $\mathbf{G}_{N_c}(i,j,\theta)$, that $1 \leq i \leq j \leq N_c$. The multiplicative action of $\mathbf{G}_{N_c}(i,j,\theta)$ (from the left) is to make a rotation of angle θ in the direct product of the i and j rowspace. Since these matrices act only on two rows and two columns at a time, their action is equivalent to that visualized in Fig. 3.16, where the weights subscripts on the axes generalize from $1,2$ to i,j. The angle θ of the Givens matrix identifies the change of angle applied to $\mathbf{W}$ as observed in Fig. 3.16. Here, the change of angle is applied to both circles simultaneously, but the right hand image remains ahead in phase by 90 degrees. In order to optimize a matrix in the Stiefel manifold by Givens rotations, the rotations must *sweep* through all i,j pairs and then repeat because the later pairwise optimizations disrupt the previously optimized pairs somewhat, just as in Jacobi diagonalization [180, Numerical

Recipes, Sec. 11.1]. This approach to maintaining the geometrical constraint of the Stiefel manifold is common to the MRMI, RADICAL, and MILCA algorithms.

Note that the Givens rotations cannot, in fact, search the entire Stiefel manifold, but only the connected manifold made by the *special orthogonal group*, *i.e.*, those unitary matrices with determinant equal +1, around the initial condition. A second portion of the manifold is disconnected from the initial condition and corresponds to the unitary matrices with determinant -1. An inversion is required to jump between the disconnected manifolds. In Fig. 3.16, this appears as changing the fixed phase relationship between the action of the two circles. As mentioned, the rotation matrices parameterize the right hand circle ahead by 90 degrees, while a non-rotation matrix will have it lag by 90 degrees. This inability to rotate through the entire Stiefel manifold is not an issue to ICA optimization, however: since the ICA solution is sign invariant, only the intersection of the Stiefel manifold with the $\mathfrak{W}$ subspaces are relevant, and such a subspace always intersects the special orthogonal group.

Practically speaking, therefore, an ICA algorithm attempts to search through matrix space to some element of $\mathfrak{W}$ under either a "hard" or "soft" normalization condition. Pre-sphering the system and then solving via Givens rotations is an example of a hard constraint because the algorithm updates attempt to remain within the Stiefel manifold. Another approach to a hard constraint is to use a "push-back" constraint on a gradient update. Here, a gradient update rule is not necessarily constrained to remain in the manifold, but the normalization constraint is forcefully applied to its result before the algorithm continues (*e.g.*, FastICA). This two-part step, however, can apply considerable computational effort in the constraint stage [171]. Other gradient descent updates can be classified as soft constraints because penalty terms in the optimization rule can stabilize the solution from drifting from the Stiefel manifold [54]. Even if tangential updates are used, this approach is recommended so that drift does not accumulate

over long trajectories. Again, it is not truly necessary to make this orthogonal matrix constraint (any member of $\mathfrak{W}$ is sufficient), but it is advantageous for ICA to enforce second-order independence, and it is rather pointless to have the algorithm move predominantly in the direction of the scaling matrix indeterminacies (*i.e.*, radially in the "Grassmann manifold" that is orthogonal to the Stiefel manifold).

3.5.4.5 Geometrical Norm

The last point of discussion for the geometry of matrix space is the definition of norm in $\mathcal{L}_{N_c}$, which is then used to define distance (closeness) between matrices. A standard norm that is practically useful is the Frobenius norm, which by definition is

$$\|\mathbf{W}\|_F^2 = \sum_{i,j} \left|W_{ij}\right|^2 \tag{3.68}$$

and therefore is equivalent to a Euclidean norm on the canonical parameterization vector (3.58). The Frobenius norm is also sub-multiplicative

$$\|\mathbf{W}_A \mathbf{W}_B\|_F \leq \|\mathbf{W}_A\|_F \|\mathbf{W}_B\|_F \tag{3.69}$$

and thus properly accounts for matrix multiplication from the perspective of operator theory [235]. On this basis, an ICA algorithm will terminate once the Frobenius norm of the update becomes smaller than a threshold. Note that under the Frobenius norm all unitary matrices in $\mathcal{L}_{N_c}$ have norm N_c. Thus the Stiefel manifold is at a constant radius from the zero matrix. Just as with Euclidean distance on a circle, however, Frobenius distance is *not* equivalent to geodesic length along the $\mathcal{L}_O$ manifold.

Now that the context of ICA algorithms has been described in theoretical and geometrical terms, some of the basic ICA update rules will be described.

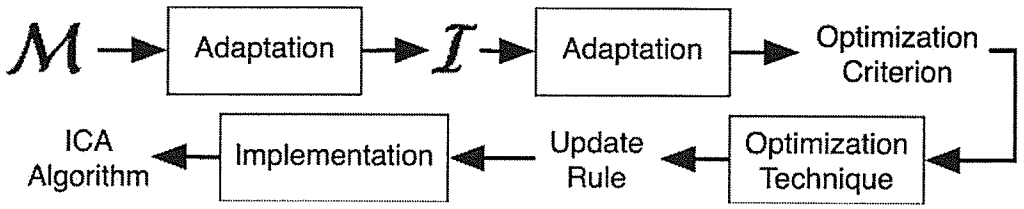


Fig. 3.17 Creating a new ICA technique.

3.5.5 ICA Theory: Basic Algorithms

As was described, ICA algorithms are all connected to the assumptions and general application of the ICA theorem, yet distinguished by the details of the cost functions and their optimization. In this section, some popular techniques are described, namely *Infomax* and *FastICA*, as well as their derivation from basic principles.

An ICA technique is typically derived from the ICA Theorem as shown by the design process of Fig. 3.17. The mutual information, $\mathcal{M}$, is adapted via approximation or theorem into another cost function $\mathcal{I}$ satisfying (or nearly satisfying) the conditions (3.52)–(3.55). The cost function is then optimized by an appropriate (nonlinear) method. ICA can be information-theoretic if the process of Fig. 3.17 is based on mutual information or other entropy considerations. It can also be directly HOS-based if the optimization criterion involves cumulants or kurtosis.

3.5.5.1 Xinfomax Technique

An implementation of ICA based on information theoretic principles is presented now. Introduced by Bell and Sejnowski in 1995 [21], *Infomax* popularized ICA among ANN researchers. It is designed around an implementation in a single-layer feed-forward ANN [25] with an unsupervised learning algorithm. Its relationship to the ICA theorem is identified next as a theorem.

Theorem 3.23 (Infomax Principle). *Let* $\underline{x} \in \mathcal{R}(\mathbb{R}^{N_c})$ *pass through an ANN with nonlinear squashing functions* η_i, $i = 1, \ldots, N_c$ *and square weight matrix* $\mathbf{W}$. *Then if* $\mathbf{W}$ *maximizes the mutual information between the input* $\underline{x}$ *and the output* $\underline{y}$ *where*

$$\underline{y_i} \stackrel{def}{=} \eta_i(\mathbf{e}_i^T \mathbf{W} \underline{x}) \tag{3.70}$$

then $\underline{y}$ *is independent. (That is, if* $\mathbf{W}$ *is updated to maximize the information between input and output, then the output converges to independence.)*

Lee *et al.* showed later [125, 1999] that the squashing functions represent the cdfs of the unknown sources, but the success of the algorithm does not depend on the finer details of the unknown pdf. Instead, the only critical factor is the non-Gaussianity of the distribution. A traditional ANN sigmoid squashing function [25] has a super-Gaussian derivative. Hence, they have shown that the information throughput of that node is maximized by super-Gaussian signals.

The *extended-infomax* (Xinfomax) algorithm by Lee *et al.* [125, 1999] extends the ANN model of ICA to sub-Gaussian signals by redefining the squashing function. A sub-Gaussian distribution will be weighted through a neuron with a sub-Gaussian cdf (*i.e.*, its related pdf has kurtosis $k < 0$). A super-Gaussian distribution will be weighted through a neuron with a super-Gaussian cdf.

Xinfomax uses the above to reformulate ICA as the optimization of information throughput in an ANN with a parameterized family of nonlinear squashing functions. An unsupervised learning algorithm is then derived from the stochastic gradient update of an anti-Hebbian [25] learning rule that maximizes the information throughput.

To summarize the Xinfomax derivation:

1. ICA Theorem;

2. Infomax Principle: Reformulate cost function as "negative of information throughput";

3. Derive stochastic gradient update;

4. Obtain minima to condition by a (suitably modified) stochastic-gradient algorithm.

3.5.5.2 FastICA Technique

A popular implementation of ICA called FastICA is presented now. Introduced by Hyvarinen and Oja in 1997 [94], it is a fixed-point algorithm intended to overcome the slow convergence and computational complexity of gradient descent algorithms. Their optimization criterion is based on the following proposition, which is a corollary to the central limit theorem.

Theorem 3.24 (Central Limit Theorem (Lyapunov)[236]). *Let $\underline{x}_i$ be a sequence of independent random variables on $\mathcal{X}$ such that $\mathcal{E}\left\{\underline{x}_i\right\} = \mu_i < \infty$ and $\mathcal{E}\left\{(\underline{x}_i - \mu_i)^2\right\} = \sigma_i^2 < \infty$ for all i. Then, provided that*

$$\sum_{i=1}^{n} \mathcal{E}\left\{\left|\underline{x}_i - \mu_i\right|^3\right\} < \infty \; \forall n \text{ and} \tag{3.71}$$

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} \mathcal{E}\left\{\left|\underline{x}_i - \mu_i\right|^3\right\}}{\sum_{i=1}^{n} \sigma_i^2} = 0 \quad \textit{(Lyapunov condition)} \tag{3.72}$$

the sum $\underline{y}_n = \sum_{i=1}^{n} \underline{x}_i$ converges to a Gaussian. In particular, the standardized variables $(\underline{y}_n - \sum_{i=1}^{n} \mu_i)/\sum_{i=1}^{n} \sigma_i^2$ converges in distribution to the standardized Gaussian $\mathfrak{g}(0,1)$ (i.e., zero-mean with unit-variance).

This theorem describes the limit of an infinite sum (*i.e.*, mixture) of variables. However, following this principle, mixtures (even finite mixtures) roughly become *more* Gaussian. This principle of *maximal non-Gaussianty in orthogonal mixing* is too complicated for the current

exposition, and so the more accessible principle of *extremal kurtosis in orthogonal mixing* is formalized below.

Corollary 3.25 (Extremal Kurtosis). *Let* $\underline{s} \in \mathcal{R}(\mathbb{R}^{N_c})$ *have independent non-Gaussian components of unit variance,* $k(\underline{s}_i) \neq 0 \ \forall i$. *Then for any unit vector* $\boldsymbol{v} \in \mathbb{R}^{N_c}$, $\|\boldsymbol{v}\| = 1$, *the unnormalized kurtosis,* k', *of the mixture* $\boldsymbol{v}^{\mathrm{T}}\underline{s}$ *satisfies*

$$\min_i k'(\underline{s}_i) \leq k'(\boldsymbol{v}^{\mathrm{T}}\underline{s}) \leq \max_i k'(\underline{s}_i) \tag{3.73}$$

with equality iff $\boldsymbol{v} = \mathbf{e}_j$ *for some* j. *That is, the kurtosis of an orthogonal independent mixture is extremized by the sources.*

Proof. Recall that the kurtosis of a mixture of independent distributions is additive (3.35) so that

$$k'(\boldsymbol{v}^{\mathrm{T}}\underline{s}) = \sum_i v_i^4 k'(\underline{s}_i) \tag{3.74}$$

First the upper bound will be proven.

Since $\|\boldsymbol{v}\|^2 = 1$,

$$|v_i|^2 \leq 1 \ \forall i \tag{3.75}$$

and equality can only be achieved in the case of a single non-zero element, *i.e.*, $|\boldsymbol{v}| = \mathbf{e}_j$ for some j. It follows that $|v_i|^4 \leq |v_i|^2 \ \forall i$ and hence $\sum_i |v_i|^4 \leq \sum_i |v_i|^2 = 1$ so that

$$\sum_i |v_i|^4 \leq 1 \tag{3.76}$$

where equality is again achieved iff $|\boldsymbol{v}| = \mathbf{e}_j$ for some j.

Now since, $k'(\underline{s}_i) \leq \max_i k'(\underline{s}_i)$ it follows that

$$v_i^4 k'(\underline{s}_i) \leq v_i^4 \max_i k'(\underline{s}_i) \tag{3.77}$$

$$\Rightarrow \sum_i v_i^4 k'(\underline{s}_i) \leq \sum_i v_i^4 \max_i k'(\underline{s}_i) = \max_i k'(\underline{s}_i)\left(\sum_i v_i^4\right) \leq \max_i k'(\underline{s}_i) \tag{3.78}$$

from which it follows that

$$k'(\boldsymbol{v}^{\mathrm{T}}\underline{\boldsymbol{s}}) \leq \max_i k'(\underline{s}_i) \tag{3.79}$$

and so the kurtosis of the mixture is bounded by the maximal source kurtosis. The mixture can achieve this maximum iff $|\boldsymbol{v}| = \mathbf{e}_{\arg\max_i k'(\underline{s}_i)}$ because the rightmost inequality of (3.78) is strict for non-cardinal $\boldsymbol{v}$.

Proof of the lower bound proceeds as the above from the inequality

$$-|v_i|^4 k'(\underline{s}_i) \leq -|v_i|^4 \min_i k'(\underline{s}_i) \tag{3.80}$$

□

FastICA uses principles like the above to formulate the ICA theorem as an optimization of non-Gaussianity (*e.g.*, negentropy or kurtosis). Since the sum of independent random variables tend to a Gaussian limit, the maximally non-Gaussian linear combination is a trivial one. This provides the FastICA optimization condition: jointly determine the n maximally non-Gaussian linear combinations of $\underline{\boldsymbol{x}}$. Using the two-step approach described earlier, $\mathbf{W}$ is orthogonal. Hence, the linear combinations to be found are orthogonal. Therefore, the linear combinations can be found one-by-one (in a deflationary sense), and constrained by Gram-Schmidt orthogonalization [40] to be (i) orthogonal to one another and (ii) normalized.

To find a single maximal non-Gaussian linear combination, FastICA uses a fixed-point implementation of nonlinear optimization. The vector $\mathbf{v}_m$ has converged when the change to it, $\Delta\mathbf{v}_m$, has become parallel; *i.e.*, $\mathbf{v}_m \propto \Delta\mathbf{v}_m$, making $\mathbf{v}_{m+1} = \mathbf{v}_m + \Delta\mathbf{v}_m = \mathbf{v}_m$ when normalized. The final component is to derive $\Delta\mathbf{v}_m$ as the gradient ascent of negentropy or an equivalent non-Gaussianity measure.

To summarize the derivation of FastICA:

1. ICA Theorem;
2. Maximum Negentropy Principle: Reformulate cost function as "negative of negentropy";
3. Replace negentropy with stable approximations of equivalent maxima;
4. Obtain minima to condition by fixed-point iteration algorithm.

3.5.6 Some Practical Considerations of ICA

The preceding theory has established how ICA can manipulate the linear demixing of a joint distribution through the optimization of cost functions that satisfy the contrast conditions. However, for the actual application of ICA in the processing of data (and not distributions), a few properties must be discussed. First, the ICA theorem was based on an assumption of continuous distributions with no point masses. Though this hypothesis does not apply directly to finite data sets, neither does it completely break down the real-life application of ICA. Instead it is understood that the data points are representatives consistent with a large family of continuous (generating) distributions. (Consider multiple Parzen windows, for example [25].) ICA applies theoretically to the members of this family, and as long as the ICA algorithm can be transcribed to account for the finite samples N, the algorithm can proceed on the data set to reveal an ICA transformation matrix $\widehat{\mathbf{W}}(N)$. This ICA matrix, however, captures conceptually the relationship between a preferred pair of continuous generating distributions. These distributions may not be an optimal or accurate selection. The following example shows the effect of ICA on a dataset with point masses in one source.

Example 3.1 (Degenerate Effects). Let $s_1(n)$ be a periodic square wave source signal. Let also $s_2(n)$ be a non-commensurate periodic signal, such as a sawtooth. A presentation of two

such signals is made in Fig. 3.18(a). The effect of the square wave is to introduce two point masses in its marginal distribution. The independent joint distribution has a similar degenerate effect appearing as two lines parallel to the axis. By introducing a mixing between the signals, the system is no longer independent and the degenerate behaviour is no longer aligned with an axis, Fig. 3.18(b). Moreover, the periodic nature of the signals is lost. An ICA algorithm (here, FastICA) can separate this mixture, however. The result is shown in Fig. 3.18(c). As can be seen, the separated signals are quite close to the original sources, with some residual artifacts still observable. These artifacts stress again the importance for the study of ICA convergence.

Now it has been shown how ICA can apply to datasets with degeneracies not considered in the original ICA theory. A second consideration when dealing with ICA in application is the importance of the *time-integrated independence* of the original signals to be estimated. In particular, statistically independent realizations of some random processes are not independent in the time-integrated sense.

Example 3.2. Let $s_1(n)$ be a Brownian noise process (*i.e.*, the time-integral of a Gaussian iid process). Let $s_2(n)$ be another realization of the same Brownian process. By construction, these are independent realizations of a *time-correlated* random process. That is, they have similar correlation in their random walks, Fig. 3.19(a). Thus, if one examines the joint time-integrated distribution of the two signals, Fig. 3.19(b), patterns that are not statistically independent (*i.e.*, factorable) are observed. Were these to be the sources to a BSS problem, ICA algorithms would not converge to the inverse of the mixing matrix $\mathbf{M}^{-1}$, but in fact, to some other mixture that is less dependent. For application to ICA, the key issue to note is that the statistical *a priori* status of the source signals is irrelevant. The behaviour of the time-integrated distribution of the *given realizations* is what matters.

In the following, the performance analysis of ICA will be considered more closely.

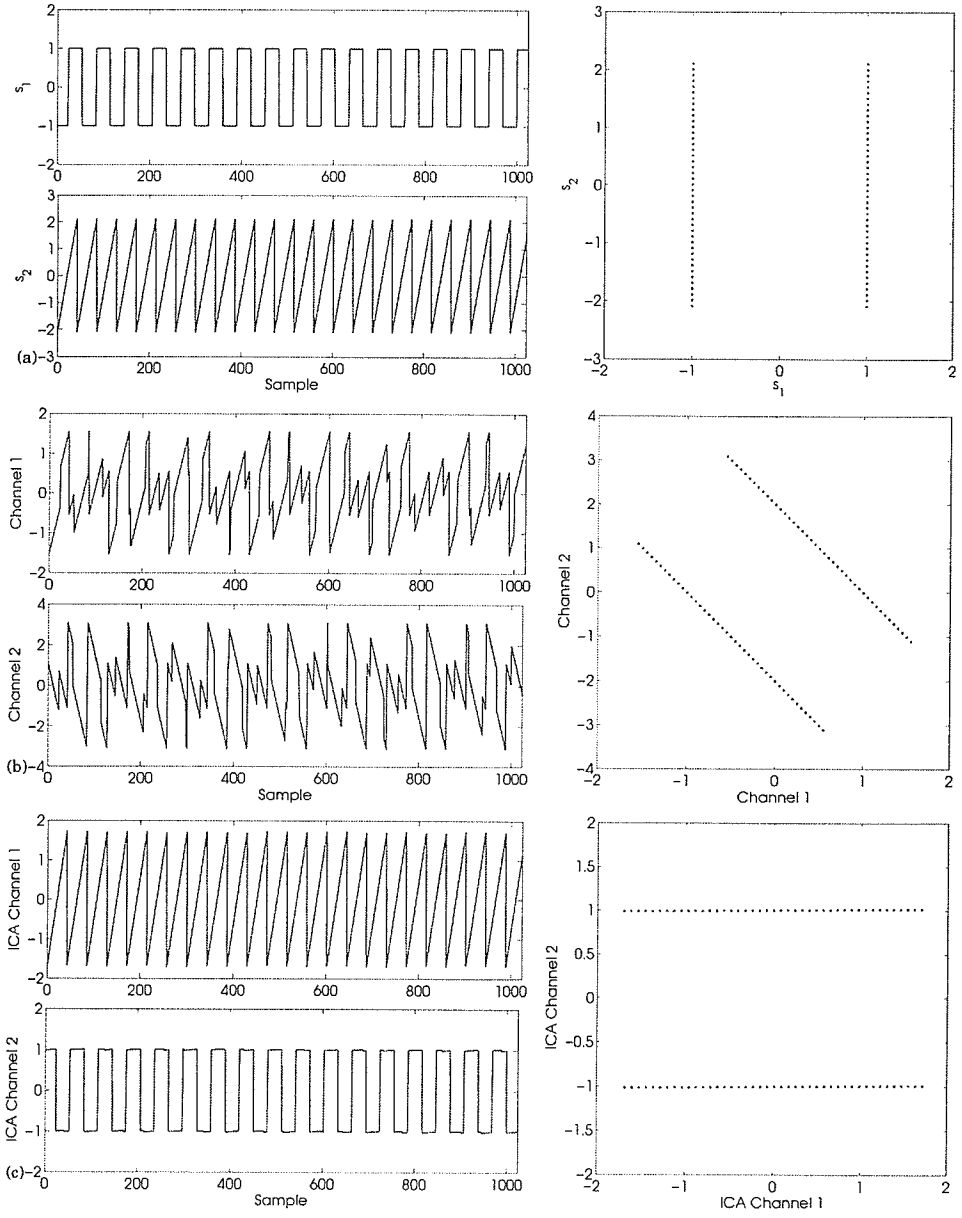


Fig. 3.18 (a) Two independent non-commensurate periodic signals with a degeneracy; (b) a mixture of these signals; (c) a FastICA separation of (b).

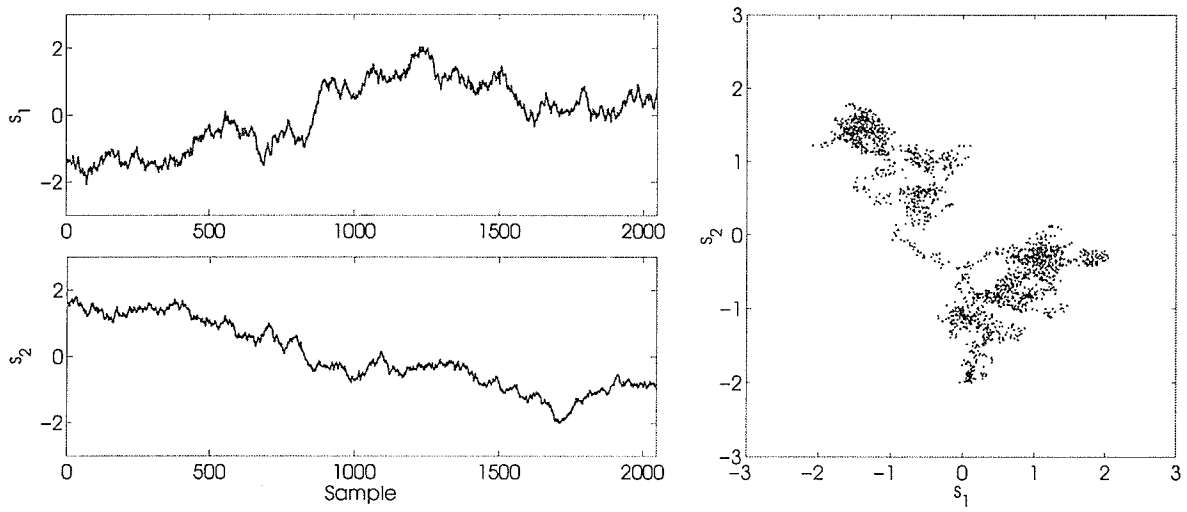


Fig. 3.19 Two independently generated Brownian motions, whose joint time-integrated pdf is not independent.

3.5.7 ICA Performance and SQMs

3.5.7.1 Performance Paradigms

ICA performance can be considered under several paradigms. First, it can be considered under the statistical estimation paradigm, and hence the estimation of $\mathbf{W}$ is essentially a statistical estimation of a multidimensional parameter of a parametric system. Under this paradigm, optimization by maximum likelihood is clearly appropriate, and the issue of estimator performance is approached by analyzing the bias and variance under finitely many points (*i.e.*, Cramer-Rao efficiency). Second, it can be considered from an algorithmic (neural network dynamical) standpoint, and hence convergence, stability, and algorithmic complexity are analyzed. Third, it can be considered from a pattern recognition standpoint. Here, since the estimates must be expressed as normalized linear combinations of the source signals, the contribution from the sources can be analyzed as confusion matrices (*i.e.*, mixture remnants). Last, ICA performance can be considered from a signal processing standpoint.

Here, this is usually approached using SNR [35], as is indicated by the preponderance of SNR based metrics in Table B.1. As was discussed in Ch. 2, SNR is a "universal" metric and analytically elegant. As such, it is actually represented within most of the paradigms just described: parametric estimator variance (Cramer-Rao bounds) and Frobenius norms are essentially power-based (mean square-error) metrics, as is SNR.

3.5.7.2 ICA Performance: The Devoted Literature

In general, the discussion of ICA performance metrics is secondary in the ICA literature and, thus, contributions tend to be disorganized and scattered. Some specific literature is devoted to the subject directly. One such consideration is the work of Vincent *et al.* (*c.f.*, [231, (2006)] and its references), which is approached from the audio processing application. The metrics proposed by Vincent are *acoustic* interpretations of multiple quadratic A-class SQMs. Philosophically, their work is in keeping with the spirit of the investigation here, because the metrics in theory are argued to represent practical acoustic properties. The limitation of their work, however, is (i) their acoustic application, which does not generalize to other BSS problems such as fECG separation, and, consequently, (ii) the consideration of quadratic SNR-style metrics. All of their proposed metrics have the form of a logarithmic ratio of a norm-square numerator to a norm-square denominator. (Note that those authors argue that their performance measures are more meaningful since, in all cases, they extend from $-\infty$ (poor quality) to ∞ (perfect quality).)

Another consideration for the measurement of ICA separation quality is a statistical characterization from the pattern recognition paradigm. Three groups (Muller [141, (2002)], Hyvarinen [89, (2003)], and Grassberger [215, (2004)]) have used a *clustering analysis* to characterize an algorithm's ability to extract sources from a mixture consistently under statistical fluctuations. This is a particularly useful idea in applications (such as fECG separation),

For an ICA solution, $z = \mathbf{W}y$, from the observations x and sphering $y = \mathbf{W}_0 x$, bootstrapping can generate surrogate results $\underline{\tilde{z}}$ for clustering analysis by:

(a) post-sphering randomization: $\underline{\tilde{y}} \stackrel{\text{def}}{=} \mathbf{U}y$ for random $\mathbf{U} \in \mathcal{L}_O$;

(b) pre-sphering randomization: $\underline{\tilde{x}} \stackrel{\text{def}}{=} \mathbf{T}x$ for random invertible $\mathbf{T}$;

(c) observation shuffling: $\underline{\tilde{x}} \stackrel{\text{def}}{=} \text{Pm}(x(t_n))$ for random permutations $\text{Pm}(\cdot)$;

(d) noise injection [77]: $\underline{\tilde{x}} \stackrel{\text{def}}{=} (\mathbf{W}\mathbf{W}_0)^{-1}\left(\mathbf{D}_\rho \mathbf{W}\mathbf{W}_0 x + \mathbf{D}_{1-\rho}\underline{n}\right)$ for random noise signals $\underline{n}$.

If the data is from simulations, different realizations of $\mathbf{M}$ and s can also be used.

Fig. 3.20 Bootstrapping methods for ICA analysis.

where multiple components extracted from the channels can be identified as a dependent multidimensional subspace and, therefore, may impact the algorithm. For example, any eight channel abdominal BSP should identify (at least) a three-channel subspace for the maternal heart. Under repeated calculations, these channels may look different (*i.e.*, being mixtures of each other), but they should always cluster together as subspace. The Hyvarinen *et al.* approach [89] considers absolute-correlation between channels as a clustering measure. Muller *et al.* [141, (2002)] uses an angular "dot-product" measure of the ICA subspaces instead. Grassberger *et al.* [215], meanwhile, evaluates MI to directly use a HOS-based clustering. Whatever clustering measure is used, it must remain invariant to the ICA degeneracies.

When available, clustering can be applied to a ground truth, or simply to the ensemble of results coming from a statistical bootstrap on the algorithm. Bootstrapping can be done by changing the initial condition of the ICA algorithm (if taken as an explicit parameter), or by modification of the data (*i.e.*, Fig. 3.20). If the algorithm is block-based or online, then the ordering of the observations can be statistically shuffled as well. (Simulations from an *a priori* statistical model, where s and $\mathbf{M}$ are available, can also simply generate different realizations

$\underline{s}$.) Muller [77, (2003)] has also considered evaluating the sensitivity of the solution (*i.e.*, robustness) to statistical changes by injecting noise under an energy constraint (*i.e.*, replacing some "source energy" with Gaussian noise energy) as a bootstrap.

Grassberger also considered evaluating the rotational-mean pairwise-MI, $\mathcal{E}_\theta\{\mathcal{M}_\theta(i,j)\}$, against the pairwise MI of the actual solution, $\mathcal{M}_{\widehat{\theta}}(i,j)$, (which is minimal over θ). Here, the greater is the difference

$$\mathcal{E}_\theta\{\mathcal{M}_\theta(i,j)\} - \mathcal{M}_{\widehat{\theta}}(i,j) \tag{3.81}$$

the more stable the solution (*i.e.*, the more independent it is). Grassberger argues that this ability to estimate easily the pairwise-independence of the solution is important in real life, where it is unlikely that all sources in an observation will be independent.

The common element of these techniques is some graphical representation (*e.g.*, an intensity plot) of the clustering-similarity of the channels. Non-diagonal contributions identify a "confusion" to the ICA system.

The last dedicated contribution to the discussion of ICA performance is a classification of algorithms by Cardoso. Some adaptive algorithms (including the maximum likelihood equivalent algorithms with a natural gradient), have the important property of BSS *equivariance* under noise-free conditions. By definition [33, Cardoso (1996)][35, Cardoso (1998)], equivariance is a property of an BSS estimating algorithm where the estimate $\widehat{\mathbf{W}}(N)$ is independent of the initial mixing matrix $\mathbf{M}$. It holds that all adaptive algorithms of the form

$$\Delta\mathbf{W} \propto \eta(\boldsymbol{y})\mathbf{W} \tag{3.82}$$

where $\boldsymbol{y} = \mathbf{W}\boldsymbol{x}$ and $\eta(\boldsymbol{y}) \in \mathcal{L}$ is some matrix function of $\boldsymbol{y}$, are equivariant, including the natural gradient (3.63). In terms of an adaptive system, this implies that the attractive dynamics (*e.g.*, stability and asymptotic statistical bias and variance under finite samples) to the optimum solutions in $\mathcal{L}$ are *only* determined by the properties of the sources $\boldsymbol{s}$, and not the mixing

matrix **M**. In these conditions, **M** simply identifies (in one sense) the initial condition to the optimization trajectory and plays no role in the "shape" of the cost surface $\mathcal{I}$. (Note, therefore, obtaining a clustering index of an equivariant algorithm by bootstrapping (*i.e.*, randomizing) the mixing matrix does not accomplish much.) This dependency of an equivariant algorithm's stability, bias, and variance on the sources, s, is a complicated relationship. Loosely speaking, however, it follows that algorithms perform better as the factorable source distribution f becomes more non-Gaussian.

Note, however, that equivariance is not possible in situations of nonzero additive noise. Here, a poorly-conditioned mixing matrix will amplify the corruption due to noise more than a well-conditioned one [35], effectively changing the cost surface $\mathcal{I}$. Thus bootstrapping with Muller's noise injection scheme [77] still has an effect. Grassberger's consideration of pairwise MI would also still be of interest, since it is analyzing the factorable source distribution $f(s)$.

The remainder of the developments in ICA performance are introduced in random papers on ICA algorithms. The most recognizable of these is Amari's performance index, which acts as scalar metric of a confusion matrix. This and related metrics used in the literature will be presented in Sec. 3.5.7.3. After this, the arsenal of SOS-based metrics will be considered.

3.5.7.3 Amari Error and Related SQMs

One of the most significant and widely quoted SQM is the *Amari performance index* (API), A_ϵ [7, (1996)]. The API is an A-class SQM requiring *a priori* information in the form of the mixing matrix **M**. Let an ICA solution be proposed using a sphering matrix $\mathbf{W_0}$ and an ICA

demixing matrix **W**. Then, according to the ICA theorem, the product matrix **R**

$$\mathbf{R} \stackrel{\text{def}}{=} \mathbf{W}\mathbf{W}_0\mathbf{M} \tag{3.83}$$

$$= \left[R_{ij}\right] \tag{3.84}$$

should be of the form

$$\mathbf{R} = \mathbf{P}\mathbf{D} \tag{3.85}$$

for an accurate solution **W**, where (as in the theorem on p. 108) **P** is an arbitrary permutation, and **D** is an arbitrary non-singular diagonal matrix. Thus only N_c elements R_{ij} should be non-zero, one in each row, and one in each column. The API is a scalar value quantifying the size of the supposed "sparse" elements of **R**. In this way, the API is invariant to the ICA degeneracies, (3.52)-(3.55), as a true ICA SQM should be. If the ICA solution is poorly chosen, then the "non-degenerate cross terms" of **R** will be nonzero, and thus it acts as a type of confusion matrix. (The analogy of a decision theory confusion matrix is made more exact if the **R** matrix was reduced into a "nearest-to-identity" form, $\mathbf{D}^{-1}\mathbf{P}^{-1}\mathbf{R}$.) The API calculates the size of the "confusion" according to

$$A_\epsilon(\mathbf{R}) \stackrel{\text{def}}{=} \sum_i \left(\frac{\sum_j |R_{ij}|}{\max_{j'} |R_{ij'}|} - 1 \right) + \sum_j \left(\frac{\sum_i |R_{ij}|}{\max_{i'} |R_{i'j}|} - 1 \right) \tag{3.86}$$

Note that each fraction in the summands are bounded below by 1 and so each summand is non-negative. (An entirely zero row or column should not exist, since only nonsingular matrices have been considered. For the sake of consistency, however, any 0/0 fraction can be considered ∞.) Here, the fraction normalizes for the scaling degeneracy, while the creative indexing compensates for the permutations. The right hand summation is essentially a column-oriented version of the left hand side, so the API is invariant under transposition; $A_\epsilon(\mathbf{T}) = A_\epsilon(\mathbf{T}^{\mathrm{T}})$.[3] For

[3]This has some interpretative benefit in the consideration of the parameterization of $\mathcal{L}$ and extensions to non-square matrices, but they needn't be discussed here.

a perfect ICA separation, each fraction is exactly 1 and hence

$$A_\epsilon(\mathbf{R}) = A_\epsilon(\mathbf{PD}) = 0 \tag{3.87}$$

since each summand is zero. For non-singular matrices, A_ϵ is maximal at $2N_c(N_c - 1)$ and achieved by an element-wise constant matrix $R_{ij} = R \neq 0 \; \forall i, j$. Sometimes the API is also used in a normalized form to bound it in the unit interval.

A very similar matrix metric called the signal-to-interference ratio, SIR, [232], considers only the left half of the API. That is,

$$\text{SIR} \stackrel{\text{def}}{=} \frac{1}{N_c} \sum_{i=1}^{N_c} \left(\frac{\sum_{j=1}^{N_c} |R_{ij}|}{\left(\max_{j'} |R_{ij'}|\right)} - 1 \right) \tag{3.88}$$

and therefore it is a row-centric formulation (and not necessarily invariant under transposition). Note also a normalization factor is applied in the above so the SIR is bounded above by 1.

A third metric very similar to API has appeared, but it penalizes the confusion matrix in a quadratic sense. Called here as the *Amari-squared index* (ASI) [166], the ASI is calculated

$$\widetilde{A}_\epsilon(\mathbf{R}) \stackrel{\text{def}}{=} \sum_i \left(\frac{\sum_j |R_{ij}|^2}{\max_{j'} |R_{ij'}|^2} - 1 \right) + \sum_j \left(\frac{\sum_i |R_{ij}|^2}{\max_{i'} |R_{i'j}|^2} - 1 \right) \tag{3.89}$$

The quadratic form of the ASI makes it correspond slightly more to power-based considerations, since the square of the matrix entries R_{ij}^2 represents the energy in the ith estimate coming from source j (loosely normalized). This is not a precise interpretation, however, since the ASI maintains the symmetric form of the original. In the opinion of the author, it would actually make better sense, from this standpoint, to only use the row-centric form

$$\widetilde{A}_\epsilon^{+}(\mathbf{R}) \stackrel{\text{def}}{=} \frac{1}{N_c} \sum_i \left(\frac{\sum_j |R_{ij}|^2}{\max_{j'} |R_{ij'}|^2} - 1 \right) \tag{3.90}$$

which would correspond to the significance of the rows as signals in the ICA paradigm.

Now the second-order statistics-based A-class SQMs used in the literature will be surveyed.

3.5.7.4 SOS Related SQMs

The classic SOS statistic for measuring the similarity of two signals is the cross-correlation function. In the case of instantaneous ICA, delays are not considered and hence the cross-correlation index (CCI) of the jth source with the ith estimates is the normalized cross-cumulant

$$\text{CCI}(i,j) \stackrel{\text{def}}{=} \frac{\sigma_{ij}}{\sigma_i \sigma_j} \tag{3.91}$$

where σ_{ij} is the cross-correlation proper

$$\sigma_{ij} \stackrel{\text{def}}{=} \left(\mathcal{E}\left\{\left(\widehat{s_i} - \mathcal{E}\left\{\widehat{s_i}\right\}\right)\left(s_j - \mathcal{E}\left\{s_j\right\}\right)\right\}\right)^{1/2} \tag{3.92}$$

and σ_i and σ_j are the standard deviations of $\widehat{s_i}$ and s_j respectively. As an ICA SQM, the CCI is appropriately invariant to the absolute scaling of the source and estimates, but the sign of the CCI still appears. It should be considered, therefore, only in absolute value, $|\text{CCI}(i,j)|$. Furthermore, the absolute magnitude is bounded by 1 (best reconstruction), which is significantly different from SNR (which is ∞ for best reconstruction). Note also that the CCI is a direct estimate between a single source and estimate channel, and therefore is not invariant to permutations. Instead the confusion matrix of such values should be considered in form that is reduced as much as possible to a "nearest-to-identity" form. This type of error measure has been used in practice in [57, (2008)] [42, (2004)].

The other classic form from an additive perspective is the SNR, previously defined in Ch. 2,

$$\text{SNR}_{ij} \stackrel{\text{def}}{=} \frac{\mathcal{E}\left\{(s_j)^2\right\}}{\mathcal{E}\left\{(\widehat{s_i} - s_j)^2\right\}} \text{ [linear units]} \tag{3.93}$$

$$\stackrel{\text{def}}{=} 10\log_{10}\frac{\mathcal{E}\left\{(s_j)^2\right\}}{\mathcal{E}\left\{(\widehat{s_i} - s_j)^2\right\}} \text{ [dB]} \tag{3.94}$$

Here, however, there is no invariance to scaling so it must be assumed that either

(a) the signals are normalized to an *a priori* standard: *e.g.*, unit standard deviation $\sigma_i = \sigma_j = 1$, or variance; or

(b) the maximal (best possible) SNR is chosen [173] over all possible scalings.

When the normalization of the former is applied to energies (not variance), the SNR in decibel units is equivalent to the signal-to-error ratio (SER) used in [9][166]:

$$n_{ij}(t) \stackrel{\text{def}}{=} s_j(t) - \widehat{s_i}\sqrt{\mathcal{E}\left\{s_j^2\right\}/\mathcal{E}\left\{\widehat{s_i}^2\right\}} \tag{3.95}$$

$$\text{SER}_{ij} \stackrel{\text{def}}{=} 10\log_{10}\frac{\mathcal{E}\left\{s_j^2\right\}}{\mathcal{E}\left\{n_{ij}^2\right\}} \tag{3.96}$$

Furthermore, note that the maximal SNR is achieved by rescaling $\widehat{s_i}$ according to

$$\widehat{s_i} \leftarrow \frac{\sigma_{ij}}{\sigma_i^2}\widehat{s_i} \tag{3.97}$$

and not the SER normalization. Both the SNR and SER must also be considered as a complete confusion matrix (*e.g.*, $[\text{SNR}_{ij}]$) to allow for permutations.

Cardoso [35] considered the pairwise interference-to-signal ratio (ISR) in an asymptotic form

$$\text{ISR}_{ij} \stackrel{\text{def}}{=} \lim_{N\to\infty} N\mathcal{E}\left\{\frac{\left|R_{ij}\right|^2\mathcal{E}\left\{\left|s_j\right|^2\right\}}{\left|R_{ii}\right|^2\mathcal{E}\left\{\left|s_i\right|^2\right\}}\right\} \tag{3.98}$$

and discussed some analytic bounds.

The remaining metrics are quadratic metrics based on the demixing matrix result $\mathbf{R} = \mathbf{W}\mathbf{W}_0\mathbf{M}$. Vincent [231] considered the inter-symbol interference (ISI) as

$$\text{ISI}_i \stackrel{\text{def}}{=} \frac{\sum_{j'}\left|R_{ij'}\right|^2 - \max_{j'}\left|R_{ij'}\right|^2}{\max_{j'}\left|R_{ij'}\right|^2} \tag{3.99}$$

(which is the simple summand of (3.90)). Principe [55], meanwhile, has considered a decibel-based signal-to-distortion ratio (SDR)

$$\text{SDR} \stackrel{\text{def}}{=} \frac{1}{N_c} \sum_{i=1}^{N_c} 10 \log_{10} \frac{\left(\max_{j'} R_{ij'}\right)^2}{\left(\sum_{j'} R_{ij'}^2\right) - \left(\max_{j'} R_{ij'}\right)^2} \tag{3.100}$$

for A-class performance measurement. This SDR estimate is equivalent to an SNR if the sources and estimates are normalized to unit variance, the signal for estimate i is the source j with maximal power contribution, and the $R_{ij} \geq 0$. The SDR summands, fundamentally, are also just $1/\text{ISI}_i$ converted to units of dB. Note, however, that SDR expresses the average dB of all channels (instead of the average of all channels in dB).

In [88, (2001)], it is declared that a "good separation is achieved at an SDR of 20 dB". While this may be loosely true when visually observing a time series, in keeping with the analysis of SNR presented in Ch. 2, any power-based global metric has inherent weaknesses. (For example, if the mECG channel has 40 dB and the fECG channel has 0 dB, yielding an SDR = 20 dB, is this a good separation?) In particular, it is of interest to study the connection between this metric and other metrics that are sensitive to (i) higher-order statistics directly, and (ii) patterns within the error distribution. All the (officially) second-order metrics above are related to HOS only indirectly through the matrix $\mathbf{W}$, which was HOS-dependent. Even though the principle of ICA directly implies that the HOSs of a signal is significant, only Grassberger has advocated that a HOS measurement of clustering or error be applied in practice (and that through MI between the channels). This work will not evaluate *all* these power and matrix-based metrics, but consider instead their relationship to the relevant HOS. That is, any of these SQMs are related to the HOS-based objective function of the ICA algorithm through the matrix $\mathbf{W}$. What should be considered is the convergence properties of the metric with that of the objective (cost) function of the algorithm.

3.5.7.5 Proposed SQM Methodology

To complete this discussion of ICA metrics, a system for structuring the ICA performance measurement for this work is now presented. In particular, a systematic approach for dealing with the degeneracies of the ICA solution must be presented (since not all metrics are invariant under them). The goal is to provide a framework by which the joint convergence of both higher-order and second-order features and metrics — both pairwise and system-wide — can be systematically evaluated.

Proposition 3.26 (ICA Solution). *For this work, it is proposed that*

(a) the sources and estimates be centralized;

(b) the variance of the sources s_j be normalized to unity;

(c) the ICA estimates be of the form $\widehat{s_i} = \mathbf{e}_i^{\mathrm{T}}\mathbf{W}\mathbf{W}_0\mathbf{M}\boldsymbol{s}$ (where the sphering matrix $\mathbf{W}_0$ is the result of a PCA, and the mixing matrix $\mathbf{M}$ is given); and

(d) the ICA demixing matrix $\mathbf{W}$ be orthogonal such that for all i,

(a) $\left|CCI(\widehat{s_i}, s_i)\right| \geq \left|CCI(\widehat{s_i}, s_j)\right| \ \forall\ j$ (resolving permutation);

(b) $CCI(\widehat{s_i}, s_i) > 0$ (resolving sign but not scale); and

(c) the variance of the estimates $\widehat{s_i}$ be normalized to unity (resolving scale).

From the traditional metrics described previously, those that will be included for study include

(a) the CCI (3.91);

(b) Amari's original API, A_ϵ (3.86); and

(c) the SNR (which, under Prop. 3.26, can be averaged to something nearly equivalent to Principe's SDR (3.100)).

To supplement these metrics, two more geometrical A-class metrics will be used, since these are consistent with the parameterizations of $\mathcal{L}_{N_c}$. The first of these is the *vector norm error* (VNE)

$$\text{VNE}_i \stackrel{\text{def}}{=} \left\| \mathbf{e}_i^{\text{T}} \left(\mathbf{W}\mathbf{W}_0 - \mathbf{M}^{-1} \right) \right\| \tag{3.101}$$

Note that there is no indeterminacy in the above under the assumptions of Prop. 3.26. The VNE represents the Euclidean vector norm of the error in determining the ith ICA vector. (A similar metric was applied in [87, (2007)].) The value of VNE_i is closely related (under a different ICA normalization than proposed here) to the sum

$$\left(\frac{\sum_j R_{ij}^2}{\max_{j'} R_{ij'}^2} \right) - 1 \tag{3.102}$$

which appeared in the ISI and (3.90).

Since the VNE is a vector norm, a full matrix version can also be defined. This is called the *matrix Frobenius error* (MFE) and is written

$$\text{MFE} \stackrel{\text{def}}{=} \left\| \mathbf{W}\mathbf{W}_0 - \mathbf{M}^{-1} \right\|_F \tag{3.103}$$

which is well defined according to Prop. 3.26. The MFE can be related to the Frobenius norm of the "non-degenerate cross terms" of $\mathbf{R}$

$$\sum_i \left(\frac{\sum_j R_{ij}^2}{\max_{j'} R_{ij'}^2} - 1 \right) \tag{3.104}$$

by a suitable normalization. It is the opinion of the author that explicitly using $\mathbf{M}^{-1}$ (instead of the product matrix $\mathbf{R}$) makes more sense from the point of view of the ICA geometry and optimization objective.

HOS will also be considered explicitly as SQM features for convergence analysis. These include

(a) the kurtosis of the estimate $k(\widehat{s_i})$;

(b) the pairwise MI of the estimate to the source, $\mathcal{M}(\widehat{s_i}, s_i)$; and

(c) the joint MI of the ICA estimate, $\mathcal{M}(\widehat{s_1}, \ldots, \widehat{s_{N_c}})$.

Calculation of the MI will be performed using the Grassberger *et al.* algorithm [120, (2004)].

3.6 Summary

This chapter has presented the relevant concepts needed to statistically characterize (i) ECG distributions and (ii) their separation by ICA. First, the ECG is an empirical signal, and hence the time-integrated distributions from discrete time series are the focus of ICA separation. Scatterplots were demonstrated to be a useful visualization of these distributions. Second, these distributions were characterized by moments and cumulants. The properties of these parameters have been described, and, in particular, the important distinction between second-order and higher-order statistics has been made. HOS are key in the characterization of non-Gaussianity. Third, the independence of a multivariate distribution has been defined, and the properties of independence under factorization, entropy, mutual information, moments, and cumulants have been discussed. The difference between second-order and higher-order independence in the presence of non-Gaussian distributions was demonstrated. These elements of independence provide the mechanism for ICA separation.

Fourth, ICA was universally defined as an extension to decorrelation (PCA) that requires a nonlinear optimization of HOS cost functions over the space of matrices. The geometry of this space and the interpretation of constrained maximization was discussed. Theorem 3.22

[45, Comon (1994)] presented the natural indeterminacies that exist in an ICA solution. Some practical considerations of ICA on data from deterministic and random processes demonstrated the significance of time-integrated distributions and the relative insignificance of point masses. (Here, ICA has not been considered from the standpoint of sparse (L^1) optimization, though some interesting literature appears on the subject (*c.f.*, [53] and references).)

Fifth, the structure of traditional metrics for ICA performance was discussed. The foundation of these tend to be either (i) an SNR-type function between an estimate channel and a source channel, or (ii) an Amari-type evaluation of the non-degenerate cross terms of the product matrix $\mathbf{WW_0M}$. Other statistical pairwise comparisons can be second-order (such as the CCI) or higher-order (such as the MI). From these, a particular scheme for the post-processing of ICA was presented whereby (Prop. 3.26)

(a) the variance of the sources s_j are normalized to unity;

(b) the ICA estimates are of the form $\widehat{s_i} = \mathbf{e}_i^{\mathrm{T}}\mathbf{WW_0M}\boldsymbol{s}$; and

(c) the ICA demixing matrix $\mathbf{W}$ is orthogonal and uniquely determined by the pairwise CCI and unit variance.

A family of performance metrics was also identified for inclusion in the experiment design; namely the

(a) cross-correlation index, $\mathrm{CCI}_i = \mathrm{CCI}(\widehat{s_i}, s_i)$;

(b) signal-to-noise ratio, $\mathrm{SNR}_i = \mathrm{SNR}(\widehat{s_i}, s_i)$;

(c) vector norm error, $\mathrm{VNE}_i(\mathbf{WW_0}, \mathbf{M})$;

(d) kurtosis of the estimate, $k_i = k(\widehat{s_i})$; and the

(e) pairwise MI of the estimate to the source, $\mathcal{M}_i = \mathcal{M}(\widehat{s_i}, s_i)$;

that apply to each ICA estimate i. Consideration hereafter will only consider the fECG component that is the interest of this work.

Three other system-wide performance metrics were also identified; namely

(a) the Amari performance index, $A_\epsilon(\mathbf{WW_0M})$;

(b) the matrix Frobenius error, MFE($\mathbf{WW_0}, \mathbf{M}$); and

(c) the joint MI of the ICA estimate, $\mathcal{M}(\widehat{s_1}, \ldots, \widehat{s_{N_c}})$.

Descriptions of the implementation of these ideas are presented in Ch. 6.

The last item of note from this chapter is the definition of the Rényi and relative Rényi entropies in Sec. 3.4.7. These will be used in the next chapter for defining scaling ECG features.

Chapter IV

Study of Scaling Invariance for ECG Features

4.1 Overview

This chapter presents a study of multifractal analysis (MFA) for ECG feature extraction. These global complexity features can then provide scaling-based SQMs for ICA. This use of scaling SQMs is a novel development in the performance analysis of ICA.

To begin the chapter, scaling as a general characterizer is motivated by some arguments from measurement theory. Then the advanced abstract mathematical concept of "measures" is defined in order to unify the field of scaling theory for sets, functions, and distributions. Fractal measures are then defined in Sec. 4.2.2.3 as non-differentiable measure limits of cascading differentiable measures.

The core background on MFA is then presented in three parts. First, the characterization of fractal measures by the scaling of fractal dimensions is motivated and explained in Sec. 4.3. The non-uniform Cantor measures are presented as analytical examples of multifractal measures. Here, *scaling* is defined mathematically, and the Hausdorff fractal dimension is presented in the context of a coarse-grained measure. Second, the preceding theme is extended in Sec. 4.4, where a general notion of fractal dimension, and the necessity of multiple dimensions to characterize multifractals. Families of fractal dimensions with respect to a coarse-grained measure's partition function are presented as suitable for MFA. For the classical Boltzmann partition function, the $\tau(q)$ family of scaling exponents is presented, and an example calculation is drawn from the non-uniform Cantor measures. Third, the families of scaling exponents

drawn from the Boltzmann partition function are extended via thermodynamic transformation to the Rényi and Mandelbrot fractal dimension spectra to gain compactness in the MFA dimensions. Here in Sec. 4.5, these spectra are demonstrated to all come from the exponential behaviour of a class of entropies extracted from the partition function. Specifically, the Rényi spectrum, Sec. 4.5.1, results from the characteristic exponents of Rényi generalized entropies, while the Mandelbrot spectrum, Sec. 4.5.3, results from the exponents of Hölder and Mandelbrot generalized entropies.

Section 4.6 applies these concepts of measure and coarse-graining to the MFA of an ECG. The general process by which an ECG time series is reconstructed into an attractor is described. The two techniques to implementing a fixed-size coarse-graining of the ECG attractor, and their merits, are also discussed. Note, however, that the implementation details of the chosen technique will be left for Ch. 6 and App. C.

One of the significant contributions developed in this work is the derivation of the direct Mandelbrot spectrum from a correlation partition. To provide context for this development, a directed review of the literature is presented in Sec. 4.7 that (i) distinguishes between direct and indirect calculations, and (ii) identifies the missing component in the fractal theory. Specifically, the required correlation-based Hölder and Mandelbrot entropies are derived in Sec. 4.7.3. An example of the direct and indirect techniques applied to the Ikeda attractor is then shown in Sec. 4.7.4. While the general concept and formulae of the direct correlation calculations are explained here, App. C contains a more detailed discussion of the correlation partition algorithm required for entropy calculation.

Last, Sec. 4.8 synthesizes the discussion of MFA spectra for the definition of SQMs. Though both "relative metrics of scaling spectra" and the "scaling spectra of relative entropies"

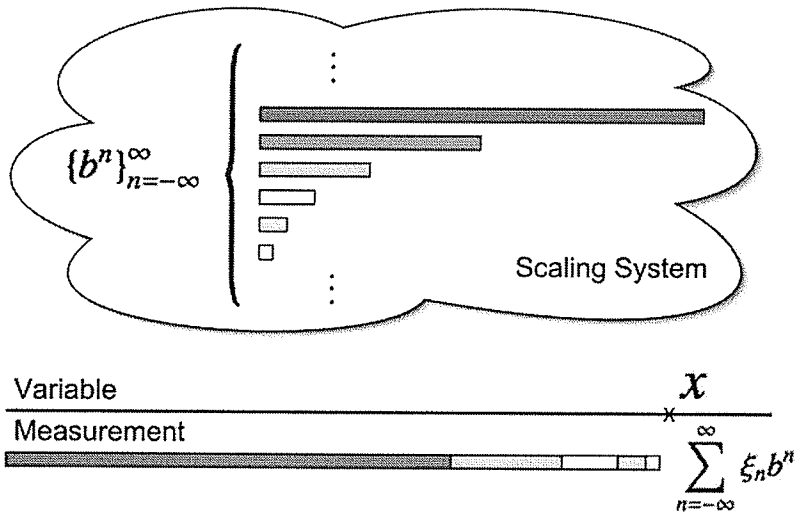


Fig. 4.1 Expressing a variable as a number is equivalent to measuring a variable by a scaling system.

are considered, it is shown here that only simple error-type metrics on the differences in spectra are available for the experimental design.

4.2 Scaling and Fractals

4.2.1 Scaling, Self-similarity, and Numbers

Expressing a variable on the real number line means that the variable inherits from the very numbers themselves an important relationship with scaling. The decimal system of number representation is fundamental and unquestioned in modern society. With the development of computers, this expression of the number line has been supplemented with the binary system as well. Both systems have in common the expression of a number as a "consecutive sum of scaled parts". Essentially, a positive unit or *base* $b > 1$ is assumed. A number $x \in \mathbb{R}$ is then expressible as an infinite sequence of integers, a sign, and a radix point of the form

$$x = \pm \ldots \xi_2 \xi_1 \xi_0 . \xi_{-1} \xi_{-2} \ldots \tag{4.1}$$

where the $\xi_n \in \{0, \ldots, \lfloor b \rfloor\}$ are integers between zero and b. This representation is unique up to infinite sequences. This formal sequence represents the number x because it satisfies the following identity

$$x = \pm \sum_{n=-\infty}^{\infty} \xi_n b^n \tag{4.2}$$

This formal representation of x is in fact the idealized form of *measuring* x as shown in Fig. 4.1. Two basic philosophical forms in this measuring process exist as mathematical functions: scaling and concatenation. The number x exists on the number line, and can be identified by comparing it with concatenated rods of length b^n. Any remaining difference can be further measured by reducing n (and thus reducing the length of the comparison rod) and repeating the measurement process. The concatenation is represented in the natural numbers ξ_n, while the scaling is represented in the changing power of the base. One can conclude, thus, that representing an object numerically means representing the object by a *scaling* system of elements. This ideal mathematical analogy of measurement is, in fact, very physical. Length in the international (metric) system of units is standardized by comparison to scaled multiples of the metre, itself defined as a concatenation of small absolute units, (*i.e.*, an integer number of wavelengths of a specific frequency of light).

From this elementary discussion of measuring numbers, it is clear that the *measuring process of a variable* is *self-similar*. The self-similarity paradigm is not a special add-on, supplementary hypothesis, or method of convenience. It is at the very root of measuring with numbers. This is not to say that every set, function, or distribution is *invariant* under scaling. Invariance is only *one* property of the scaling behaviour of the said object. It does argue, however, that the scaling behaviour of these objects *exists* and can be characterized; be it by invariance or other more complicated forms. The application of scaling (or multifractal) analysis is too often bogged down in the mathematical idealism of historical invariant fractal

objects and true multiplicative processes. It suffices to note that these classical fractal objects and multiplicative processes are specifically well-characterized by multifractal analysis. They "fulfill each other" in much the same way the characterization of a random variable is fulfilled by its distribution. However, as was mentioned at the beginning of Ch. 3, distributions can be analyzed and characterized whether they came from a random variable or not. Similarly, in the development of the multifractal analysis theory, it will be shown how the power-law scaling behaviour of partition functions can be characterized. This may be ideal only for multiplicative processes, but this does not inhibit anyone from using these methods in characterizing the scaling behaviour of partition functions from empirical processes [70, (1996)][117, (1994)]: here, the ECG.

4.2.2 Measure Theory and Fractals

In the popular imagination, *fractals* are geometrically "rough" natural objects like clouds, mountains, and trees, Fig. 4.2. They were coined historically as such by Mandelbrot after *fractus* ("broken", the same root as "fracture") only as of 1975 [133]. Since then, economists consider the stock trends as fractals. Geometers consider intriguing pathological "monster" sets from iterated equations as fractals. To dynamical theorists, fractals are the "stable but abnormally distributed" attractors of nonlinear systems. The defining element in these very different fields is that "fractal objects" can confuse the eye with a loss of characteristic scale. Essentially, it is difficult to tell whether the object is being analyzed with the naked eye (at normal scale) or a zoom lens (high magnification). This demonstrates that these diverse examples of fractals have a *self-similar complexity* at different scales. This property of interdisciplinary objects captures the universal spirit of MFA. However, it is very difficult to systematically express an MFA in a mathematical formalism without violating its universality [132, Mandelbrot (1999)].

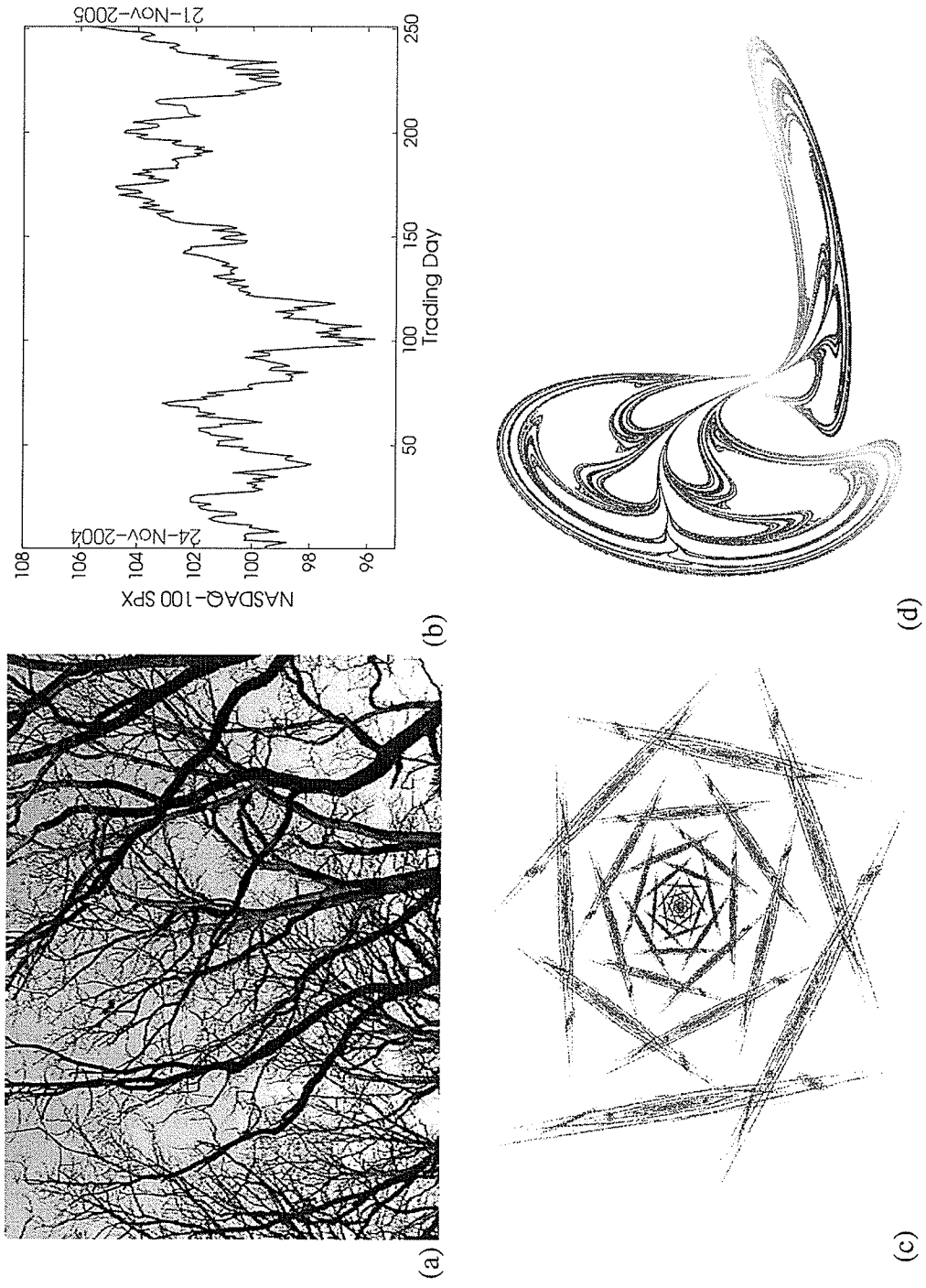


Fig. 4.2 Examples of fractals across disciplines: (a) Natural fractal (From [214]‡); (b) Economics fractal (Data from [158]); (c) Geometric fractal (From [212]‡); (d) Dynamical fractal (From [213]‡).
‡Image is copyright-free for educational use.

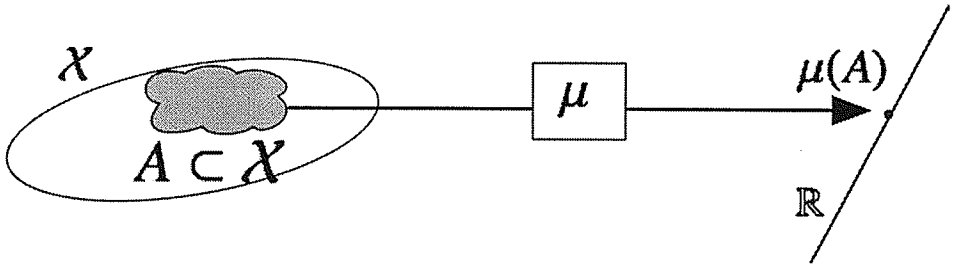


Fig. 4.3 Measures quantify the abstract "mass" of a set.

One helpful tool in defining fractals for engineering efficiently in a broad way is the use of measure theory and partition functions. Measure theory is a mathematical method that unifies geometry and probability in a common treatment. Therefore the geometrical scaling concerned with fractal sets, and the probabilistic scaling concerned with multifractal distributions can both be expressions of measure-theoretic constructs. To this end, a quick discussion of measures is provided, which will enable the treatment of both sets and functions in a uniform way. It will be shown, then, that measures can be systematically transformed to partition functions over coarse-grained scales. The partition functions act as different measurements of distribution "complexity". That is, the meaning of fractal *self-similar complexity* is defined by the partition function being used. When *scaling* exists intuitively (as described in the previous paragraph), it will be shown that power-law relationships are approximated in the appropriate partition functions over scales.

4.2.2.1 Measures Defined

Definition 4.1 (Measure). A *measure* μ is a set function to $\mathbb{R}$ that quantifies the "mass" of the set in abstract terms, Fig. 4.3. (This abstraction is important in the theory of integration.) If one

considers a space $\mathcal{X}$, then for "every"[1] subset $A \subset \mathcal{X}$, the mass of A is written $\mu(A) \in \{\mathbb{R} \cup \pm\infty\}$,

$$\mu : \{A \subset \mathcal{X}\} \mapsto \{\mathbb{R} \cup \pm\infty\} \tag{4.3}$$

Very often, some restrictions to positive and/or finite mass ($0 \leq \mu(A) < \infty$) are made.

4.2.2.2 Sets and Functions

Note now the representation of sets in function space, and vice-versa.

Definition 4.2 (Characteristic Function)**.** Let $\Omega \subset \mathcal{X}$ be a set. Then there exists a unique function $\chi_\Omega \in L^\infty(\mathcal{X})$ such that

$$\chi_\Omega(\boldsymbol{x}) \stackrel{\text{def}}{=} \begin{cases} 1 & : \boldsymbol{x} \in \Omega \\ 0 & : \boldsymbol{x} \notin \Omega \end{cases} . \tag{4.4}$$

This function contains all the information of the set and is called the *set characteristic function.*

Definition 4.3 (Support)**.** The *support* of a function $g : \mathcal{X} \mapsto \mathbb{R}$ is the subset of the domain $\mathcal{X}$

$$\Omega_1 \stackrel{\text{def}}{=} \{\boldsymbol{x} \in \mathcal{X} \mid g(\boldsymbol{x}) \neq 0\} \tag{4.5}$$

It follows that the support and characteristic function are nearly "inverse operations" between sets and functions, since $\Omega_1(\chi_\Omega) = \Omega$.

Measures are set functions. From the above, however, one understands that sets can be expressed as functions as well. This implies that measures can also be expressed as *operators* on functions. The "natural" measures on sets in $\mathbb{R}^{N_e}$, (*i.e.*, the natural mass or "size" of a set), are related to geometrical notions, such as distance. The usual measure on $\mathbb{R}$ is the idea of *interval length.* In higher dimensions, $\mathbb{R}^{N_e}$, this is the notion of N_e-dimensional *volume.* The

[1] This thesis will not concern itself with the distinction of the sigma-algebra of measurable sets from the non-measurable sets, and simplify by saying "every" subset can be measured [31].

volume measures are formally called the *Lebesgue measures* on $\mathbb{R}^{N_e}$ and the notation μ_L will be used here. It is possible to express now how these measures act as operators of characteristic functions. Recall that the Lebesgue integral of a function $g : \Omega \subset \mathbb{R}^{N_e} \mapsto \mathbb{R}$ is defined as

$$I(g, \Omega) = \int_{\Omega} g(\boldsymbol{x})\, d\boldsymbol{x} \tag{4.6}$$

where the dependence on both the function g and the domain Ω have been written explicitly. As is common in physics and engineering, the volume of a set is in fact the integral

$$\mu_L(\Omega) = \int_{\Omega} d\boldsymbol{x} \tag{4.7}$$

$$= I(1, \Omega) \tag{4.8}$$

By the use of the characteristic function, the set function (4.8) can also be expressed as a mapping from L^{∞} to $\mathbb{R}$:

$$\mu_L(\Omega) = \int_{\Omega} d\boldsymbol{x} = \int_{\mathbb{R}^{N_e}} \chi_{\Omega}(\boldsymbol{x})\, d\boldsymbol{x} = \int_{\mathbb{R}^{N_e}} \chi_{\Omega}(\boldsymbol{x}) \cdot 1\, d\boldsymbol{x} \tag{4.9}$$

$$= I(\chi_{\Omega}, \mathbb{R}^{N_e}) \tag{4.10}$$

which depends on functions instead of sets. This transformation from set expression to function expression is diagrammed in Fig. 4.4.

The above equations identify (i) a representation of the Lebesgue measure in the form of an integral of χ_{Ω}, and (ii) a uniform weighting function (namely 1) in the integral of (4.9). This presents an obvious way to define arbitrary *integral measures* by introducing a nontrivial weighting function $m(\boldsymbol{x}) \neq 1$. This extension introduces a large class of measures. For completeness, the following theorem, which demonstrates a sufficient requirement on a *mass density function* m, is included. For a proof, see [31].

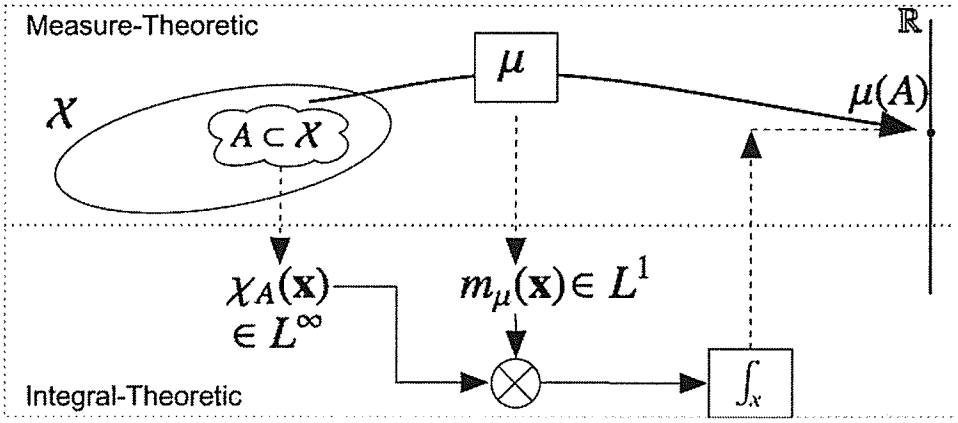


Fig. 4.4 The relationship of the action of an integral measure μ to the set characteristic function, the mass density function, and integration.

Theorem 4.4 (Integral Measures). *Let μ_L be the classical Lebesgue (volume) measure on $\mathbb{R}^{N_e}$. Let also that m, as a* mass density function, *be continuous or a step function. Then*

$$\mu_m(A) = \int_{\mathbb{R}^{N_e}} \chi_A(\boldsymbol{x})\, m(\boldsymbol{x})\, d\mu_L(\boldsymbol{x}) = \int_{\mathbb{R}^{N_e}} \chi_A(\boldsymbol{x})\, m(\boldsymbol{x})\, d\boldsymbol{x} \tag{4.11}$$

is a well-defined measure μ_m. Further, if m is absolutely integrable (i.e., $m \in L^1$), then μ_m is finite; and if $m \geq 0$, then μ_m is positive.

The complexity of the measure μ_m will come to depend very importantly on whether the weighting of m is uniform or not. Figure 4.5 demonstrates the equivalence of uniform measures and the characteristic functions of their support.

4.2.2.3 Fractal Measures

Probabilities are a special case of measures [165]. If the measure μ is (i) positive, and (ii) normalized over the entire space $\mathbb{R}^{N_e}$ (*i.e.*, $\mu(\mathbb{R}^{N_e}) = 1$), then the measure acts as a *probability measure*, assigning a numerical value between 1 and 0 (the *probability*) to subsets of $\mathbb{R}^{N_e}$ (the *event*). (Note, in fact, that the *event* is a *set* and not necessarily a point.) The pdf f_μ is the most common way of dealing with probabilities in engineering. It is perfectly acceptable

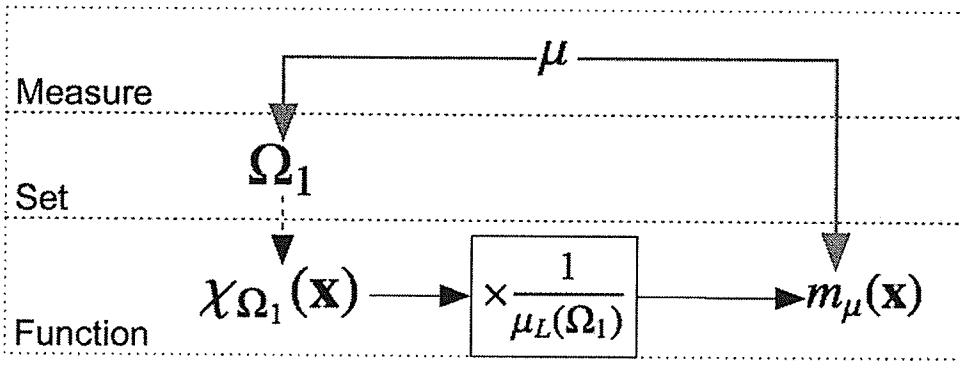


Fig. 4.5 The degenerate relationship of a uniform measure μ with its characteristic function.

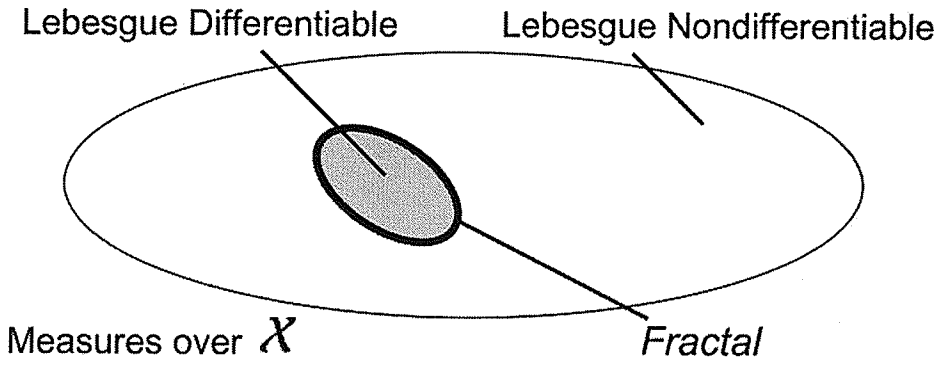


Fig. 4.6 Fractal measures exist as limits beyond the set of differentiable measures.

to reduce probability measures to the pdf function (and much more expedient than a complete measure formulation) as long as the measures can be described by integrals of the type

$$\mu(A) = \int_A f_\mu(\boldsymbol{x})\, d\boldsymbol{x} \tag{4.12}$$

as in in Theorem 4.4. *This is not always the case.* If a point-mass exists in the measure μ, similar to a *Dirac Delta function*, δ_∞, [31], then any "density" would not be a true function. Even though μ is a true measure, the ratio $\mu(A)/\mu_L(A)$ can diverge. This non-differentiable property of a measure is the trademark of fractals.

Fractal measures are *not* characterized among the weighted Lebesgue measures that have

just been described, Fig. 4.6. Instead, they are often characterized by a multiplicative *cascade* of weighted Lebesgue measures. It is from this cascade rule that a fractal acquires its scaling properties. It is also this cascading process that makes fractals occur in nature. No single definition of a fractal has been universally accepted. For the purposes of this thesis, therefore, the definition of fractal will follow Kinsner [117] but be phrased in the terms of the current measure discussion.

Definition 4.5 (Fractal Measure)**.** A *Fractal measure* μ is a measure attainable as the limit of an iterative cascade of weighted Lebesgue measures μ_n, that cannot be represented as a weighted Lebesgue measure itself. That is,

$$\mu(A) = \lim_{n\to\infty} \mu_n(A) \neq \int_{\mathbb{R}^{N_e}} \chi_A(\boldsymbol{x}) m_\infty(\boldsymbol{x})\, d\boldsymbol{x} \tag{4.13}$$

such that

$$\mu_{n+1}(A) = \mu_n(A) \cdot \Phi(\mu_n) = \int_{\mathbb{R}^{N_e}} \chi_A(\boldsymbol{x}) m_{n+1}(\boldsymbol{x})\, d\boldsymbol{x} \tag{4.14}$$

where Φ is the fractal iteration, or *generator*, process. If the measures μ_n are uniform measures (*i.e.*, with density function related to a set characteristic function, $m_n \propto \chi_n$), than μ is *monofractal* and its support is a *fractal set*. Otherwise μ is multifractal.

The two important features of this definition to be recognized are that (i) the μ_n are weighted Lebesgue integral measures, while their limit is not, and (ii) that the convergence of the cascade is regularized through the repeated application of the fractal iteration process Φ. Thus, it does well to write the limit multifractal measure as μ_Φ.

In the usual treatment of fractals [117][132], *fractal sets* are described, and then the necessity of multifractal measures is introduced by example. Here, a more general mathematical treatment has been used which demonstrates first the more descriptive capability of multifractals. It is clear from the definition that monofractals are only the special cases of cascading

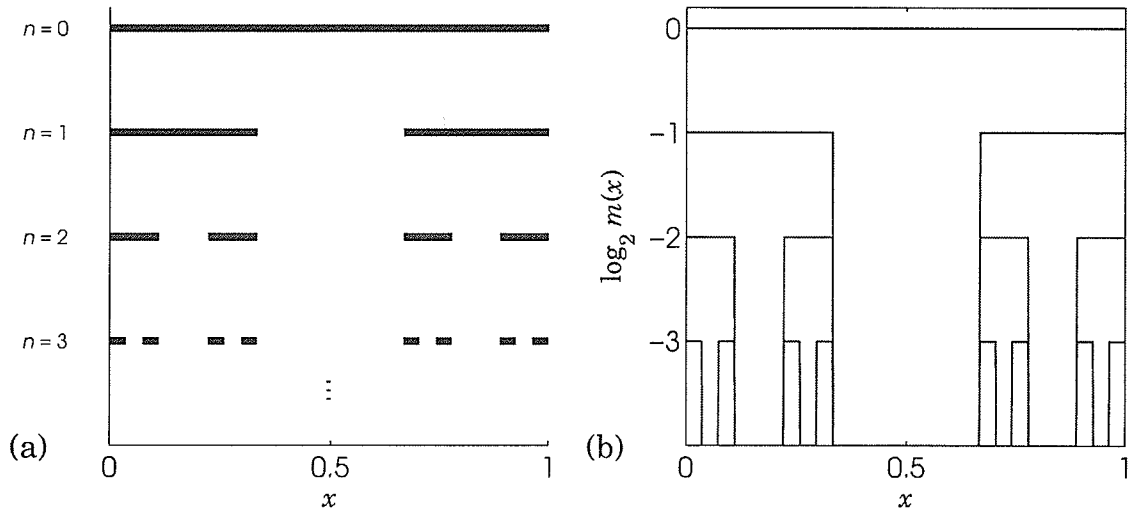


Fig. 4.7 The Cantor set: (a) the traditional "deleted middle thirds" construction is geometric, $C = \cap C_n$; (b) the measure-theoretic approach using uniform density measures, $C = \Omega_1(lim_{n\to\infty}\mu_{m_n})$.

characteristic functions (*i.e.*, uniform μ_n). It follows from this, that any example of a fractal set has an equivalent expression in a measure-theoretic terms. The *Cantor set* C is a famous traditional fractal often presented in introductory treatises on fractals [117]. This pathological set of Cantor (1883) is an example of a monofractal embedded in $\mathbb{R}$. Here the definition of C will be presented in both the set- and measure-theoretic terms.

Example 4.1 (Cantor Set: Set-Theoretic). Take the unit interval (*i.e.*, of length 1). Delete the open middle third from the set, leaving two closed equal lengths of 1/3. By recursively deleting the open middle thirds from the remaining segments of the set, Fig. 4.7(a), an uncountable number of points remain, each an accumulation point, but none of them an interior point. This sparse closed set has zero length. Note that at every stage, the number of segments double and are scaled to 1/3 the size of the previous stage.

Example 4.2 (Cantor Set: Measure-Theoretic). Begin with the characteristic function of the

unit interval, $\chi_{[0,1]}$. Define the generator Φ as the linear transformation

$$\Phi(g(x)) = \frac{1}{2} g\left(\frac{x}{3}\right) + \frac{1}{2} g\left(\frac{x+2}{3}\right) \tag{4.15}$$

This transformation has a characteristic 1/3 scaling in the right hand arguments and 2× replication (*i.e.*, g appears twice in a sum). By defining a sequence of integral measures μ_n from the mass density functions $m_n(x)$ that satisfy the cascade

$$m_{n+1}(x) = m_n(x) \cdot \Phi(m_n(x)) \tag{4.16}$$

one can define the limit density, $\mu_\Phi = \lim_{n\to\infty} \mu_n$. Several stages of mass density functions are shown in Fig. 4.7(b). The visual equivalence of both sides of Fig. 4.7 is clear: the sets on the left are the supports of the equivalent m_n on the right. Also, since the m_ns are uniform, they can be normalized to characteristic functions of the intervals on the left. The extra information of the density height on the right-hand side is useful, though. Note that the non-zero values of m_n is proportional to 2^{-n} (the height of Fig. 4.7(b) is in log-scale). If one tries to express the limit measure μ_Φ as an integral measure, $\lim_{n\to\infty} m_n = 0$, and the "measure" would be trivial and have empty support. This is a demonstration of the non-differentiability of μ_Φ. If one examines how the height of m_n scales with the length of the individual segments, the measure is assigning 1/2 of the previous mass to 1/3 of the length (Lebesgue measure). A rough calculation then shows that the ratio of the measures increases by 3/2 at every stage. As n goes to infinity, this ratio diverges and Dirac Delta functions begin appearing in the "derivative of μ_n with respect to Lebesgue measure". Though the m_ns converge to zero, their ratio to Lebesgue measure diverges, and so the true limit measure cannot be an integral measure. It is a fractal measure, the *uniform Cantor measure* whose support is the Cantor set C.

Example 4.3 (Non-uniform Cantor Measures). As stated in the definition, monofractals such as the Cantor set are special cases of the general formalism. A non-uniform Cantor measure

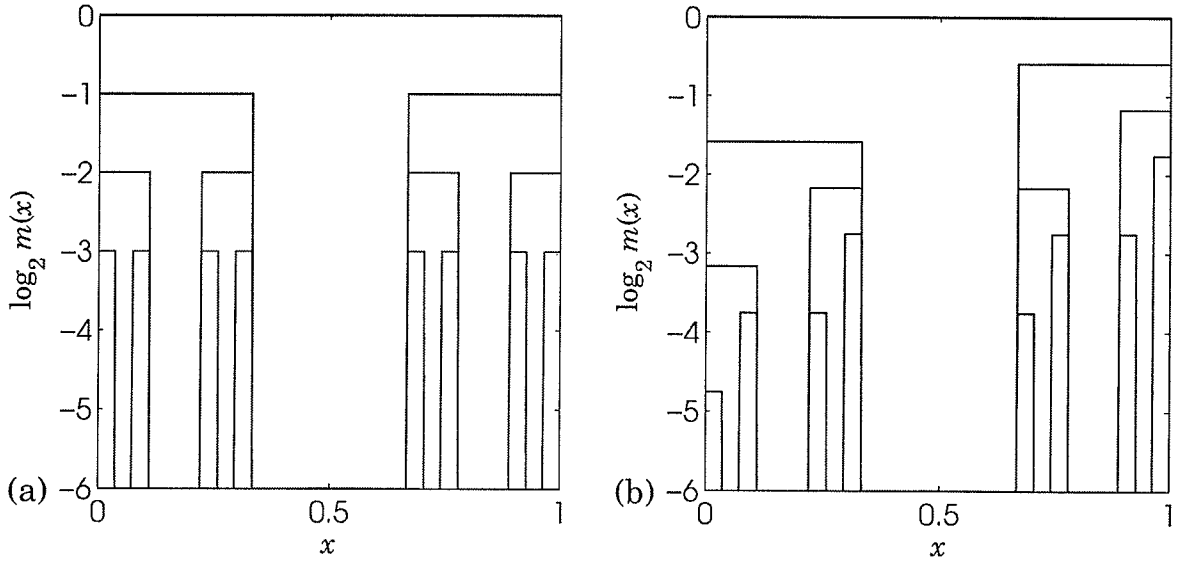


Fig. 4.8 Uniform and Non-uniform Cantor Measures: (a) the uniform Cantor measure cascade, $\varphi = [1/2, 0, 1/2]$; (b) the $\varphi = [1/3, 0, 2/3]$-weighted non-uniform Cantor measure cascade.

defined on the Cantor set can be defined by introducing inhomogeneities in the generator Φ. Consider now the generator

$$\Phi(g(x)) = \frac{1}{3} g\left(\frac{x}{3}\right) + \frac{2}{3} g\left(\frac{x+2}{3}\right) \tag{4.17}$$

Note that the weighting (*i.e.*, constant prefactor to g) of each replication is now different. This weighting can be described by the sequence of prefactors $\varphi = [1/3, 0, 2/3]$ if it is written formally as

$$\Phi(g(x)) = \sum_{i=1}^{\|\varphi\|} \varphi(i)\, g\left(\frac{x + (i-1)}{\|\varphi\|}\right) \tag{4.18}$$

where $\|\varphi\|$ is by definition the length (or number of elements) of φ, and $\sum_i |\varphi(i)| = 1$. In this formal sense, the uniform Cantor measure is defined by $\varphi = [1/2, 0, 1/2]$. Figure 4.8 shows the inhomogeneities in the measure cascade introduced by the non-uniform φ weightings.

Remark 4.1 (Fractal Terminology). This thesis uses the word "fractal" as a generic term of

regularized scaling, meaning multifractal and monofractal behaviour inclusively. In the case of a specific behaviour, it will be identified as multi- or monofractal as required.

This completes the introduction of measure theory and its inter-relationship with sets and density functions. The thesis will proceed by describing the characterization of measure fractality through the regularized scaling of the coarse-grained partition functions.

4.3 Multifractal Analysis: Scaling and Dimensionality

Multifractal analysis (MFA) is concerned with measuring the regularized complexity of a multifractal measure. It is *not* the *inverse fractal problem*, which tries to determine the fractal iteration process (Φ in (4.13)) from observation of the data. This section begins with the transition from a measure to coarse-grainings over scales. From these coarse-grainings, a critical exponent, called a *fractal dimension*, is defined. In later sections, these exponents are extended to the scaling properties of partition functions and entropies defined from the measure's coarse-graining. The end goal of this pedagogy is to present the two canonical expressions of MFA dimensions: the *Rényi fractal dimension spectrum* and the *Mandelbrot fractal dimension spectrum*. It will be assumed throughout this section that a measure is well-defined and fractal. However, as described in the beginning of this chapter, the MFA method can be extended to study the scaling behaviour of any empirical measure, multiplicative or not.

4.3.1 Dimensions and Coarse-Graining

4.3.1.1 Coverings and Coarse-Grainings

MFA originated with the mathematical study of pathological sets and functions. Here, the elementary concepts of MFA will be introduced with fractal sets, which inspires the original

ideas and terminology. For completeness, however, extensions to full measures are done where appropriate. To begin, some mathematical machinery is formalized.

Definition 4.6 (ϵ-Covering)**.** Let $A \subset \mathbb{R}^{N_e}$. An *$\epsilon$-covering*, $\mathcal{V}_\epsilon$, of A is an ensemble of open balls V_r of radius $r \leq \epsilon$ such that

$$A \subset \bigcup_{V_r \in \mathcal{V}_\epsilon} V_r \tag{4.19}$$

A *minimal ϵ-covering* is the ϵ-covering with the fewest members.

Hereafter, the r subscript on the members of an ϵ-covering will be dropped when convenient, and indexing over all the members $V \in \mathcal{V}_\epsilon$ will be written shorthand as indexing over $\mathcal{V}_\epsilon$.

Now given a measure μ with support Ω_1, and also given an ϵ-covering of Ω_1, it follows that $\mu(V)$ exists for every $V \in \mathcal{V}_\epsilon$. This helps the following be defined:

Definition 4.7 (Coarse-Graining)**.** Let μ be a measure on $\mathbb{R}^{N_e}$ and $A \subset \mathbb{R}^{N_e}$. Let $\mathcal{V}_\epsilon$ be an ϵ-covering of A. Then the *coarse-graining* of (μ, A) by $\mathcal{V}_\epsilon$ is the sum

$$\sum_{\mathcal{V}_\epsilon} \mu(V) \tag{4.20}$$

Thus by (4.19), the coarse-graining is an upper bound for $\mu(A)$ if $\mu \geq 0$.

Definitions 4.6 and 4.7 can be recapped by saying that an ϵ-covering is an object extracted from a set A, while a coarse-graining is an object extracted from a measure and a set.

4.3.1.2 Lebesgue Scaling and Fractional Dimensions

A coarse-graining is a mutual property of the measure μ, the set A and the covering $\mathcal{V}_\epsilon$. The Lebesgue measures (normal volume measures) that were introduced in the previous section satisfy the following proposition regarding the calculus of coarse-grainings:

Proposition 4.8 (Lebesgue Measure [117][70]). *Let $A \subset \mathbb{R}^{N_e}$. Let $\mathcal{V}_\epsilon$ be an ϵ-covering of A. The Lebesgue measure μ_L of A can be characterized as the limit lower bound of all coarse-grainings. That is,*

$$\mu_L(A) = \lim_{\epsilon \to 0} \inf \sum_{\mathcal{V}_\epsilon} \mu_L(V) \tag{4.21}$$

Consider now the importance of the dimension of the *embedding space*, N_e, to Lebesgue measure. This is done by defining the notion of local scaling.

Definition 4.9 (Local Scaling). A function $g : \mathbb{R} \to \mathbb{R}$ *locally scales as* ϵ^d *at* x_0, written $g(x)\big|_{x_0} \sim \epsilon^d$, if g is "proportional to ϵ^d in the limit of small ϵ". That is,

$$\lim_{\epsilon \to 0} \frac{g(x_0 + \epsilon) - a_1}{\epsilon^d} = a_2 \in (0, \infty) \text{ for some } a_1 \in \mathbb{R}. \tag{4.22}$$

This implies that, for $\epsilon > 0$, the local behaviour of $g(x_0 + \epsilon)$ and ϵ^d are proportional.

This definition implies that if g is differentiable at x_0, then $g\big|_{x_0} \sim \epsilon^d$ with $d \geq 1$, and $a_1 = g(x_0)$. Similarly, it is an often used property that all differentiable extrema are locally quadratic: *i.e.*, all differentiable g scale as ϵ^2 in the neighbourhood of an extremum x_0. However, local scaling is a relevant characteristic for non-differentiable g as well.

Consider now the local scaling of the function $g(r) = \mu_L(V_r)$. It is a property of the Lebesgue metrics that $\mu_L(V_r) \sim r^{N_e}$: that is, Lebesgue measure has a characteristic exponent equal to the embedding dimension (and this is irrespective of the metric being used measuring r). Examine now the scaling behaviour of coarse-grainings: that is, in the neighbourhood of the limit $\epsilon \to 0$. As before, let $\mathcal{V}_\epsilon$ be an ϵ-covering of a bounded set $A \subset \mathbb{R}^{N_e}$. There is a finite number of sets in the covering, which is identified here as $N_\epsilon(A)$. Now for each of the $N_\epsilon(A)$ sets in the covering,

$$\mu_L(V_r) \sim r^{N_e} \leq \epsilon^{N_e} \tag{4.23}$$

Applying this upper bound to the coarse-graining scaling above, one gets

$$\sum_{\mathcal{V}_\epsilon} r^{N_e} \leq \sum_{\mathcal{V}_\epsilon} \epsilon^{N_e} = N_\epsilon(A)\epsilon^{N_e} \tag{4.24}$$

Since

$$\mu_L(A) \sim \sum_{\mathcal{V}_\epsilon} \mu_L(V_r) \sim g(\epsilon) \tag{4.25}$$

Eq. (4.24) then provides an upper bound to the local scaling of the Lebesgue measure as

$$g(\epsilon) \leq N_\epsilon(A)\epsilon^{N_e} \tag{4.26}$$

That is, the scaling of $N_\epsilon(A)\epsilon^{N_e}$ is an upper bound to the scaling of $\mu_L(A)$. It is of particular interest that the bound in (4.26) has one factor, $N_\epsilon(A)$, which is characteristic of the set A, while the other, ϵ^{N_e}, is the characteristic scaling of only the embedding space. The balance between the scaling of these two factors contribute to the possible outcomes for $\mu_L(A)$. If the scaling of the geometric factor is written as

$$N_\epsilon(A) \sim \epsilon^{-d} \tag{4.27}$$

then if $d < N_e$, the upper bound product $N_\epsilon(A)\epsilon^{N_e} \sim \epsilon^{-d}\epsilon^{N_e} = \epsilon^{N_e-d}$ converges to zero, thus forcing $\mu_L(A)$ to zero as well. Equivalently, a *necessary condition* for the set A to have positive volume is that $N_\epsilon(A)$ must scale (and diverge) in (4.27) as $d \geq N_e$.

Since the scaling of the bounding factor $N_\epsilon(A)$ is characteristic of the set A, it raises the question, "Can the scaling of $N_\epsilon(A)$ be extracted as a geometrical feature of the set A?" Historically, this is the origin of fractal analysis by a *fractal dimension.*

A side effect of measuring the geometry of A by this scaling of $N_\epsilon(A)$ is that, by identifying the scaling exponent of $N_\epsilon(A)$ as the fractal dimension of A, fractal dimensions are no longer constrained like Lebesgue measures to be integer. *Fractional* dimensions occur, as can be

shown by analyzing the Cantor set. Consider the set construction of C by the intersection of deleted middle thirds, Ex. 4.1 and Fig. 4.7(a). At each stage n, one can identify a covering of C by its construction at that stage. Thus one can observe that at each stage there is 2^n intervals, such that $N_\epsilon(C) = 2^n$. Note, however, that the interval length at that stage is 3^{-n}, and so $\epsilon = 3^{-n}$. By eliminating n, it can be written

$$N_\epsilon(C) = 2^{\frac{\log \epsilon}{-\log 3}} = \epsilon^{-\log 2/\log 3}. \tag{4.28}$$

Thus, by comparing the above to (4.27), one can see the Cantor set has non-integer fractal dimension $d = \log 2/\log 3$.

The formalized method for fractional dimensions that follows is the result of Hausdorff's extension to Lebesgue measure, the Hausdorff measure and dimension of a set [70, Gouyet (1996)].

Definition 4.10 (Hausdorff(-Besicovitch) Measure of Fractal Sets)**.** Let $A \subset \mathbb{R}^{N_e}$. Let $\mathcal{V}_\epsilon$ be an ϵ-covering of A. The *Hausdorff measure* of dimension d, μ_{Hd}, is the set measure defined by:

$$\mu_{Hd}(A) = \lim_{\epsilon \to 0} \inf \sum_{V_r \in \mathcal{V}_\epsilon} r^d \tag{4.29}$$

The unique positive real number d making $\mu_{Hd}(A)$ finite but nonzero is called the *Hausdorff dimension* of the set A. In this way,

$$d = \sup \left\{ d' \,\middle|\, \mu_{Hd'}(A) = \infty \right\} \tag{4.30}$$

$$= \inf \left\{ d' \,\middle|\, \mu_{Hd'}(A) = 0 \right\} \tag{4.31}$$

The traditional fractal sets covered in fractal texts are examples of nontrivial Lebesgue measure zero sets embedded in $\mathbb{R}$ or $\mathbb{R}^2$.

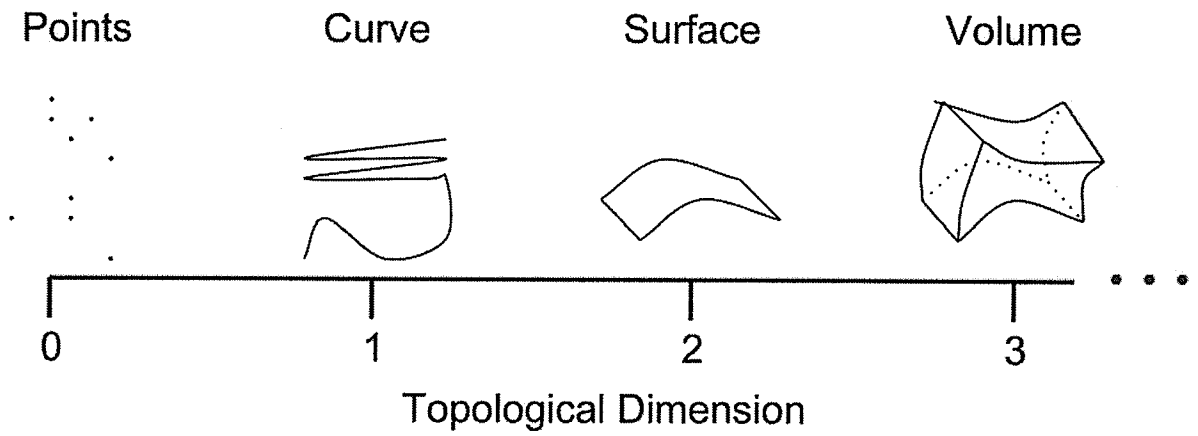


Fig. 4.9 The topological dimension of objects are integer dimensions describing their intrinsic manifold behaviour.

As can be seen from the definitions, the Hausdorff measure generalizes the Lebesgue measures by allowing the mass of the ϵ-balls to scale in a fractional way, and not under the Euclidean (integer) dimension of the embedding space, N_e. As an example, the Lebesgue mass of a simple curve (*e.g.*, line, circle) in $\mathbb{R}^2$ is necessarily zero because it has no area. These curves (as sets) have an intrinsic *topological dimension* irrespective of the embedding space, Fig. 4.9 [117, Kinsner (1994)]. In fact, any set with a topological dimension less than the embedding dimension will have zero Lebesgue measure (in the embedding space). By introducing the scaling as a *real-valued* parameter, d, things become more flexible, as shown in Fig. 4.10. The geometric dimension of the set is now intrinsic to the *set* and not the embedding space. It can now take non-integer values.

The important components of this analysis are the *coarse-graining* of the geometry by the ϵ-coverings, and the (Hausdorff) *critical exponent* d, which is the "natural" scaling exponent of the covering sets V_r. Coarse-graining is the key to measuring a fractal process at a given scale. By changing ϵ, it is as if the resolution, or zoom, of the fractal can be changed. This is the self-similar process of measurement discussed at the beginning of this chapter, as applied

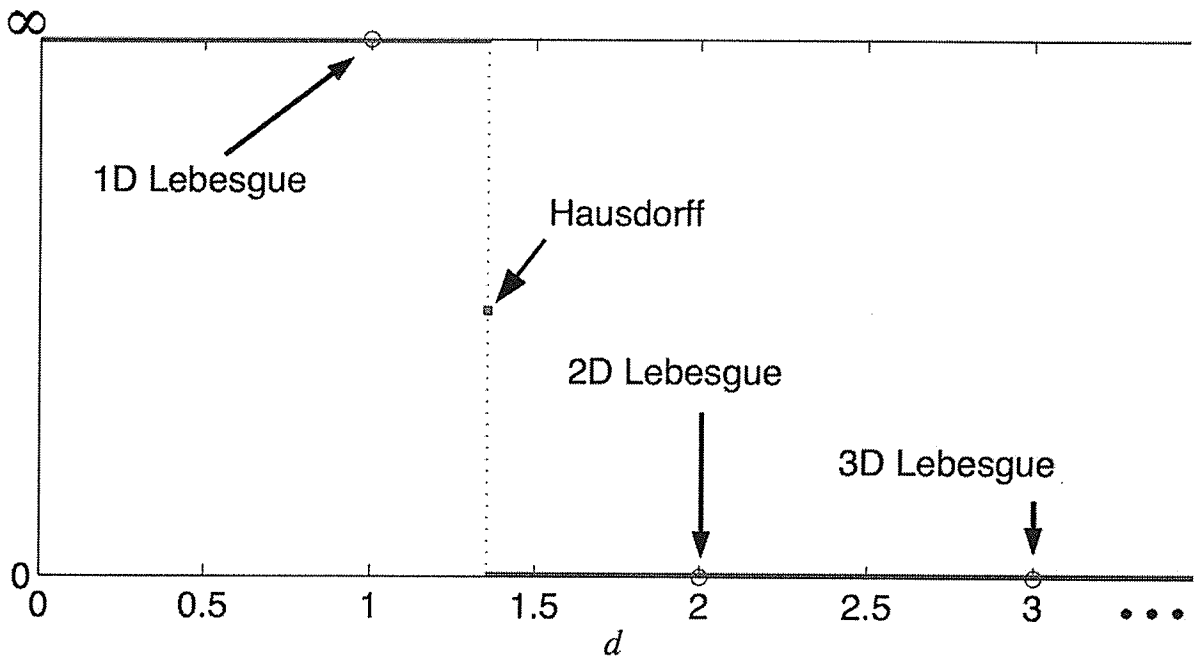


Fig. 4.10 The theoretical limit of the coarse-grained measures of a fractal set with varying dimension d, and its relationship to Lebesgue measures and integer dimensions.

to sets.

4.4 Multifractal Analysis: Scaling and Partition Functions

It has been shown how the intrinsic geometry of *sets* can be characterized by scaling through the Hausdorff dimension. This analysis by scaling should now be extended to measures — or, more precisely, to non-uniform measures — by the set and measure formalism that was introduced previously. Here, the goal is extend the coarse-graining methodology to define scaling partition functions and entropies suitable to characterize multifractal measures.

First, observe that the Hausdorff "critical-exponent" formalism (involving an inf or sup of a continuous interval of dimensions) is impossible in real circumstances. Instead it is more

common, following the logic of the Cantor example on p. 161, to express the Hausdorff dimension d of the set A as a ratio of the form

$$d(A) \approx -\frac{\log N_\epsilon(A)}{\log \epsilon} \tag{4.32}$$

For experimental data this would be expressed as the negative slope of two variables on a double-logarithmic plot. That is, a family of different ϵ_i are chosen, and for each ϵ_i, one ϵ-covering is expressed, from which $N_{\epsilon_i}(A)$ is calculated. If the points $(\log \epsilon_i, -\log N_{\epsilon_i}(A))$ follow a linear relationship, $d(A)$ can be estimated from the slope. This does not satisfy the formal definitions of scaling because it omits the infimum over many open coverings, and avoids the limit as $\epsilon \to 0$. However, the estimate (4.32) is tractable, and a linear relationship between $-\log N_\epsilon(A)$ and $\log \epsilon$ over a finite interval is sufficient to calculate $d(A)$, though not necessary for the Hausdorff dimension to exist. (For example, if $N_\epsilon(A) \sim \log \epsilon$, then A has Hausdorff dimension $d = 0$ but no power-law or linear scaling region in a log-log plot.) The estimate of (4.32), however, is not unbiased, and therefore it is often given the name of *box-counting dimension.* Falconer reports that the box-counting dimension is always greater than the Hausdorff dimension [56, Falconer (1990)]. In this work, the estimation of a dimension by (4.32) will be called the *box-counting formalism.*

Encapsulating the current discussion for the extension to the multifractal case, it is clear that:

(a) The Lebesgue measure of open balls $\mu_L(V_r)$ scale with integer dimension N_e;

(b) The Hausdorff dimension of a set exists as the critical exponent of the Hausdorff measures. Specifically, this captures the scaling behaviour of $N_\epsilon(A)$ by analyzing $\sum_{\mathcal{V}_\epsilon} r^d$. The dimension d may be non-integer;

(c) The box-counting dimension examines the scaling behaviour of $N_\epsilon(A) = \sum_{\mathcal{V}_\epsilon} |\mu(V_r)|^0$

directly. A linear relationship between $-\log N_\epsilon(A)$ and $\log \epsilon$ over a finite interval of scale is required.

In the above, the first item remarks on the scaling of a measure. The second item remarks on the scaling of a set. The final item represents a scaling as the slope of a linear relationship between a function of the $\{\mu(V_r)\}$ over the logarithm of scale, $\log \epsilon$. The following definition is a unification of these principles: a general notion of a *fractal dimension.*

Definition 4.11 (Arbitrary Fractal Dimension)**.** Let μ be a measure with support Ω_1. Let $\mathcal{V}_\epsilon$ be an ϵ-covering of Ω_1. Suppose g is a positive scalar functional on the coarse-grained measure. Then if g satisfies a *power-law* with respect to ϵ (that is, if

$$\log g(\{\mu(V_r)\}) \text{ and } \log \epsilon \tag{4.33}$$

have a linear relationship on a finite interval of scale), then the scaling exponent is the "g-dimension" of μ. That is,

$$d_g \approx \frac{\log g(\{\mu(V_r)\})}{\log \epsilon} \tag{4.34}$$

First it will be demonstrated how this definition conforms to the aforementioned examples.

(a) Lebesgue Measure: Here, μ is given as Lebesgue measure. Let $g(\{u_i\}) \stackrel{\text{def}}{=} \sum_i u_i$ (*i.e.*, g is a simple sum). Therefore $g(\{\mu_L(V_r)\}) = \sum_{\mathcal{V}_\epsilon} \mu_L(V_r)$. Then by (4.21), $g(\{\mu_L(V_r)\}) \sim \epsilon^{N_e}$, the desired dimension.

(b) Hausdorff Measure: Here, μ is given as the d'-dimensional Hausdorff measure. As above, let g be a simple sum. Then $g(\{\mu_{H_{d'}}(V_r)\}) = \sum_{\mathcal{V}_\epsilon} \mu_{H_{d'}}(V_r)$ scales as d'. This satisfies the definition of fractal dimension used here. Note: The Hausdorff dimension *of a set* is well-defined because *only one* $d \in \{d'\}$ results in g converging to a finite and non-zero limit.

(c) Box-Counting Dimension: Here, μ is arbitrary. Let $g(\{u_i\}) \stackrel{\text{def}}{=} (\sum_i (u_i)^0)^{-1}$, where $0^0 = 0$ by assumption, so the sum behaves as a counting function. Then $g(\{\mu_L(V_r)\}) = (N_\epsilon(A))^{-1}$ and satisfies the definition.

There exist many other dimensions for the characterization of fractals in the literature (Gouyet (1996) [70], for example). These are seen to conform with the general principle of Def. 4.11 by the appropriate choice for the function g.

What is special to note is the role of g in taking the finitely many real values $\{u_i = \mu(V_r)\}$ and mapping them to a positive real number. This real number does not necessarily characterize an entire measure, however. For if one considers two measures μ_1 and μ_2 that have identical support Ω_1, and then calculate the box-counting dimension of these measures, one should observe that their box-counts $N_\epsilon(A)$ are identical for every ϵ. Since the values of the measures $\{u_i = \mu(V_r)\}$ do not appear, only their counts, they will have the same dimension. This degeneracy is clear from the presented examples of uniform and non-uniform Cantor measures, Ex. 4.2 and 4.3. This observation is extended now into a formal proposition.

Proposition 4.12 (Limitations of Box-Counting Dimension)**.** *The box-counting dimension of a measure μ depends only on the support of the measure, $\Omega_1(\mu)$. Thus all measures with the same support have the same box-counting dimension.*

To characterize a measure by its scaling, then, it is clear that many dimensions are needed [117, Kinsner (1994)]. As mentioned above, the main property of g is that it maps the finitely many $\{u_i = \mu(V_r)\}$ of a given coarse-graining into a positive real value

$$g : \{u_i\} \mapsto \mathbb{R}^+ \tag{4.35}$$

Instead, if g were extended to map into a function (*i.e.*, an infinite family of values) depending on some parameter, an entire family of dimensions could be extracted from a power-law

analysis such as (4.34). Now enters the thermodynamical formalism to multifractal analysis.

Definition 4.13 (Partition Function)**.** Let μ be a measure with support Ω_1. Let $\mathcal{V}_\epsilon$ be an ϵ-covering of Ω_1. Then define a positive functional $g : l^0 \mapsto C^2(\mathbb{R})$; that is which takes the finite sequence $\{u_i = \mu(V_r)\}$ to a differentiable function $g(\{u_i\}, q) \geq 0$ over the free parameter q. Then the *partition function* of a coarse-graining of μ is

$$Z_g(\mu, q \mid \mathcal{V}_\epsilon) = g(\{\mu(V_r)\}, q) \tag{4.36}$$

As an example of the partition function, consider a predominant model in classical theory, the Boltzmann partition.

Example 4.4 (Boltzmann Partition Function [239]). The *Boltzmann partition function* of μ is

$$Z_B(\mu, q \mid \mathcal{V}_\epsilon) = \sum_{\mathcal{V}_\epsilon} |\mu(V_r)|^q \tag{4.37}$$

That is, $g(\{u_i\}, q) = \sum_i |u_i|^q$. This is classically defined for all $q \geq 0$, but may be extended to $q < 0$ as desired.

From the above example, one can observe that the box-count $N_\epsilon(\Omega_1)$ of a coarse-grained measure μ with support Ω_1 is a special case of the Boltzmann partition:

$$N_\epsilon(\Omega_1) = Z_B(\mu, q = 0 \mid \mathcal{V}_\epsilon) \tag{4.38}$$

Remark 4.2 (Interpretation of the Boltzmann Partition). A coarse-grained Boltzmann partition $Z_B(\mu, q \mid \mathcal{V}_\epsilon)$ can be interpreted in analogy to digital images. The choice of ϵ changes the *resolution* of the measure, in much the same effect as pixel size in a digital image. The smaller the choice of ϵ, the finer the partition becomes. The q parameter is a non-linear effect on the partition that behaves like image *contrast*. It controls how the different values of the pixel elements contribute to the partition sum. For $q = 0$, all elements contribute equally, regardless

of individual intensity, like a black-and-white image. For large $q >> 0$, the largest values dominate. While for $q << 0$, the smallest non-zero values dominate. In this way, q serves to enhance and quantify the inhomogeneities in the measure.

The power of a partition function is that it enables an extension to Def. 4.11 to intrinsic families of fractal dimensions, through the following proposition.

Proposition 4.14 (Partition Function Scaling). *Let μ be a measure with support Ω_1. Let $\mathcal{V}_\epsilon$ be an ϵ-covering of Ω_1. Suppose $Z_g(\mu, q \mid \mathcal{V}_\epsilon)$ is a partition function on the coarse-grained measure. Then if $Z_g(\mu, q \mid \mathcal{V}_\epsilon)$ satisfies a power-law with respect to ϵ for every fixed q (that is, if $\log Z_g(\{\mu(V_r)\}, q)$ and $\log \epsilon$ have a linear relationship on a finite interval of scale), then the scaling exponent $\tau(q)$ is the "qth multifractal dimension" of $Z_g(\mu, q)$. That is a family of dimensions*

$$\tau(q) \approx \frac{\log Z_g(\{\mu(V_r)\}, q)}{\log \epsilon} \tag{4.39}$$

characterize the partition function $Z_g(\mu, q)$, since

$$Z_g(\mu, q \mid \mathcal{V}_\epsilon) \sim \epsilon^{\tau(q)} \tag{4.40}$$

In particular, one finds for a Boltzmann partition function,

$$\tau(q) \approx \frac{\log \sum_{\mathcal{V}_\epsilon} |\mu(V_r)|^q}{\log \epsilon} \tag{4.41}$$

By characterizing the partition function of μ, therefore, a level of characterization of μ greater than a single fractal dimension is achieved, capturing the influence of the non-uniformity. This, by definition, is *multifractal analysis* (MFA). The scaling properties of the measure is now expressed by the functional relationship of $(q, \tau(q))$, derived from the choice of partition function, $Z_g(\mu, q)$.

Now, an example of $\tau(q)$ will be evaluated.

Example 4.5 (Cantor set MFA). Consider the weighted Cantor measure defined by the normalized finite vector φ. As before, the generator of the measure μ_Φ is

$$\Phi(g(x)) = \sum_{i=1}^{\|\varphi\|} \varphi(i)\, g\left(\frac{x+(i-1)}{\|\varphi\|}\right) \tag{4.42}$$

where $\|\varphi\|$ is by definition the length (or number of elements) of φ, and $\sum_i |\varphi(i)| = 1$. Therefore the cascading mass density functions are the

$$m_{n+1}(x) = m_n(x)\,\Phi(m_n(x)) = \Phi \circ \overset{n}{\cdots} \circ \Phi(x) \tag{4.43}$$

an example of which is shown in Fig. 4.11(a). To calculate the MFA of this general form, it is assumed for convenience (as is always done in the literature examples [70]) that the ϵ-coverings $\mathcal{V}_\epsilon$ matches the set of natural intervals of length $\epsilon = \|\varphi\|^{-n}$ from the cascade to some level n. By construction, then, all the points in a covering set, $x' \in V_{r(n)}$, begin with the same real expansions $(.\xi_1\xi_2 \ldots \xi_n)$, after which they differ. Also by construction, it then follows that the value of $m_n(x')$ is constant over $V_{r(n)}$, such that

$$\mu(V_{r(n)}) = \int_{V_{r(n)}} m_n(x')\,dx' = m_n(x') \int_{V_{r(n)}} dx' = m_n(x')\mu_L(V_{r(n)}) \tag{4.44}$$

$$= m_n(x')\,\|\varphi\|^{-n} \tag{4.45}$$

By the cascading construction of (4.42), $m_n(x') = \prod_{i=1}^{n} \varphi(\xi_i)$, and consequently the equation above reduces to

$$\mu(V_{r(n)}) = \left(\prod_{j=1}^{n} \varphi(\xi_j)\right) \|\varphi\|^{-n} \tag{4.46}$$

which takes the "product of fractions" form

$$= \prod_{j=1}^{n} \left(\varphi(\xi_j)/\,\|\varphi\|\right) \tag{4.47}$$

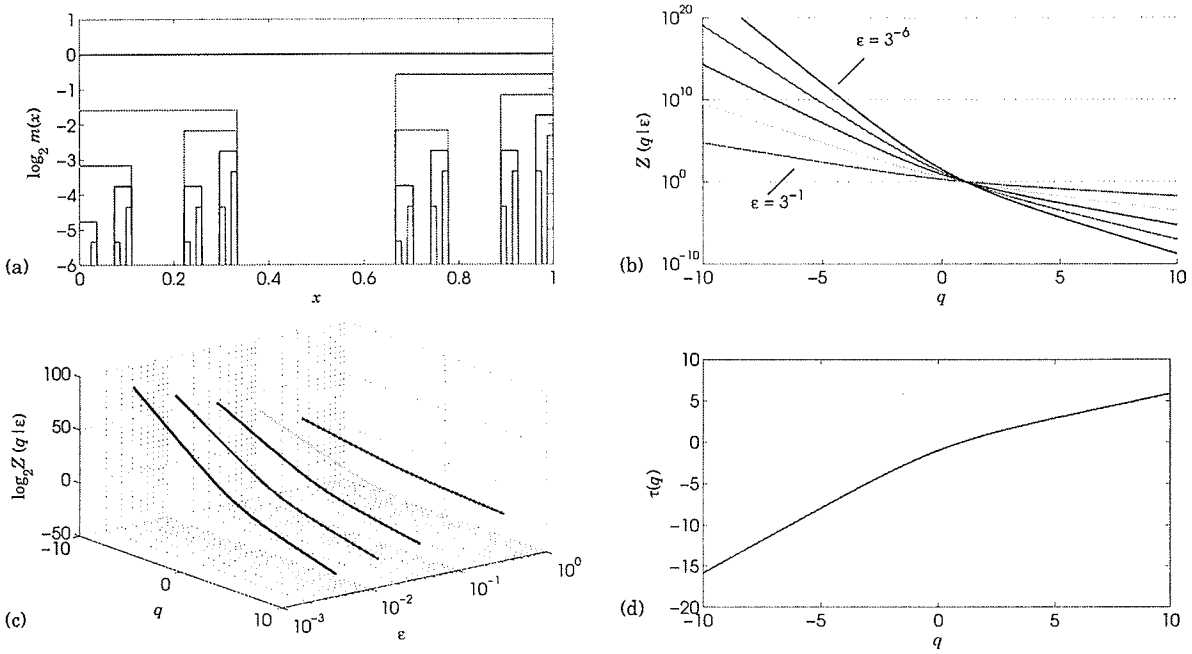


Fig. 4.11 The calculation of $\tau(q)$ for the non-uniform Cantor measure, $\varphi = [1/3, 0, 2/3]$: (a) the cascading m_n; (b) Boltzmann partition function for several ϵ_i; (c) $\log Z_g(\mu, q \mid \mathcal{V}_{\epsilon_i})$ vs $\log \epsilon_i$; (d) $\tau(q)$ vs. q.

That is, the measure of the ϵ-covering set is determined by its "position" in the Cantor set cascade, as indexed by the integers $\xi_1, \xi_2, \ldots, \xi_n$ in the common radix expansion of the set members.

To establish the MFA of the Cantor measure, the partition function is then evaluated as

$$Z_B(\mu_\varphi, q \mid \mathcal{V}_\epsilon) = \sum_{\mathcal{V}_\epsilon} |\mu(V_r)|^q = \sum_{\xi_1} \cdots \sum_{\xi_n} \prod_{j=1}^{n} \left(\frac{\varphi(\xi_j)}{\|\varphi\|} \right)^q \tag{4.48}$$

which can still undergo a combinatorial simplification. This simplification arises because (4.47) is dependent on the product of a certain set of $\varphi(i)$, and not their ordering. Specifically, the factor $\varphi(i)$ contributes the same no matter from which scale (*i.e.*, which ξ_j) it originated. Thus at any finite level n there is a finite amount of combinations that are achieved. This can be observed by the few recurring values assumed by $m_n(x)$ in the interior "columns" of

Fig. 4.11(a). If the multiple sum in (4.48) is collected over the common values, then the sum over the entire covering is, in fact, summing over all orderings to give the closed form

$$Z_B(\mu_\varphi, q \mid \mathcal{V}_\epsilon) = \sum_{\sum_i c_i = n} \frac{n!}{c_1!c_2!\cdots c_n!} \|\varphi\|^{-qn} \prod_{i=1}^{n} \varphi(i)^{qc_i} \tag{4.49}$$

This new sum is one over all "algebraic decompositions of n as a sum of n parts". Specifically, all possible combinations of $\{c_1, \ldots, c_n\} \in \mathbb{Z}_n^n$ such that $\sum_{i=1}^n c_i = n$

The exponential form of $Z_B(\mu_\varphi, q \mid \mathcal{V}_\epsilon)$ is shown in Fig. 4.11(b) for several choices of ϵ. Now fixing q, the behaviour of $\log Z_B(\mu_\varphi, q \mid \mathcal{V}_\epsilon)$ over $\log \epsilon$ is shown in Fig. 4.11(c). Here, the scale variable $\log \epsilon$ is equivalent to $-n \log \|\varphi\|$ in (4.49). The scaling exponent $\tau(q)$ can be extracted by fixing q and computing the slope across $\log \epsilon$. This derivative (for fixed q) is linear and dominated by Stirling's approximation for large factorials. The resulting $\tau(q)$ as a function of q is finally shown in Fig. 4.11(d). Even in the general case, $\tau(q)$ is a monotonic function [75].

Remark 4.3. Note that the seminal paper on this partition formalism [75, Halsey *et al.*, (1986)] uses the same "inferred critical exponent" method of the Hausdorff dimension formalism. A simpler presentation of the same is also shown on page 36 of Gouyet [70, (1996)]. Here, however, the scaling of the partition function has been expressed more directly in analog to the "box-counting" formalism, since this is more useful in practice. The main point to keep in mind is the sequence of analysis:

(a) Begin with a given measure μ;

(b) Extract the support, Ω_1;

(c) For a sequence of ϵ_j, coarse-grain the measure with an ϵ-covering, $\{\mathcal{V}_{\epsilon_j}\}$: the finite sequence $\{u_i = \mu(V_r)\}_j$ is calculated;

(d) A partition function then expresses each coarse-graining as a smooth functional $Z_g(\mu, q \mid \mathcal{V}_{\epsilon_j})$;

(e) The scaling of the partition function can be characterized by a set of fractal dimensions estimated by the box-counting formalism ($\approx$) and assigned to μ.

4.5 Multifractal Analysis: Partition Entropies and Thermodynamic Transformations

In the preceding section, an MFA of a fractal measure was defined in terms of the scaling exponents of a partition function defined from the coarse-graining of the measure. It is a property of the thermodynamics of partition functions, however, that many different and relevant, but dependent, variables can be extracted from the partition function. By using this change of coordinate system, the scaling properties of μ can be expressed in any of several convenient forms. In fact, the previous expression using the $(q, \tau(q))$-pairs is one of the least convenient since it unbounded along both axes. Among these other thermodynamically conjugate variables are

(a) The Rényi generalized dimensions, D_q;

(b) The Hölder exponents, α; and

(c) The Mandelbrot entropy, $D_0(\alpha)$,

which are defined and used in the coming sections.

The end goal of the following subsections is to formalize the two canonical expressions of MFA spectra: the *Rényi fractal dimension spectrum* and the *Mandelbrot fractal dimension spectrum.* It will be assumed throughout this section that a measure is well-defined and fractal.

However, as described in the beginning of this chapter, the MFA method can be extended to study the scaling behaviour of any empirical measure, multiplicative or not.

4.5.1 The Rényi Spectrum

As shown in the previous subsection, MFA is the characterization of a multifractal measure by the intrinsic scaling of a partition function as a family of fractal dimensions. This family of dimensions is defined through the behaviour of coarse-graining and a partition function analysis. From Ex. 4.5, it is clear that $\tau(q)$ vs. q can be very simple (*i.e.*, linear), but unbounded along both axes. In that sense, the $(q, \tau(q))$ "spectrum" is poorly behaved, because it is not compact. The Rényi spectrum is an entropy-based manipulation of the Boltzmann partition function to create compaction along one axis. Historically, this method by Hentschel and Procaccia [85, (1983)] actually predates the thermodynamic formalism. It is the first multifractal characterization that unified many of the previously defined fractal dimensions into a well-defined smooth family.

To develop the Rényi spectrum, recall the Rényi generalized entropy, (3.42) from p. 95, which is presented again here in discrete form

$$H_q(\underline{\boldsymbol{u}}) = \frac{1}{1-q} \log \sum_i u_i^q \tag{4.50}$$

Identifying the probability $u_i = |\mu(V_r)|$ for comparison to the Boltzmann partition function, (4.37), one gets,

$$H_q(\mu, \mathcal{V}_\epsilon) = -\frac{1}{q-1} \log Z_B(\mu, q \mid \mathcal{V}_\epsilon) \tag{4.51}$$

or

$$\log Z_B(\mu, q \mid \mathcal{V}_\epsilon) = -(q-1)\, H_q(\mu, \mathcal{V}_\epsilon) \tag{4.52}$$

Now, by applying the definition of arbitrary fractal dimensions, Def. 4.11, the following one-parametric family of dimensions is defined:

Definition 4.15 (Rényi generalized dimensions, D_q)**.** The *Rényi generalized dimensions* are the scaling exponents (*i.e.*, fractal dimensions) D_q defined from the scaling of e^{-H_q}; that is,

$$\exp\left(-H_q(\mu, \mathcal{V}_\epsilon)\right) \sim \epsilon^{D_q} \tag{4.53}$$

In the box-counting formalism, this is expressed as the slope of $\left(\log \epsilon_i, -H_q(\mu, \mathcal{V}_{\epsilon_i})\right)$ pairs over a linear region. In the notation used here, this is expressed by

$$D_q \approx \frac{-H_q(\mu, \mathcal{V}_\epsilon)}{\log \epsilon} \tag{4.54}$$

(Note that the exponential in the definition suppresses the log on the vertical axis/numerator.)

This definition for Rényi generalized dimensions provides the first canonical MFA spectrum.

Definition 4.16 (Rényi Fractal Dimension Spectrum)**.** Let μ be a measure with support Ω_1. Let $\mathcal{V}_\epsilon$ be an ϵ-covering of Ω_1. Then the *Rényi spectrum* (RS) of μ is the function $D_q : \mathbb{R} \mapsto \mathbb{R}$ matching the Rényi generalized entropies to the partition function variable q: D_q vs. q. The variable D_q is compact, bounded between,

$$D_{-\infty} = \left(\lim_{q \to -\infty} D_q\right) \geq D_q \text{ , and} \tag{4.55}$$

$$D_{\infty} = \left(\lim_{q \to \infty} D_q\right) \leq D_q \tag{4.56}$$

and is monotonic decreasing in q. These dimensions are related to the Boltzmann critical exponent $\tau(q)$ as

$$\tau(q) = -D_q(1 - q) \tag{4.57}$$

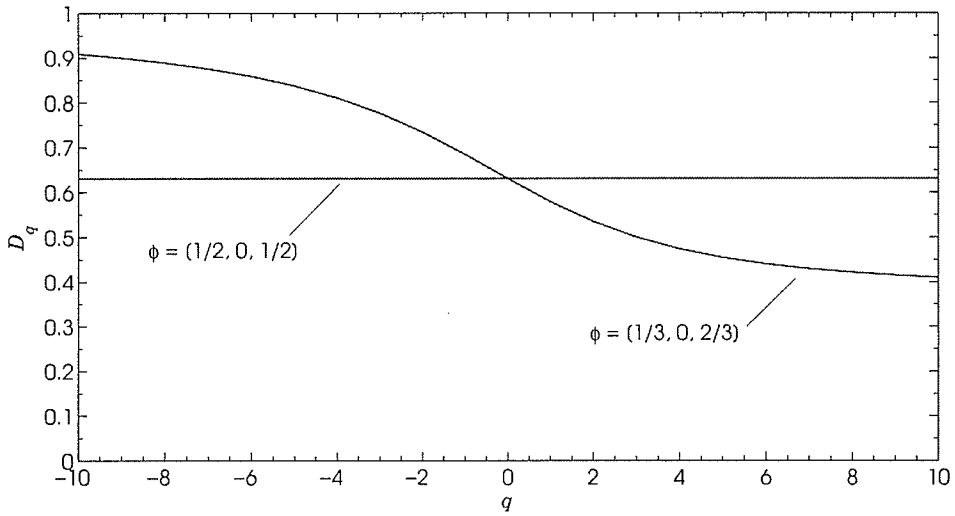


Fig. 4.12 Some analytic Rényi spectra. The monofractal (horizontal line) is the uniform Cantor measure ($d\mu = \chi_C\, dx$), and the curve is the $[1/3, 0, 2/3]$-weighted Cantor measure. They match at the box-counting dimension, ($q = 0$).

thus replacing $\tau(q)$ with a compact variable. In the case of uniform measures, such as the Cantor set, the RS reduces to a single scaling, $D_q = D_0$, equal to the box-counting dimension [117, Kinsner (1994)][70, Gouyet (1996)].

The RS represents the "scaling spectrum of entropies". Examples of monofractal and multifractal Rényi spectra are shown in Fig. 4.12. The RS has been a highly used and effective tool within the University of Manitoba Signal and Data Compression research group.

4.5.2 Changing Variables: Singularities and Hölder Entropies

As mentioned, the goal of reformulating the scaling information of a measure's coarse-graining partition function is to present a more compact form than $\tau(q)$ vs. q. The Rényi spectrum achieves compaction in one variable by transforming $\tau(q)$ into D_q. The goal of the Mandelbrot spectrum is to achieve a completely bounded representation (*i.e.*, along both axes). The transformation to the Mandelbrot spectrum is governed by the common thermodynamical

relationships of partition functions: namely, the Legendre transform [70][10].

The goal of this current section is to define α_Z as the Legendre transform equivalent of q (*i.e.*, the domain variable), determine a form of generalized entropy that can be used to calculate α_Z directly, and provide it an appropriate fractal (*i.e.*, scaling) interpretation.

From the definition of being "globally conjugate to q" in the Boltzmann partition formalism, the variable α_Z is defined, for given q, as [70, Gouyet (1996)]

$$\alpha_Z = \frac{\partial}{\partial q}\tau(q) \tag{4.58}$$

$$= -\frac{\partial}{\partial q}D_q(1-q) \tag{4.59}$$

Theoretically speaking, this process is a change of variable $\alpha_Z = \alpha_Z(q)$ applied to the $\tau(q)$ vs. q relationship after the scaling limit to small ϵ has been taken. A *direct* approach, however, would apply the derivative change of variable before the limit. To derive such a form, consider the partition scaling $Z_B \sim \epsilon^{\tau(q)}$, and by taking logarithms and invoking the box-counting ratio, one can write

$$\alpha_Z(q) \approx \frac{1}{\log \epsilon}\frac{\partial}{\partial q}\log Z_B(\mu, q \mid \mathcal{V}_\epsilon) \tag{4.60}$$

Now, recall that the direct form for the Rényi spectrum derived in the previous section is

$$D_q \approx -\frac{1}{\log \epsilon}H_q(\epsilon) \tag{4.61}$$

where the Rényi generalized entropies of the Boltzmann partition function are $H_q(\epsilon) = -(q-1)^{-1}\log Z_B$. (Here the shorthand $Z_B = Z_B(\mu, q \mid \epsilon)$ is introduced to save space, and will be used consistently throughout the remainder of the chapter.) A similar representation can be designed into the new formalism if it is written that

$$\alpha_Z(q) \approx -\frac{1}{\log \epsilon}\gamma_q(\mu, q \mid \mathcal{V}_\epsilon) \tag{4.62}$$

where $\gamma_q(\mu, q \mid \mathcal{V}_\epsilon)$ is defined as below.

Definition 4.17 (Hölder Generalized Entropies, γ_q)**.** Let a partition $Z(\mu, q \mid \epsilon)$ be the result of a coarse-graining of a measure μ. Let q be an arbitrary real number. Then the *Hölder generalized entropy* of order q is the scalar value

$$\gamma_q(\mu, q \mid \epsilon) \stackrel{\text{def}}{=} -\frac{\partial}{\partial q} \log Z(\mu, q \mid \epsilon) \tag{4.63}$$

For a Boltzmann partition Z_B, this can be calculated by passing the derivative through (4.37) as

$$\gamma_q(\mu, q \mid \epsilon) = \sum_{\mathcal{V}_\epsilon} \frac{|\mu(V_r)|^q}{\sum_{\mathcal{V}_\epsilon} |\mu(V_r)|^q} \log |\mu(V_r)| \tag{4.64}$$

Units of γ_q are measured by the base of the logarithm, just as with Shannon or Rényi generalized entropies. Thus the use of $\log_2$ will imply the units of *bits*.

The express purpose of the Hölder generalized entropies is to provide an "entropy" for the box-counting formalism in keeping with the pattern of the Rényi spectrum formula. The use of the term entropy here is very loose and for convenience only. (The consistent feature of these "entropies" is they are functions of $\log Z_B$.) Now, to help determine the fractal interpretation of both α_Z and γ_q, a special case of local scaling in a measure is introduced: *local measure singularity*, or Hölder exponents.

Definition 4.18 (Measure Singularity, α [70])**.** A measure μ has a local *singularity* of strength α at the point $\boldsymbol{x}$ if the mass of the local neighborhood V_ϵ scales with exponent $\alpha(\boldsymbol{x})$, *i.e.*,

$$|\mu(V_\epsilon \ni \boldsymbol{x})| \sim \epsilon^{\alpha(x)} \tag{4.65}$$

The exponent $\alpha(\boldsymbol{x})$ is called the *Hölder exponent* of μ at $\boldsymbol{x}$.

As can be seen by comparing (4.65) with (4.34), the Hölder exponent at $\boldsymbol{x}$ is essentially a fractal dimension of the set $A = \{\boldsymbol{x}\}$ under the measure μ and the partition function $g(u) =$

$|u|$. In this sense the dimension α is *local* to the point $\boldsymbol{x}$ and contains no information about other points. Note, however, that for all Lebesgue measures μ_L, $\alpha(\boldsymbol{x})$ would be the integer constant N_e everywhere. The Hölder exponent is a quantitative interpretation of the Lebesgue differentiability of μ at the point $\boldsymbol{x}$. If $\alpha(\boldsymbol{x}) > N_e$, then μ is Lebesgue differentiable at $\boldsymbol{x}$. It is clear that any measure will have a family of singularities $\{\, \alpha(\boldsymbol{x}) \mid \boldsymbol{x} \in \Omega_1 \,\}$ (though $\alpha(\boldsymbol{x})$ may be unbounded). It is a property of truly multiplicative processes that $\{\alpha\}$ is in fact an interval in $\mathbb{R}$ [75, Halsey *et al.*, (1986)].

From the definition of Hölder singularities, it is now possible to derive an interpretation for α_Z and γ_q.

Proposition 4.19. *The thermodynamic conjugate to q, α_Z, is a* weighted average of the set of Hölder exponents *achieved by μ.*

Proof. Substituting (4.64) into (4.62), one then arrives at

$$\alpha_Z(q) \approx \frac{1}{\log \epsilon} \sum_{\mathcal{V}_\epsilon} \frac{|\mu(V_r)|^q}{\sum_{\mathcal{V}_\epsilon} |\mu(V_r)|^q} \log |\mu(V_r)| \tag{4.66}$$

which when rearranged becomes

$$\approx \sum_{\mathcal{V}_\epsilon} \frac{|\mu(V_r)|^q}{\sum_{\mathcal{V}_\epsilon} |\mu(V_r)|^q} \frac{\log |\mu(V_r)|}{\log \epsilon} \tag{4.67}$$

$$\approx \sum_{\mathcal{V}_\epsilon} \frac{|\mu(V_r)|^q}{\sum_{\mathcal{V}_\epsilon} |\mu(V_r)|^q} \alpha(\boldsymbol{x}) \tag{4.68}$$

□

Since, for a given q, the variable α_Z has the interpretation of a *global average* Hölder singularity, the subscript is dropped hereafter. Note that just as the value of q shifted the sensitivity of the Rényi spectrum from small to large probabilities, here by ranging over all

q, the weighting function shifts from the most singular to the least singular. Furthermore, the extreme conditions

$$\max_q \alpha(q) = \alpha(-\infty) \tag{4.69}$$

$$\min_q \alpha(q) = \alpha(+\infty) \tag{4.70}$$

are satisfied.

The global Hölder exponent α is useful for the reparameterization of the Rényi spectrum because (in useful cases) the extreme values above are bounded. Note that the lower bound $\alpha(+\infty)$ is effectively greater than $-N_e$, which is the Hölder exponent of the Dirac singularity in $\mathbb{R}^{N_e}$. In practical cases, the upper bound will also be finite $\alpha(-\infty) < \infty$, though Mandelbrot has demonstrated the limitation of the Halsey *et al.* Legendre formalism by designing a self-similar measure with $\alpha(0) = \infty$ and with τ and α undefined for $q < 0$ [134, Mandelbrot (1990)]. (This was done in criticism of reserving the term "multifractal" for only the Halsey *et al.* (*i.e.*, restricted) sense.) For the purposes of this thesis, however, these pathological conflicts are irrelevant. Just as the box-counting formalism is sufficient for the characterization of a fractal dimension, the "coarse-graining Boltzmann-partition" formalism is sufficient (but perhaps not necessary) for the characterization of a multifractal.

4.5.3 The Mandelbrot Spectrum

The final requirement to compactify an MFA is to derive a thermodynamical conjugate variable to $\tau(q)$. This new conjugate variable is, at once, both a new and old idea. As with the Hölder singularities, the conjugate variable is obtained by a Legendre transform, and hence

$$D_0 = q\,\alpha - \tau(q) \tag{4.71}$$

$$= q\,\frac{\partial \tau(q)}{\partial q} - \tau(q) \tag{4.72}$$

Note that the parameterization of D_0 is most commonly expressed in terms of $\alpha = \alpha(q)$ and so is written

$$D_0 = D_0(\alpha) \tag{4.73}$$

for strong reasons soon to be explained. Parametrically, however, D_0 can also be a function of q.

The power of this "new" variable $D_0(\alpha)$ is its interpretation in the multifractal formalism. It is, in fact, not a new feature requiring a new interpretation, but simply a Hausdorff dimension of a very particular subset of Ω_1 [75, Halsey *et al.* (1986)]. These special subsets are the subject of the next definition.

Definition 4.20 ((Hölder) Equisingular Sets)**.** Let μ be a measure with support $\Omega_1 \subset \mathbb{R}^{N_e}$. Let $\alpha \in \mathbb{R}$ be a Hölder singularity value of μ. Then the (Hölder) *equisingular subset*, A_α, is the set of all points in Ω_1 with Hölder exponent equal to α: *i.e.*,

$$A_\alpha = \{\, \boldsymbol{x} \in \Omega_1 \mid \alpha(\boldsymbol{x}) = \alpha \,\} \tag{4.74}$$

By the details of its construction [75] in the multifractal formalism, then, D_0 is in fact the scaling exponent of $(N_\epsilon(A_\alpha))^{-1}$: *i.e.*, the box-counting dimension of the equisingular set A_α. Since $D_0(\alpha)$ is a Hausdorff dimension, no new symbol is introduced and the Hausdorff dimension symbol D_0 (consistent with the Rényi dimensions, $D_0 = D_q\big|_{q=0}$), is reused. This rather involuted relationship of D_0 and A_α is shown in Fig. 4.13.

Recall that the objective here is to replace $\tau(q)$ with a conjugate bounded variable. The variable D_0 is therefore suitable. For, as a Hausdorff dimension, D_0 is positive, so that $D_0 \geq 0$.

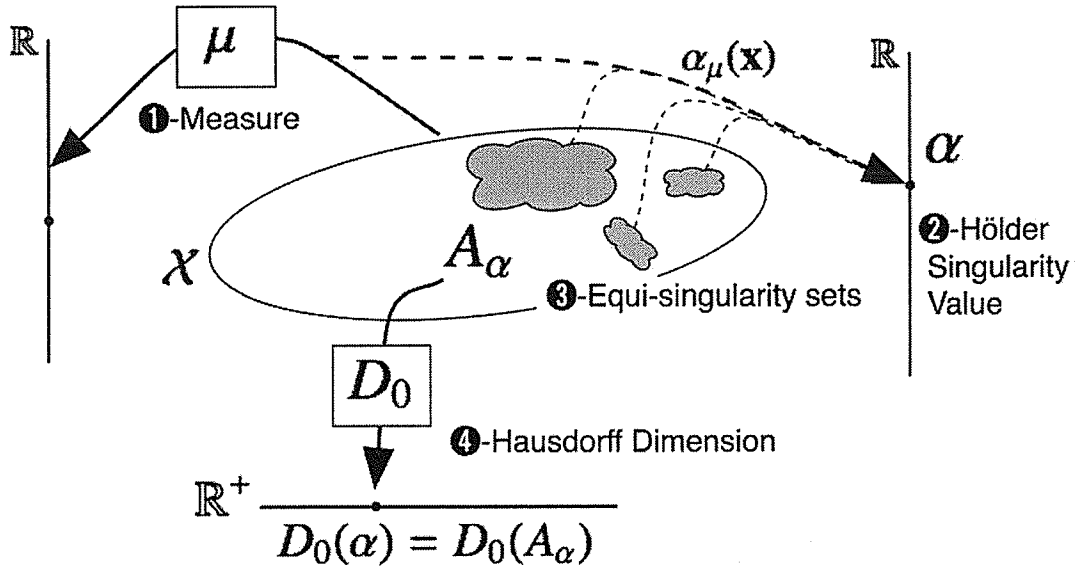


Fig. 4.13 The relationship of the Mandelbrot spectrum to the Hausdorff dimension of the singularity subsets.

Furthermore, since $A_\alpha \subset \Omega_1$, it follows that, [70, Gouyet (1996)]

$$\mu_{Hd}(A_\alpha) \leq \mu_{Hd}(\Omega_1) \quad \forall d \in \mathbb{R}^+ \tag{4.75}$$

$$\Longrightarrow \left\{ d \,\middle|\, \mu_{Hd}(A_\alpha) = 0 \right\} \supset \left\{ d \,\middle|\, \mu_{Hd}(\Omega_1) = 0 \right\} \tag{4.76}$$

$$\Longrightarrow \inf\left\{ d \,\middle|\, \mu_{Hd}(A_\alpha) = 0 \right\} \leq \inf\left\{ d \,\middle|\, \mu_{Hd}(\Omega_1) = 0 \right\} \tag{4.77}$$

$$\Longrightarrow D_0(A_\alpha) \leq D_0(\Omega_1) \leq N_e \tag{4.78}$$

which demonstrates that $D_0(\alpha)$ is bounded above by N_e.

This completes the machinery for transforming the $\tau(q)$ vs. q MFA into a completely bounded multifractal formalism. Here, this fully bounded spectrum is called the Mandelbrot spectrum[2] [117, Kinsner (1994)].

Definition 4.21 (Mandelbrot Spectrum [117, Kinsner (1994)]). Let μ be a measure with support $\Omega_1 \in \mathbb{R}^{N_e}$. Let also A_α be the (Hölder) equisingular set of strength α. Then the *Mandelbrot*

[2]In the literature, the Mandelbrot spectrum is most often called the $f(\alpha)$ spectrum (which is unfortunately inconsistent with the symbol used here for a pdf).

spectrum (MS) of μ is the continuous function $D_0 : \mathbb{R} \mapsto [0, N_e]$ satisfying

$$D_0(\alpha) = d \quad \text{such that} \quad 0 < \mu_{Hd}(A_\alpha) < \infty \tag{4.79}$$

(The obvious dependence on μ is suppressed to keep the notation of the left-hand side simple.) The MS is completely bounded since $\{ \alpha(\boldsymbol{x}) \mid \boldsymbol{x} \in \Omega_1 \}$ for μ will (effectively) be a finite interval. The MS captures the inhomogeneous scaling of μ by expressing the relative "density" (as measured by the box-counting dimension) of the singularities of strength α (as measured by the Hölder exponent), and is considered the "spectrum of scaling indices" [222].

This *Mandelbrot spectrum* [117, Kinsner (1994)] was originally conceived for multiplicative cascades by Mandelbrot (1974) [70][154] and is the multifractal characterization derived by Halsey *et al.* [75, (1986)]. It is complementary to the Rényi spectrum and can be calculated from D_q and q by the Legendre transform [222].

As with the previous change of variables, the change of variable to $D_0(\alpha)$ from $\tau(q)$ or D_q is, theoretically speaking, on the MFA relationships produced after the scaling limit to small ϵ has been taken. As before, a *direct* approach, is desirable, whereby the change of variable occurs on the partition function itself, before applying the limit. To maintain a symmetry with both the definition of D_q and α, a generalized entropy is desired, whose scaling can be measured by the box-counting formalism. This is the subject of the next definition.

Definition 4.22 (Mandelbrot Generalized Entropies, Υ_q)**.** Let a partition $Z(\mu, q \mid \epsilon)$ be the result of a coarse-graining of a measure μ with support Ω_1. Let q be an arbitrary real number, and let γ_q be the Hölder generalized entropy of order q. Then the *Mandelbrot generalized entropy* of order q is the scalar value

$$\Upsilon_q(\mu, q \mid \epsilon) \stackrel{\text{def}}{=} q\gamma_q(\mu, q \mid \epsilon) - \log Z(\mu, q \mid \epsilon) \tag{4.80}$$

Units of Υ_q are measured by the base of the logarithm, just as with Shannon, Rényi, and Hölder generalized entropies. Thus the use of $\log_2$ will imply the units of *bits*.

In particular, the Boltzmann form of the Mandelbrot generalized entropies is quite convenient, as presented in the following proposition.

Proposition 4.23. *For a Boltzmann partition Z_B and real q, the Mandelbrot generalized entropies Υ_q have the explicit form of a Shannon entropy.*

Proof. This can be calculated directly by the substitution of (4.64) and (4.37) into (4.80) and the proper collection of terms. (Again, the shorthand $Z_B(\mu, q \mid \epsilon) = Z_B$ is used to save space.) Begin with

$$\Upsilon_q(\mu, q \mid \epsilon) = q\left(\sum_{\mathcal{V}_\epsilon} \frac{|\mu(V_r)|^q}{Z_B} \log|\mu(V_r)|\right) - \log Z_B \tag{4.81}$$

$$= \left(\sum_{\mathcal{V}_\epsilon} \frac{|\mu(V_r)|^q}{Z_B} (q \log|\mu(V_r)|)\right) - \log Z_B \tag{4.82}$$

$$= \left(\sum_{\mathcal{V}_\epsilon} \frac{|\mu(V_r)|^q}{Z_B} (\log|\mu(V_r)|^q)\right) - \log Z_B \tag{4.83}$$

By adding and subtracting $\log Z_B$ within the summand's second factor, as

$$\Upsilon_q(\mu, q \mid \epsilon) = \left(\sum_{\mathcal{V}_\epsilon} \frac{|\mu(V_r)|^q}{Z_B} (\log|\mu(V_r)|^q - \log Z_B + \log Z_B)\right) - \log Z_B \tag{4.84}$$

$$= \left(\sum_{\mathcal{V}_\epsilon} \frac{|\mu(V_r)|^q}{Z_B} \left(\log \frac{|\mu(V_r)|^q}{Z_B} + \log Z_B\right)\right) - \log Z_B \tag{4.85}$$

a symmetry with $|\mu(V_r)|^q / Z_B$ can be designed in the first term. Expanding out the second term gives

$$\Upsilon_q(\mu, q \mid \epsilon) = \left(\sum_{\mathcal{V}_\epsilon} \frac{|\mu(V_r)|^q}{Z_B} \log \frac{|\mu(V_r)|^q}{Z_B}\right) + \left(\sum_{\mathcal{V}_\epsilon} \frac{|\mu(V_r)|^q}{Z_B}\right) \log Z_B - \log Z_B \tag{4.86}$$

which reduces by the definition of Z_B reduces to

$$= \left(\sum_{\mathcal{V}_\epsilon} \frac{|\mu(V_r)|^q}{Z_B} \log \frac{|\mu(V_r)|^q}{Z_B} \right) + (1) \log Z_B - \log Z_B \tag{4.87}$$

and deletes all but the first term

$$= \sum_{\mathcal{V}_\epsilon} \frac{|\mu(V_r)|^q}{Z_B} \log \frac{|\mu(V_r)|^q}{Z_B} \tag{4.88}$$

Thus Υ_q is an explicit Shannon entropy for all q. Note however, that the discrete probability measure analyzed here is $\widetilde{\underline{u}}_q$ such that

$$\left\{ \widetilde{u}_q(i) = \frac{u^q(i)}{\sum_i u^q(i)} \,\middle|\, u(i) = |\mu(V_i)| \;\; \forall V_i \in \mathcal{V} \right\} \tag{4.89}$$

and so $\widetilde{\underline{u}}_q$ is q-dependent. □

As with the Hölder generalized entropies, the purpose of the Mandelbrot generalized entropies is to provide a direct box-counting formalism for the estimation of D_0 from a Boltzmann partition. This can be established by considering the partition scaling $Z_B \sim \epsilon^{\tau(q)}$, and by taking logarithms and invoking the box-counting ratio, so that one can write

$$\tau(q) \approx \frac{1}{\log \epsilon} \log Z_B \tag{4.90}$$

and using (4.62)

$$\alpha(q) \approx -\frac{1}{\log \epsilon} \gamma_q(Z_B) \tag{4.91}$$

$$\Rightarrow q\alpha(q) \approx -q \frac{1}{\log \epsilon} \gamma_q(Z_B) \tag{4.92}$$

which, when combined as in (4.71), produce

$$D_0(q) \approx -\frac{1}{\log \epsilon} \left(q\gamma_q(\mu, q \mid \epsilon) - \log Z_B \right) \tag{4.93}$$

$$\approx -\frac{1}{\log \epsilon} \Upsilon_q(\mu, q \mid \epsilon) \tag{4.94}$$

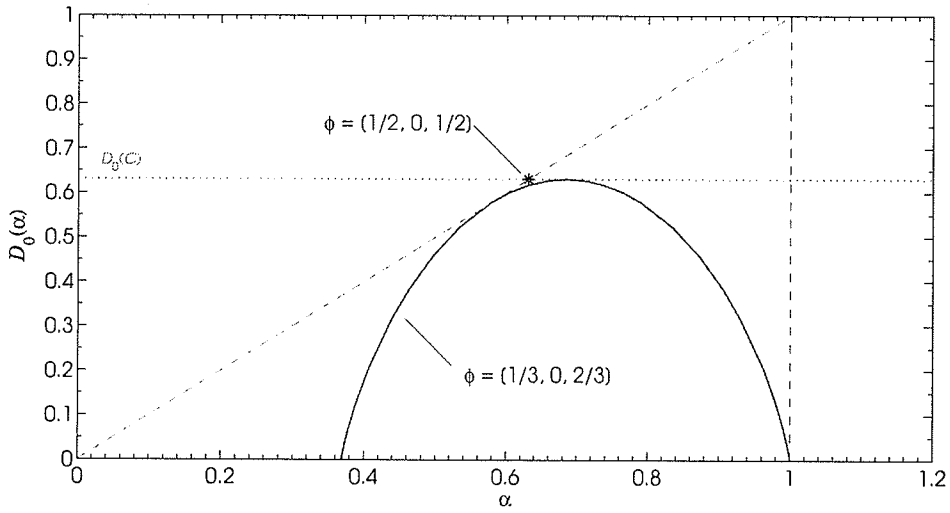


Fig. 4.14 Some analytic Mandelbrot spectra. The single point (star) is the monofractal Cantor set, and the curve is the spectrum of the multifractal $\varphi = [1/3, 0, 2/3]$-weighted non-uniform Cantor measure.

Thus a complete formalism for the estimation of the MS has been presented using the Mandelbrot generalized entropies. The analytic Mandelbrot spectra of some monofractal and multifractal examples is shown in Fig. 4.14. As can be seen, the uniform Cantor measure appears as a point in $(\alpha, D_0(\alpha))$ space, since it is a monofractal (*i.e.*, defined from a cascade of uniform measures). A multifractal measure has a cap-convex curve, which, in the case shown, has endpoints on the $D_0(\alpha) = 0$ line. The (complete) boundedness of the MS is apparent.

With both the Rényi and Mandelbrot spectra for the characterization of an MFA defined, the discussion now turns to the application of MFA to ECG.

4.6 Application of MFA: ECG Time Series

The preceding discussion has presented a complete overview of MFA. The question remains, however, "How do these features apply to the ECG?" From the preceding, it should be clear that this question is, effectively, "How should a coarse-graining be applied to an ECG?"

There are two possible paradigms for the MFA of an ECG. The first uses embedding theory [3] [107] to define an ECG *pseudo-attractor* from which a measure is characterized by MFA (Sec. 4.6.1). This approach is consistent with the discussion in Ch. 2 about nonlinear systems and the variability that can be sustained on their attractors. The second paradigm uses the continuous wavelet transform to analyze the scaling properties of the ECG time-series as a function over time. This so-called *wavelet-transform modulus-maxima* (WTMM) formalism was investigated and developed for inclusion in the thesis design, however, its full implementation was deferred to future work due to time constraints. To assist in this future work, a description of the WTMM approach to MFA is included for completeness in App. D.

Consequently, the remainder of this chapter will focus on the analysis of an ECG pseudo-attractor.

4.6.1 ECG Pseudo-Attractor as a Fractal Measure

As discussed in Ch. 2, the ECG is not a fully stochastic signal, yet its time-integrated distributions can be statistically characterized (Ch. 3). Equivalently, the ECG does not have a straightforward dynamical model, yet the dynamical method of attractor reconstruction and characterization by MFA can produce useful results for ECG signals [107, Kantz and Schreiber (1998)][207, Small *et al.* (2000)][211, Small *et al.* (2002)]. Note that only the conceptual requirements for attractor reconstruction will be addressed here. The many details required in a complete implementation will be left for Ch. 6.

The ECG $\boldsymbol{x}$ is, foremost, a random signal realized as a multidimensional time series. Its cyclostationary structure, however, enables the definition of an ECG pseudo-attractor. That is, the second-instance joint-pdf $f(\boldsymbol{x}_1, t_1; \boldsymbol{x}_2, t_2)$ has a very strong structure if the t_i are in the same cycle. In particular, it can be observed in Fig. 4.15 that the constraints on the structure are very nonlinear if $t_2 - t_1$ is on the order of a cycle (*i.e.*, not so small that $\boldsymbol{x}(t_1)$ approximates

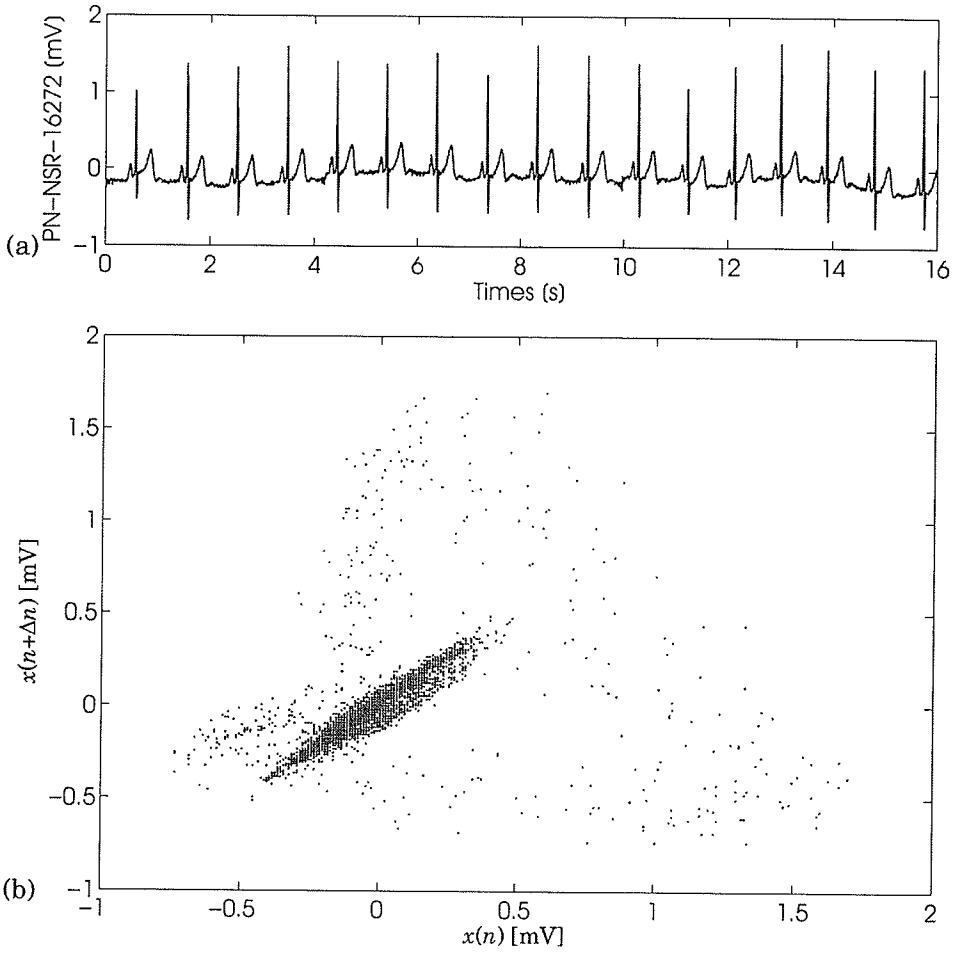


Fig. 4.15 A scatterplot of the second-instance joint-pdf of an ECG at time positions within the one adult heart cycle (*i.e.*, here a delay of 15.6 ms).

$\boldsymbol{x}(t_2)$). The constrained structure of non-zero $f(\boldsymbol{x}_1, t_1; \boldsymbol{x}_2, t_2)$ can be modelled as a manifold in $\mathbb{R}^{N_e} \times \mathbb{R}^{N_e}$. This manifold is the support of the probability measure defined by f. It follows that μ_f and its manifold is dependent on t_1, t_2, and the signal $\boldsymbol{x}(t)$. The goal, however, is to extract something intrinsic to the dynamics of the system, something *invariant*, rather than so explicitly variable. To illustrate, the *Ikeda attractor* will be used as an example of a true

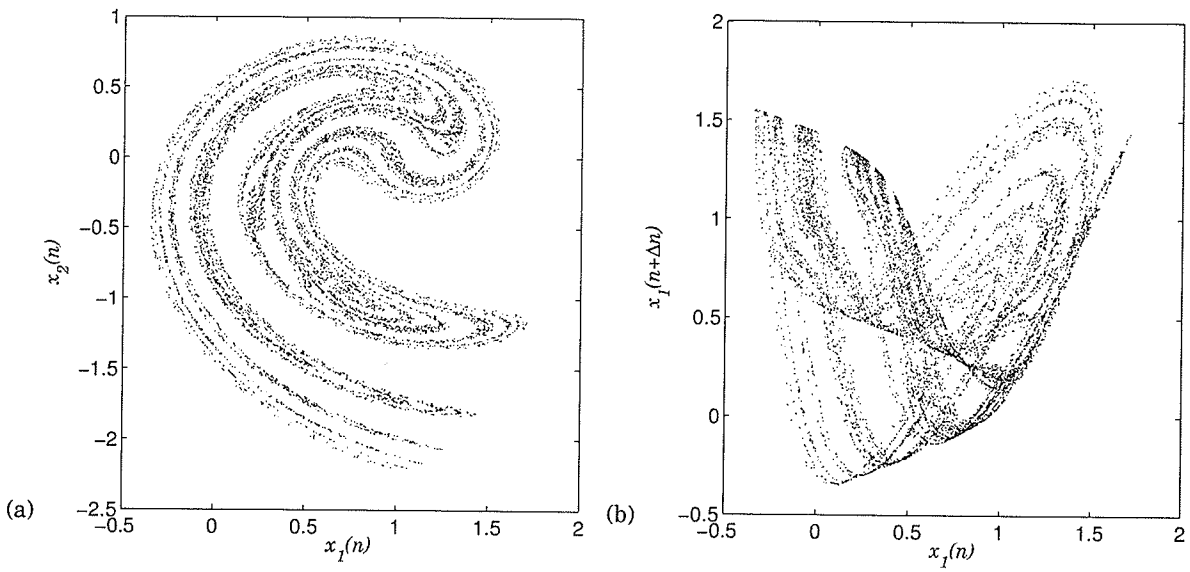


Fig. 4.16 Attractor Reconstruction: (a) Ikeda attractor; (b) Ikeda pseudo-attractor from the $\{x_1\}$ with $N_e = 2$ and $\Delta n = 1$.

attractor for characterization. Defined by the nonlinear system of equations,

$$x_1(n+1) = 1 + 0.9x_1(n)\cos\left(0.4 - \frac{6}{1 + x_1^2(n) + x_2^2(n)}\right) \tag{4.95}$$

$$x_2(n+1) = 0.9x_2(n)\sin\left(0.4 - \frac{6}{1 + x_1^2(n) + x_2^2(n)}\right) \tag{4.96}$$

the Ikeda system represents a model of nonlinear laser dynamics. More importantly here, it is observed that a sequence of points $\boldsymbol{x}(n)$ satisfying the above can be observed as a scatterplot in $\mathbb{R}^2$ by ignoring the well-ordering by n, Fig. 4.16(a). This scatterplot represents the dynamics of the system equations, since no extraneous variables have entered into it. It is a theorem in nonlinear dynamical theory that the measure μ_f defined from this attractor $\{\boldsymbol{x}(n)\}$ can also be reconstructed from the observation of a single observational variable [3]

$$\{y_0 = y_0(\boldsymbol{x}(n)) \mid \forall n\} \tag{4.97}$$

More precisely, *Takens' theorem* [219, Takens (1981)][197] states that an equivalent (diffeomorphic) measure μ'_f can be defined by an embedded pseudo-attractor manifold of the form

$$\mathbf{y}(n) = \begin{bmatrix} y_0(n) \\ y_0(n+\Delta n) \\ \vdots \\ y_0(n+(N_e-1)\Delta n) \end{bmatrix} \in \mathbb{R}^{N_e} \tag{4.98}$$

where N_e is the *embedding dimension* of the pseudo-attractor manifold, and Δn is a *lag* (here, *sample lag*) of the embedding. Specifically, Takens showed that for an attractor with a manifold of Hausdorff dimension D_0, a diffeomorphic equivalent μ'_f exists with an embedding in $\mathbb{R}^{N_e}$ such that $N_e \geq 2D_0 + 1$ [219, Takens (1981)]. Specifically, an MFA of the Takens pseudo-attractor is the same as the MFA of the real attractor. This existence theorem, however, does not specifically provide the means of determining N_e and Δn. Further work in embedding theory has devised reasonable and tractable forms for estimating both N_e and Δn [3]. For comparison, a pseudo-attractor from the observation of the $\{y_0(n) = x_1(n)\}$ of the Ikeda attractor is shown in Fig. 4.16(b) with $N_e = 2$ and lag $\Delta n = 1$. As can be seen, significant overlap of the pseudo-attractor still exists, and a higher embedding dimension is required for the pseudo-attractor than the attractor itself. This *unfolding* of the attractor is important, and is related to the *trajectory* through the points on the manifold (*i.e.*, their well-ordering by time).

Thus, in imitation of this approach, ECG pseudo-attractors can be defined from a time series $x(n)$ by deriving some sample lag Δn and embedding dimension N_e by embedding theory and thereby define a manifold equivalent to the N_eth-instance pdf at times $t_i = t+(i-1)\Delta n$. The measure on this particular manifold can be characterized using MFA as a scaling characterization of ECG [90, Huang (2004)]. According to the work of Huang, the ECG has embedding dimension of $N_e = 7$, making visualization of the manifold impossible.

4.6.2 Attractor Measure Coarse-Graining

Now that the measure μ is implied on $\mathbb{R}^{N_e}$ by the set of sample points $\{\boldsymbol{y}\}$, it is desirable to describe two implementations for defining the coarse-grained elements $u_i = \mu(V_r)$ required for the partition function from these sample points. The first is the *histogram technique*, and the second is the *correlation technique*. The main difference between histogram based methods and correlation-based methods is the pattern of the coarse-graining ϵ-covering: be it homogeneously distributed in space (*i.e.*, histogram-based), or "naturally distributed" according to the measure (*i.e.*, correlation-based).

4.6.2.1 Histogram Coarse-Graining

The histogram coarse-graining of an implied pseudo-attractor is straightforward to conceptualize and implement. Cover the pseudo-attractor manifold with a single N_e-dimensional hypercube. Subdivide, for $j = 1, \ldots, J$ the hypercube into $2^{N_e j}$ equal hypercubes indexed by i. Thus, for each j, a non-overlapping ϵ-covering is created by the hypercubes, such that $\epsilon \propto 2^{-j}$. Then for each hypercube $V_j(i)$, an estimate of $u_j(i) = \mu(V_j(i))$ is made by relative frequency, that is

$$\widehat{u_j}(i) = \frac{N_j(i)}{N} \quad \forall V_j(i) \tag{4.99}$$

where N is the total number of sample points implying the pseudo-attractor, and $N_j(i)$ is the number of sample points in the hypercube $V_j(i)$. It is clear that $\sum_i N_j(i) = N$ for all j. The histogram coarse-grained estimate of a Boltzmann partition function is therefore

$$\widehat{Z_B}(\mu, q \mid \epsilon(j)) = \sum_{i=1}^{2^{N_e j}} \widehat{u_j}^q(i) \tag{4.100}$$

$$= \sum_{i=1}^{2^{N_e j}} \left(\frac{N_j(i)}{N}\right)^q \tag{4.101}$$

A visualization of histogram coarse-graining on the true Ikeda attractor is shown in Fig. 4.17.

As mentioned, histogram coarse-graining is straightforward to implement. However, it is not necessarily the most robust, particularly because the estimate (4.99) is sensitive for small hypercubes and large embedding dimensions [222]. It is preempted in the thesis design for a more robust approach using the correlation-integral, as presented next.

4.6.2.2 Correlation Coarse-Graining

Another method of coarse-graining an implied pseudo-attractor from data points $\{\boldsymbol{y}\}$ relies on correlation between the sample points. Extending previous work by Grassberger and Procaccia [72, (1983)], Pawelzik and Schuster [167, (1987)] described a method of estimating $Z_B(\mu, q \mid \epsilon)$ by some ergodic "slight-of-hand". Assume that μ is positive, so that, from the definition (Def. 4.4), the Boltzmann sum can be decomposed as

$$Z_B(\mu, q \mid \mathcal{V}_\epsilon) = \sum_{\mathcal{V}} |\mu(V)|^q = \sum_{\mathcal{V}} \mu(V)\,\mu^{q-1}(V) \tag{4.102}$$

The final form has the interpretation of an average of the function $\mu^{q-1}(V)$ weighted by μ. A more exact calculation with this interpretation can be written as [85][167]

$$Z_B(\mu, q \mid \mathcal{V}_\epsilon) = \mathcal{E}\left\{\mu^{q-1}\left(V_\epsilon\right)\right\} \tag{4.103}$$

$$= \int \mu^{q-1}\left(V_\epsilon(\boldsymbol{y})\right) d\mu \tag{4.104}$$

where the $V_\epsilon(\boldsymbol{y})$ in (4.104) are specifically the open balls centered on $\boldsymbol{y}$. The form, in fact, is well designed for an actual calculation by estimating $\mu\left(V_\epsilon(\boldsymbol{y})\right)$ by the frequency of neighbours around $\boldsymbol{y}$. As such, this estimation of Z_B is based on the clustering, or *correlation*, of the points $\boldsymbol{y}(n)$ in $\mathbb{R}_e^N$. Since every point $\boldsymbol{y}$ must be considered as a centre, however, the covering sets $V_\epsilon(\boldsymbol{y})$ must naturally overlap.

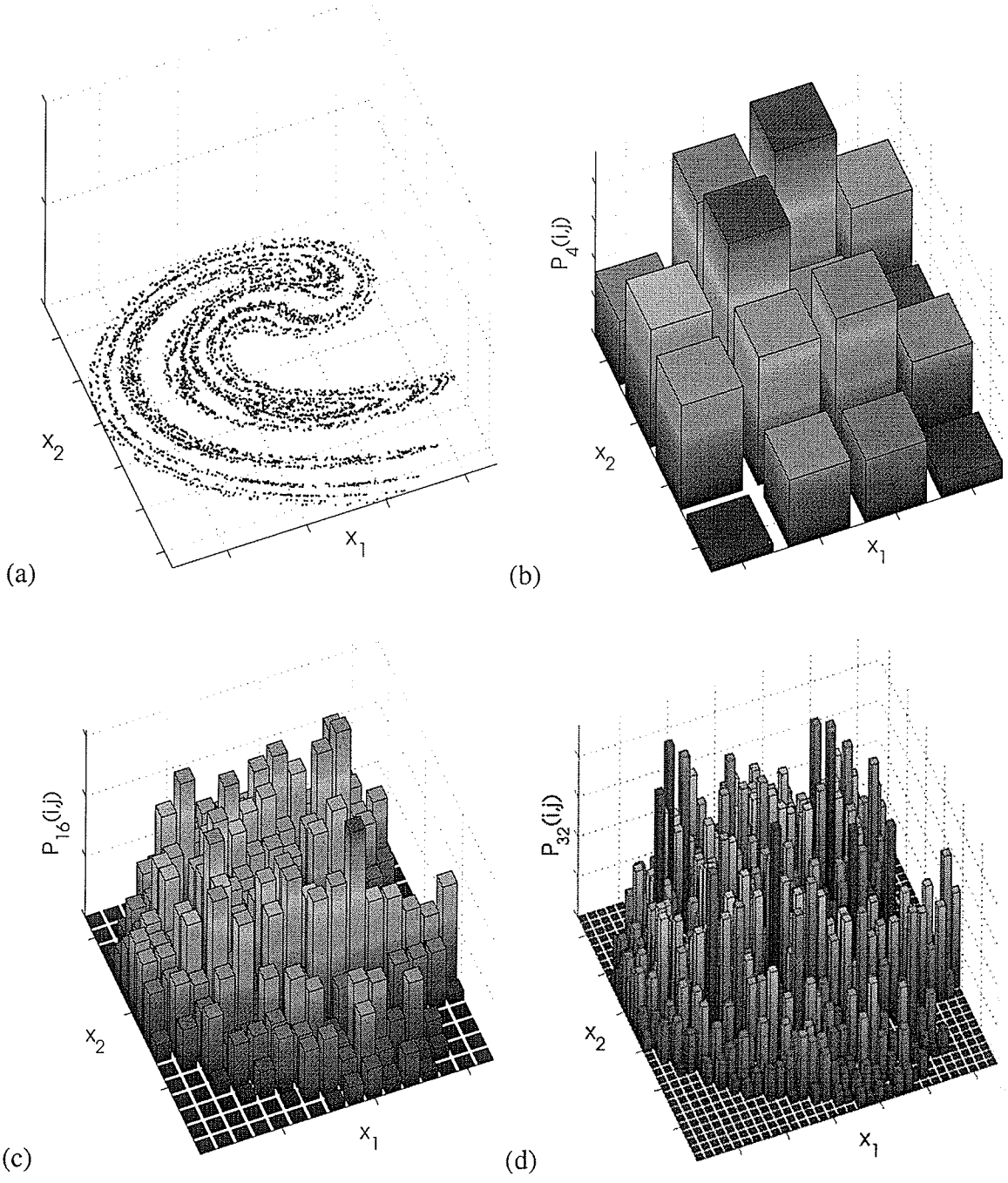


Fig. 4.17 Histogram Coarse-Graining: (a) actual Ikeda attractor in $\mathbb{R}^2$; (b) $j = 4$ coarse-graining; (c) $j = 16$ coarse-graining; (d) $j = 32$ coarse-graining.

Thus coarse-graining by correlation is different in principle from the histogram technique only because the expectation of (4.103) is coarse-grained by fixed-size cells of a different type: correlation uses cells that overlap and are nonuniformly distributed in space (but naturally distributed on the measure support), while the histogram technique uses cells without overlap that are homogeneously distributed in space. Specifically, the correlation method in application leaves the coarse-graining balls centred on the *data-points themselves*, and so the distribution of the ϵ-covering sets is the same as the ECG measure itself. This equivalence of distributions means each ball is *equally likely* under an average, so that for all N points in the attractor, $\mu(V_\epsilon(\mathbf{y})) = \Pr(V_\epsilon(\mathbf{y})) = 1/N$. This simplifies (4.104) to the form

$$\widetilde{Z}_B(\mu, q \mid \mathcal{V}_\epsilon) = \frac{1}{N} \sum_j \widetilde{u}_j^{q-1} \tag{4.105}$$

where the $\widetilde{u}_j$ are the mass of $V_\epsilon(\mathbf{y}(j))$, and a $\widetilde{\cdot}$ is used to identify the correlation form of the Boltzmann partition function. With finitely many points, the $\widetilde{u}_j$ can be estimated by the ratio of points in the ball to those outside, effectively

$$\widehat{\widetilde{u}}_j = \frac{\widetilde{N}_j}{N-1} \tag{4.106}$$

$$= \left(\frac{1}{N-1} \sum_{i \neq j} \Xi(\epsilon - \|\mathbf{y}(i) - \mathbf{y}(j)\|) \right) \tag{4.107}$$

where Ξ is the Heaviside function,

$$\Xi(r) = \begin{cases} 0 & : \quad r \leq 0 \\ 1 & : \quad r > 0 \end{cases} \tag{4.108}$$

(Note that the frequency ratio is normalized by $N-1$ to avoid the self-counting of $\mathbf{y}(j) \in V_\epsilon(\mathbf{y}(j))$.)

All together, the correlation-based formula is

$$\widehat{\widetilde{Z}}_B(\mu, q \mid \mathcal{V}_\epsilon) = \frac{1}{N} \sum_j \left(\frac{1}{N-1} \sum_{i \neq j} \Xi(\epsilon - \|\mathbf{y}(i) - \mathbf{y}(j)\|) \right)^{q-1} \tag{4.109}$$

which is an $O(N^2)$ calculation. This can be improved, however, because (i) $\Xi(\epsilon - \|\mathbf{y}(i) - \mathbf{y}(j)\|)$ is zero with high probability, and so contributes only rarely; and (ii) $\Xi(\epsilon - \|\mathbf{y}(i) - \mathbf{y}(j)\|)$ is symmetric in i and j. Therefore, wise algorithmic constructs can reduce the amount of computation/comparison, generally at the cost of increased memory use.

Note that for the case $q = 2$ (Grassberger-Procaccia *correlation dimension* [72, Grassberger and Procaccia (1983)]), the power $(q - 1) = 1$ in (4.109) and the equation simplifies into a double sum that counts all the pairs $(\mathbf{y}(i), \mathbf{y}(j))$ of distance less than ϵ. In that case, the efficiency of the algorithm can be greatly improved because, even as double counting is removed, only the partial *correlation sum* (for every scale ϵ) needs to be stored in memory.

Correlation coarse-graining is, in fact, preferred for ECG attractor analysis over histogram-based coarse-graining, since it performs better in higher-dimensional embeddings. This better performance is a result of superior dynamic range [222, Theiler (1990)], meaning the coarse-grained measure is approximated better at smaller scales and lower densities with the finitely many points available.

4.6.2.3 Theiler Correction to Correlation Coarse-Grainings

From the previous sections on the two different coarse-grainings, it should be clear that the Boltzmann partition function can be calculated from either a *histogram partition* (HP) or *correlation partition* (CP). To review, the Boltzmann partition function estimate for an attractor represented by the points $\mathbf{y}(n) \in \mathbb{R}^{N_e}$ is

$$\begin{cases} \text{HP} & : \widehat{Z}_B(\mu, q \mid \mathcal{V}_\epsilon) = \sum_{i=1}^{N(\mathcal{V})} \widehat{u}_i^q(\epsilon) \quad : \widehat{u}_i(\epsilon) = \frac{N_i}{N} \\ \text{CP} & : \widehat{\widetilde{Z}}_B(\mu, q \mid \mathcal{V}_\epsilon) = \frac{1}{N} \sum_{i=1}^{\widetilde{N}(\mathcal{V})} \widehat{\widetilde{u}}_i^{q-1}(\epsilon) \quad : \widehat{\widetilde{u}}_i(\epsilon) = \frac{\widetilde{N}_i}{N-1} \end{cases} \tag{4.110}$$

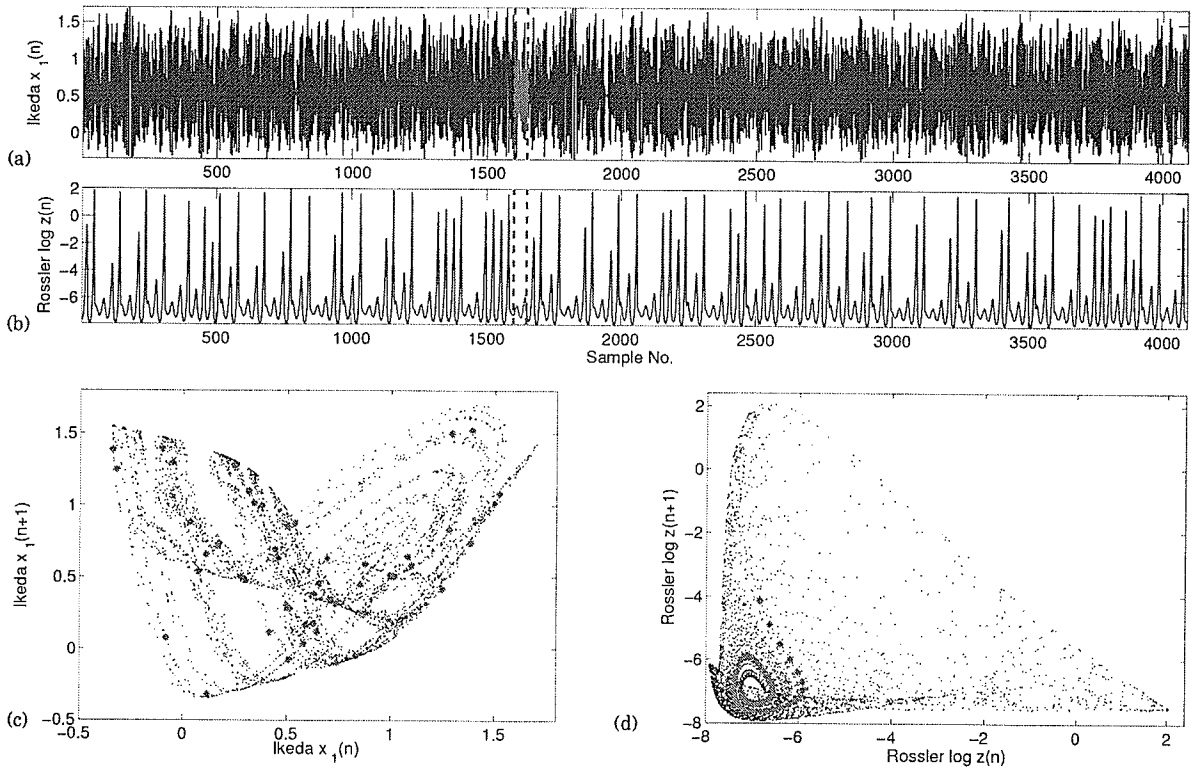


Fig. 4.18 Time-correlated neighbours: (a) trajectory of Ikeda first component; (b) trajectory of a chaotic Rössler z-component (in log scale); Both trajectories have a small time interval marked to identify time neighbours. (c) 2D Ikeda reconstruction; The time neighbours (stars) are scattered throughout the attractor. (d) 2D Rössler reconstruction; The time neighbours (stars) are correlated and along the trajectory flow.

where $\widetilde{N}(\mathcal{V}) = N$ is constant for all ϵ, while $N(\mathcal{V})$ increases to N as $\epsilon \to 0$, and

$$N_i = \|\{\boldsymbol{y}(j) \in V_i\}\| \tag{4.111}$$

$$\widetilde{N}_i = \|\{\boldsymbol{y}(j) \in V_i \mid j \neq i\}\| \tag{4.112}$$

represent the number of points appearing in the ith set in the ϵ-covering for the HP and CP respectively. Note the exclusion made in the CP case of (4.112) to avoid the self-counting of $\boldsymbol{y}(i) \in V_i(\epsilon)$.

Depending on the natural correlation of the time series sequence $y(j)$ that reconstructs the

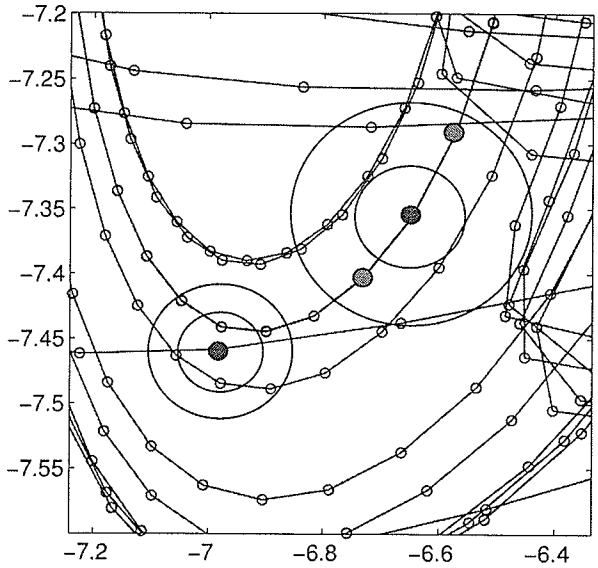


Fig. 4.19 Theiler correction for time-correlation: All points in the balls surrounding the left centre point (dark solid) contribute to the CP, because they are not time correlated. The two grey solid points in the largest ball around the right centre point (dark solid) should not contribute to the CP because they are adjacent to the centre point in time. (*i.e.*They are neighbours on the flow.) Setting the Theiler window $W_T \geq 2$ will exclude them from the calculation.

embedded $\{\boldsymbol{y}\}$, the avoidance of *only* self-counting may degrade the CP estimates. Theiler [221, (1986)] analyzed the correlation integral (CP under $q = 2$) scaling and determined a bias is introduced by the counting of points $\boldsymbol{y}(j) \in V_i(\epsilon)$ that are neighbours in the time series itself; *i.e.*, where $|j - i|$ is small. This natural clustering of time series neighbours is dependent on the amount of natural correlation in the times series, as shown in Fig. 4.18. Here it can be observed that the embedding of an Ikeda trajectory scatters time series neighbours widely throughout the attractor without clustering, while neighbours from a chaotic Rössler trajectory do cluster in the embedding. This clustering property of a dynamical *flow* [3] introduces the bias in the CP. The solution recommended by Theiler [221][222] is to increase the exclusion criteria in (4.112) to apply to neighbours and not just self-counting. That is, a *Theiler window*,

W_T, is specified so that

$$\widetilde{N}_i = \|\{\, \boldsymbol{y}(j) \in V_i \mid |j-i| \geq W_T \,\}\| \tag{4.113}$$

It is clear that $W_T \geq 1$, and the original estimate is achieved by the equality. References [221, Theiler (1986)][222, Theiler (1990)][108, Kantz and Schreiber (1997)] can provide insight into proper selection of W_T. Using a Heaviside form like that of (4.107), this adjusted form of $\widetilde{N}_i$ can be expressed by limiting the sum as

$$\widetilde{N}_i(\epsilon) = \sum_{|j-i| \geq W_T} \Xi(\epsilon - \|\boldsymbol{y}(j) - \boldsymbol{y}(i)\|) \tag{4.114}$$

A schematic of the motivation and effect of the Theiler window is shown in Fig. 4.19.

As a consequence of the excluded points in $\widetilde{N}_i$, however, the normalization in the relative frequency estimate of $\widetilde{u}_i(\epsilon)$ requires further modification, as well, yielding

$$\widetilde{\widetilde{u}}_i(\epsilon) = \frac{\widetilde{N}_i}{N - N_{W_T}} \tag{4.115}$$

where

$$N_{W_T} = 2W_T - 1 \tag{4.116}$$

Before the next section, where the consequences of all these adaptations to the correlation partition are presented in the context of the Rényi, Hölder, and Mandelbrot generalized entropies, the equations for the calculation of Boltzmann partition function from a CP are

summarized as

$$N_{W_T} = 2W_T - 1 \tag{4.117}$$

$$\widetilde{N}_i(\epsilon) = \sum_{|j-i|\geq W_T} \Xi(\epsilon - \|y(j) - y(i)\|) \quad \forall i \tag{4.118}$$

$$\widehat{\widetilde{u}}_i(\epsilon) = \frac{\widetilde{N}_i}{N - N_{W_T}} \quad \forall i \tag{4.119}$$

$$\widehat{Z}_B(\mu, q \mid \mathcal{V}_\epsilon) = \frac{1}{N} \sum_{i=1}^{\widetilde{N}(\mathcal{V})} \widehat{\widetilde{u}}_i^{q-1}(\epsilon) \tag{4.120}$$

where, as before, $\widetilde{N}(\mathcal{V}) = N$ and N is the number of points in the embedding $\{y\} \subset \mathbb{R}^{N_e}$.

4.7 MFA Spectra: Direct and Indirect Forms

The preceding sections of this chapter have presented a unified and rather straightforward pedagogy for the MFA of an attractor measure by observing the scaling of a Boltzmann partition function or some related entropies. The historical development of these ideas is much less straightforward. Monofractal approaches for different fixed values of q were developed independently using HP, CP [72, Grassberger and Procaccia (1983)], and other techniques [70, Gouyet (1996)]. Legendre transform relationships between thermodynamical variables, such as the Mandelbrot spectrum, existed early on in the literature [70], but their value were not recognized in the literature until much later after a "unification of fractal dimensions" under the Rényi spectrum formalism [85, Hentschel and Procaccia (1983)]. This unification, however, assumed an underlying histogram partition. The Rényi spectrum formalism was later generalized to estimates from correlation partitions [167, Pawelzik and Schuster (1987)]. The seminal paper on the Mandelbrot spectrum as an experimental approach came slightly earlier, [75, Halsey *et al.* (1986)]. Last of all, Chhabra and Jensen [39, (1989)] discussed the estimation of the MS directly from (4.62) and (4.94). A recommended review on MFA that

covers most of this literature, (and a great source of references, mathematical notation, and introduction to fractal analysis) is [222, Theiler (1990)].

The pedagogical approach taken in this exposition of MFA theory, however, emphasizes the elements of (i) partitions (coarse-grainings) of a measure, and (ii) estimation of MFA by the box-counting relationship of a function of the Boltzmann partition function, $g(Z_B)$, informally called an "entropy". In this synthesis of the literature, it becomes clear that a missing formalism, namely a *direct MS* from a *correlation partition*, has not been discussed. The following sections discuss further the distinction between direct and indirect methods, the mosaic of possible formalisms, and derives the novel *direct correlation partition Mandelbrot spectrum* (DCPMS).

4.7.1 2×2 Mosaic of MFA Formalisms

Previously, this chapter has covered the two equivalent multifractal formalisms that are used for MFA characterization: (i) the Rényi spectrum formalism, D_q, [85] which is based on the generalized dimensions derived from a family of q-th order Rényi entropies; and (ii) the Mandelbrot singularity spectrum formalism, $D_0(\alpha)$, [75] which is based on equisingular subsets of Hölder singularity α.[3] The spectra are significant for different reasons.

The D_q formalism is important because it (i) is the landmark unification and extension of the various historical fractal dimensions defined previously by Mandelbrot and others, (ii) has an intuitive interpretation for positive integral q regarding q-tuple correlations (Grassberger, 1983) [222] [116], (iii) has the nice property of existing in the $L^\infty(\mathbb{R})$ function space, and (iv) provides a template for the estimation of an MFA by the box-counting formalism of a generalized entropy.

[3]Note that in the literature, the more common notation for $D_0(\alpha)$ is $f(\alpha)$.

The $D_0(\alpha)$ formalism is preferred, however, since it (i) describes a multifractal as a union of interwoven monofractal sets, and therefore has a "naturally intuitive" interpretation [39], and (ii) also provides a functional form with compact support (in practical situations[4]).

As mentioned, these two spectra are thermodynamically related via a *Legendre transform* identity [75, Halsey *et al.* (1986)][222, Theiler (1990)]. Though the $D_0(\alpha)$ formalism has the preferred interpretation, Halsey *et al.* [75, (1986)] only considered these variables indirectly attainable from experimental data and calculated them through the D_q formalism (4.54) via the Legendre transform identity [222],

$$\alpha(q, D_q) = -\frac{\partial}{\partial q}(1-q)D_q \tag{4.121}$$

$$D_0(q, D_q) = q\alpha + (1-q)D_q \tag{4.122}$$

instead. Applying these equations to derive $(\alpha, D_0(\alpha))$ from (q, D_q) (or the reverse), is considered an *indirect* calculation of the MS (RS, respectively). Due to the nonlinear nature of the transform, however, it is difficult to transfer confidence intervals or uncertainty estimates between the spectra when numerical data are used.

The previous sections have also described *direct* methods for MFA, whereby the MFA of an attractor is estimated by

(a) a scaling collection of coarse-grainings (or *partitions*), which defines in turn

(b) a scaling collection of partition functions (or scaling collection of entropies), whose

(c) scaling exponents are extracted to, ultimately,

define a fractal dimension spectrum. A *partition*, as it used here, is a coarse-graining with open sets of equal size, V_ϵ, and thus is a more practical form than a general coarse-graining. The two

[4] *c.f.* [134] for a contrived counterexample.

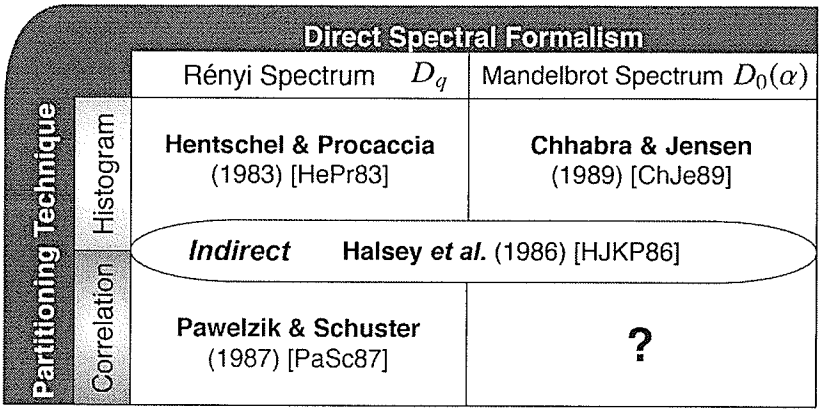


Fig. 4.20 MFA theory "mosaic": the organization of traditional multifractal spectra thermodynamic formalisms.

distinct techniques to defining partitions on the attractor were also reviewed: (i) a histogram partitioning (HP) of fixed-size boxes that is non-overlapping and uniform in space; and (ii) a correlation partitioning (CP) of overlapping fixed-size balls nonuniformly distributed in space, but naturally distributed on the attractor (*i.e.*, equally likely). The simplicity and speed of HP technique naturally lends it to the origin and instruction of multifractal methods. It also has a simple interpretation as a histogram estimate (of size ϵ) of the attractor's probability density function. However, the convergence of HP tends to become inefficient for attractors embedded in high-dimensional spaces [74, Greenside *et al.* (1982)]. CP, on the other hand, is preferred for high-dimensional embeddings, because it has better dynamic range than HP [222, Theiler (1990)]. (This means the attractor density is approximated better at smaller scales and lower densities with finitely many points.) Thus, though the CP is computationally expensive with a native $O(N^2)$ algorithm complexity [167], it is the more popular technique in applied research when the attractors are embedded in higher dimensional spaces.

Between these two spectral formalisms and these two partitioning techniques, a "mosaic" of multifractal theory is generated, as shown in Fig. 4.20. The literature, however, has a gap

in their theoretical combination. CP was originally developed [85] (and extended [167]) for the D_q formalism, but can also be indirectly transformed to $D_0(\alpha)$. A direct formalism for the $D_0(\alpha)$ singularity spectrum was defined from first principles [39], but only in the case of a HP. As such, the theory of a direct $D_0(\alpha)$ calculation from a CP remains incomplete.

In the course of this work, the novel contribution of a *direct* $D_0(\alpha)$ formalism under the assumption of a CP is made, filling the missing piece in Fig. 4.20. This novel technique maintains the sound interpretation of the Mandelbrot spectrum, provides an efficient estimation of numerical uncertainty, but also allows the dynamical range of the correlation integral to enhance the calculation from finite data in higher-dimensional embeddings.

To proceed, a more in depth review of Chhabra and Jensen's direct MS under HP will be presented to motivate the derivation that follows.

4.7.2 Motivation: Direct MS from Histogram Partitions

Using the Legendre transform (4.121) and (4.122) [75] or more numerically robust forms [222], it is possible to calculate the singularity spectrum $D_0(\alpha)$, with its elegant multifractal interpretation, from an experimental D_q spectrum. This process, however, introduces further numerical error, can obscure phase-transitions, and also makes it very difficult to estimate uncertainty estimates. Chhabra and Jensen [39, (1989)] derived a direct formula in order to specifically define uncertainty estimates on empirical $D_0(\alpha)$ spectra. Their formula and its interpretation was derived from an independent set of first principles based on the Shannon entropies of a q-ordered family of measures $\mu(q,\epsilon) = \left\{u_i^q(\epsilon)/\sum_i u_i^q(\epsilon)\right\}$. From this *a priori* definition, it followed that their result must be consistent with the Legendre transformations of Hentschel and Procaccia's D_q formalism. Their direct result is equivalent to the set of parametric equations (4.62) and (4.94) reproduced below in the formal sense with entropy

substitutions

$$Z_B(\epsilon) = \sum_i u_i^q(\epsilon) \tag{4.123}$$

$$\alpha(q) = \lim_{\epsilon \to 0} \frac{-\gamma_q(\epsilon)}{\log \epsilon} \tag{4.124}$$

$$= \lim_{\epsilon \to 0} \frac{1}{\log \epsilon} \sum_i \frac{u_i^q(\epsilon)}{Z_B(\epsilon)} \log u_i \tag{4.125}$$

$$D_0(q) = \lim_{\epsilon \to 0} \frac{-\Upsilon_q(\epsilon)}{\log \epsilon} \tag{4.126}$$

$$= \lim_{\epsilon \to 0} \frac{1}{\log \epsilon} \sum_i \frac{u_i^q(\epsilon)}{Z_B(\epsilon)} \log \frac{u_i^q(\epsilon)}{Z_B(\epsilon)} \tag{4.127}$$

The singularity spectrum $D_0(\alpha)$ is thus the parametric relationship of α and $D_0(\alpha)$ via q.

Note that, notwithstanding the mathematical complexity of scaling "$\sim$" in the formal sense, (which includes the limit $\epsilon \to 0$ and an inf), or its awkward circumvention by the box-counting estimation "$\approx$", the Legendre transform can *effectively* be applied to both sides of the relationships (4.54), (4.62), and (4.94). That is, the Legendre transform (with its derivatives) of the formal limits (τ, α, and D_0) can be estimated by the box-counting formalism of the Legendre transform of the entropies (H_q, γ_q, and Υ_q). This calculation process can also be exploited to obtain from (4.120) the direct (parametric) form of $D_0(\alpha)$ under the assumption of a correlation partition.

4.7.3 Derivation: Direct MS from Correlation-Integral Partitions

Now the novel direct $D_0(\alpha)$ formulation from the correlation partition will be derived. The same procedure of applying the Legendre transforms (4.121) and (4.122) to both the partition entropies and the spectral variables is used, but the process begins from (4.120) on p. 198.

Thus, using the shorthand $\widehat{\widetilde{Z}}_B$ for the complete Boltzmann partition function $\widehat{\widetilde{Z}}_B(\mu, q \mid \mathcal{V}_\epsilon)$

of (4.120), it is clear that (4.121) requires from (4.120) that

$$\widehat{\alpha}(q) \approx \frac{1}{\log \epsilon} \frac{\partial}{\partial q} \log \left(\frac{1}{N} \sum_{i=1}^{\widetilde{N}(\mathcal{V})} \widehat{\widetilde{u}}_i^{(q-1)}(\epsilon) \right) \tag{4.128}$$

$$\approx \frac{1}{\log \epsilon} \frac{1}{N} \sum_{i=1}^{\widetilde{N}(\mathcal{V})} \frac{\widehat{\widetilde{u}}_i^{q-1}(\epsilon)}{\widehat{Z}_B(\epsilon)} \log \widehat{\widetilde{u}}_i(\epsilon) \tag{4.129}$$

Furthermore, dropping the obvious ϵ dependency, and substituting (4.129) and (4.120) into (4.122) produces

$$\widehat{D_0}(q) \approx \frac{1}{\log \epsilon} q \left(\frac{1}{N} \sum_{i=1}^{\widetilde{N}(\mathcal{V})} \frac{\widehat{\widetilde{u}}_i^{q-1}}{\widehat{\widetilde{Z}}_B} \log \widehat{\widetilde{u}}_i \right) - \log \widehat{\widetilde{Z}}_B \tag{4.130}$$

$$\approx \frac{1}{\log \epsilon} \frac{1}{N} \sum_{i=1}^{\widetilde{N}(\mathcal{V})} \frac{\widehat{\widetilde{u}}_i^{q-1}}{\widehat{\widetilde{Z}}_B} \left(q \log \widehat{\widetilde{u}}_i \right) - \log \widehat{\widetilde{Z}}_B \tag{4.131}$$

$$\approx \frac{1}{\log \epsilon} \frac{1}{N} \sum_{i=1}^{\widetilde{N}(\mathcal{V})} \frac{\widehat{\widetilde{u}}_i^{q-1}}{\widehat{\widetilde{Z}}_B} \left(\log \widehat{\widetilde{u}}_i^{q} \right) - \log \widehat{\widetilde{Z}}_B \tag{4.132}$$

and by adding and subtracting $\log \widehat{\widetilde{Z}}_B$ in the summand's second factor

$$\approx \frac{1}{\log \epsilon} \frac{1}{N} \sum_{i=1}^{\widetilde{N}(\mathcal{V})} \frac{\widehat{\widetilde{u}}_i^{q-1}}{\widehat{\widetilde{Z}}_B} \left(\log \widehat{\widetilde{u}}_i^{q} - \log \widehat{\widetilde{Z}}_B + \log \widehat{\widetilde{Z}}_B \right) - \log \widehat{\widetilde{Z}}_B \tag{4.133}$$

$$\approx \frac{1}{\log \epsilon} \frac{1}{N} \sum_{i=1}^{\widetilde{N}(\mathcal{V})} \frac{\widehat{\widetilde{u}}_i^{q-1}}{\widehat{\widetilde{Z}}_B} \left(\log \frac{\widehat{\widetilde{u}}_i^{q}}{\widehat{\widetilde{Z}}_B} + \log Z_B \right) - \log \widehat{\widetilde{Z}}_B \tag{4.134}$$

which expands to the three terms

$$\approx \frac{1}{\log \epsilon} \left(\frac{1}{N} \sum_{i=1}^{\widetilde{N}(\mathcal{V})} \frac{\widehat{\widetilde{u}}_i^{q-1}}{\widehat{\widetilde{Z}}_B} \log \frac{\widehat{\widetilde{u}}_i^{q}}{\widehat{\widetilde{Z}}_B} \right) + \left(\frac{1}{N} \sum_{i=1}^{\widetilde{N}(\mathcal{V})} \frac{\widehat{\widetilde{u}}_i^{q-1}}{\widehat{\widetilde{Z}}_B} \right) \log Z_B - \log \widehat{\widetilde{Z}}_B \tag{4.135}$$

which then reduces by the definition of Z_B to

$$\approx \frac{1}{\log \epsilon} \left(\frac{1}{N} \sum_{i=1}^{\widetilde{N}(\mathcal{V})} \frac{\widehat{\widetilde{u}}_i^{q-1}}{\widehat{\widetilde{Z}}_B} \log \frac{\widehat{\widetilde{u}}_i^{q}}{\widehat{\widetilde{Z}}_B} \right) + (1) \log Z_B - \log \widehat{\widetilde{Z}}_B \tag{4.136}$$

$$\approx \frac{1}{\log \epsilon} \frac{1}{N} \sum_{i=1}^{\widetilde{N}(\mathcal{V})} \frac{\widehat{\widetilde{u}}_i^{q-1}}{\widehat{\widetilde{Z}}_B} \log \frac{\widehat{\widetilde{u}}_i^{q}}{\widehat{\widetilde{Z}}_B} \tag{4.137}$$

These new results can be cast in the same box-counting entropy form as the HP direct results, yielding

$$\widehat{\alpha}(q) \approx -\frac{1}{\log \epsilon}\widehat{\widetilde{\gamma}}_q(\epsilon) \tag{4.138}$$

$$\widehat{D_0}(q) \approx -\frac{1}{\log \epsilon}\widehat{\widetilde{\Upsilon}}_q(\epsilon) \tag{4.139}$$

where the entropies specific to the CP Boltzmann partition function are

$$\widehat{\widetilde{\gamma}}_q(\mu, q \mid \epsilon) \stackrel{\text{def}}{=} \frac{-1}{N}\sum_{i=1}^{\widetilde{N}(\mathcal{V})} \frac{\widehat{\widetilde{u}}_i^{q-1}(\epsilon)}{\widehat{Z}_B(\epsilon)} \log \widehat{\widetilde{u}}_i(\epsilon) \tag{4.140}$$

$$\widehat{\widetilde{\Upsilon}}_q(\mu, q \mid \epsilon) \stackrel{\text{def}}{=} \frac{-1}{N}\sum_{i=1}^{\widetilde{N}(\mathcal{V})} \frac{\widehat{\widetilde{u}}_i^{q-1}(\epsilon)}{\widehat{\widetilde{Z}}_B(\epsilon)} \log \frac{\widehat{\widetilde{u}}_i^{q}(\epsilon)}{\widehat{\widetilde{Z}}_B(\epsilon)} \tag{4.141}$$

The equations (4.140) and (4.141) are the new results. It is particularly important to notice the asymmetry in the exponents of the $\widehat{\widetilde{u}}_i$ in (4.141) as compared to the completely symmetric HP case (4.127). As a result of this loss of symmetry, (4.141) is no longer an explicit Shannon entropy. It is recommended for future work that more analysis be done on the effects and interpretation of the asymmetric form of $\widetilde{\Upsilon}_q$ in (4.141). Recall that the benefit of using these new direct forms is that uncertainties from the slope-fitting of the box-counting estimation can be applied to the Mandelbrot spectrum variables.

The application of these techniques, including a simultaneous estimation of the Rényi, Hölder, and Mandelbrot generalized entropies, is shown in the next section.

4.7.4 Example of Complete CP-based Attractor MFA

To demonstrate the validity of the new direct CP-based MS, and also to demonstrate its benefits, an experiment with the Ikeda attractor is now presented. A sequence of $N = 96 \times 2^{10}$ points were drawn from the Ikeda map. The x_1 component (4.95) was then lag-embedded into $\mathbb{R}^5$ with a lag of $\Delta n = 1$. A CP was applied to the reconstructed attractor for $\epsilon \in [2^{-2}, 2^{-9}]$ with

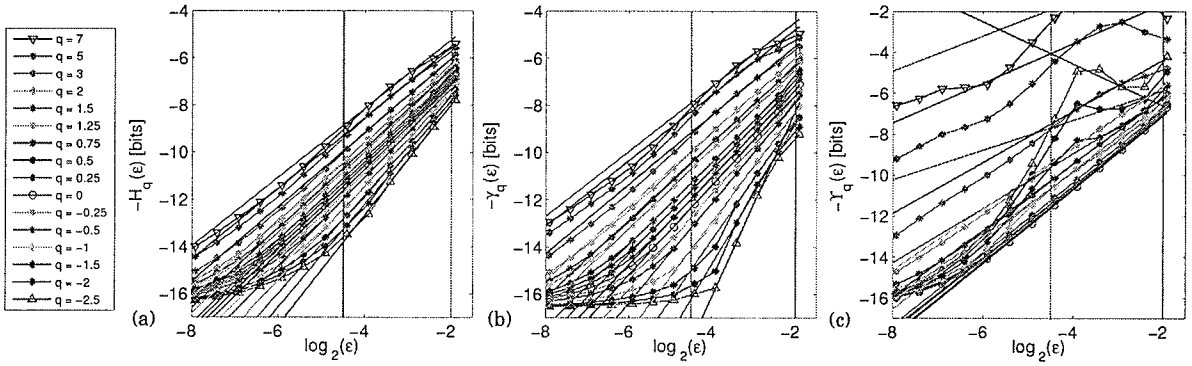


Fig. 4.21 Entropy scaling of 5D Ikeda pseudo-attractor reconstruction: scaling plots of (a) $-H_q(\epsilon)$ vs. $\log\epsilon$, (b) $-\gamma_q(\epsilon)$ vs. $\log\epsilon$, and (c) $-\Upsilon_q(\epsilon)$ vs. $\log\epsilon$ over the scaling region $\epsilon \in [2^{-4.5}, 2^{-2}]$. Approximating the D_q and $D_0(\alpha)$ spectra by fitting the slopes in (a)-(c), the calculated spectra are shown in Fig. 4.22.

$W_T = 1$ to prevent self-counting only, since the Ikeda map trajectories are highly uncorrelated. From the CP, the generalized entropies were calculated and are shown in Fig. 4.21. Numerical line-fitting to the equations produced a good slope estimate for $q = (-2, \ldots, 7)$ over a scaling region of about 2.5 octaves. The slopes of the linear fits in Fig. 4.21 provide the experimental estimates of the D_q and $D_0(\alpha)$ multifractal spectra, and also estimates of their uncertainty. The slope value is taken from a minimum absolute deviation fit [180, Sec. 15.7], but the uncertainty in the slope is measured using the least-squares estimate [180, Sec. 15.2] for simplicity. The line was fit to the functions in the scaling region common to all three direct entropies of Fig. 4.21, and the red lines in the figure indicate the scaling interval endpoints. Note that for $q \in [0, 2]$, the scaling of the entropies persist below the chosen interval. The limited scaling range for other q is consistent with the increased statistical sensitivity that occurs for small and large Boltzmann probabilities $\widetilde{u}_i$ that are weighted by these q. In order to provide a sufficient scaling region for the large $|q|$ used in this experiment, the number of samples N used in defining the Ikeda trajectory was chosen quite large. (For example, an estimation of only the correlation dimension, $D_q\big|_{q=2}$, could use a much larger scaling region, or many fewer

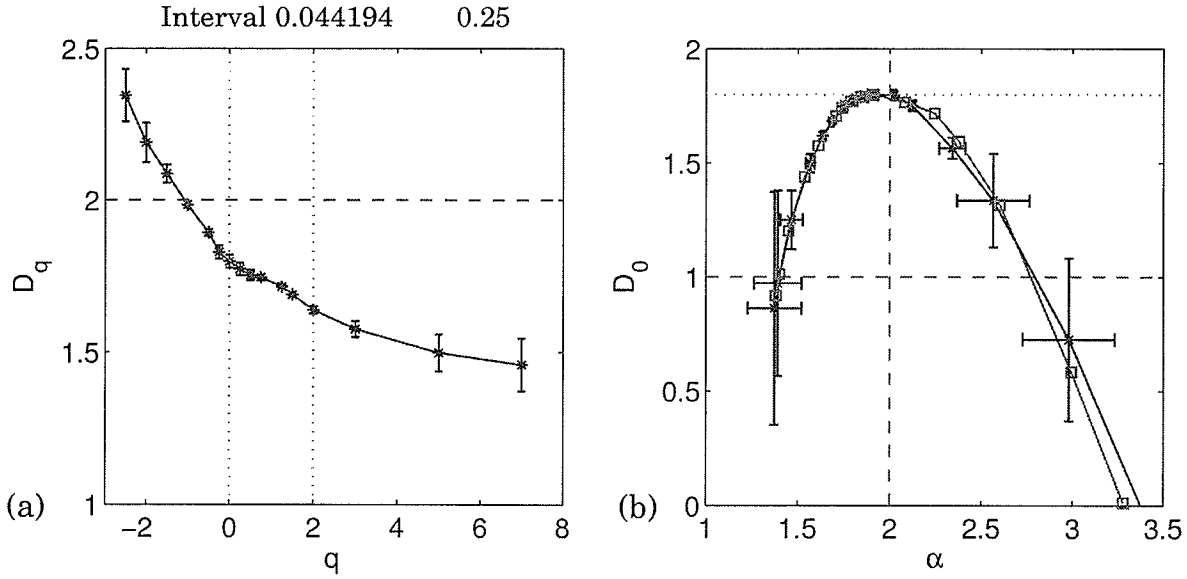


Fig. 4.22 MFA of 5D Ikeda pseudo-attractor reconstruction: (a) Rényi spectrum direct from the CP with errors; (b) Mandelbrot spectrum direct from the new CP formulas with errors (dark solid). The indirect form of $D_0(\alpha)$ through the Legendre transform of panel (a) is also shown (grey squares). The new direct MS technique is in agreement with the classical indirect technique. Note that errors are estimated from the linear fits in Fig. 4.21 and are not absolutely representative. They do indicate, however, the non-uniformity of the uncertainty in the spectrum values.

points.) Further research is required to identify the minimal number of points $N = N_{\min}(\epsilon, q)$ that would be required to maintain a scaling region. (In all likelihood, the effective functional form of $N_{\min}(\epsilon, q)$ is not universal, but unique to each attractor.)

The Rényi fractal dimension spectrum D_q is shown in Fig. 4.22(a). The least dense portions of the attractor, weighted more strongly for $q < 0$, exhibit a scaling dimension D_q that surpasses 2. At the other end, D_q drops near 1.5. Since the Ikeda attractor has a fractal support in the plane, it should have a Hausdorff dimension strictly less than 2. Thus the estimates $D_q\big|_{q=0}$, roughly around 1.8, and $D_q\big|_{q=2}$, roughly 1.64, are consistent with the theory. The smallest uncertainties are in the interval $q \in [1, 2]$, and as $|q|$ increases, so does the uncertainty in the dimension estimate.

The Mandelbrot singularity spectrum $D_0(\alpha)$ is shown in Fig. 4.22(b). It has the traditional cap-convex form, and is supported by Hölder exponents $\alpha \in [1, 3.5]$. The maximal exponent $D_0(\alpha)$ is in agreement with $D_q\big|_{q=0}$, around 1.8. Since the MS is estimated parametrically, error bars are applied both horizontally and vertically. As with the RS, the smallest errors occur in the middle of the curve, near the cap. The most singular portions of the attractor ($q > 1$) appear to the left (small α), and notwithstanding large errors, this has positive dimension $D_0(\alpha)$. This is not a conclusive feature however, since it is not taken in the limit of $q \to \infty$, ($\max q = 7$). Estimates for least singular sets A_α appear on the right ($q < 0$) and are actually reported as negative, with very large error bars, which may indicate that $D_0(\max \alpha)$ would be near zero.

To validate the new direct MS calculation, the D_q curve in Fig. 4.22(a) is transformed using the Legendre transform (4.121) and (4.122). This indirect estimate for the $D_0(\alpha)$ spectrum is shown in Fig. 4.22(b) as grey squares. This curve does not have a simple uncertainty estimate as does the direct technique. Note that the $D_0(\alpha)$ estimate provided by the new formulae is consistent with the indirect transform approach.

It should be emphasized however, that the error bars in Fig. 4.22 are a convenience and do not mathematically represent an *absolute error* for the fractal dimensions. The box-counting formalism, which measures an exponent by the slope of a linear scaling region (in a double-log plot), is known to have a complicated and significant bias as an estimator of the "true" dimensions (using either HP or CP) [71, Grassberger (1988)][108, Kantz and Schreiber (1997) Sec. 11.3.1].[5] Instead, the error bars shown here are a useful representation of the *goodness-of-fit* of the box-counting slope. Since the multifractal spectra consist of many slopes, and the curvature properties of the scaling entropies in Fig. 4.21 are not uniform, confidence in these estimates is not homogeneous. It is emphasized in the literature [108][222] that a *scaling*

[5] Other estimators, such as the Hill estimator, are considered better power-law estimators in a statistical sense, but these apply to *samples from a power-law distribution*, and not values of a power-law *function* itself.

region must exist for any credence to be placed in an MFA, yet the practice in research to justify and display this region is less than inspiring. This work will attempt to justify a scaling by a visual inspection of all entropies. Since the number of these figures can be overwhelming when convergence is being considered, it is a convenience to at least represent some useful measure of confidence in the spectrum itself. Thus, an indicator (in the spectrum) of a poor scaling region (in the entropies) would be consistently large error bars throughout the entire MFA spectrum. These error bars are, therefore, useful in a *relative* sense. This is particularly important for an analysis of convergence in an MFA.

The next section discusses how the convergence of MFA spectra can be evaluated quantitatively to provide SQMs for ICA.

4.8 MFA-Based Separation Quality Metrics

The goal for defining these ECG scaling features is to characterize the convergence of ICA estimates to the original signals, which requires *relative* measurements of these features. As described in Ch. 2, the key features under investigation are those of the fetal ECG (*i.e.*, the source component of the weakest observed signal). This thesis is specifically, therefore, looking to compare the fractal spectra of the true fetal ECG signal (RS_s, MS_s) to that of the ICA processed signal ($RS_{\widehat{s}}$, $MS_{\widehat{s}}$). This is approached in two ways: (i) relative metrics of MFA scaling, and (ii) the MFA scaling of relative metrics.

4.8.1 Relative Metrics of Scaling Spectra

The Rényi and Mandelbrot spectra have been defined as multifractal features. The goal here, however, is to use these features as performance metrics for the separation of a signal into its independent sources. This requires *relative scaling metrics* to be defined from these features.

The simplest relative multifractal metrics involve the use of error metrics on the *difference* between the spectra. That is a "relative metric of the MFA scaling spectra", such as

$$\Lambda_p\left[D_q\right] = \left|\widehat{D_q} - D_q\right|^p \tag{4.142}$$

as a real function over q with $p \geq 1$, or as a scalar

$$\overline{\Lambda_p}\left[D_q\right] = \left(\int_{\mathbb{R}} \left|\widehat{D_q} - D_q\right|^p dq\right)^{\frac{1}{p}} \in \mathbb{R}^+,\ p \geq 1 \tag{4.143}$$

among even others such as

$$\overline{\Lambda_\infty}\left[D_q\right] = \sup_q \left|\widehat{D_q} - D_q\right| \in \mathbb{R}^+ \tag{4.144}$$

Since the consideration here is the convergence of error metrics, any of these behave equivalently on the smooth Rényi spectra.[6] Different values of p simply change the shape of the convergence profile. For simplicity sake, $\overline{\Lambda_2}\left[D_q\right]$ will be chosen for its Euclidean character and $\overline{\Lambda_\infty}\left[D_q\right]$ will be chosen since that is the natural norm for the space of Rényi spectra (which are bounded functions on $\mathbb{R}$). These are used in Ch. 6 as MFA-based SQMs under the acronyms MSE-RS and CHE-RS respectively.

How to express the convergence of the Mandelbrot spectrum is more ambiguous, since its support in α is variable. One could use the parametric form in q and thus describe

$$\Lambda_p\left[\alpha, D_0\right] = \frac{1}{2}\left|\widehat{D_0}(q) - D_0(q)\right|^p + \frac{1}{2}\left|\widehat{\alpha}(q) - \alpha(q)\right|^p \tag{4.145}$$

as a real function over q with $p \geq 1$, or the scalar

$$\overline{\Lambda_p}\left[\alpha, D_0\right] = \frac{1}{2}\left(\int_{\mathbb{R}} \left|\widehat{D_0}(q) - D_0(q)\right|^p dq\right)^{\frac{1}{p}} + \frac{1}{2}\left(\int_{\mathbb{R}} \left|\widehat{\alpha}(q) - \alpha(q)\right|^p dq\right)^{\frac{1}{p}} \in \mathbb{R}^+,\ p \geq 1 \tag{4.146}$$

[6] Theoretically speaking, that is. Estimated RS may not be as smooth.

It is also possible to have

$$\overline{\Lambda_\infty}\left[\alpha, D_0\right] = \frac{1}{2}\sup_q\left|\widehat{D_0}(q) - D_0(q)\right| + \frac{1}{2}\sup_q\left|\widehat{\alpha}(q) - \alpha(q)\right| \in \mathbb{R}^+ \tag{4.147}$$

This is much less intuitive than the RS formulation.

4.8.2 Relative Multifractal Spectra

Another method for the calculation of MFA convergence is the "scaling spectra of *relative* entropies": that is, the scaling of a *relative partition function* is analyzed. This is feasible considering the relationship of Rényi entropies H_q and dimensions D_q to partition functions. The reader has already seen the generalized Kullback-Leibler divergence (Def. 3.20), here written in discrete form

$$\Re_{\mathrm{KL}q}\left(u_1 \| u_2\right) = \frac{1}{1-q}\log\sum_i u_1(i)\left(\frac{u_1(i)}{u_2(i)}\right)^{q-1} \tag{4.148}$$

If the relationship of Rényi generalized entropy with the Boltzmann partition function,

$$H_q(\mu, \mathcal{V}_\epsilon) = \frac{1}{1-q}\log Z_B(\mu, q \mid \mathcal{V}_\epsilon) \tag{4.149}$$

is extended to a relative case, namely

$$\Re_{\mathrm{KL}q}\left(\mu_1 \| \mu_2, \mathcal{V}_\epsilon\right) = \frac{1}{1-q}\log\Re_{ZB}(\mu_1, \mu_2, q \mid \mathcal{V}_\epsilon) \tag{4.150}$$

then between (4.148) and (4.150), a "relative Boltzmann partition" $\Re_{ZB}(\mu_1, \mu_2, q \mid \mathcal{V}_\epsilon)$ can be defined as

$$\Re_{ZB}(u_1, u_2, q) = \sum_i u_1(i)\left(\frac{u_1(i)}{u_2(i)}\right)^{q-1} \tag{4.151}$$

By analyzing the scaling behaviour of coarse-grainings $\Re_{ZB}(\mu_1, \mu_2, q \mid \mathcal{V}_\epsilon) \sim \epsilon^{\tau_\Re(q)}$ as has been described in this chapter, an entire class of relative scaling metrics can be defined.

Definition 4.24 (Relative Rényi Spectrum [48, Dansereau and Kinsner (2001)]). The *relative Rényi dimension spectrum* is the characteristic function $(q, \Re_{D_q})$ of μ_1 and μ_2 defined from the scaling (linear log-log relationship)

$$\Re_{D_q} \approx \frac{-1}{1-q} \frac{\log \Re_{ZB}(\mu_1, \mu_2, q \mid \mathcal{V}_\epsilon)}{\log \epsilon} \quad \forall\, q \in \mathbb{R} \tag{4.152}$$

Since the MS can also be estimated directly from entropies under a histogram partition, a similar approach can be made for a *relative Mandelbrot spectrum.*

Definition 4.25 (Relative Mandelbrot Spectrum). The *relative Mandelbrot dimension spectrum* is the characteristic function $(\Re_\alpha, \Re_{D_0})$ of μ_1 and μ_2 defined from the scaling (linear log-log relationship)

$$\Re_\alpha(q) \approx \frac{1}{\log \epsilon} \sum_i \mu_1(V_\epsilon(i)) \frac{\left(\frac{\mu_1(V_\epsilon(i))}{\mu_2(V_\epsilon(i))}\right)^{q-1}}{\Re_{ZB}(\mu_1, \mu_2, q \mid \mathcal{V}_\epsilon)} \log \frac{\mu_1(V_\epsilon(i))}{\mu_2(V_\epsilon(i))} \tag{4.153}$$

$$\Re_{D_0}(q) \approx \frac{1}{\log \epsilon} \sum_i \mu_1(V_\epsilon(i)) \frac{\left(\frac{\mu_1(V_\epsilon(i))}{\mu_2(V_\epsilon(i))}\right)^{q-1}}{\Re_{ZB}(\mu_1, \mu_2, q \mid \mathcal{V}_\epsilon)} \log \frac{\left(\frac{\mu_1(V_\epsilon(i))}{\mu_2(V_\epsilon(i))}\right)^{q-1}}{\Re_{ZB}(\mu_1, \mu_2, q \mid \mathcal{V}_\epsilon)} \tag{4.154}$$

for all $q \in \mathbb{R}$

For the case of identical measures, $\Re_{ZB}(\mu_1, \mu_2, q)|_{\mu_i = \mu} = 1$, and hence $\Re_{D_q} = 0\ \forall\ q$ or $(\Re_\alpha, \Re_{D_0}) = (0, 0)$.

For a scalar measure, norms can be used to express the magnitude of these functions, such as

$$\sup_q |\Re_\alpha(q)| + \left|\Re_{D_0}(q)\right| \tag{4.155}$$

or

$$\sup_q \left|\Re_{D_q}\right| \tag{4.156}$$

As powerful and interesting as these methods may be, they are directly enabled by the *common partitioning* of the two measures. That is, the same ϵ-covering is applied to coarse-grain both measures simultaneously. Since HP is established in the context of the embedding space, this is easily applied. Correlation partitioning, however, is not uniform in the embedding space, but follows the manifold of the attractor itself. While this has the many benefits described in the earlier sections of this chapter, it makes it very difficult for the partitioning to be common to both measures. Since the ϵ-covering is centred on the samples of the attractor themselves, then considerable expense must be made to simultaneously reconstructed the attractors, and count the probabilities $\widetilde{u_i}$ from the first attractor's point $\boldsymbol{y}(i)$, from the points $\boldsymbol{y}'(j)$ in the second attractor that are near it. It is beyond the scope of this work to develop such an algorithm. It is recommended, however, that this be explored in future work. To assist in this, App. C, contains the details of the current developed CP algorithm for the calculation of the Boltzmann partition $\widetilde{u_i}$ as a starting point. (The code in electronic form is also available for download from `www.ee.umanitoba.ca/~kinsner/projects/`.)

A consequence, however, of (i) the high-embedding dimensions required for ECG MFA, and (ii) the central focus of CP in this work, entails that the application of histogram-based relative MFA spectra is inappropriate. Thus, at least at the current level of investigation, SQMs for the convergence of ICA using MFA must focus on relative metrics of scaling spectra such as $\Lambda_p[\cdot]$ and $\overline{\Lambda_p}[\cdot]$. Further discussion of the application of these metrics will be presented in Ch. 6.

4.9 Summary

This chapter has presented the background required to define and understand multifractal features of ECG for the use as SQMs. Attention has focused here on the application of

correlation-based partitions to reconstructed ECG attractors. (Another time-frequency approach to ECG MFA is briefly reviewed in App. D, but is not implemented in the experiment design.) This attractor formalism emphasizes the cyclostationary variability of the ECG dynamics.

One contribution of this chapter is its non-conventional approach to the synthesis of MFA theory. In general, the synthesis of the background is pedagogical and not historical. First, the scaling properties of general measurement provides a motivation for the consideration of characterizing scaling. Second, all discussion of fractal scaling was presented in a measure-theoretic sense. Third, the concept of MFA was defined as generally as possible through the scaling exponents of a partition function of a coarse-grained measure. From this perspective, the similarities and differences of the Rényi and Mandelbrot fractal dimension spectra are most clear. Specifically, the Hölder and Mandelbrot entropies are introduced to amplify the consistent formalism for direct dimension calculations. Fourth, the application of MFA to the characterization of ECG is described in the context of analysis of a reconstructed attractor from a time series.

Another contribution of this chapter is the derivation of the direct Mandelbrot spectrum from a correlation-based partition function. A specific review of the literature to identify this missing element of MFA theory is included. In the context of the previous background, the derivation is a simple adaptation to the definition of the Hölder and Mandelbrot generalized entropies. It is of particular interest that the correlation-based Mandelbrot entropy, $\widetilde{\Upsilon}_q(\epsilon)$, is not an explicit Shannon entropy (though its histogram counterpart $\Upsilon_q(\epsilon)$ is).

In practical terms, this chapter contributes to the experimental work the mechanics for the calculation of a complete correlation-based MFA. Summarized in Table 4.1, these formulae involve the calculation of entropies from point-correlations on the reconstructed attractor.

Fractal dimensions are estimated by the slope of a linear scaling region, and not in the limit of small ϵ.

The second practical contribution of this chapter is the identification of the important relative measures of MFA convergence, $\Lambda_p[\cdot]$ (a function), $\overline{\Lambda_p}[\cdot]$ (a scalar), and specifically $\overline{\Lambda_\infty}[\cdot]$. These will be the foundation for MFA-based SQMs for the measurement of feature convergence under ICA (namely the MSE error-norm, $\overline{\Lambda_2}[\cdot]$, and the CHE error-norm, $\overline{\Lambda_\infty}[\cdot]$ on p. 304).

The next chapter considers the use of surrogate data techniques to provide an "Occam's razor" for the analysis and interpretation of MFA. Specifically, these surrogate data techniques are designed to obscure the nonlinear dynamics that are present in the ECG attractor. By performing MFA on the surrogate data, one can begin to develop intuition for the contribution of nonlinear dynamics to the scaling of the MFA entropies.

Table 4.1 Practical summary of CP formulae for the estimation of MFA*

Correlation partition (CP) and parameters:

$$N_{W_T} = 2W_T - 1 \tag{4.157}$$

$$N_i(\epsilon) = \sum_{|j-i| \geq W_T} \Xi(\epsilon - \|y(j) - y(i)\|) \tag{4.158}$$

$$u_i(\epsilon) = \frac{N_i}{N - N_{W_T}} \tag{4.159}$$

$$\mathcal{P} = \{ u_i(\epsilon) \mid i = 1, \ldots, N; \ \forall \epsilon \} \tag{4.160}$$

CP partition function and entropies:

$$Z_B(\epsilon) = \frac{1}{N} \sum_{i=1}^{N} u_i^{q-1} \quad \forall q \tag{4.161}$$

$$H_q(\epsilon) = -\frac{1}{q-1} \log Z_B \quad \forall q \tag{4.162}$$

$$\gamma_q(\epsilon) = -\frac{1}{N} \sum_{i=1}^{N} \frac{u_i^{q-1}}{Z_B} \log u_i \quad \forall q \tag{4.163}$$

$$\Upsilon_q(\epsilon) = -\frac{1}{N} \sum_{i=1}^{N} \frac{u_i^{q-1}}{Z_B} \log \left(\frac{u_i^q}{Z_B} \right) \quad \forall q \tag{4.164}$$

Direct box-counting scaling estimates:

$$D_q \approx \frac{-H_q(\epsilon)}{\log \epsilon} \quad \forall q \tag{4.165}$$

$$\alpha \approx \frac{-\gamma_q(\epsilon)}{\log \epsilon} \quad \forall q \tag{4.166}$$

$$D_0 \approx \frac{-\Upsilon_q(\epsilon)}{\log \epsilon} \quad \forall q \tag{4.167}$$

The correlation sum is one value of the CP Boltzmann partition function, which is particularly easy to calculate since it collapses to the linear form

$$Z_C(\epsilon) \stackrel{\text{def}}{=} Z_B(\epsilon)|_{q=2} = \frac{1}{N(N - N_{W_T})} \sum_{i=1}^{N} N_i(\epsilon) \tag{4.168}$$

$$= \frac{2}{N(N - N_{W_T})} \sum_{i=1}^{N} \sum_{j \geq i + W_T} \Xi(\epsilon - \|y(j) - y(i)\|) \tag{4.169}$$

* Here, the estimation modifier $\widehat{\cdot}$ and variant modifier $\widetilde{\cdot}$ that appear in the chapter text are suppressed.

Chapter V

Study of Surrogate Data for MFA Validation

5.1 Overview

This chapter presents a study of surrogate data for the validation of the multifractal features described in Ch. 4. Surrogate data can improve the interpretation of MFA results substantially, and thus is a significant element in the "proper" application of MFA. By justifying an MFA through surrogate data, fractal-based SQMs becomes more objective, and their value as performance metrics increases.

To begin the chapter, the origins of surrogate data are reviewed as a means to distinguish between stochastic and chaotic deterministic systems (*i.e.*, chaos or noise). Then the three canonical hypotheses for nonlinear surrogate data are reviewed in Sec. 5.3. Included in Sec. 5.3.4 are examples of these surrogate models on deterministic, stochastic, and natural HRV data. Section 5.3.5 presents some simple but important considerations about avoiding periodicity artifacts in the application of surrogate data techniques based on the *fast Fourier transform* (FFT). Since ECGs are clearly quite distinct from canonical stochastic models, Sec. 5.4 reviews two techniques in the literature that are specifically applicable for cyclostationary signals. Using the techniques of App. A, a novel surrogate data approach specific to synthetic ECG is described in Sec. 5.4.3. Specifically, this approach provides an evaluation of the influence of HRV on an ECG's MFA. Last, Sec. 5.5 explains how the surrogate data models are used in the experimental design for the validation of MFA.

5.2 Time Series Dynamics, Null-Hypotheses, and Surrogate Data

The techniques for multifractal analysis described in the previous chapter are useful in the characterization of nonlinear systems. There is ambiguity, however, when these techniques are applied to real data recorded from sources with unknown dynamics. As estimators, an MFA using the Rényi fractal dimension spectrum or Mandelbrot singularity spectrum produces numerical values with very complicated contributions from bias and variance. Thus the interpretation of the MFA values is somewhat circumspect in absolute terms and are best considered in a relative way. In particular, the MFA dimensions estimated by log-log slopes require scaling regions where the relationship is linear. In both mathematical and real data, the linearity of an appropriate scaling region is subject to user interpretation due to either oscillation or curvature. Some true mathematical multifractals have entropy scaling with discontinuities as in Fig. 5.1, and the dimension estimates are in fact the average linear slope of the curve. Here, the correct interpretation is that observed scaling relationships are composed of a linear baseline with sawtooth-type oscillations, Fig. 5.1(b). In other scenarios with dynamical data, the log-log scaling relationship of entropies is often subject to curvature [102], such as shown in Fig. 5.2. Considering these artifacts appear regularly in MFA estimation, how is one to know whether MFA is an appropriate characterization of the data? Is variability in the signal from chaotic dynamics, or simply noise corruption? [119]

During the 1990s, this question was posed in earnest, and the dominant solution drew from the null-hypothesis approach used in statistical inference [225][226]. In statistical inference, the conformance of any real data to a hypothesized model could be quantified in a statistical way using (i) a discriminating statistic and (ii) bootstrapping techniques for data resampling, otherwise known as generating *surrogate data*. More specifically, any given dataset can be

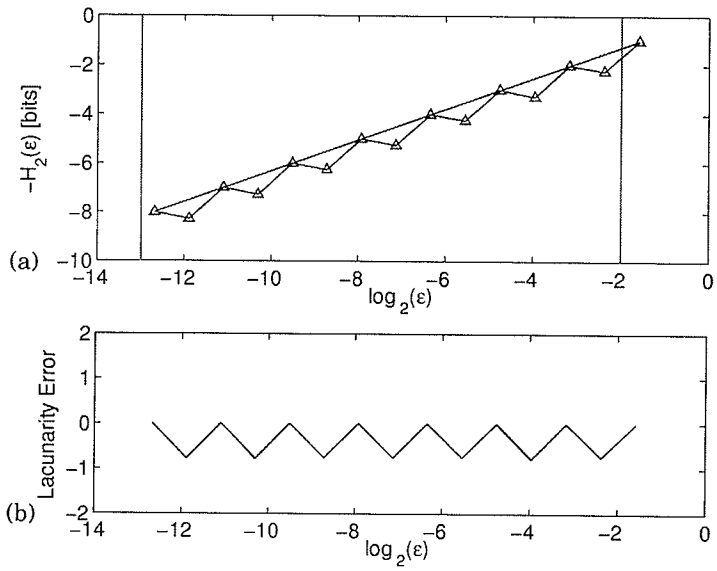


Fig. 5.1 Scaling entropy nonlinearities — oscillation: (a) Scaling of quadratic entropy (H_2) of the Cantor set C has oscillatory behaviour because of the "zero-mass" (*i.e.*, deleted) sections of the fractal. The proper fractal dimension is the slope of a linear fit; (b) The "sawtooth" residual from the linear fit represents an oscillatory *lacunarity error*.

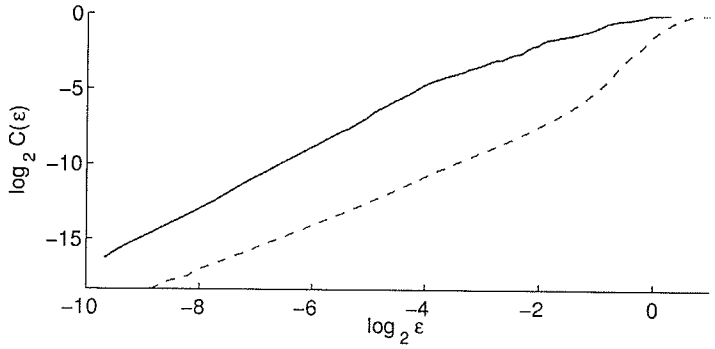


Fig. 5.2 Scaling entropy nonlinearities — curvature: Correlation sum scaling of 7-dimensional reconstructed embedding of the Rössler attractor in chaotic motion (solid) and the Ikeda attractor (dashed). At larger scales, both characterizations exhibit curvature, notwithstanding their noise-free mathematical origins. In conditions with realistic noise contamination and uncertain dynamics, curvature effects can become even stronger. (Note the Rössler time series here is the z-component with control parameter $c = 5.0$, while the Ikeda trajectory is its real component with standard parameters.)

compared (with a discriminating statistic) to a set of "typical realizations" (the surrogate data) of a particular uninteresting model (the null-hypothesis model). This provides a statistical method for the application of Occam's razor to the problem. If the test statistic calculated from the real data differed "strongly" from the distribution of statistics calculated from the surrogate data, then it is "unlikely" that the real data conforms to the null-hypothesis model. Under these circumstances, it is reasonable to "reject" the null-hypothesis, which, while not a proof that the null-hypothesis is false, provides a justification for the consideration of more complex models. A failure to reject the null-hypothesis, on the other hand, does not imply that the null-hypothesis is true, but simply that the features of the data (specifically, the discriminating statistic) are reasonably expected under the simple model. This could be because (i) the null-hypothesis model is appropriate, or (ii) the test statistic is not sufficiently discriminating. In either case, a failure to reject the null-hypothesis demonstrates the feature of interest is present in the context of the null-hypothesis, and more complex models should be considered unjustified. Further arguments about quantifying the level of inference (*i.e.*, the size and power) of the null-hypothesis test need not be considered for this analysis, but may be found in the literature [226][203][204].

Theiler *et al.* considered this approach for the justification of nonlinear time series analysis. Here, the distinction of interest is whether variability in the time series is a result of dynamical or stochastic influence. Therefore, the generic null-hypothesis in time series analysis concerns stochastic models. For discriminating statistics, the nonlinear features extracted from time series can be considered. The canonical surrogate data method that is discussed in the next section describes three Monte-Carlo techniques for "constrained realizations" for nonlinear time series analysis [226][204].

5.3 Canonical (Linear) Surrogate Data Models

The canonical null-hypothesis models for nonlinear surrogate data are three stochastic systems, namely

Type-0 surrogate: an iid stochastic process;

Type-1 surrogate: a time-invariant Gaussian process (*i.e.*, filtered Gaussian noise);

Type-2 surrogate: a static monotonic nonlinear mapping of a time-invariant Gaussian process (*i.e.*, warped filtered Gaussian noise).

For each model, an algorithm is required by which surrogate time series consistent with the aforementioned models can be generated, while also preserving the surrogate's "likeness" to the original data. In its original formulation [226], these algorithms are Monte-Carlo methods that generate "constrained realizations", rather then "typical realizations", consistent with the model. Though the detailed distinction between the two is left for the reader in [226][204], it suffices to say that the finite "constrained realizations" preserve the likeness of the original finite time series data better, and do not rely on asymptotic arguments. Therefore, the surrogate data algorithms are effectively distinguished by which specific stochastic feature of the original time series is chosen to be preserved (the likeness), and what randomization is applied to destroy any deterministic relationships that exist. In fact, it becomes very difficult to write in closed form what the null-hypothesis model class actually is as more general surrogate techniques are considered [204]. Thus the focus is on distinguishing the null-hypothesis models practically by the constraints and randomizations of the techniques instead. According to this format, the canonical surrogate data algorithms are distinguished by the following

Type-0 surrogate: preserves the distribution (*i.e.*, statistical rank) and randomizes time-order;

Type-1 surrogate: preserves the power-spectrum (*i.e.*, linear dependencies) and randomizes phase;

Type-2 surrogate: preserves the distribution and power-spectrum while randomizing phase and time-order.

The algorithms are discussed in greater detail in the following subsections.

5.3.1 Type-0 Surrogates

The type-0 surrogates of a time series $x(n)$, written $\pi_0[x](n)$, are surrogates under a stochastic iid null-hypothesis. In order to keep the distribution of the surrogates identical to that of the $x(n)$, they must have the same rank order. This is achieved simply by taking a permutation of the values of $x(n)$. That is, letting Pm be a permutation operator on N elements, where N is the length of the time series x,

$$\pi_0[x] = \mathrm{Pm}\{\, x(n) \mid n = 1, \ldots, N \} \tag{5.1}$$

5.3.2 Type-1 Surrogates

The type-1 surrogates of a time series $x(n)$, written $\pi_1[x](n)$, are surrogates under a stochastic linear Gaussian process null-hypothesis. It is equivalent to linear dynamics represented in a stochastically driven linear system, or *auto-regressive moving-average* (ARMA) model. In order to keep the linear properties of the system the same, the constraints are applied in the Fourier domain. Specifically, the power spectrum of the surrogates will be identical to that of the $x(n)$ (and thus, so is the autocorrelation function). This is achieved by taking an FFT of the data $x(n)$ and representing it in polar form. Representing its Fourier magnitudes as $|X(\omega_n)| = |\mathcal{F}_x(\omega_n)|$ and its phase as $\exp(i\theta_n)$, the common constraint on the surrogates should be that $\left|\mathcal{F}_{\pi_1[x]}(\omega_n)\right| = |X(\omega_n)|$. Under the null-hypothesis model, the surrogate's phase θ'_n is

consistent with a uniformly distributed random variable over $[0, 2\pi)$, and several techniques for randomization have been suggested. First, the original phase sequence $\{\theta_n\}$ can be simply thrown away and replaced with realizations from a uniform random variable on $[0, 2\pi)$ [204]. Second, the original phase elements can be randomized by shuffling the values of θ_n [206]. In either case, care has to be taken to preserve the skew-symmetry required in the phase of real valued time series [181]. A third technique preserves the phase skew-symmetry in a natural way by considering the phase of a real signal. Here, the Fourier magnitudes $|X(\omega_n)|$ are matched with the phase $\exp(i\theta'_n)$ of a type-0 surrogate. With all the techniques, the type-1 surrogate time series is then defined by the inverse FFT of the element-wise product $|X(\omega_n)| \exp(i\theta'_n)$.

In this work, the algorithm of choice uses the third option, matching a surrogate phase to the natural power spectrum so that the surrogate is defined as

$$\pi_1[x] = \mathcal{F}^{-1}\{|X(\omega_n)| \exp(i\theta_n)\} \tag{5.2}$$

where

$$X(\omega_n) = \mathcal{F}_x(\omega_n) \tag{5.3}$$

$$Y(\omega_n) = \mathcal{F}_{\pi_0[x]}(\omega_n) = |Y(\omega_n)| \exp(i\theta_n) \tag{5.4}$$

A few things should always be considered when the type-1 surrogate algorithm is applied. First, since the discrete Fourier transform (as an FFT) is applied to the finite time series $x(n)$, the Fourier representation $X(\omega_n)$ is actually for the periodic extension of $x(n)$ and not an estimate of the Fourier transform. (It is also possible to consider $x(n)$ as wrapped on the unit circle.) Specifically, windowing is not used to suppress end-point mismatch, since an inverse FFT is also applied. Therefore, any discontinuity or mismatch at the end-points introduces

spurious high frequencies into the surrogates. That is, the type-1 surrogate technique preserves the frequency contribution of the end point mismatch, but also distributes it throughout the time series $\pi_1[x](n)$, and does not isolate it as a mismatch at the end-points. When the end-point mismatch is considerable, this means the surrogates may not properly represent the linear features of $x(n)$, as usually interpreted in the asymptotic Fourier transform sense. A discussion on how to mitigate this problem is presented in a later section of this work.

The second thing to consider with the application of a type-1 algorithm is that, irrespective of the original time-integrated distribution of $x(n)$, the time-integrated distribution of $\pi_1[x]$ is essentially Gaussian. This is a result of the inverse FFT operation on random phases, and is essentially a byproduct of the central limit theorem.

Neither of these two considerations undermines the application of the type-1 algorithm, however. It simply limits its proper application to signals $x(n)$ that are at least somewhat consistent with the Gaussian null-hypothesis model. A trivial rejection of this null-hypothesis (without surrogate data) is possible if (i) the time-integrated distribution of $x(n)$ is strongly non-Gaussian, or (ii) $x(n)$ is too smooth to be drawn from a linear stochastic process. If $x(n)$ is sufficiently nonstationary, however, then mitigation techniques (such as interval selection) might be key in reducing the end-point mismatch artifact. Alternatively, an extension to the surrogate technique could be considered, where only the phases in the higher frequency band are randomized, thus preserving the low-frequency nonstationary behaviour.

5.3.3 Type-2 Surrogates

One of the limitations of the type-1 surrogates is that the original time series must have a sufficiently Gaussian time-integrated distribution. Since this applies to only a small class of interesting signals, it would be suitable to have an algorithm that allows for an arbitrary distribution (such as with the type-0 surrogates). The type-2 surrogate is a hybrid of both

the type-0 and type-1 algorithms, and theoretically allows for a wide class of distribution. Suppose a static instantaneous warping function $\eta : \mathbb{R} \mapsto \mathbb{R}$, were to map a linear Gaussian process $s(t)$ into a new process $x(t) = \eta(s(t))$. Then the time-integrated distribution of $x(t)$ would not be Gaussian. Surrogates in this class can be generated by an iterative process of two stages, where (i) the Fourier magnitudes of the surrogate are constrained to the original $|X(\omega_n)|$, and (ii) the rank amplitude values of the surrogate are constrained to the original values $\{\xi_n\} = \text{sort}\{x(n)\}$. Action (i) is consistent with a type-1 surrogate, while action (ii) is a constraint equivalent to a type-0 surrogate. This iterative process continues until the reordering required for rank sorting no longer changes [202]. Details on the convergence properties of this *iterated amplitude adjusted Fourier Transform* (IAAFT) technique are discussed in [204].

Once converged, however, there is debate in the literature on which of the two actions in the algorithm should be performed last. Since the constraints applied by the actions are "perfect", the real issue here is which of the constraints should accept a residual error. Upon convergence, the "type-2.0 surrogate" finishes with the amplitude constraint, and thus preserves the distribution of the original data but leaves residual error in the power spectrum. Otherwise, the "type-2.1 surrogate" can apply the Fourier constraint one final time. Since the algorithm has converged, the rank order of the surrogate values is unchanged, but their values may have changed slightly (*i.e.*within the tolerance produced by its nearest rank values). Kugiumtzis [121, 1999] characterized some of the differences between the two alternatives. In general, nonlinear metrics are more sensitive to changes in linear correlations than in distribution [224]. Consequently, with most distinguishing nonlinear statistics, type-2.1 surrogate data would be the most appropriate.

In this work, the type-2.1 algorithm takes the Fourier magnitudes $\{|X(\omega_n)|\}$ and the ranked values $(\xi_1, \ldots, \xi_N)$ of the original data as the constraints. Let z represent the surrogate time

series of length N, where z is initialized as a type-0 surrogate to destroy any dynamics. Then the FFT of z,

$$Z(\omega_n) = \mathcal{F}_z(\omega_n) \tag{5.5}$$

$$= |Z(\omega_n)| \exp(i\theta_n) \tag{5.6}$$

has its magnitudes replaced by the desired values and is updated in the time domain, as

$$Z(\omega_n) \leftarrow |X(\omega_n)| \exp(i\theta_n) \tag{5.7}$$

$$z(n) \leftarrow \mathcal{F}^{-1} Z(\omega_n) \tag{5.8}$$

which reintroduces the linear dynamics of the surrogate model. Then the surrogate amplitudes $z(n)$ are rank ordered, achieved by the unique permutation Pm_r, such that

$$\mathrm{sort}(z(1), \ldots, z(n)) = \mathrm{Pm}_r(z(1), \ldots, z(n)) \tag{5.9}$$

The surrogate amplitudes are then updated by the ranked values of the data,

$$(z(1), \ldots, z(N)) \leftarrow \mathrm{Pm}_r^{-1}(\xi_1, \ldots, \xi_N) \tag{5.10}$$

The process of (5.6)-(5.10) is continued until the Pm_r permutation becomes the identity, at which point the algorithm has converged to the double solution represented by the outputs of (5.8) and (5.10). Since the type-2.1 surrogate is desired, the calculations (5.6)-(5.8) are performed once again and then the algorithm exits returning $\pi_{2.1}[x] = z$. For long time series many iterations may be required to achieve convergence (*e.g.*, 1600 iterations for 30K points).

5.3.4 Examples of Canonical Surrogates

Three examples representing different levels of dynamics will now be presented. First, the chaotic trajectory of the Ikeda map (discussed in Ch. 4) will be analyzed. Although an

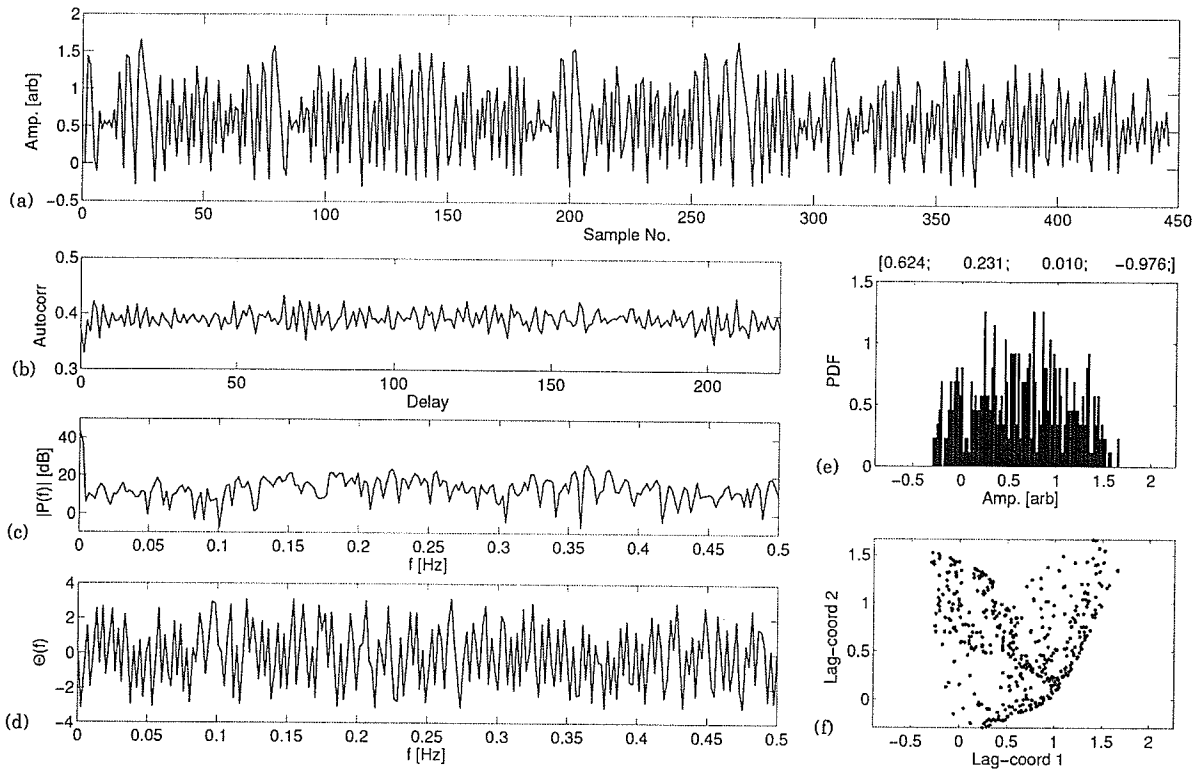


Fig. 5.3 Features of a chaotic data sequence: (a) sequence of values from the chaotic Ikeda map exhibit stationary but irregular oscillations; (b) (time-domain) autocorrelation sequence is statistically constant with non-constant variability; (c) power spectrum (assuming sampling is at "1 Hz") is statistically fairly constant, with a slight depression on the [0.010.15] Hz interval; (d) Fourier phase appears as a random variable; (e) time-integrated distribution is fairly symmetric, but sub-Gaussian; (f) return map (lag=1) features a characteristic nonlinear pattern.

essentially white process, it has strong nonlinear dynamics. Second, surrogates of a time series of pink noise will be presented, all of which have an equal lack of dynamics. Last, the surrogate models are applied to a recording of HRV, whose dynamics, though unknown, would have a weak nonlinear contribution at best.

5.3.4.1 Ikeda Chaos

The Ikeda map was used as an illustration in both Ch. 2 and 4 for a complicated nonlinear dynamical system. The nonlinear recursive equations that define the Ikeda map are written in (4.95)-(4.96) on p. 188. Here, the real (x_1) component of the Ikeda trajectory for an arbitrary initial condition is analyzed after discarding a transient buffer of 2K iterations. 450 points of an Ikeda trajectory and its basic properties are shown in Fig. 5.3, where it is assumed (since the trajectory is discrete) that the sampling frequency is 1 Hz. It can be seen that the trajectory has an essentially white spectrum, sub-Gaussian time-integrated distribution, and a nonlinear relationship in the return map (lag=1). The properties of the different surrogate algorithms can be inferred from the equivalent visualization of their properties shown in Figs. 5.4-5.7.

The type-0 surrogate of Fig. 5.4 has the identical time-integrated distribution as the original signal, and very similar autocorrelation function, power spectrum, and fourier phase. The surrogate signal itself appears less oscillatory and slightly more random than the true signal data, which is clearly distinguished in the randomized and independent return map (as expected for an iid model).

The type-1 surrogate of Fig. 5.5 has a more Gaussian time-integrated distribution than the original signal, but has a very similar autocorrelation function, power spectrum, and fourier phase. The autocorrelation function and power spectrum in panels (b) and (c) are not point-by-point exact matches to Fig. 5.3, because these displays better represent the spectral content of the time series (as an extraction of a larger observation). The autocorrelation sequence in panel (b) uses the unbiased time domain algorithm, while the power spectrum in panel (c) uses Hann windowing. Consequently, the expected contribution of the FFT periodic artifacts are measured in these panels and point-by-point matches are not expected, even though point-by-point constraints were applied (to the FFT) in the surrogate algorithm. Notwithstanding these

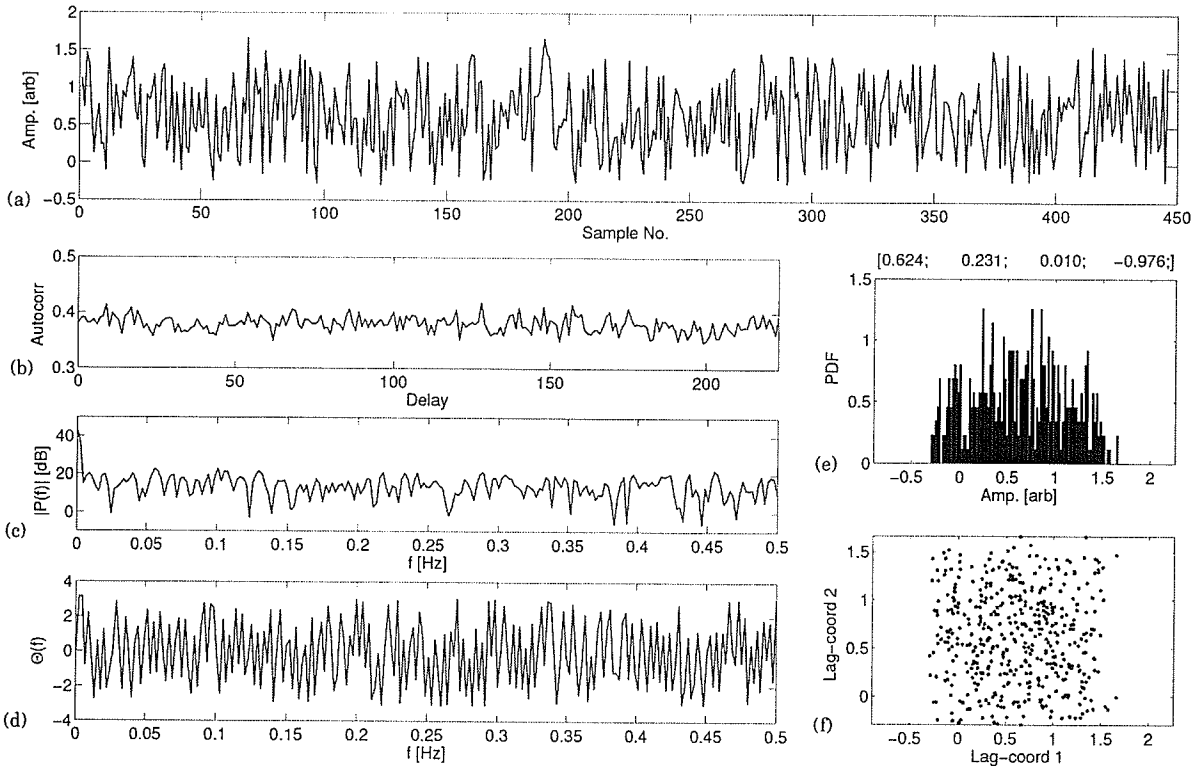


Fig. 5.4 Features of a type-0 surrogate sequence of the Ikeda mapping: (a) sequence of values are less oscillatory and more stochastic; (b) (time-domain) autocorrelation sequence is statistically constant; (c) power spectrum is statistically fairly constant, with no depression; (d) Fourier phase appears as a random variable; (e) time-integrated distribution is identical to Fig. 5.3(e); (f) return map (lag=1) appears random and independent (*i.e.*, marginal factorization).

point-by-point differences, panels (b), (c), and (d) of Fig. 5.5 preserve the statistical character of the originals in Fig. 5.3. The return map in panel (e), however, is clearly more stochastic, Gaussian, and isotropic, in keeping with the Gaussian nature of the linear null-hypothesis model for this data.

The type-2.0 surrogate shown in Fig. 5.6 is very similar in practice to the type-0 model for the Ikeda data. Again, the distribution is preserved exactly, and the return map appears stochastic, independent, and non-Gaussian. Compared to Fig. 5.4, however, the properties of

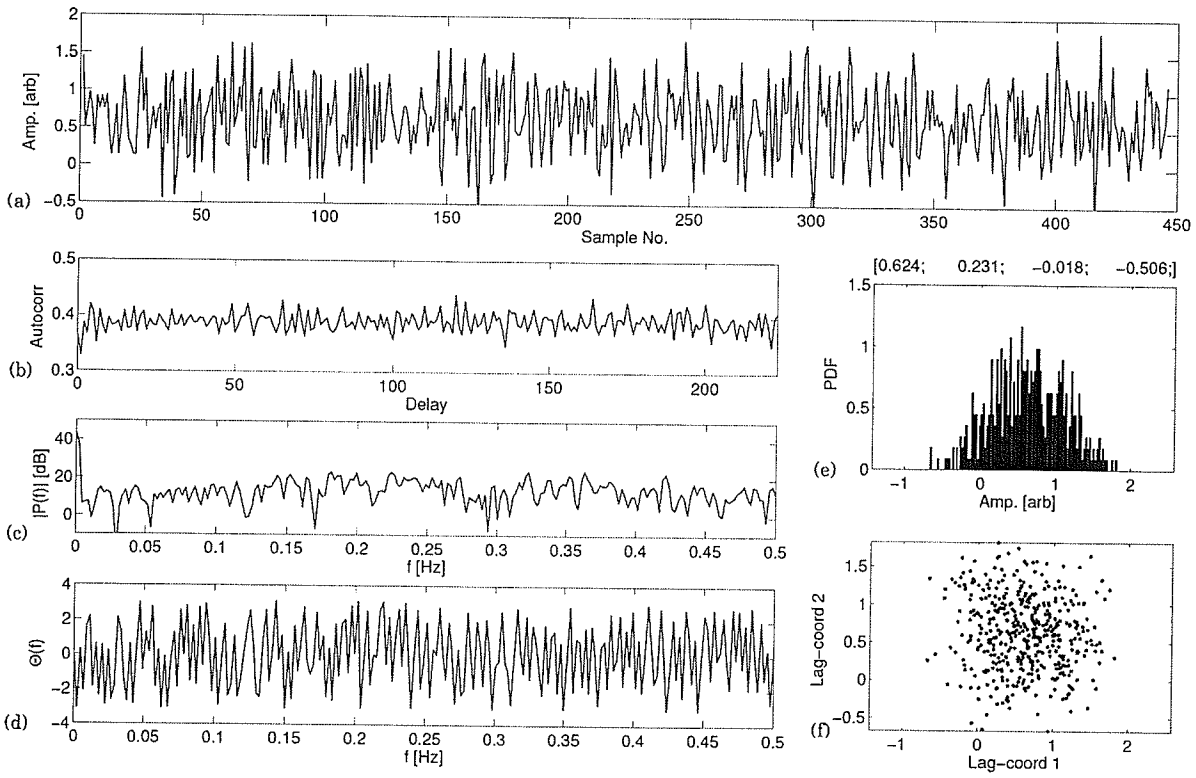


Fig. 5.5 Features of a type-1 surrogate sequence of the Ikeda mapping: (a) sequence of values is quite like the original; (b) (time-domain) autocorrelation sequence is statistically equivalent to Fig. 5.3(b); (c) power spectrum is is statistically equivalent to Fig. 5.3(c); (d) Fourier phase appears as a random variable; (e) time-integrated distribution is more Gaussian than the original; (f) return map (lag=1) appears random and more isotropic than either Fig. 5.3 or Fig. 5.4.

panels (b) and (c) are more representative of the variability in the original data, since a spectral constraint has been applied.

The strong similarities between the type-2.1 surrogate shown in Fig. 5.7 and its counterpart in Fig. 5.6 faithfully represents the commonalities in the algorithms. Their differences in panel (e) also demonstrates the difference induced by the choice of final constraint. Given that the type-2.0 surrogate gives a very good representation of the (nearly white) Ikeda spectrum, that algorithm would be preferred since it preserves the time-integrated distribution as well.

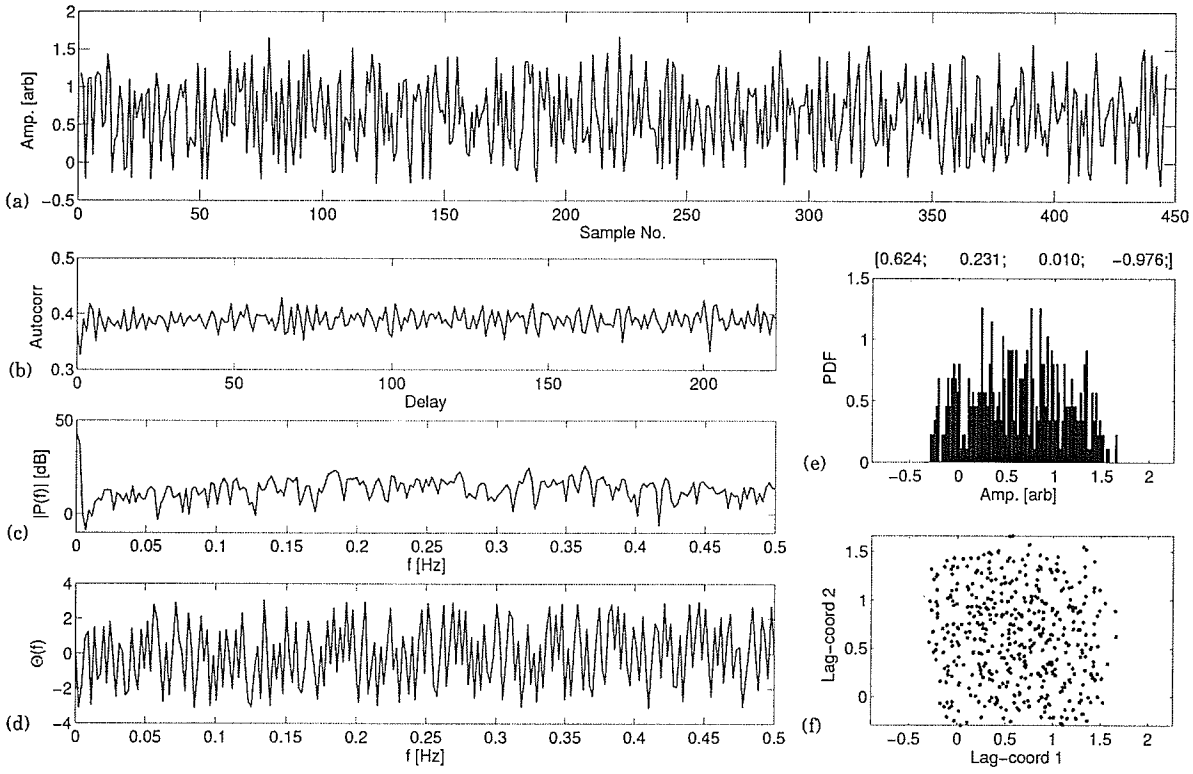


Fig. 5.6 Features of a type-2.0 surrogate sequence of the Ikeda mapping: (a) sequence of values is quite like the original; (b) (time-domain) autocorrelation sequence is statistically equivalent (and almost point-wise equal) to Fig. 5.3(b); (c) power spectrum is is statistically equivalent to Fig. 5.3(c); (d) Fourier phase appears as a random variable; (e) time-integrated distribution is identical to Fig. 5.3(e); (f) return map (lag=1) appears random but slightly less independent than Fig. 5.4(f).

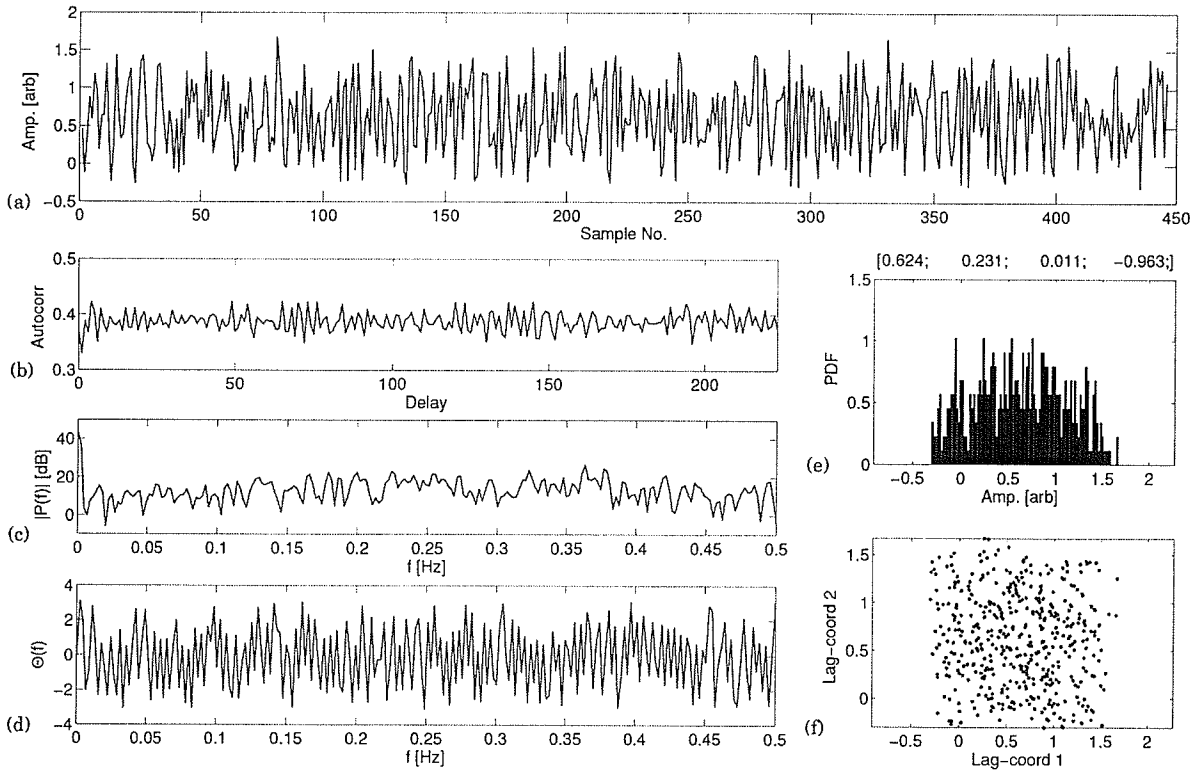


Fig. 5.7 Features of a type-2.1 surrogate sequence of the Ikeda mapping: (a) sequence of values is quite like the original and Fig. 5.6; (b) autocorrelation sequence is statistically equivalent (and almost point-wise equal) to Fig. 5.3(b) and corresponds well to Fig. 5.6(b); (c) power spectrum is is statistically equivalent to original and has better low frequency matching than Fig. 5.6(c); (d) Fourier phase appears as a random variable; (e) time-integrated distribution is slightly less sub-Gaussian than original; (f) return map (lag=1) appears random and consistent with Fig. 5.6(f).

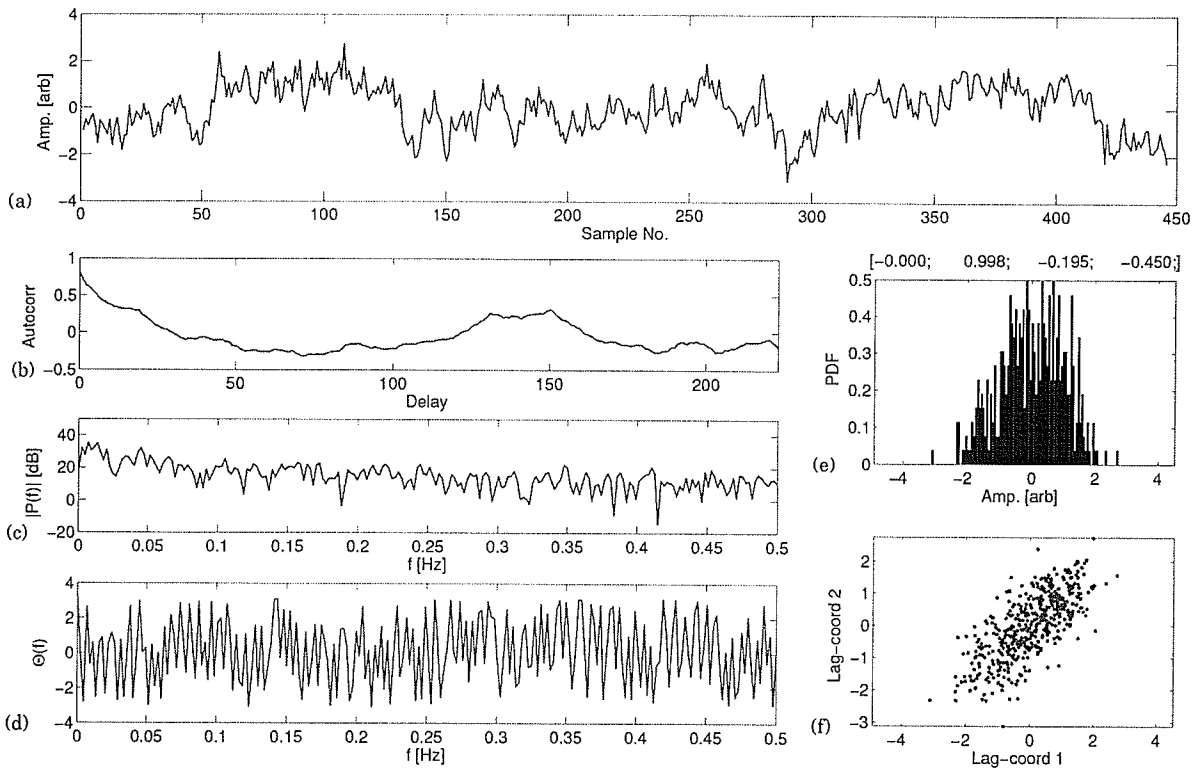


Fig. 5.8 Features of a pink noise sequence: (a) sequence of values from a linear stochastic system with power spectrum power-law decay exponent $\beta = 1$ is more nonstationary and less rough than white noise; (b) autocorrelation sequence is rather smooth and has a local maximum in the neighbourhood of a delay of 140 samples, and negative minima at 70 and 200 samples; (c) power spectrum (assuming sampling is at "1 Hz") is a decay over 20 dB with statistical fluctuations; (d) Fourier phase appears as a random variable; (e) time-integrated distribution is zero-mean, unit variance, and mildly sub-Gaussian; (f) return map (lag=2) exhibits a correlated scattering.

5.3.4.2 Pink Noise

As previously mentioned in Ch. 2, any broadspectral behaviour of a signal (*e.g.*, power-law decay) makes the signal very difficult to segment into frequency bands. These correlations over frequency bands, however, can come from nonlinear or linear dynamics. Coloured noises are a class of stochastic signals with "linear dynamics" (*i.e.*, linear dependencies) that produces a power-law in the power-spectrum. Pink noise has a spectral decay of ω^{-1}, and is common in

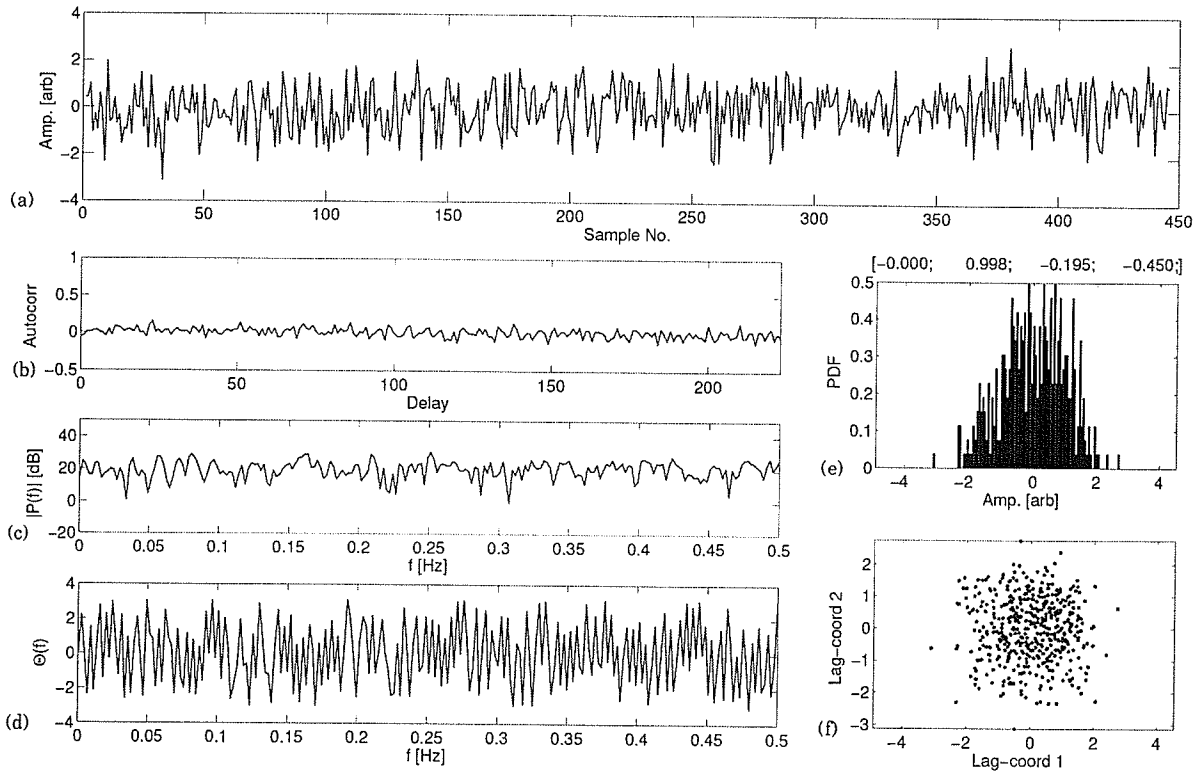


Fig. 5.9 Features of a type-0 surrogate sequence of pink noise: (a) sequence of values are stationary and more rough; (b) autocorrelation sequence is statistically constant; (c) power spectrum is statistically constant; (d) Fourier phase appears as a random variable; (e) time-integrated distribution is identical to Fig. 5.8(e); (f) return map (lag=2) appears random and independent (*i.e.*, marginal factorization).

biological systems and some relaxation processes [153]. As shown in Fig. 5.8, pink noise is mildly non-stationary but fairly rough. Its power spectrum has a slow decay, dropping about 20dB over 0.5 Hz (where the original series is sampled at 1 Hz). The phase is significantly random and the time-integrated distribution is fairly Gaussian. The return map (lag=2) shows a strong linear relationship, but a fairly significant scatter and no nonlinear patterns.

A type-0 surrogate is shown in Fig. 5.9, and the randomization of amplitudes, while keeping the time-integrated distribution invariant, removes all nonstationarity and correlation in the samples.

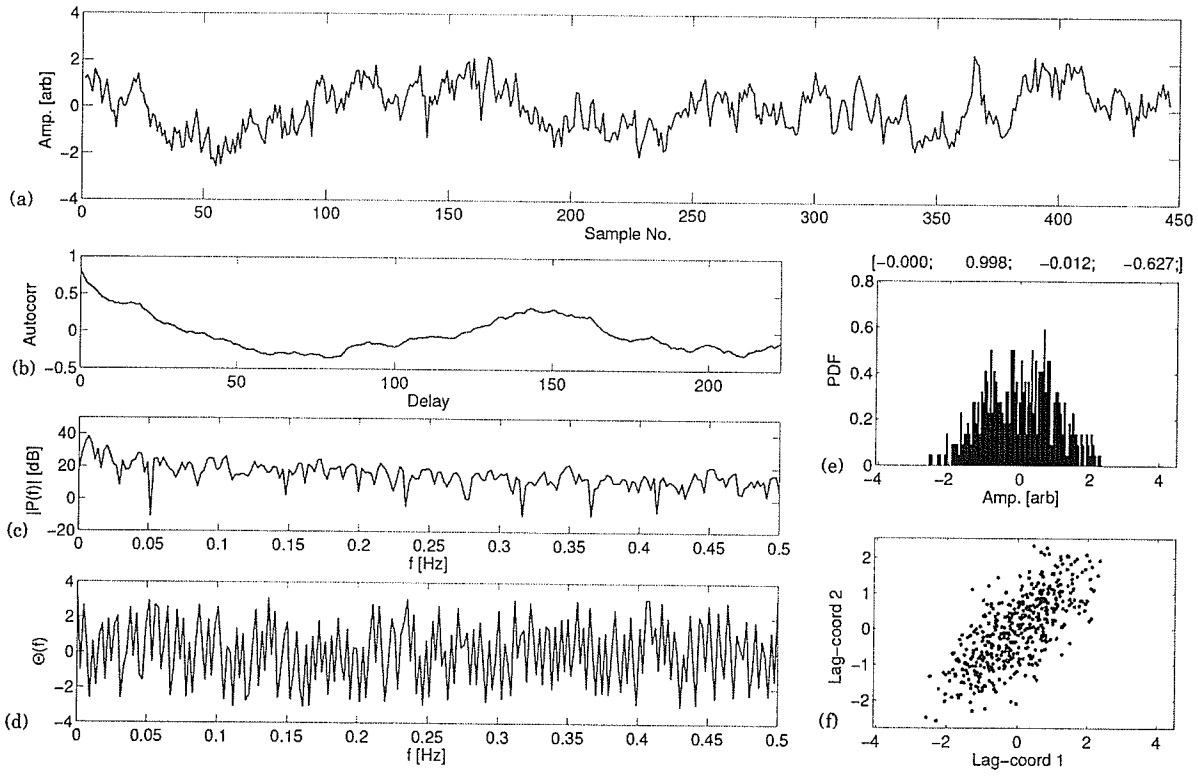


Fig. 5.10 Features of a type-1 surrogate sequence of pink noise: (a) sequence of values is quite like the original; (b) autocorrelation sequence has same smoothness and profile as original; (c) power spectrum is statistically equivalent to original; (d) Fourier phase appears as a random variable; (e) time-integrated distribution is statistically equivalent to the original; (f) return map (lag=2) appears statistically equivalent to the original.

Since the type-1 null hypothesis model matches the model of the actual pink noise, it is expected that the type-1 surrogate of Fig. 5.10 has essentially the same features as the original signal. As expected, there is a consistent matching in all panels with the features of the original in Fig. 5.8.

There is little extra benefit with a type-2.0 surrogate, but, as shown in Fig. 5.11, the time-integrated distribution of the surrogate exactly matches the original pink noise.

In Fig. 5.12, the features of the type-2.1 surrogate show no improvement in reconstruction of the original features when compared to the type-2.0.

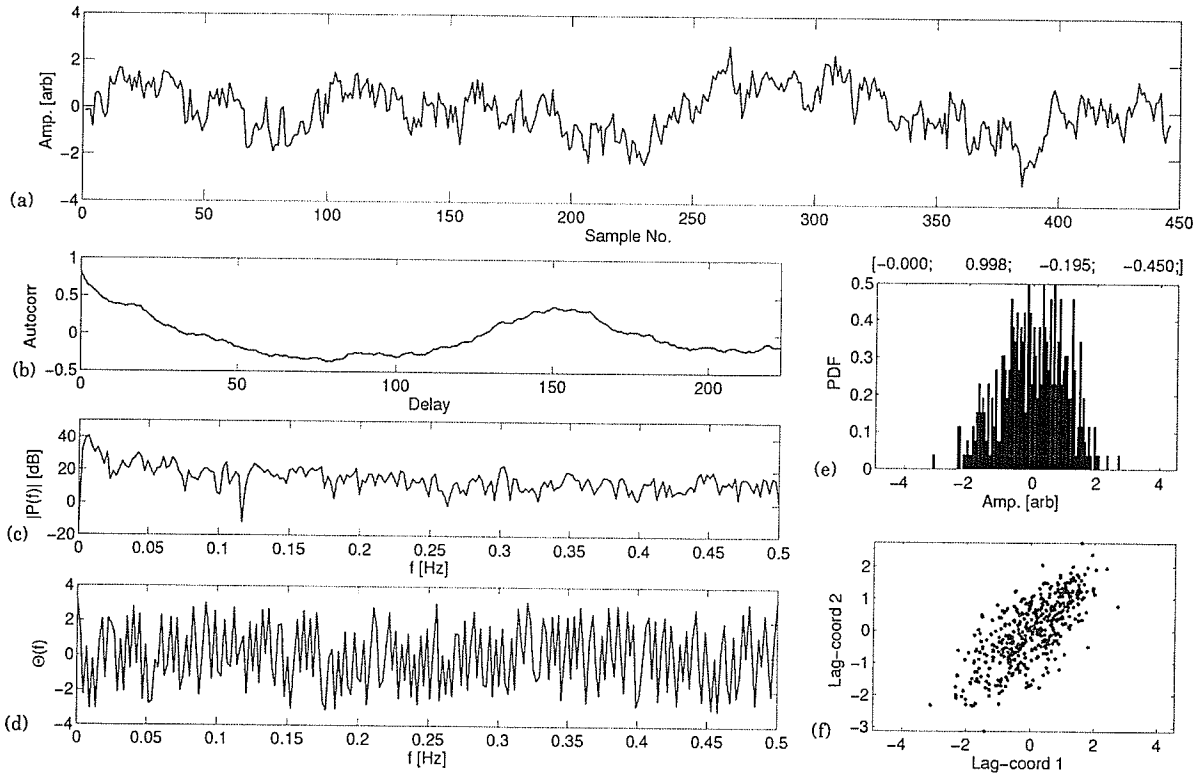


Fig. 5.11 Features of a type-2.0 surrogate sequence of pink noise: (a) sequence of values is quite like the original; (b) autocorrelation sequence has same smoothness and profile as original; (c) power spectrum is statistically equivalent to original; (d) Fourier phase appears as a random variable; (e) time-integrated distribution is exactly equal to the original; (f) return map (lag=2) appears statistically equivalent to the original.

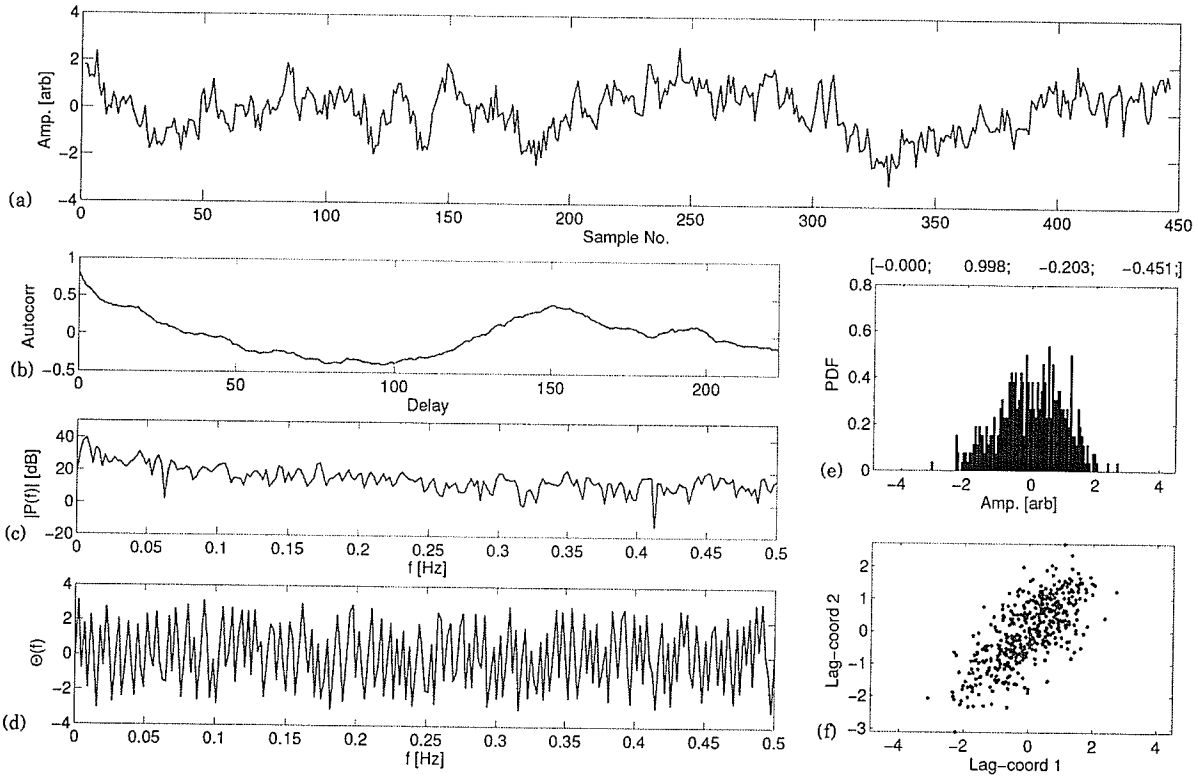


Fig. 5.12 Features of a type-2.1 surrogate sequence of pink noise: (a) sequence of values is quite like the original; (b) autocorrelation sequence has same smoothness and similar profile as original; (c) power spectrum is statistically equivalent to original; (d) Fourier phase appears as a random variable; (e) time-integrated distribution is better match to the original than the type-1 surrogate; (f) return map (lag=2) appears statistically equivalent to the original.

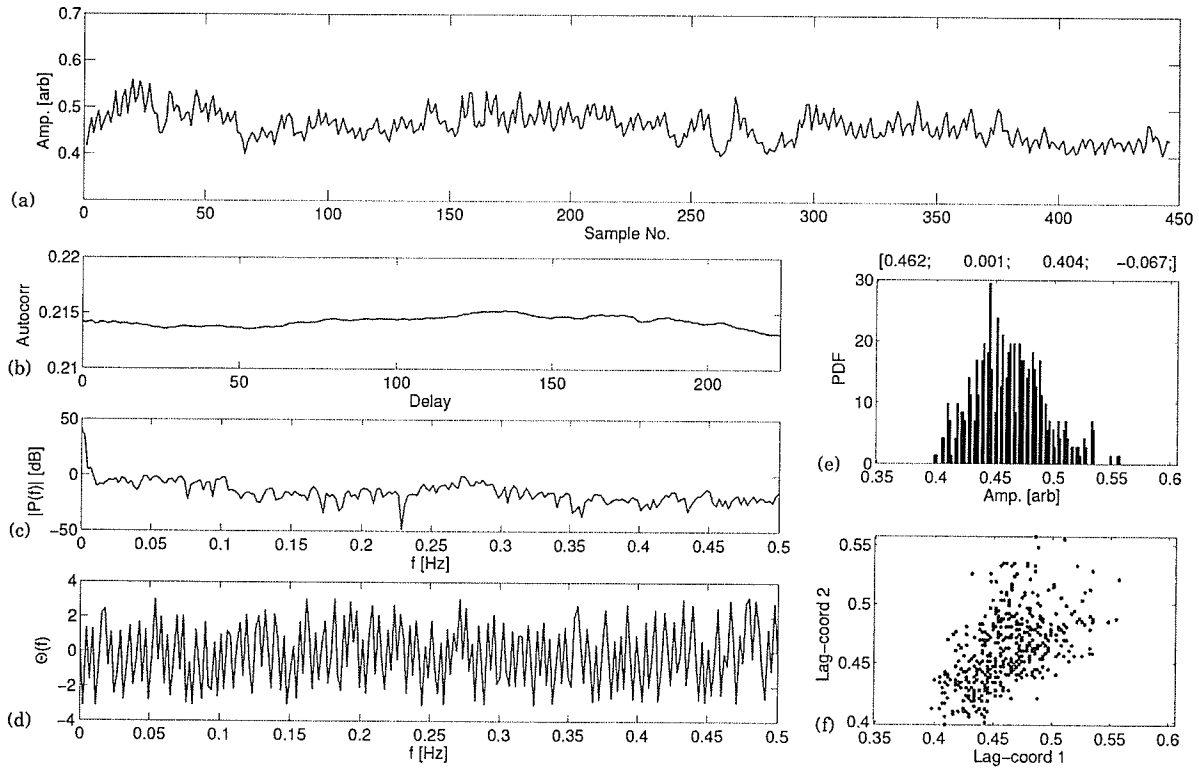


Fig. 5.13 Features of a fetal RR-interval sequence: (a) sequence of RR-interval values appears somewhat like pink noise; (b) autocorrelation sequence is rather smooth and has a local minimum at a lag of 50 samples, but has a very constant value overall; (c) power spectrum (assuming sampling is at "1 Hz") has a local maximum at 0.28 Hz and a slow decay; (d) Fourier phase appears as a random variable; (e) time-integrated distribution is skewed Gaussian; (f) return map (lag=2) exhibits a correlated radial scattering.

5.3.4.3 HRV

Now an example of surrogates from a natural signal is presented. Notwithstanding that a sequence of RR-intervals is not a true time series (the same applies to the Ikeda trajectory), the surrogate data techniques may be applied to provide null-hypotheses about the structure within the sequence. The more than 400 beats of Fig. 5.13(a) represent the RR-intervals of an isolated fetal ECG used in the later portions of this work for the generation of fetal ECG. (This dataset is not, however, directly recorded from fetal ECG since that is unavailable. The

data is in fact a rescaling of a recorded sequence of adult RR-intervals. In the context here, this amounts to nothing more than a "linear measurement function", and the dynamics of the recorded sequence is preserved here.) The sequence exhibits some nonstationary elements over an otherwise homogeneous pink nature. The power spectrum of the sequence has a background of gradual decay with a slight mound at 0.28 Hz. (Of course, the units of Hz is rather loose in this application, since it is not precisely a time series. More precisely it is units of "cycles per beat".) The time-integrated distribution is nearly Gaussian, but slightly skewed. The return map (lag=2), exhibits a linear correlation, but with a greater spread at higher RR-interval values.

The type-0 surrogate of HRV is shown in Fig. 5.14, and the loss of correlation is distinctly apparent. While the time-integrated distribution is identical, the autocorrelation and power spectrum are flat, and the sequence itself appears more stationary. The return map now approximates independence (marginal factorization), instead of the correlation of the original. This type of null-hypothesis can clearly be rejected.

The type-1 surrogate of Fig. 5.5 represents a better null-hypothesis, but it is still not sufficient. The surrogate sequence has a visual likeness to the original, and its power spectrum is consistent with the local maximum at 0.28 Hz (as expected by design). The autocorrelation over 200 lags is somewhat weaker, however, and the time-integrated distribution has less skew (which is, of course, expected). The correlation in the return map of panel (e), however, is clearly less radial and more Gaussian than the original in Fig. 5.13. This type of null-hypothesis can likely be rejected as well.

An example of a type-2.0 surrogate for the HRV sequence is shown in Fig. 5.16, and improves the likeness of the surrogate to the original over the type-0 surrogate. In particular, a proper correlation appears in the return map, autocorrelation, and power spectrum. However,

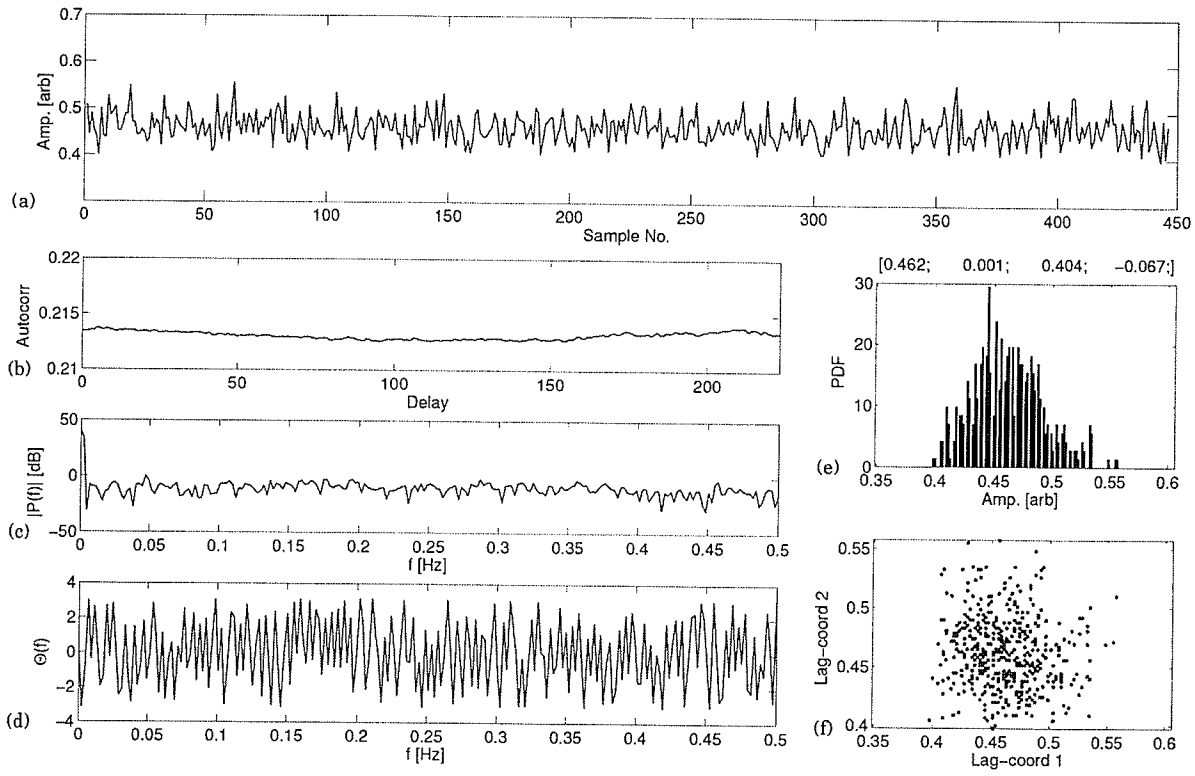


Fig. 5.14 Features of a type-0 surrogate sequence of fetal HRV: (a) sequence of values are stationary and more rough; (b) autocorrelation sequence is statistically constant; (c) power spectrum is statistically constant; (d) Fourier phase appears as a random variable; (e) time-integrated distribution is identical to Fig. 5.13(e); (f) return map (lag=2) appears random and almost independent (*i.e.*, marginal factorization).

the variability in the surrogate sequence itself appears less homogeneous than the original. (This effect is rather strong in this surrogate realization, but persists to a lesser degree in other realizations.) This would represent a good null-hypothesis, specifically because the return map pattern exhibits a more radial pattern than the type-1 surrogate.

Last, an example of a type-2.1 surrogate for the HRV sequence is shown in Fig. 5.17. This surrogate sequence has the best likeness to the original over all the other types. The autocorrelation, power spectrum, and return maps appear statistically equivalent. The time-integrated

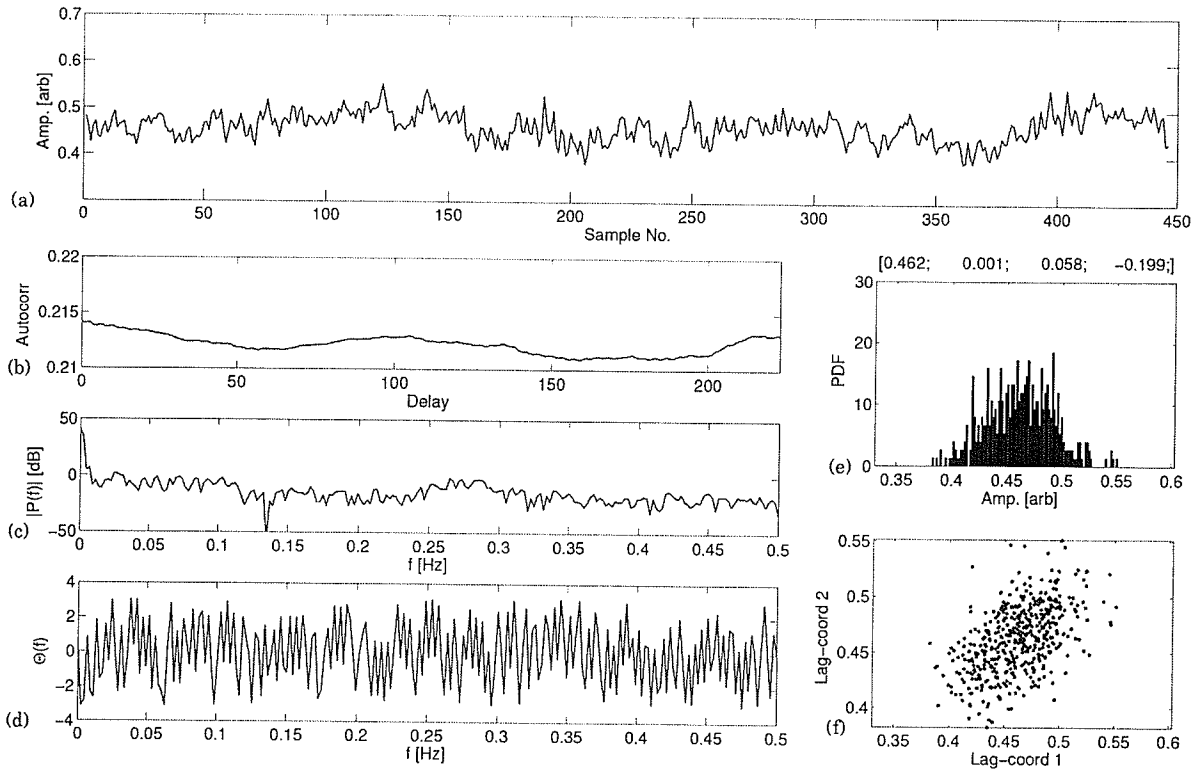


Fig. 5.15 Features of a type-1 surrogate sequence of fetal HRV: (a) sequence of values is quite like the original; (b) autocorrelation sequence has the same smoothness but lower values than the original; (c) power spectrum is statistically equivalent to original; (d) Fourier phase appears as a random variable; (e) time-integrated distribution is less skewed and slightly more sub-Gaussian than the original; (f) return map (lag=2) appears scattered and correlated with no radial effects.

distribution retains its non-Gaussianity, though it is slightly more symmetrical than the original. This surrogate would represent an appropriate null-hypothesis to provide context for nonlinear and multifractal metrics. The discrimination power of a nonlinear statistic, however, might be low in this case, due to the relatively few points available. The reader would find that studies of nonlinearity in the HRV literature normally use datasets extending up to 24 hours in length, or about 98K beats, instead of the 450 shown here.

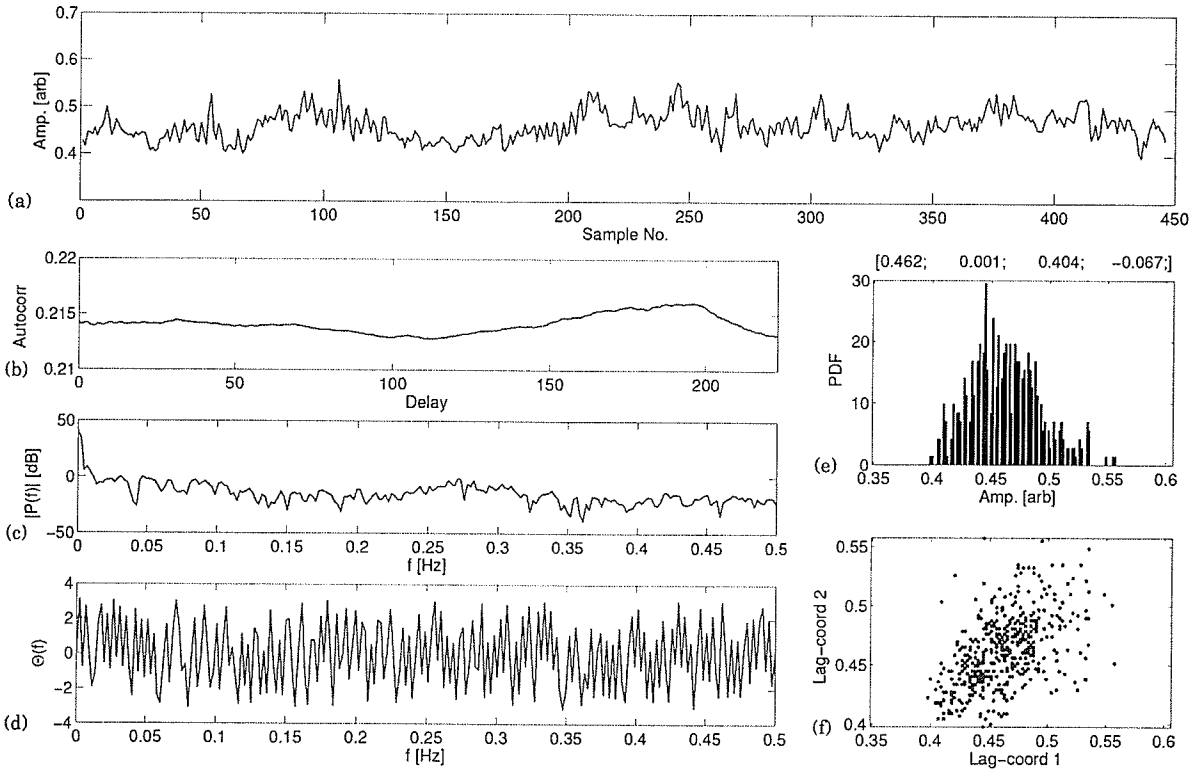


Fig. 5.16 Features of a type-2.0 surrogate sequence of fetal HRV: (a) sequence of values is not homogeneous like the original, but improves over the type-0 surrogate; (b) autocorrelation sequence has same smoothness as the original but has a depression at a delay of 100 beats; (c) power spectrum is statistically equivalent to original; (d) Fourier phase appears as a random variable; (e) time-integrated distribution is exactly equal to the original; (f) return map (lag=2) has a better statistical match to the original over the type-1 surrogate.

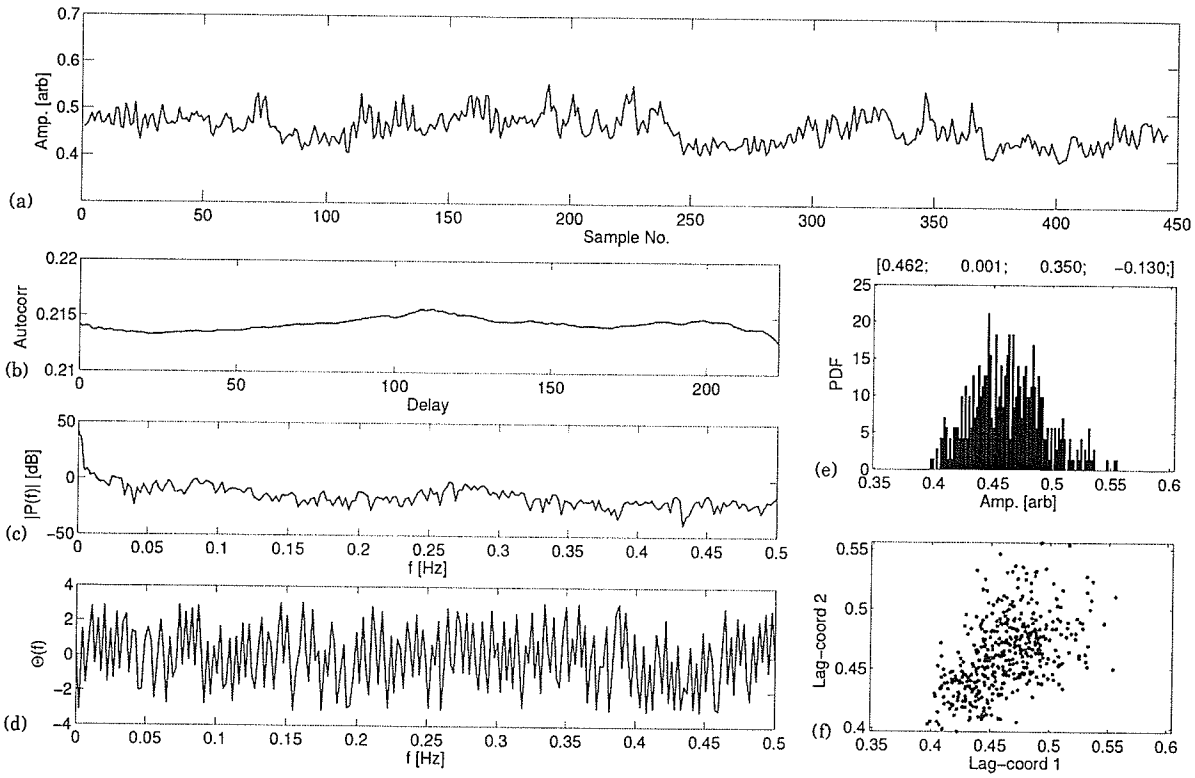


Fig. 5.17 Features of a type-2.1 surrogate sequence of fetal HRV: (a) sequence of values is less homogeneous than the original, and consistent with a type-2.0 surrogate; (b) autocorrelation sequence has great match in smoothness and profile to the original; (c) power spectrum is statistically equivalent to original; (d) Fourier phase appears as a random variable; (e) time-integrated distribution is statistically equivalent to the original; (f) return map (lag=2) appears statistically equivalent to the original, including radial effects.

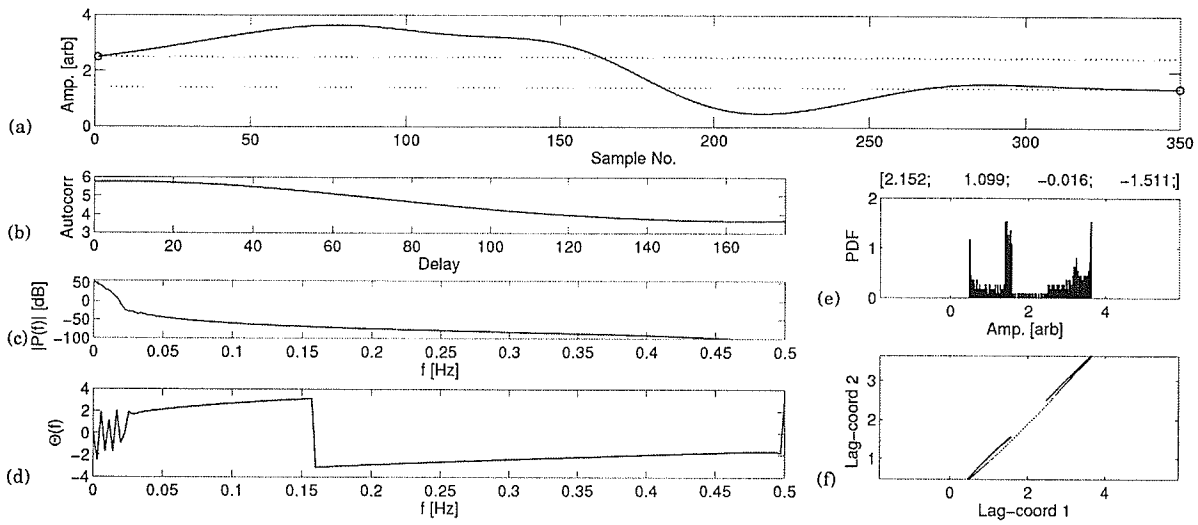


Fig. 5.18 Features of a smooth time series with endpoint mismatch: (a) sequence of values is smooth and nonstationary (*e.g.*, representative of a baseline wander) with endpoint mismatch (circles and dotted); (b) autocorrelation sequence is smooth decay beginning at 30 samples; (c) Hann filtered power spectrum: power is isolated in low frequencies (below 0.02 Hz); (d) Fourier phase appears as a random variable only in low frequencies; (e) time-integrated distribution is multimodal; (f) return map (lag=2) has strong cyclic and correlated structure.

5.3.5 Managing Periodicity Artifacts in Type 1 and 2 Surrogates

Consider a recorded time series x_n, for $n = 1, \ldots, N$. As was mentioned in Sec. 5.3.2, the use of the unwindowed FFT in the type 1 and 2 surrogate algorithms essentially realizes the constraint of the original data's linear dynamics on a periodic support. Consequently, the difference between the first and last point of the time series has a significant effect on the spectral properties of the surrogates. As shown in Fig. 5.18 and Fig. 5.19, this implies that smooth time series (*e.g.*, from nonlinear dynamical systems) can develop (i) noise-like high frequency distortion, and (ii) changes to the low frequency profile. Both of these artifacts are introduced by the delocalization of the characteristic properties of the endpoint mismatch (*i.e.*, the discontinuity). In [204], Schreiber suggests that these artifacts should be managed by careful selection of a representative time series segment that provides better endpoint matching.

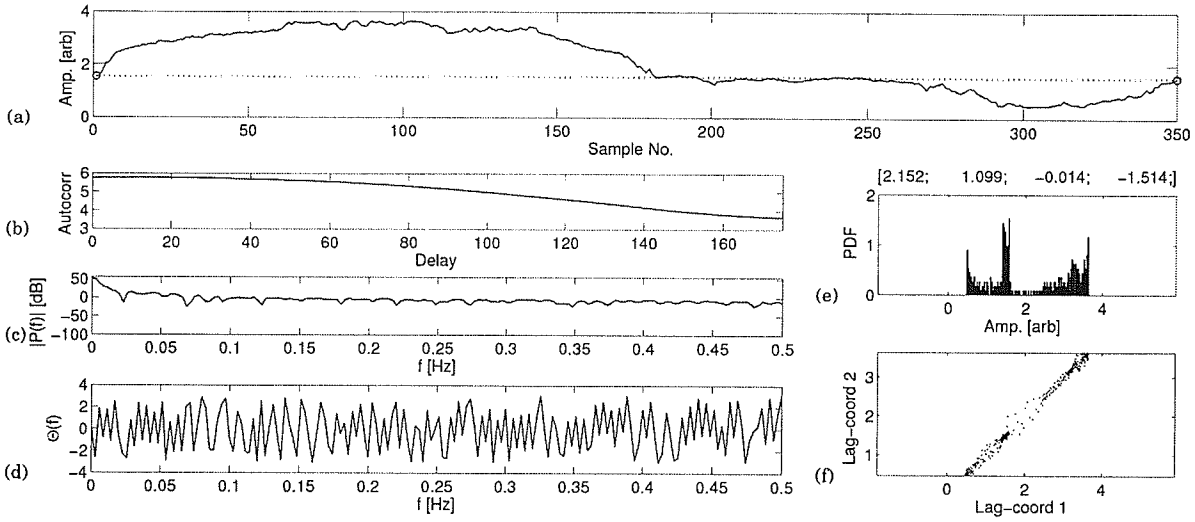


Fig. 5.19 Features of a type-2.1 surrogate sequence of Fig. 5.18(a) exhibit periodicity artifacts: (a) sequence of values has high frequency corruption and reduced endpoint mismatch (circles and dotted); (b) autocorrelation sequence is still smooth, but decay begins at 50 samples; (c) Hann filtered power spectrum: dominant low frequency band is equivalent to the original, but high frequency power is increased by over 50 dB; (d) Fourier phase appears as a random variable, as designed; (e) time-integrated distribution is a strong match to the original; (f) return map (lag=2) maintains correlation, but increases scatter. Power from periodic endpoint mismatch artifact is spread throughout the surrogate.

In effect, Schreiber suggests that the relative power of the endpoint mismatch in amplitude (*i.e.*, "jump") and first derivative (*i.e.*, "slip") be systematically measured and a subinterval of the data be chosen to minimize these measures. It is the opinion of the author, however, that the objective of these measures should not be minimization (*i.e.*, a limit to 0), but "typicalization". Essentially, the artificial sample interval $[x_N, x_1]$ arising from the periodic support should *a priori* be *typical* of the true time series, not constant. For smooth sequences, such as in Fig. 5.18, these objectives coincide. For rough sequences such as HRV (Fig. 5.13), however, typicalization implies that the periodic extension sequence is more representative of the actual extension, rather than more smooth. To achieve a measure of typicalization, the jump

energy proposed by Schreiber [204]

$$(x_N - x_1)^2 \tag{5.11}$$

is compared (i) to the amplitude variance

$$\mu_2\{ x_n \mid n = 1, \ldots, N\} \tag{5.12}$$

which represents the average internal energy per point (*i.e.*, around the mean), and (ii) to the *distribution* of the set of squared-differences

$$\left\{ \Delta_x^2(n) = (x_n - x_{n-1})^2 \mid n = 2, \ldots, N \right\} \tag{5.13}$$

which describes the population of jump energies between natural neighbours. Thus, if the correspondence between (5.11), (5.12), and the bulk of (5.13) is strong, there will be few periodic artifacts, since the transition to the periodic extension of $x(n)$ is indistinguishable from the properties of the time series itself. If, however, the value (5.11) is not representative of either (5.12) or the bulk of (5.13) (*e.g.*, the value of (5.11) is in the tail of (5.13)), then subintervals should be analyzed for better correspondence to avoid artifacts.

As an example, Fig. 5.20 shows the application of these metrics to a sequence of RR-intervals. Suppose it is desired to have a subinterval of about 100 beats for surrogate analysis. The subinterval of Fig. 5.20(a) marked A yields the analysis of panels (b) and (c), while the subinterval marked B is analyzed in panels (d) and (e). For the two subintervals, panels (b) and (d) show a histogram estimate of the distribution of (5.13) as well the value of the amplitude variance (5.12) (solid line), and the jump energy (5.11) (star) along the horizontal axis. The correspondence between these values is stronger in panel (d), implying that subinterval B in Fig. 5.20(a), is less likely to provoke periodicity artifacts in FFT surrogates. Furthermore, a demonstration of the cyclic extension of the subintervals is provided in panels (c) and (e).

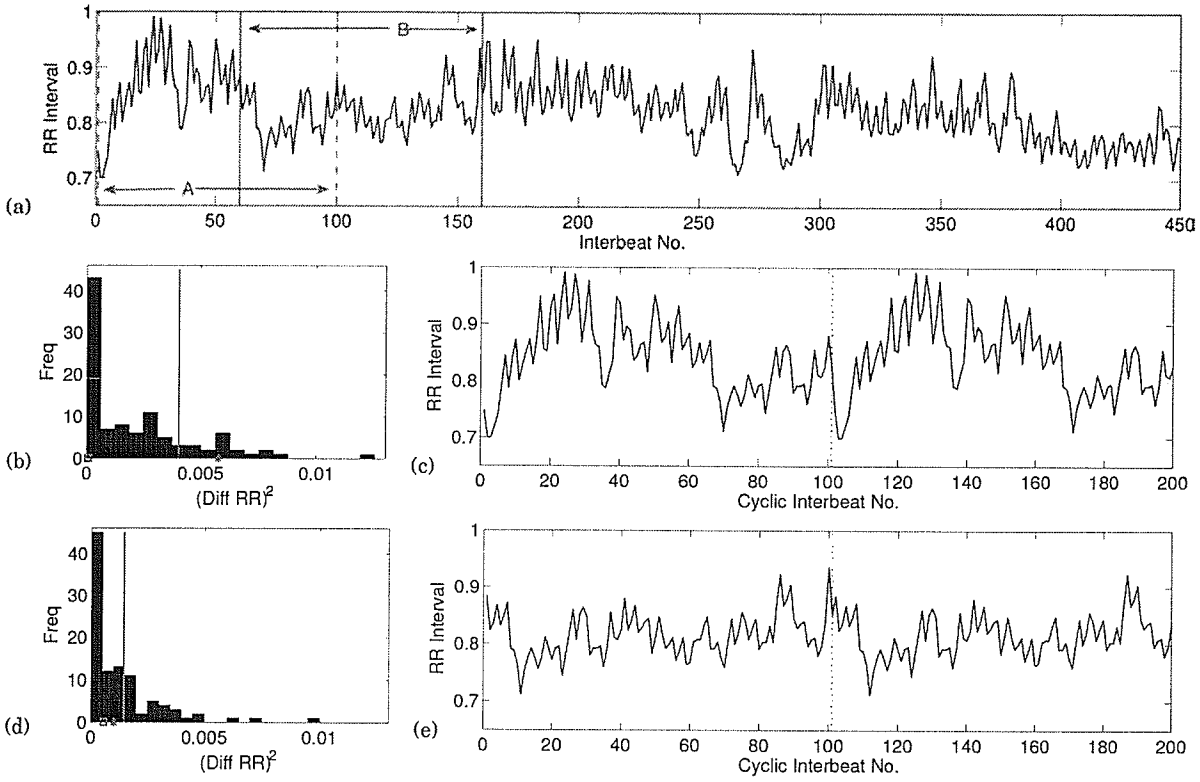


Fig. 5.20 Measuring typicalization: (a) 450 beats of natural HRV with two subintervals of 100 beats marked A and B; (b) analysis of periodic artifact for subinterval A: the jump energy (star) is above the mean internal energy (line) and in the tail of the distribution of $\Delta_x^2(n)$; (c) periodic extension of subinterval A: local neighbourhood of the periodic transition (dashed line) is distinct; (d) analysis of periodic artifact for subinterval B: the jump energy (star) is near the mean internal energy (line) and the bulk of the distribution of $\Delta_x^2(n)$; (c) periodic extension of subinterval B: the periodic transition (dashed line) neighbourhood blends nicely with the remainder. Subinterval B is preferred to mitigate periodic artifacts in FFT surrogates.

These sequences represent the data constraints applied to the FFT surrogates in practice. If the behaviour of the neighbourhood of the 100th interbeat number (dashed line) is distinct from the rest of the series, then periodic artifacts should be anticipated. Here, the transition behaviour of panel (d) appears more visually consistent with the remainder of the sequence, in agreement with the analysis of (5.11)-(5.13).

Schreiber has also described a simulated annealing algorithm for surrogate data generation which has benefits for non-periodic constraints [200]. Here cost functions based on the properties of the observed time series are used, and a simulated annealing procedure used to optimize the cost functions over pairwise permutations of the x_n values. In particular, the mean square error between the (time-domain) autocorrelation functions of the data and the potential surrogate can be used as a cost function (*i.e.*, panel (b) in the figures of Sec. 5.3.4.3). Since the autocorrelation functions are calculated over the time samples, no underlying periodic structure is introduced. By his own admission [204], however, the extra computational intensity required by this algorithm is only justified when simpler techniques (such as the sub-interval approach discussed here) cannot be applied.

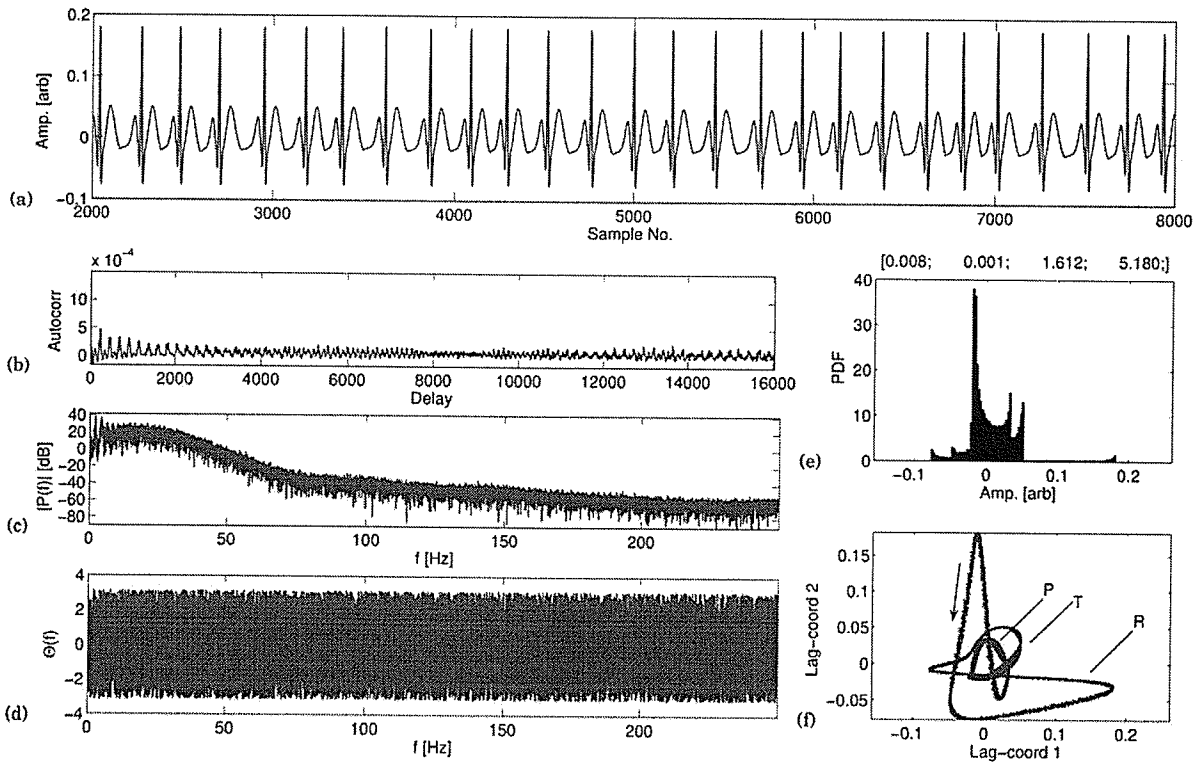


Fig. 5.21 Features of synthesized fetal ECG: (a) time series is characteristic of ECG waveform with variability in the interbeat interval (sampled at 500 sps); (b) ECG cyclostationarity produces a sequence of decaying harmonics in the autocorrelation sequence; (c) power spectrum has peaks at fundamental cyclostationary frequency, falloff at 20 Hz, and an inflection point near 70 Hz; (d) Fourier phase appears as a random variable; (e) time-integrated distribution has a strong peak at the baseline and a large tail from the R-wave; (f) return map (lag=18) consists of multiple loops induced by the P-, R-, and T-waves as marked. The flow induced by time ordering is also shown by the arrow.

5.4 Cyclic and ECG Surrogate Data Models

5.4.1 Cycle-Shuffle Surrogates

Despite the significance and utility of the canonical surrogate models in general, they represent poor null-hypothesis alternatives for ECG signals. Clearly a type-0 surrogate does not apply, since the ECG does not have a white spectrum. Neither does a type-1 surrogate apply, since the time-integrated distribution of ECG is super-Gaussian. Figure 5.22 shows the

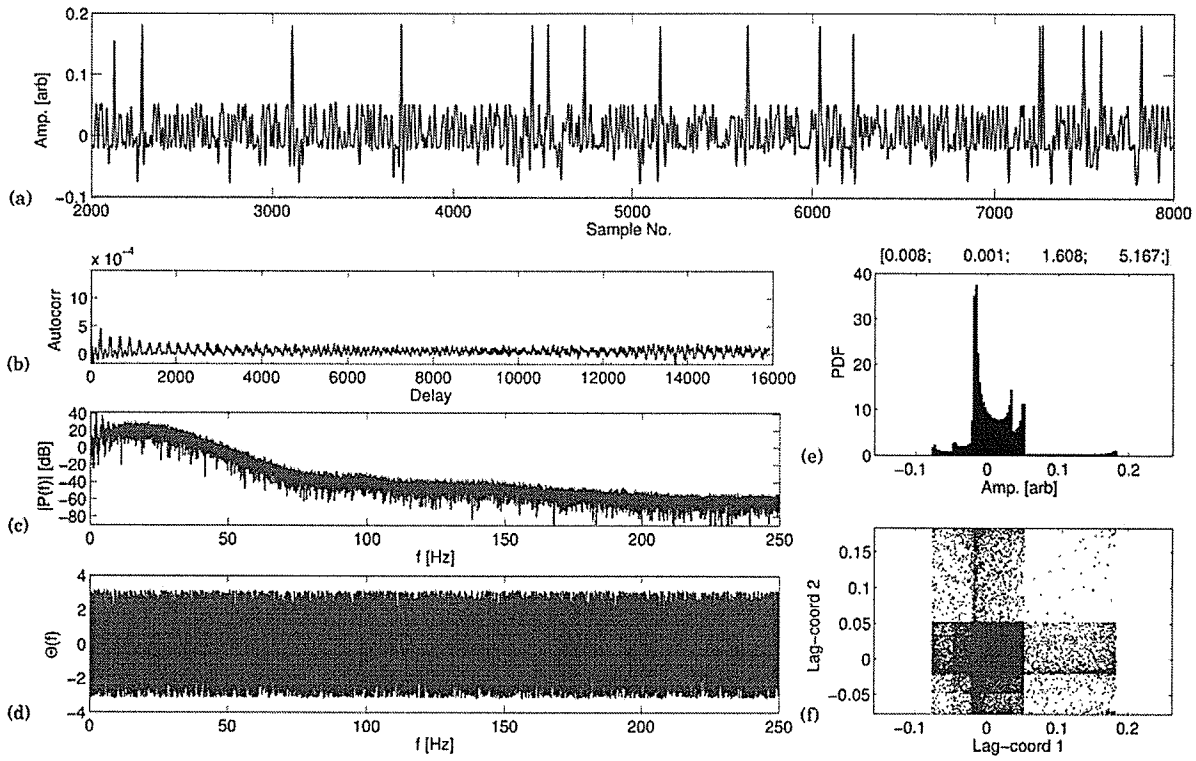


Fig. 5.22 Features of type-2.1 surrogate of fetal ECG: (a) time series is not cyclostationary, but exhibits fast oscillations and irregular spikes; (b) autocorrelation sequence is a very strong match to the original; (c) power spectrum is a strong match to the original; (d) Fourier phase appears as a random variable; (e) time-integrated distribution is a strong match to the original; (f) return map (lag=18) appears random and independent (marginal factorization) with no flow.

features of type-2.1 surrogate on the synthetic ECG time series of Fig. 5.21. As is easily seen, the high frequency spikes that in natural ECG are restricted to the QRS-complexes are replicated throughout the surrogate time series. Thus the cyclostationary nature of the signal is destroyed in the surrogate model. A more suitable approach for "ECG surrogates" is to maintain the local determinism of the ECG beat, the cyclostationary regularity, but destroy the interbeat dynamics, the cyclostationary variability. A surrogate approach to cyclic signals, called *cycle-shuffle surrogates* (CSS), was proposed by Theiler [223, (1995)] whereby the actual *cycles* in the signal are shuffled just as in a type-1 scenario. ECGs in sinus rhythm

in this sense are theoretically quite suitable for application, specifically because convenient breakpoints can be identified (such as the QRS-complex maxima, or the interbeat interval). Application in practice, however, is subject to certain artifacts. As discussed by Small in [208, (2002)], the CSS algorithm is truly only applicable to stationary and (most likely) long time series. The first limitation is because a slow drifting amplitude baseline makes it difficult to join the shuffled cycles and maintain a likeness with the original. Here, the choice is either (i) not to adjust the baseline heights of the cycles, and thus introduce sharp transitions that are not present in the original, or (ii) to adjust the baseline heights of the cycles and produce nonstationary baselines that are not consistent with the original. The second limitation applies to certain discriminating statistics, such as MFA dimensions based on attractor reconstruction. This limitation applies because only the neighbourhoods of the cycle transition "cut points" are actually changed in the attractor. If these neighbourhoods are infrequent in the attractor, then the MFA dimensions will not discriminate sufficiently between the surrogates and the originals. By increasing the relative number of cycles in the time series (by increasing embedding dimension of the attractor, reconstruction lag, and by increasing the length of the time series while lowering its sampling frequency) a better discrimination is possible, but more likely a different nonlinear metric should be chosen (*i.e.*, one specifically geared for CSS).

Since this work uses a model for synthetic ECG, the nonstationary limitations of the CSS algorithm are not significant. The features of a synthetic ECG with natural HRV is shown in Fig. 5.21 and the features of a corresponding CSS are shown in Fig. 5.23. Only two elements in the features shown could discriminate between the CSS and original. First, note that the power spectrum decay above 50 Hz is much weaker in the CSS (Fig. 5.23(c)) than in the original. Second, the differences in the return maps in panels (e) of Fig. 5.21 and Fig. 5.23 should be considered. These plots are identical for the smaller loops because, at an embedding

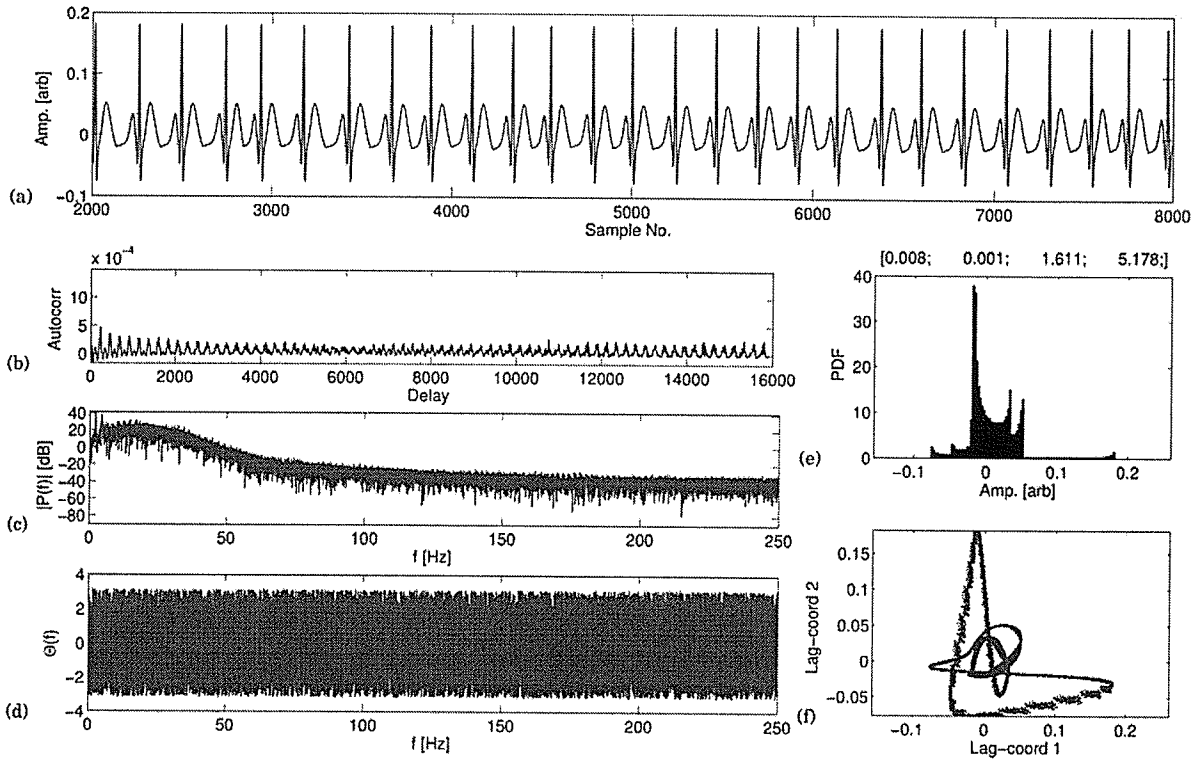


Fig. 5.23 Features of fetal ECG CSS: (a) time series is cyclostationary and resembles an ECG waveform; (b) autocorrelation sequence has slightly stronger harmonics than the original, but is otherwise a good statistical match; (c) power spectrum has the same falloff as the original, but the decay is weaker in higher frequencies (*e.g.*, 20 dB higher at 250 Hz); (d) Fourier phase appears as a random variable; (e) time-integrated distribution is a strong match to the original; (f) return map (lag=18) preserves the flow of the original, and the small loops for the P- and T-waves are identical to the original. Only an induced patchwork effect on the R-wave loop distinguishes the map from the original.

dimension of 2 and lag of 18, these points were unaffected by the shuffling at the R-wave cut points. However, the broad loop representing the R-wave undergoes a distinct statistical change from the CSS process. The spread of the band increases and the connection of the loop decreases, becoming patchwork. In a higher-dimensional embedding space, these changes will occur on a larger portion of the attractor, but probably not enough to provide discrimination by MFA. Since the sample lag determines the width of the CSS effect, the 17 distinguishable

patches in Fig. 5.23(f) would correspond to the chosen lag of 18 samples. Discontinuities occur where the samples cross the R-wave fiducial point used for the CSS cut point. (*i.e.*, The CSS process with a lag of 18 gives the 2D lag-vector 17 possible phases with the cut point: namely (R,17), (1+R,16), ..., (16+R,1) where the letter R represents the position of the R-wave cut point.) Further experimentation has shown that the number of patches changes with the lag.

Though often neglected in general because of the difficulties with their application, CSSs are a suitable null-hypothesis model for evaluating the nonlinear dynamics of the fetal ECG signals in this work. It is expected, however, that the discrimination context for MFA provided by CSSs is limited because so little of the attractor is affected. On the other hand, a consideration of the MFA of a fetal ECG CSS can essentially evaluate the significance of the fractal contribution of that piece of the attractor most influenced by the CSS process. In the remainder of this work, this will be the R-wave (because of its implementation convenience). For future work, however, it is recommended that an analysis be made of the TP-interval, where the HRV tends to be isolated and the mode of the time-integrated distribution is located.

Now another recent null-hypothesis model for ECG surrogates, which more profoundly changes the statistics of the attractor while preserving its geometry, will be considered.

5.4.2 Pseudo-Periodic Surrogates

In general, the failure of the canonical surrogate models with ECG is due to the strong nuisance linearities and nonlinearities represented by the geometry of the ECG attractor[1]. Here, the interest in surrogate data would be to evaluate the *complexity* of this geometry, (*i.e.*,

[1] In general, note that the stronger the surrogate constraints become, the less likely the algorithm will achieve them. *e.g.*, The type-0 surrogates have the simplest constraints, and they also always have the highest quality representation of the null-hypothesis model.

the variability of statistics on the manifold itself), and not the underlying manifold looping (observed in Fig. 5.21(f)). The regular features of the ECG are direct indicators of nonlinear relationships which are essentially beyond the representation of the canonical surrogate techniques. These regular features, however, do not present much indication of dynamical variability, which would be the interest of an MFA. Consequently, it makes sense to redefine the null-hypothesis in order to allow stronger nuisance nonlinearities to account for the attractor manifold, yet keep it distinct from the dynamical variability. The CSS model, however, does not fully satisfy this because, though it reproduces the cyclostationary geometry of the attractor, too little randomization of the variability has been induced.

Another alternative is to consider the underlying nonlinearity of the system uninteresting if it is *periodic*. Following this, Small *et al.* proposed a *pseudo-periodic surrogate* (PPS) model [210, (2001)] which is designed to faithfully represent any gross periodic dynamics, but destroy by randomization all other dynamics. This is accomplished by (i) constructing the attractor by time delays and (ii) reconstructing a time series from the attractor with a dynamical noise contribution to destroy dynamics.

Under the null-hypothesis PPS model, a point in the attractor is essentially of the form $\boldsymbol{x}_0 = \boldsymbol{s}_0 + \boldsymbol{n}_0$, where the point $\boldsymbol{s}_0 \in \mathbb{R}^{N_e}$ is from periodic dynamics and $\boldsymbol{n}_0$ is a noise component. Thus a neighbour, $\boldsymbol{x}_1$, of $\boldsymbol{x}_0$ is essentially

$$\boldsymbol{x}_1 = \boldsymbol{s}_1 + \boldsymbol{n}_1 \approx \boldsymbol{s}_0 + \boldsymbol{n}' \tag{5.14}$$

since the error between $\boldsymbol{s}_1$ and $\boldsymbol{s}_0$ determined by the periodic dynamics must be small, and thus can be grouped into the noise component $\boldsymbol{n}'$. This indicates that the null-hypothesis is effectively invariant to the permutation of local neighbours on the attractor.

In this work, the PPS algorithm is implemented by Small's own Matlab implementation, and proceeds by permuting the successive points in the trajectory by a local neighbour of the

attractor, as follows. The time series $x_1, \ldots, x_n$ are reconstructed into an attractor using the lag Δn and embedding dimension N_e, which is represented as the trajectory $s_n \in \mathbb{R}_e^N$, called the *natural trajectory*. The PPS algorithm then defines a *surrogate trajectory* $z_n \in \mathbb{R}_e^N$ on the attractor iteratively, element-by-element, using random sampling with replacement scheme that requires *a priori* a noise level parameter that tunes the "scale" of the neighbourhood sampling. First the iterative process defining the surrogate trajectory will be defined, followed by a discussion on the choice of *noise radius*, σ_{PPS}.

To begin, the surrogate trajectory is initialized at a random point on the attractor z_1. The iterative construction of the surrogate trajectory begins from the nth point z_n as follows. Its neighbours are weighted according to an exponential probability distribution, *i.e.*

$$\Pr_{z_n}(\widehat{z_n}) \propto \exp \frac{-\left\|\widehat{z_n} - z_n\right\|}{\sigma_{\text{PPS}}} \tag{5.15}$$

suitably normalized to 1. From this weighted set, the alternative point $\widehat{z_n}$ is chosen at random. According to the natural trajectory, $\widehat{z_n} = x_k$ for some k and so has a natural successor x_{k+1}. The next element of the surrogate trajectory z_{n+1} is then defined as the surrogate successor, *i.e.*,

$$z_{n+1} \leftarrow x_{k+1} \tag{5.16}$$

and the algorithm is iterated, $n \leftarrow n + 1$.

Once the surrogate trajectory is completely defined, the surrogate time series

$$\pi_{\text{PPS}}[x](n) = \mathbf{e}_1^{\mathrm{T}} z_n \tag{5.17}$$

is extracted as the first coordinate projection of the surrogate trajectory.

Since the PPS process introduces dynamical noise into the surrogate trajectory, any structure beyond the general geometry (and flow) of the attractor is corrupted. Note the lag embedding of $\pi_{\text{PPS}}[x]$ does not reproduce the surrogate trajectory z_n from which it was defined,

and thus its reconstructed attractor is distinct from that of the original series x_n. Thus any deterministic nonperiodic intercycle dynamics, such as chaos, that appeared in the original is reduced in the surrogate. However, it is not only deterministic dynamics that are eliminated in the surrogate. If the underlying geometry of (5.14) had coloured noise, than the PPS will have whitened the noise, eliminating the stochastic dynamics.

To properly distinguish between the fine scale dynamics and the gross periodicity, however, the PPS model requires the noise radius, σ_{PPS}, to be properly chosen. This value defines the horizon between the large geometries that are preserved and the small dynamics that are randomized. If the value is too large, the underlying geometry of the attractor will not be preserved. Whereas if the value is too small, the surrogate is too similar to the original. Small *et al.* [210] recommends that σ_{PPS} be chosen to maximize the number of 2-point intervals that are coincident between the surrogate series and the original. (That is, σ_{PPS} should be chosen to maximize the likelihood that the surrogate's horizon of predictability is the next point only.) Their experiments demonstrated that such a value was easily determined, and provide an algorithm to calculate it from data. Using a shortened sequence of the original time series ($N' = 3000$ points), all $N' \times N'$ transition probabilities between the points are calculated. A binary search is initiated on σ_{PPS} in order to make the average transition probability (average of the off-diagonals) equal to $P = 0.5$, which maximizes the probability of a 2-point interval since Pr(2 pt-interval) $= P(1-P)$. Provided the restricted number of points N' covers multiple periods of the attractor flow, this approach is sufficient. Using a subsequence in the form of every nth point (for small n, *e.g.*, $n = 2, 3$) to increase the number of revolutions of the attractor was tried, but did not improve the performance of the σ_{PPS} selection.

The features of a PPS of the fetal ECG signal of Fig. 5.21 is shown in Fig. 5.24. The times series of the PPS maintains the cyclostationary form of an ECG, with an added noise

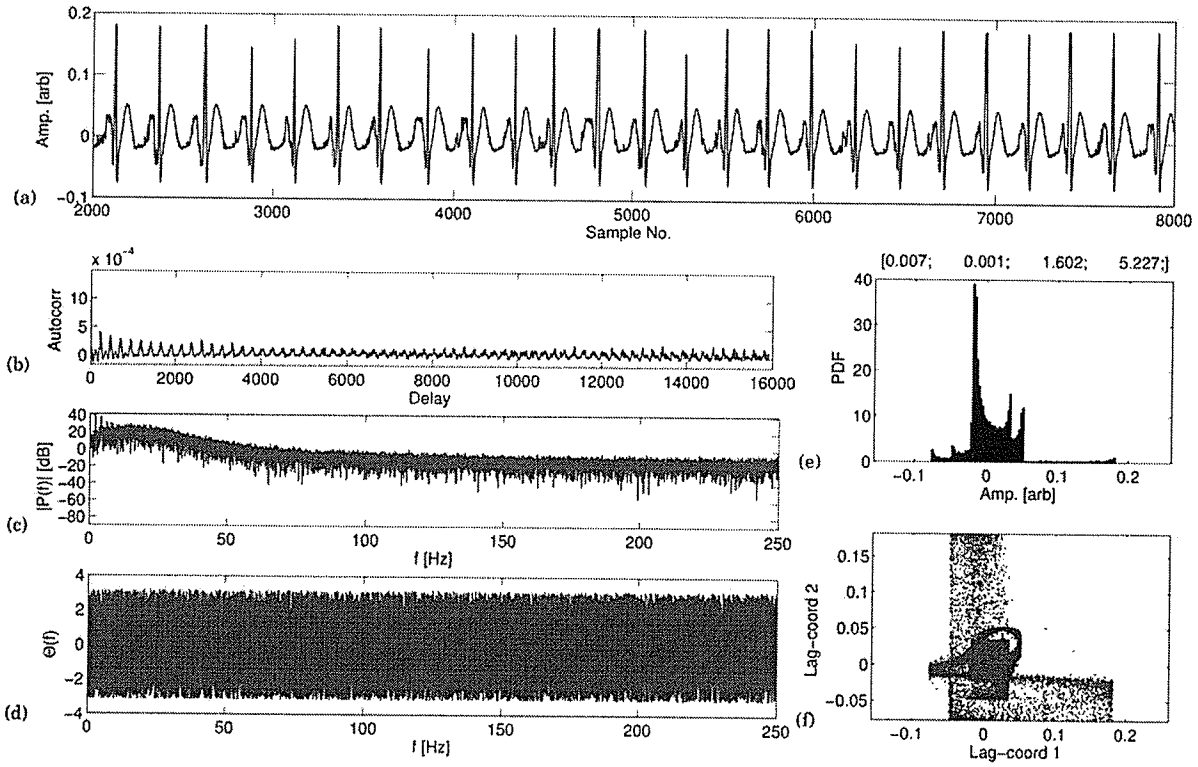


Fig. 5.24 Features of fetal ECG PPS: (a) time series is cyclostationary and resembles an ECG waveform with a noise component; (b) autocorrelation sequence has slightly stronger harmonics than the original, but is otherwise a good statistical match; (c) power spectrum has the same falloff as the original, but has significantly stronger higher frequencies (above 50 Hz); (d) Fourier phase appears as a random variable; (e) time-integrated distribution is a strong match to the original; (f) return map (lag=18) preserves the flow of the original, but the loops broaden under the stochastic influence. PPS surrogate is good surrogate model for ECG.

component. One could imagine that under the "periodic plus noise" null-hypothesis model, the time series shown here would be indistinguishable. From the dynamical view, however, the most important observation of the time series is that the RR-intervals measurable from the surrogate are different from the original. Thus the most significant indicator of ECG dynamics verifies its disruption. The autocorrelation sequence in Fig. 5.24(b) is very similar to the original, but the harmonics appear slightly stronger. Consistent with the observed noise in panel (a), the power spectrum in panel (c) has significantly stronger higher frequencies (specifically,

a white noise floor 45 dB below the peak power). Notwithstanding all the dynamical changes, the time-integrated distribution of PPS, Fig. 5.24(e), is nearly identical to the original. The changes in the return map clearly indicate the success of the algorithm's design. All the loops are maintained in Fig. 5.24(f), but their spread has greatly increased. In particular, the R-wave loop has widened and lost some definition. The squareness of the R-wave is an indication of the independence of the noise component induced by the PPS model. The changes to the other loops are significant also, however, which was not achieved through the CSS algorithm. Considering these results, the PSS model is an important surrogate model for the validation of MFA.

5.4.3 Synthetic ECG: HRV Surrogates

Since this work uses the novel technique for ECG synthesis from RR-intervals, a new and unusual opportunity for surrogate generation is presented: a synthetic ECG based on HRV surrogates. By this method, the effect of the HRV properties on the MFA characterization of the resulting ECG can be analyzed. Considering the synthetic ECG system of App. A as an operator ECGfm on the sequence of ECG event-intervals $\mathtt{I}$, the HRV-surrogate of the synthetic ECG time series $x = \text{ECGfm}(\mathtt{I})$ is the respective synthetic ECG

$$\pi_{\text{HRV}}[x] = \text{ECGfm}(\pi_{2.1}[\mathtt{I}]) \tag{5.18}$$

The features of an HRV-surrogate of the fetal ECG signal of Fig. 5.21 is shown in Fig. 5.25. Since its time series is generated from the same dynamical process as the original, panel (a) is itself a synthesized ECG signal, but its RR-interval pattern is distinct. The autocorrelation sequence in panel (b) is very similar to the original, but the harmonics appear slightly stronger. The remainder of the surrogate's features, including power spectrum, time-integrated distribution, and return map, are virtually indistinguishable from the figures. This is to be anticipated,

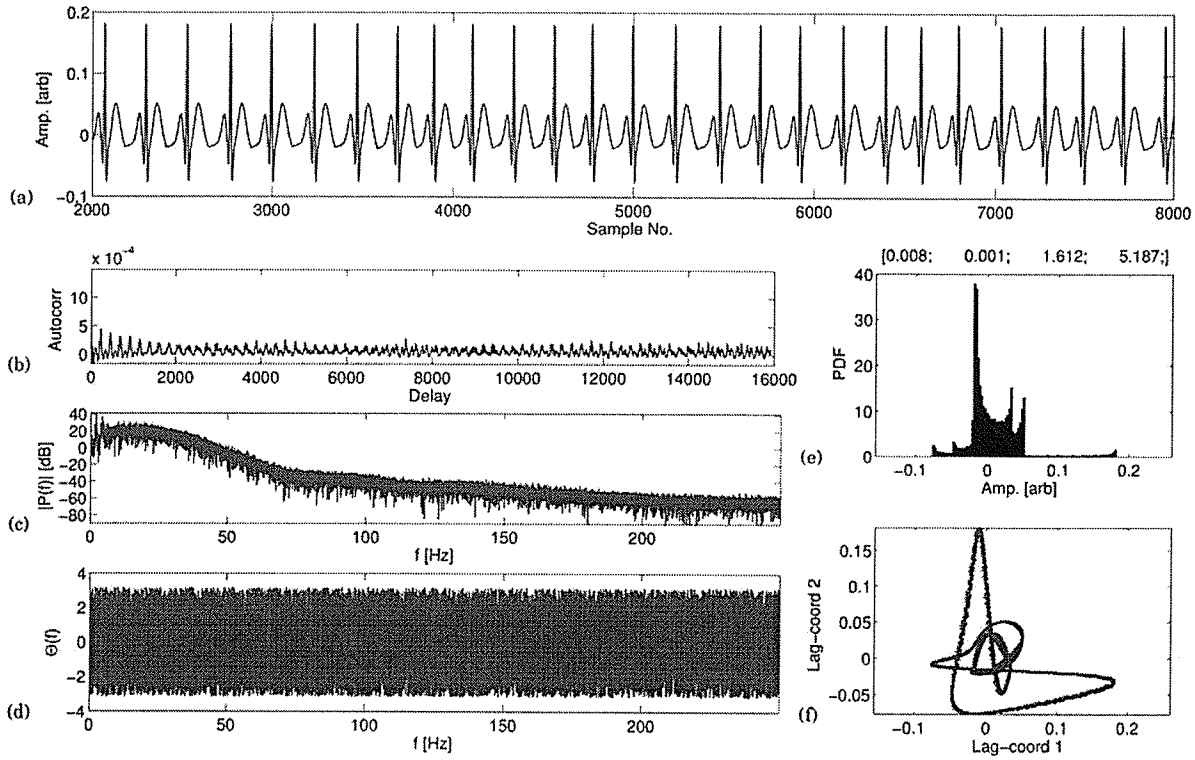


Fig. 5.25 Features of fetal ECG HRV-surrogate: (a) time series is a valid ECG waveform; (b) autocorrelation sequence has slightly stronger harmonics than the original, but is otherwise a good statistical match; (c) power spectrum is a strong match to the original; (d) Fourier phase appears as a random variable; (e) time-integrated distribution is a strong match to the original; (f) return map (lag=18) is virtually indistinguishable from the original.

considering the similarity between the HRV-surrogate process and the source of the original synthesized fetal ECG signal. The interest, here, is not the gross similarities, but whether any significant differences can be discerned between them. Since the original signal was defined from natural HRV while the HRV-surrogate is only stochastic, any distinction between the two can be considered as evidence of some sensitivity in the characterization to the nonlinearity dynamics of HRV. This approach to surrogate synthesis considers the ECGsyn dynamical system effectively as a "nuisance nonlinearity" and can provide insight into the relative significance of HRV vs. the ECGsyn synthesis model (A.6) to the end result.

With that end in mind, there is another test sequence that can be used to provide some MFA context for ECG. Using the same ECG synthesis method previously used, an ECG template can be constructed *without HRV*. Replacing the RR-interval sequence $\mathtt{I}.RR(n)$ that defined the time series x with its constant average $\mathtt{I}_0.RR(n) \stackrel{\text{def}}{=} \mathcal{E}\{\mathtt{I}.RR\}$, the constant-RR surrogate, $\pi_{\text{RRC}}[x]$, can be defined

$$\pi_{\text{RRC}}[x] \stackrel{\text{def}}{=} \text{ECGfm}(\mathtt{I}_0.RR) \tag{5.19}$$

By removing the HRV from the ECG altogether, another perspective on the relative contributions of HRV vs. the ECGsyn synthesis model can be analyzed. Unlike the HRV-surrogates, however, this is not a stochastic representation, and therefore the RRC-surrogate is unique and does not require multiple realizations.

The features of the RRC-surrogate of the fetal ECG signal of Fig. 5.21 is shown in Fig. 5.26. In panel (a), it is difficult without a direct overlay comparison to see the difference that true periodicity makes. The autocorrelation sequence in panel (b), however, is easily distinguished because of the strong regular spikes that occur without decay. The power spectrum at high frequencies (*i.e.*, above 120 Hz) is essentially the same as the original. At lower frequencies however, there is no modulation widening the peaks of the harmonics, and so the harmonic frequencies dominate the reduced floor. The time-integrated distribution of the surrogate, however, remains essentially unchanged. The return map (lag=18) has very little change to it except near the origin, where the portion of the manifold representing the TP-interval has narrowed.

This approach to surrogate synthesis considers the ECGsyn dynamical system as the only effective source of nonlinearity and can help identify artifacts in the MFA due to the gross geometry of the attractor manifold alone.

Now the entire complement of surrogate data alternatives has been reviewed. The next

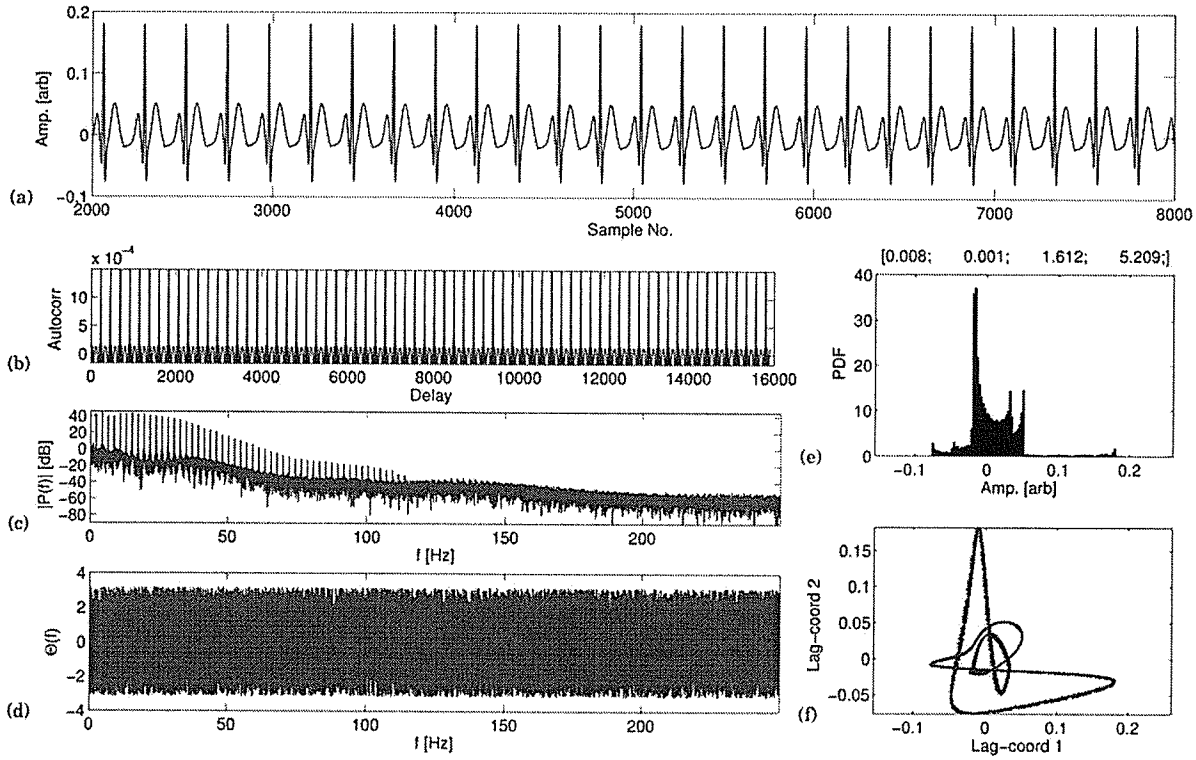


Fig. 5.26 Features of fetal ECG RRC-surrogate: (a) time series is an unusually periodic ECG waveform; (b) autocorrelation sequence exhibits strong harmonics with no decay; (c) power spectrum above 120 Hz is a strong match to the original, while strong harmonics dominate a lowered floor of low frequencies; (d) Fourier phase appears as a random variable; (e) time-integrated distribution is a strong match to the original; (f) return map (lag=18) through the R-wave loop is strong match to the original, but the loop segments closest to the origin representing the TP-interval has a marked narrowing.

section discusses how these techniques are used for the validation of MFA.

5.5 Using Surrogate Data Models in MFA Validation

The use of surrogate data in the literature is often a form of "sanity check" for the validation of a fractal dimension. That is, provided the null-hypothesis has been rejected, a nonlinear dynamical feature (*e.g.*, the correlation dimension) can be considered an accurate characterization of the dynamics. More elaborate studies, specifically of the surrogate techniques on known data models, will measure the size and power of the discriminating nonlinear statistic, which requires the generation of 19 surrogates (at minimum) or more.

This work uses surrogates in the first sense, but in two capacities. First, the basic correlation sum (4.169) of the surrogate signal is used to identify any artifacts limiting the scaling interval. If linear (or nonlinear) regions of the correlation sum log-log plot exist in both stochastic surrogates and fECG, these regions should be excluded from the scaling interval, since they do not represent the dynamical nonlinear properties of the fECG signal that are to be featured. Second, once a chosen scaling region has been chosen, the scaling of the MFA entropies of the surrogates should be examined for comparison to that of the data. The scaling properties should clearly be different if any credibility is to be applied to the MFA. In some circumstances, no linear fit is possible on surrogates. If it is possible to fit the scaling of the MFA entropies of the surrogates (perhaps on a sub-interval), the MFA can be compared directly to the MFA of the data. Hence surrogates are useful in providing context for (i) the scaling interval, and (ii) a sanity check for the MFA dimensions.

Note when MFA is applied as an SQM for the measurement of ICA separation, however, the surrogate algorithms are not being applied to a synthetic ECG signal, but a mixture of ECG signals. This implies that CSS and HRV-surrogates are not directly applicable the way they are with isolated ECG signals. Thus, at the synthesis level of the experimental method, these supplementary surrogate algorithms can shed insight on the contribution of the R-wave

and the HRV to the MFA that is measured. At the SQM level of the experimental method, however, the only surrogate algorithms available for analysis are the *type-2.1* (T21) surrogate and the PPS.

5.6 Summary

This chapter has presented the background required to define and understand surrogate data for the validation of nonlinear features extracted from a time series. The canonical surrogate data hypotheses and their limitations to cyclostationary time series, such as ECGs, were described. Alternatives from the literature, including cycle-shuffle surrogates (CSS) and pseudo-periodic surrogates (PPS), were reviewed. Also, using the new technique for ECG synthesis provided in App. A, a novel approach to surrogate data for ECGs based on HRV randomization was presented.

This chapter also identified two applications of surrogate data for the upcoming experimental design. First, based on the correlation sum (4.169), surrogate data provide context for the selection of scales to be used in MFA. Second, the multifractal scaling entropies of surrogate data can assist in identifying a proper scaling region for multifractal characterization and context for the resulting fractal dimensions.

This concludes the general background required for the description of the thesis experiment. The next chapter defines in more detail the implementations of all the statistical, fractal, and surrogate techniques as required for the experiment design.

Chapter VI

Design of Experiments

6.1 Overview

This chapter describes the design and implementation of experiments to achieve an analysis of feature convergence under the ICA processing of simulated fetal ECG. It is broken into several large ideas:

(a) ICA parameterization and convergence profiles;

(b) Features and SQMs;

(c) Signal synthesis; and

(d) Overall procedure of experiments.

The discussion proceeds by first describing in Sec. 6.2.1 the use of convergence profiles and sampling of the Stiefel manifold to provide convergence control. Thereafter, the several classes of features that will be analyzed for convergence are discussed categorically by (i) simple higher-order statistics (Sec. 6.3), and (ii) nonlinear features (*i.e.*, attractors, MFA, and surrogates in Sec. 6.4). From these features, the construction of the ICA error space by SQMs is described in Sec. 6.5. Although the stand alone algorithm for surrogate ECG generation is presented in App. A, a description of the methodology for the synthesis of noise-free ECG is also discussed. In particular, Sec. 6.6 covers the selection of event-interval templates, the sampling frequency for the synthesizer, and the length of time series to be analyzed. Finally, Sec. 6.7 discusses the overall procedure of experiments.

6.2 Parameterized ICA: Convergence Profiles

6.2.1 Convergence Profiles Defined

Since the goal of this work is to compare the convergence properties of nonlinear features under the action of ICA, control over convergence sampling is an important consideration of this experiment design. To achieve sufficient control over the ICA process, a methodology co-developed with N. Gadhok [63, (2007)] will be used, whereby the secondary optimization component of the ICA algorithm is discarded and replaced by a *parameterization of the Stiefel manifold.* That is, the ICA algorithms are represented solely by their cost function, $\mathcal{I}$, in matrix space, but the values of the cost functions need only be evaluated at specified samples of $\mathbf{W}$ on the Stiefel manifold. As discussed in Sec. 3.5.4, a simple approach to parameterizing the Stiefel manifold is with Givens angles. Since the consideration of multiple features already makes the ICA error space difficult to analyze, only the simplest possible ICA *convergence profile* will be used. Thus only two source signals, one maternal and one fetal, are considered, and consequently, the Stiefel manifold can be parameterized by a single Givens angle θ. By the geometric character of the ICA degeneracies discussed in Sec. 3.5.4, the angle θ, furthermore, only needs to parameterize an eighth of a circle to find an ICA solution. Without loss of generality, this will be assumed to be the interval $\theta \in [0, \pi/4]$ and the optimal solution will be preserved at (or near) $\theta = 0$.

Let $\boldsymbol{x}$ be a sphered linear mixture $\boldsymbol{x}(n) = \mathbf{W_0 M} \boldsymbol{s}(n)$ observed as a time series. Then under this design, any SQM, $\aleph$, can be characterized by its convergence profile

$$\aleph(\theta) \equiv \aleph(\mathbf{W}(\theta), \boldsymbol{x}) \tag{6.1}$$

on the parameterized Stiefel manifold $\mathbf{W}(\theta)$. Examining joint convergence of multiple SQMs,

therefore, amounts to the joint evaluation of the parameterized system

$$[\aleph_1(\theta), \ldots, \aleph_{N_\aleph}(\theta)]^{\mathrm{T}} \tag{6.2}$$

while holding the sources s, mixing matrix $\mathbf{M}$, and sphering matrix $\mathbf{W_0}$ constant.

To populate the SQM vector, 3 classes of convergence profiles are considered. The first class to consider are those from the ICA algorithms themselves. For an ICA cost function $\mathcal{I}(\mathbf{W}(\theta), \boldsymbol{x})$, its convergence profile can be calculated by

$$z = \mathbf{W}(\theta)\mathbf{W_0}\mathbf{M}s = \mathbf{W}(\theta)\boldsymbol{x} \tag{6.3}$$

$$\aleph^{[\mathcal{I}]}(\theta) \stackrel{\text{def}}{=} \mathcal{I}(z) = \mathcal{I}(\mathbf{W}(\theta)\boldsymbol{x}) \tag{6.4}$$

This approach to representing ICA numerically is slightly unusual, but frames the "practical algorithms" of ICA in a more theoretical way, and so is consistent with this work's objective to bridge theory and practice.

The other classes of convergence profiles to be analyzed against the ICA cost functions are feature-based SQMs. The implementation of these profiles will be discussed in Sec. 6.5, after a discussion of the feature extraction techniques themselves.

6.2.2 ICA Convergence Profiles

This section presents the objective functions $\mathcal{I}$ that represent the ICA algorithms and, therefore, represent B-class SQMs of the form $\aleph^{[\mathcal{I}]}(\theta)$. It is of interest to mention that these objective functions are rarely presented in the explicit forms used here. Instead, they are derived very generally (with many unspecified parameters) which then take on specific values only in the implementation of a practical algorithm. For this work, the default parameter values used in practice are assumed (unless specified), but they appear in the context of the original ICA cost function. Four ICA cost functions are considered in this work for their differences and significance:

(a) Xinfomax [125] (nonlinear maximum-likelihood objective);

(b) FastICA [91] (non-Gaussianity objective);

(c) RADICAL [123] (robust marginal maximum-likelihood objective);

(d) MILCA [215] (robust MI maximum-likelihood objective).

6.2.3 Xinfomax Cost Function

The extended-infomax (Xinfomax) algorithm is based on a maximum-likelihood extension consistent with the original infomax algorithm [21, (1995)]. The maximum-likelihood cost function is formulated as the mutual information of a mixture from a *parametrically formulated* non-Gaussian distribution [125]. Though derived in this parametric context, Xinfomax can be considered as a non-parametric optimization of *arbitrary distributions* because, effectively, the cost function acts as a "nonlinear moment" of the arbitrary distribution. This is consistent with the several levels of approximation that ICA algorithms generally undergo in their derivation (Sec. 3.5.5).

Let $\langle\cdot\rangle$ represent the sample mean operator. When expressed with default parameters in negative log-likelihood form (for consistency with a cost minimization), the Xinfomax cost function is [125]

$$\mathcal{I}^{\text{Xinfomax}}(\mathbf{W}, \boldsymbol{x}) \stackrel{\text{def}}{=} -\log|\det(\mathbf{W})| - \sum_{j=1}^{N_c} \left\langle \log \eta\left(\mathbf{e}_j^{\mathrm{T}}\mathbf{W}\boldsymbol{x}(n), \mathbf{e}_j^{\mathrm{T}} k(\mathbf{W}\boldsymbol{x})\right)\right\rangle \tag{6.5}$$

where the nonlinear function $\eta\,(y(n), k)$ is defined piecewise as

$$\eta(y_j(n), k_j) \stackrel{\text{def}}{=} \begin{cases} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y_j^2(n)\right) \operatorname{sech}^2 y_j(n) & : k_j \geq 0 \\ \frac{1}{(2\sqrt{2\pi})} \exp\left(-\frac{1}{2}(y_j(n)-1)^2\right) + \frac{1}{(2\sqrt{2\pi})} \exp\left(-\frac{1}{2}(y_j(n)+1)^2\right) & : k_j < 0 \end{cases} \tag{6.6}$$

Here, the piecewise nonlinearity $\eta(y_j(n), k_j)$ is based on the kurtosis $\boldsymbol{k} = k(\mathbf{W}\boldsymbol{x})$ to account for the proper optimization of both super-, $k_j \geq 0$, and sub-Gaussian distributions, $k_j < 0$. Here, (6.6) is a strange formula, but it is consistent with the stochastic natural gradient descent formula

$$\boldsymbol{y} = \mathbf{W}\boldsymbol{x} \tag{6.7}$$

$$\Delta\mathbf{W} \propto \left(\mathbf{I} - \mathbf{K}\tanh(\underline{\boldsymbol{y}}(n))\underline{\boldsymbol{y}}^{\mathrm{T}}(n) - \underline{\boldsymbol{y}}(n)\underline{\boldsymbol{y}}^{\mathrm{T}}(n)\right)\mathbf{W} \tag{6.8}$$

as a final byproduct. The kurtosis switching appears here as the matrix $\mathbf{K}$, which is a diagonal unitary matrix with entries $K_{jj} = \operatorname{sign} k(\boldsymbol{y}_j)$.

Thus the B-class SQM, $\aleph$, drawn from the Xinfomax cost function is algorithmically computed from the sphered observations $\boldsymbol{x}(n)$ as

$$\boldsymbol{y}(n) \leftarrow \mathbf{W}(\theta)\boldsymbol{x}(n)\ \forall n \tag{6.9}$$

$$\boldsymbol{k} \leftarrow k(\boldsymbol{y}) \tag{6.10}$$

$$z_j(n) \leftarrow \eta(y_j(n), k_j)\ \forall n,\ \forall j \tag{6.11}$$

$$\aleph^{\text{Xinfomax}}(\theta) \leftarrow -\log|\det\mathbf{W}(\theta)| - \sum_{j=1}^{N_c}\left(\frac{1}{N}\sum_{n=1}^{N}\log z_j(n)\right) \tag{6.12}$$

where the nonlinearity η in the definition of $z_j(n)$ is as defined in (6.6). This SQM estimate is implemented in the function `objectiveXinfomax.m`.

6.2.4 FastICA Cost Function

The FastICA objective is based on maximizing functions that estimate non-Gaussianity of a single channel by measuring the "norm" of some nonlinear feature with respect to a Gaussian standard [91], as in

$$\mathcal{I}^{\text{FastICA}}(\mathbf{e}_j^{\mathrm{T}}\mathbf{W}, \boldsymbol{x}) \stackrel{\text{def}}{=} \left(\langle\eta(\mathfrak{g})\rangle - \left\langle\eta\left(\mathbf{e}_j^{\mathrm{T}}\mathbf{W}\boldsymbol{x}\right)\right\rangle\right)^2 \tag{6.13}$$

Here, observations $\boldsymbol{x}$ are assumed to be zero-mean and sphered, and the first term is an equivalent zero-mean sphered Gaussian variable $\mathfrak{g}$. If the nonlinearity η is chosen as a polynomials, results are equivalent to traditional higher-order cumulants, while in the case of $\eta = \log f_{s_j}$, using the true source pdf, the result is a true negentropy.

Of the different nonlinearities presented in [91], studies have demonstrated [63] that the "Gauss" nonlinearity is the most robust and therefore it is considered here. This is the specific case that sets

$$\eta(y(n)) \stackrel{\text{def}}{=} -\exp\left(-\frac{1}{2}y^2(n)\right) \tag{6.14}$$

and, therefore, has an insensitivity to large values.

From this, the B-class SQM, $\aleph$, drawn from the FastICA cost function is algorithmically computed from the sphered observations $\boldsymbol{x}(n)$ as

$$\boldsymbol{y}(n) \leftarrow \mathbf{W}(\theta)\boldsymbol{x}(n)\ \forall n \tag{6.15}$$

$$\aleph^{\text{FastICA}}(\theta) \leftarrow -\sum_{j=1}^{Nc}\left[\frac{1}{\sqrt{2}} - \left(\frac{1}{N}\sum_{n=1}^{N} e^{-\frac{1}{2}y_j^2(n)}\right)\right]^2 \tag{6.16}$$

Note that the foremost negative sign in the cost function makes it suitable for minimization (versus a maximal non-Gaussianity measure). This SQM estimate is implemented in the function `objectiveFastICA.m`.

6.2.5 Radical Cost Function: Spacings-Estimates for Entropy

Entropy is a very important HOS, and its estimation is a field of research in its own right. If a model for the statistical distribution is assumed *a priori*, *parametric* formulas for the entropy can be used and would typically require the knowledge of certain moments (*c.f.*, (3.38)). *Non-parametric* estimation of the entropy can be approached from several standpoints, and a survey such as [20, Beirlant *et al.* (1997)] is recommended. One such approach is the *plug-in*

estimates, whereby a semi-parametric estimate of the pdf f is made first, and then the integral of (3.37) is evaluated (either as an analytic function of the resulting pdf formula, or through numerical integration/quadrature). The pdf estimation technique can range from the simple (*e.g.*, regular histograms [149, (1989)]) to the complex (*e.g.*, adaptive recursive partitioning [49], adaptive AR-models [23]). Other approaches to entropy estimation can be done through *spacings-estimates* [20]. These essentially use the characterization of a distribution by its cdf (*c.f.*, (3.4)), and can be effective on unidimensional distributions [20, Beirlant *et al.* (1997)][123, Miller and Fisher (2003)].

The *Robust, Accurate, Direct ICA ALgorithm* (RADICAL) approach to ICA [123, (2003)] is based on a spacings-estimate of the marginal entropies of the data $\boldsymbol{y}(n) = \mathbf{W}\boldsymbol{x}(n)$. A spacings estimate is based on the properties of the cumulative distributions function (cdf). Specifically, the property is that the *order statistics*

$$z_j(n') \stackrel{\text{def}}{=} \text{sort}\, y_j(n) \tag{6.17}$$

roughly represent the inverse of the cdf,

$$F(z_j(n')) \approx n'/N \tag{6.18}$$

in a direct way [20]. This is equivalent to the property that the likelihood of an *m-spacing*

$$\overline{z}^{[m]}(n') \stackrel{\text{def}}{=} z_j(n'+m) - z_j(n') \geq 0 \tag{6.19}$$

is expected to be $m/(N+1)$ for any underlying continuous distribution [123]. That is, for a fixed m, each m-spacing $\overline{z}^{[m]}(n')$ is equally likely and independent of n'. From this, an m-spacings estimate for the entropy can be estimated as [123, Eq. 8]

$$\widehat{H}^{[m]}(y_j) \stackrel{\text{def}}{=} \frac{1}{N-m}\sum_{n'} \log \frac{N+1}{m}\overline{z}^{[m]}(n') \tag{6.20}$$

$$= \frac{1}{N-m}\sum_{n'=1}^{N-m} \log \frac{N+1}{m}\left(z_j(n'+m) - z_j(n')\right) \tag{6.21}$$

If the MI of the data $\boldsymbol{y}(n) = \mathbf{W}\boldsymbol{x}(n)$ is considered, then it follows that

$$\widehat{\mathcal{M}}(\boldsymbol{y}) = \left(\sum_{j=1}^{N_c} \widehat{H}(y_j)\right) - \widehat{H}(\mathbf{W}\boldsymbol{x}) \tag{6.22}$$

$$= \left(\sum_{j=1}^{N_c} \widehat{H}(y_j)\right) - \log\det\mathbf{W} - \widehat{H}(\boldsymbol{x}) \tag{6.23}$$

of which the last term is independent of $\mathbf{W}$. Thus the estimate of a cost function can be made purely on one dimensional entropy estimates

$$\mathcal{I}^{\text{RADICAL}}(\mathbf{W}, \boldsymbol{x}) \stackrel{\text{def}}{=} -\log\det\mathbf{W} + \sum_{j=1}^{Nc} \widehat{H}^{[m]}(\mathbf{e}_j^{\mathrm{T}}\mathbf{W}\boldsymbol{x}) \tag{6.24}$$

From this, the B-class SQM, $\aleph$, drawn from the RADICAL cost function is algorithmically computed from the sphered observations $\boldsymbol{x}(n)$ as

$$\boldsymbol{y}(n) \leftarrow \mathbf{W}(\theta)\boldsymbol{x}(n)\ \forall n \tag{6.25}$$

$$\aleph^{\text{RADICAL}}(\theta) \leftarrow -\log\det\mathbf{W} + \sum_{j=1}^{Nc} \widehat{H}^{[m]}(\mathbf{e}_j^{\mathrm{T}}\mathbf{W}\boldsymbol{x}) \tag{6.26}$$

with the entropy estimate $\widehat{H}^{[m]}$ actually simplified from (6.21) by ignoring all common normalization as

$$\widehat{H}^{[m]}(y_j) \stackrel{\text{def}}{=} \sum_{n'=1}^{N-m} \log\left(z_j(n'+m) - z_j(n')\right) \tag{6.27}$$

This SQM estimate is implemented in the `objectiveRadical.m` function.

6.2.6 MILCA Cost Function: Estimating Mutual Information

The previous section presented an application of spacings-estimates to the marginal entropies of the multidimensional system. This is essentially a one dimensional approach, and it is quite difficult to extend this process to multidimensional supports [146, Miller (2003)]. A related approach developed by Grassberger *et al.* [215, (2004)] counts neighbours in the

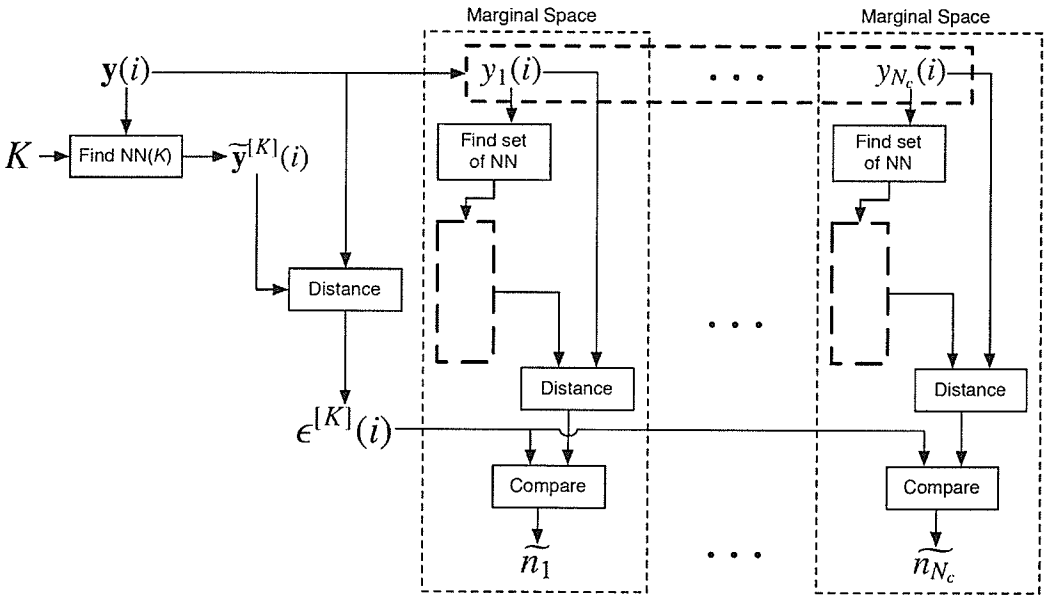


Fig. 6.1 Grassberger MI algorithm: Beginning from a multidimensional sample $\boldsymbol{y}(i)$, the Kth nearest-neighbour defines a distance $\epsilon^{[K]}(i)$. This distance defines intervals in each marginal space, j, and the number of neighbours $\widetilde{n}_j(\epsilon^{[K]}(i))$ in these intervals define the contribution of that point $\boldsymbol{y}(i)$ to the MI estimate as $\sum_j \mathring{\psi}(\widetilde{n}_j(\epsilon^{[K]}(i)))$ through the digamma function $\mathring{\psi}$. The algorithm then averages over all i, (6.29).

marginals on spacings defined by the nearest neighbours in the joint space. Since the behaviour of the joint entropy is implicitly included in this technique, it actually estimates the MI and is distinct from the RADICAL cost function which ignored the joint-entropy term. Thus, this section primarily presents an estimator of the MI, $\widehat{\mathcal{M}}$, as an HOS feature for convergence analysis, but secondarily, also presents an ICA cost function $\mathcal{I}^{\text{MILCA}} \stackrel{\text{def}}{=} \widehat{\mathcal{M}}$.

The Grassberger technique [120, Grassberger *et al.* (2004)] estimates the average difference between the sum of the marginal entropies and the joint entropy by a comparison of nearest-neighbours between the joint-space (N_c-dimensional) and the marginal spaces (Fig. 6.1). That is, the K-th nearest-neighbour $\widetilde{\boldsymbol{y}}^{[K]}(i)$ to the point $\boldsymbol{y}(i)$ in the joint space produces a distance $\epsilon^{[K]}(i) = \left\|\widetilde{\boldsymbol{y}}^{[K]}(i) - \boldsymbol{y}(i)\right\|_\infty$ which will vary statistically for all the $i = 1, \ldots, N$ points.

This variable distance $\epsilon^{[K]}(i)$ defines, on each marginal axis j, an interval around the marginal value $y_j(i)$. Then, letting $\widetilde{n}_j(\epsilon^{[K]}(i))$ be the number of marginal neighbours in that interval, *i.e.*

$$\widetilde{n}_j(\epsilon^{[K]}(i)) \stackrel{\text{def}}{=} \sum_{i' \neq i} \Xi\left(\epsilon^{[K]}(i) - \left|y_j(i') - y_j(i)\right|\right) \tag{6.28}$$

it follows that [120, Eq. 23] an estimate for the MI can be expressed in the form

$$\widehat{\mathcal{M}}(\underline{\boldsymbol{y}}) \stackrel{\text{def}}{=} \mathring{\psi}(K) + (N_c - 1)\mathring{\psi}(N) - \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{j=1}^{N_c} \mathring{\psi}(\widetilde{n}_j(\epsilon^{[K]}(i)))\right) \tag{6.29}$$

where the digamma function $\mathring{\psi}$ provides the suitable nonlinearity for estimating an HOS. Here, the properties of $f_{\boldsymbol{y}}$ drive the statistical average in the final term, while the first two terms simply reduce the bias from the estimator under the finite number of samples N and parameter K. A diagram of the algorithmic flow of the MI estimator for a single sample $\boldsymbol{y}(i) \in \mathbb{R}^{N_c}$ is shown in Fig. 6.1.

A few extra processing steps are also used to assist the speed and quality of the algorithm. First, the data is translated to make $\min_i \boldsymbol{y}(i) = 0$ (*i.e.*, *nonzero*-mean). It is also scaled to have unit variance along each marginal. Second, the algorithm spends some overhead to speed up calculation. The primary element here is to use a box-assist technique similar to the one used in App. C to find the nearest neighbours of point $\boldsymbol{y}(i) \in \mathbb{R}^{N_c}$ more easily. Sorted marginal arrays are also used to identify the neighbours in the marginal spaces. A supplementary overhead used to speed up calculation is the calculation of $\mathring{\psi}(i)$ from $i = 1$ to N (via a recursion relationship), which is then stored in an array for later use. Third, a small uniform random variable is added to the input $\{\boldsymbol{y}\}$ to increase the "continuity" of the distribution. Consequently, (i) this technique can only provide a fractal analysis if the noise amplitude is set to zero (or to levels below the scaling region to be studied), and (ii) a non-deterministic element in the calculation may introduce some small variability in the MI output.

The MI estimator is implemented in `MIxnyn.m` with an underlying C executable and can be processed without user intervention.

From this, the B-class SQM, $\aleph$, drawn from the MILCA estimate of MI is algorithmically computed from the sphered observations $\boldsymbol{x}(n)$ as

$$\boldsymbol{y}(n) \leftarrow \mathbf{W}(\theta)\boldsymbol{x}(n)\ \forall n \tag{6.30}$$

$$\aleph^{\text{MILCA}}(\theta) \leftarrow \mathring{\psi}(K) + (N_c - 1)\mathring{\psi}(N) - \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{j=1}^{N_c}\mathring{\psi}(\widetilde{n}_j(\epsilon^{[K]}(i)))\right) \tag{6.31}$$

just as in (6.29). Similar to the other algorithms, default parameters are used here, which implies that $K = 6$.

This completes a discussion of the measurement of ICA convergence by the ICA objective functions themselves. The novel component of this experimental work is the comparison of *feature-based* SQMs under ICA convergence, which is discussed in the following sections.

6.3 Feature Class 1: Simple Statistics

This section describes the implementation of the simple statistical features to be used in the standardization of signals and the estimation of SQMs (in particular, SQMs based on kurtosis and mutual information). As described in Ch. 3, all statistical features represent the time-integrated distribution of the time series $\boldsymbol{x}(n) \in \mathbb{R}^{N_c}$ with samples $n = 1, \ldots, N$, and, hence, are modelled by iid estimators. Furthermore, all statistical features can be calculated without user input.

6.3.1 Second-Order (Nuisance) Statistics and Standardization

As described in Ch. 3, the mean and variance are essentially considered as nuisance parameters in ICA. These will be calculated normally for the purposes of standardization. Specifically, the mean and variance are calculated from the time series $\boldsymbol{x}(n)$ as

$$\widehat{\mu_1}(\boldsymbol{x}) = \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}(n) \tag{6.32}$$

$$\widehat{\mu_2}(\boldsymbol{x}) = \frac{1}{N}\sum_{n=1}^{N}\left(\boldsymbol{x}(n) - \widehat{\mu_1}(\underline{\boldsymbol{x}})\right)^2 \tag{6.33}$$

through calls to the built-in `Matlab` functions `mean.m` and `var.m` under the form `var(·,1)`. Note that the usual unbiased variance estimate is not used here, since no Gaussian modelling is being considered, only the measurement of moments.

A time series is considered *standardized* if it is zero-mean and unit variance. Any time series can be standardized by the transformation

$$\boldsymbol{x}(n) \leftarrow \frac{\boldsymbol{x}(n) - \widehat{\mu_1}(\boldsymbol{x})}{\sqrt{\widehat{\mu_2}(\boldsymbol{x})}} \tag{6.34}$$

which is implemented in the `standardized.m` function.

6.3.2 Second-Order Features

Two other second-order statistical features are calculated from the source estimate $\widehat{s}(n)$ and the original $s(n)$ as introduced in Sec. 3.5.7.5. After both signals are standardized, the cross-correlation index CCI is calculated

$$\mathrm{CCI}(\widehat{s}, s) \stackrel{\text{def}}{=} \frac{1}{N}\sum_{n=1}^{N}\widehat{s}(n)s(n) \tag{6.35}$$

Last, the SNR of the standardized signals is expressed in dB and calculated

$$\mathrm{SNR}^*(\widehat{s}, s) \stackrel{\text{def}}{=} 10\log_{10}\frac{\sum_{n=1}^{N} s^2(n)}{\sum_{n=1}^{N}\left(\widehat{s}(n) - s(n)\right)^2} \tag{6.36}$$

where the * explicitly identifies the SNR as being standardized by arguments of unit variance. Standardization is an important requirement for SNR measurement in consideration of ICA and its scaling degeneracy (*c.f.*, Secs. 3.5.3 and 3.5.7.4).

6.3.3 Higher-Order Features

An estimator for MI has already been presented in Sec. 6.2.6. The remaining higher-order statistic considered as a time series feature is the kurtosis. This is estimated

$$\widehat{k}(x) = \frac{1}{N\widehat{\mu}_2^2(x)}\left(\sum_{n=1}^{N}(x(n)-\widehat{\mu}_1(x))^4\right) - 3 \tag{6.37}$$

and implemented in `kurt.m`. This calculation is non-parametric, but may not be robust in the presence of outliers. Since outlier robustness is not the focus of this thesis, however, alternatives will not be developed here. Interested readers should consult [63, Gadhok (2007)] and references.

Now the second class of features for SQM consideration will be presented.

6.4 Feature Class 2: Nonlinear Features

A contribution of this work is the consideration of nonlinear features for convergence analysis. As discussed in Chs. 2 and 4, this is being approached through the MFA of a correlation partition from a reconstructed attractor. Recall that the benefit of attractor reconstruction is that the dynamical relationship between points in time can be evaluated through statistical means. As discussed in Sec. 4.6.1, the ECG attractor is a multidimensional representation of the underlying dynamics in a unidimensional time series, $x(n)$ for $n = 1, \ldots, N$, constructed through the lag embedding procedure [219, Takens (1981)][197, Yorke *et al.* (1991)]

$$\mathbf{y}^{[\Delta n, N_e]}(n) = \sum_{i=1}^{N_e} \mathbf{e}_i\, x(n + (i-1)\Delta n) \tag{6.38}$$

where the sample lag Δn and the embedding dimension N_e are the embedding parameters to be determined. This section will describe the practical procedures used to (i) estimate the embedding parameters $[\Delta n, N_e]$ and construct the attractor

$$\mathcal{A} \stackrel{\text{def}}{=} \left\{ \boldsymbol{y}^{[\Delta n, N_e]}(n) \,|\, n = 1, \ldots, N_o \right\} \tag{6.39}$$

(ii) estimate the Theiler window W_T for a correlation partition (CP) (*c.f.*, Sec. 4.6.2.3), (iii) calculate a CP from the embedding $\boldsymbol{y}^{[\Delta n, N_e]}$ (*c.f.*, Table 4.1), and (iv) estimate an MFA from the CP (*c.f.*, Table 4.1).

6.4.1 Sample-Lag

The optimum sample-lag for attractor reconstruction (in nonlinear time series analysis jargon) is one that *fully unfolds* the attractor. By this, it is meant that too small a sample-lag forces the attractor manifold to remain strongly correlated and, hence, bound to a diagonal of the embedding space. In such conditions, the attractor's pattern and statistics are not fully expressed. Too large a sample-lag, however, will begin to accentuate curvature and twisting in the attractor manifold, and, hence, over-folding confuses the attractor's pattern and statistics. Figure 6.2 shows the effect of changing sample-lag on the two-dimensional projection of (i) the Rössler attractor and (ii) a synthetic ECG attractor.

In the original (continuous) mathematical theory of Takens, the lag is a nuisance parameter (*i.e.*, without much effect). The experience of researchers in practice, however, has shown some numerical significance[1], and lag selection is part art as well as science. To properly unfold the attractor, the rule-of-thumb is to promote independence and decorrelation between the lag coordinates (which promotes the attractor away from the diagonal), but penalize large sample-lags (which will over-complicated the manifold). In the literature, two applications of

[1] *c.f.*, [108, NTSA (1997)][209, Small and Tse (2004)] and references

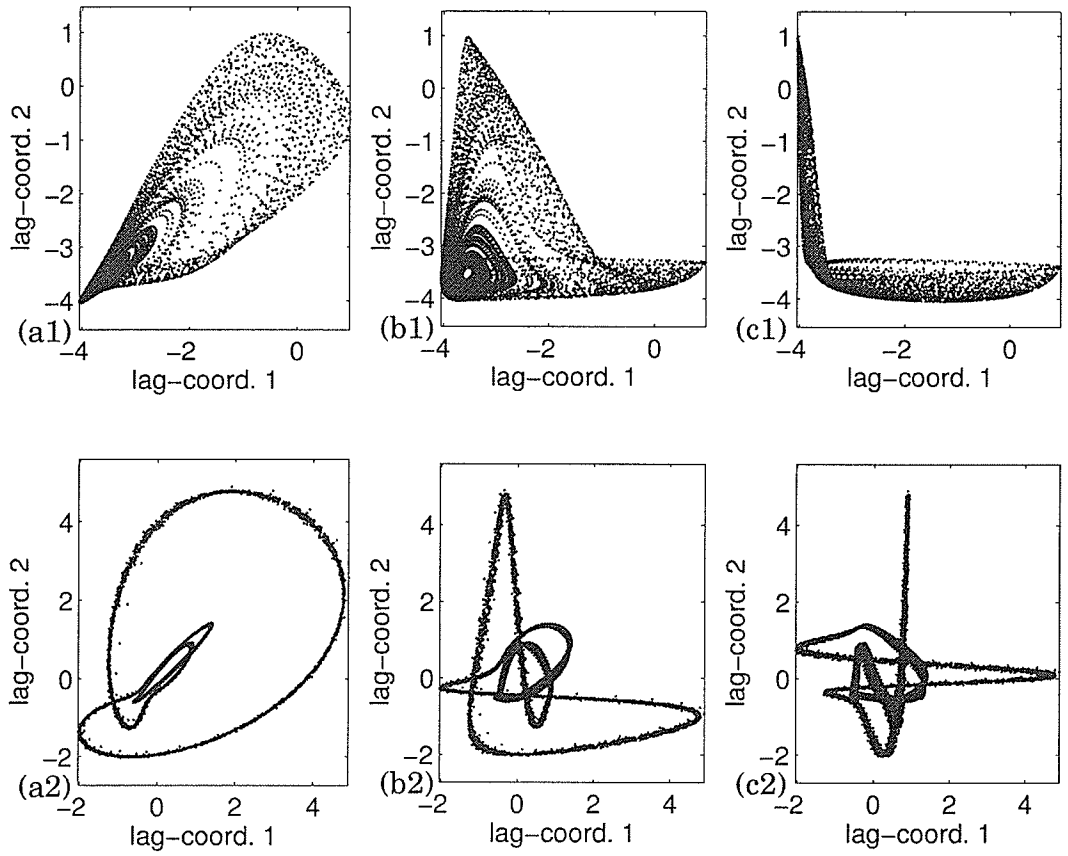


Fig. 6.2 Two-dimensional projections of the Rössler attractor (top) and synthetic fetal ECG (bottom) with different sample lags: (a) too small, (b) near optimal, and (c) too large. (Here, the Rössler attractor is observed under a logarithmic ($\log_2$) measurement.)

this rule-of-thumb are common for defining the reconstruction sample-lag. The first uses an autocorrelation measurement, and the second a measurement of mutual information.

6.4.1.1 Lag by Autocorrelation

The autocorrelation function of a random signal, $R_x(t_1, t_2)$, was introduced on p. 50. Under the assumption of a wide-sense stationary process (*c.f.*, p. 51), R_x becomes a function only of the lag $t_2 - t_1$. Considering the time series $x(n)$ as an iid process, it follows that the autocorrelation function (2.8) can be estimated at interval samples of the sampling interval as

[163]

$$N'(\Delta n) \stackrel{\text{def}}{=} N - \Delta n \tag{6.40}$$

$$R_x(\Delta n) \stackrel{\text{def}}{=} \frac{1}{N'(\Delta n)} \sum_{i=1}^{N'(\Delta n)} x(i)x(i+\Delta n) \tag{6.41}$$

where the normalization by $N'(\Delta n)$ is the unbiased estimate for finite N. Two significant features of $R_x(\Delta n)$ have been considered for estimation of the embedding sample-lag. One such consideration is the *autocorrelation time*, λ_n,

$$\lambda_n \stackrel{\text{def}}{=} \min \arg \min_{\Delta n} \left| R(\Delta n) - \frac{1}{e} R(0) \right| \tag{6.42}$$

By definition, λ_n is the time (in samples) it takes for the autocorrelation function to decay from its maximum at $\Delta n = 0$ to $\exp(-1)R_x(0)$ [108]. Another consideration for the definition of Δn is the first zero of $R_x(\Delta n)$

$$\lambda_0 \stackrel{\text{def}}{=} \min \arg \min_{\Delta n} |R(\Delta n)| \tag{6.43}$$

Both of these conditions fit the rule-of-thumb for unfolding an attractor. Since they both occur on the downslope from the maximum at zero lag, "large lags" are penalized (*i.e.*, effectively never considered[2]). Although the degree of "unfolding" represented by either selection is not made quantitatively explicit, it has been argued [224, Theiler *et al.* (1993)][108, NTSA (1997)] that the presence of strong linear correlation in an attractor can inhibit the accurate analysis of nonlinear behaviour. From this standpoint, a zero of R_x would stand as the preferred unfolding, since the linear relationship between the time-integrated distributions of the lag-coordinates is minimized.

Notwithstanding which R_x feature is considered, the estimation of Δn occurs relative to the maximum that occurs at zero lag, $R_x(0)$, (which is effectively a measurement of the power

[2]Some care, however, must be taken in conditions of white process, where $R_x(\Delta n) \approx R_x(0)$ and so both features are effectively "null". White processes can use $\Delta n = 1$.

(mean energy) in the signal). It is convenient, therefore, to apply the normalization

$$R_x(\Delta n) \leftarrow \frac{R_x(\Delta n)}{R_x(0)} \tag{6.44}$$

so that $\max_{\Delta n} R_x(\Delta n) = R_x(0) = 1$ prior to an analysis of λ_n or the zero λ_0.

6.4.1.2 Lag by Mutual Information

Instead of evaluating the autocorrelation of the joint-distribution of lag-coordinates as a function of the lag, $R(\Delta n)$, a supplementary approach to estimate the sample-lag [61, Fraser *et al.* (1986)] is to evaluate the MI, $\mathcal{M}_x(\Delta n)$, of the second-instance distribution

$$y(n) \stackrel{\text{def}}{=} \begin{bmatrix} x(n) \\ x(n+\Delta n) \end{bmatrix} \tag{6.45}$$

$$\mathcal{M}_x(\Delta n) \stackrel{\text{def}}{=} \mathcal{M}(y(n)) \tag{6.46}$$

and the Grassberger algorithm described previously may be used for the calculation. In all likelihood, $\mathcal{M}_x(\Delta n)$ approaches zero only asymptotically in the infinite limit $\Delta n \rightarrow \infty$, and therefore a zero cannot be taken as the determining feature for Δn. Instead, Fraser and Swinney [61, (1986)] suggest using the first minimum of $\mathcal{M}_x(\Delta n)$ to identify Δn [3, Ch. 7]. (Hence it is usually presumed that the MI is generally smooth or prohibits the identification of spurious noise minima by denoising to some degree.) As with the autocorrelation technique, this follows the rule-of-thumb of penalizing large lags while increasing the unfolding of the attractor in the embedding space. At a local minimum of $\mathcal{M}_x(\Delta n)$, it can be argued that the lag-coordinates are minimally redundant (*i.e.*, as independent as possible under small changes) and, thereby, maximally filling the embedding space. Certainly it suits the nature of this work to consider the HOS of the joint-distribution of the lag coordinates and not only their linear effects. The nature of the ECG attractor, however, makes things slightly complicated.

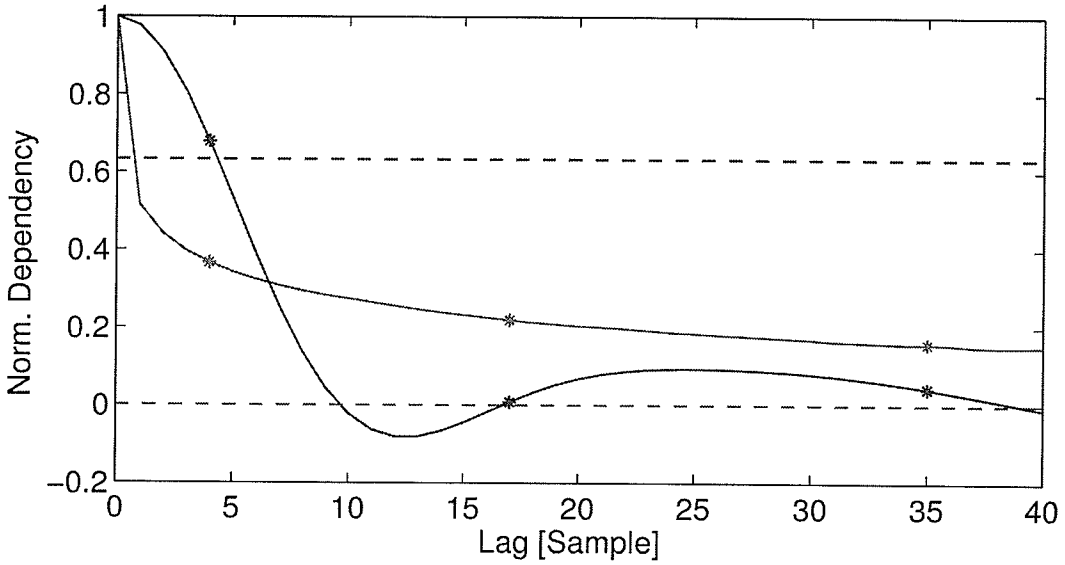


Fig. 6.3 Estimation of fECG embedding sample-lag: Autocorrelation function $R_x(\Delta n)$ (dark solid) and MI $\mathcal{M}_x(\Delta n)$ (grey) as a function of the lag Δn. The R_x curve has two early zeros related to the QRS complex (which has an average width of 20 samples). MI decreases steadily with increasing lag. The lags used for the fECG scatterplots in Fig. 6.2 are marked (stars). This work proposes to use the second zero of R_x as the embedding lag for fECG attractor reconstruction.

6.4.1.3 Lag Selection for ECG Reconstruction

Figure 6.3 shows the concurrent plots of the $R_x(\Delta n)$ and $\mathcal{M}_x(\Delta n)$ for a synthetic fECG signal. As described in [108, NTSA (1997)], the autocorrelation is dominated by the high-power and short-width QRS-complex. With these two features so linked, it follows that at λ_0, the first zero of $R_x(\Delta n)$, the QRS loop is separated well (as in Fig. 6.2(a2)). The smaller (low-power) waves, however, are still tightly bundled near the baseline and not very well-expressed. At larger lags, where the $\mathcal{M}_x(\Delta n)$ has decreased, these smaller loops are unfolded better, but the QRS loop now begins to overlap itself (as in Fig. 6.2(c2)). With this in mind, this work proposes to use the *second zero of the autocorrelation function*, λ_{0^2}, for determining the embedding sample-lag as a compromise. This still follows the rule-of-thumb of keeping

the lag small[3], encouraging unfolding, and minimizing linear correlations. In fact, since the first two zeros of the autocorrelation occur on the downslope of the MI, this technique should encourage greater unfolding than the first zero of $R_x(\Delta n)$. While this technique may not apply in general to non-ECG signals, it should generalize to other ECG morphologies (*i.e.*, lead configurations) since the expression of the QRS complex persists in all leads (in either a polar or bipolar form). Further work would be required to establish this as a general principle, but for the limited purposes of this work (which uses a fixed morphology), it seems to suffice. The fECG attractor represented in the bottom images of Fig. 6.2 correspond to the values $\Delta n = 4$ (left), the second zero of R_x, $\Delta n = 17$ (centre), and $\Delta n = 35$ (right) respectively. These points are also identified in Fig. 6.3 (stars).

This autocorrelation-based estimate is implemented in a `Matlab` program `bestecglag.m`. The second zero is identified by being the "first increasing zero" in $R_x(\Delta n)$. This algorithm iterates until either the optimum lag is found or it reaches the maximum lag length specified in the argument. As such, no user intervention is required, but if the algorithm exits unsuccessfully (at the maximum lag), nothing is assigned and user intervention may be required. The results of the lag estimate can then be displayed using the `ViewLagAutoCorr.m` function (as shown in Fig. 6.3).

Remark 6.1. More can be said on the issue of lag-embedding models, but more complicated forms [103, Mees *et al.* (1998)][209, Small and Tse (2004)] will not be considered in this work. These references consider the effect of multiple characteristic times that occur in frequency-modulated signals, and therefore would be of interest in the analysis of ECG (*c.f.*, App. A).

Although the embedding dimension is the last significant parameter required for the attractor reconstruction, the determination of the Theiler window will be discussed next, since

[3] In fact, effectively smaller than an MI-based lag.

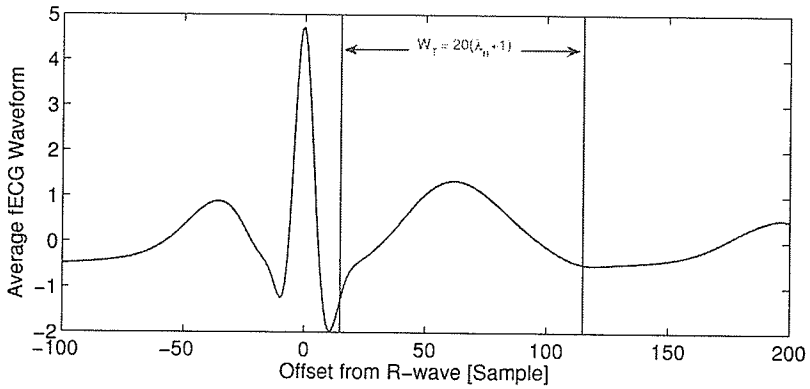


Fig. 6.4 The Theiler window for the cyclostationary ECG should exclude members of a single ECG wave. For the average fECG beat here, a $W_T = 20(\lambda_n + 1)$ is sufficient (as shown).

it will be used in the estimation technique of N_e.

6.4.2 Theiler Window

The Theiler window, W_T, is a parameter required for the determination of a correlation partition. Although it would seem out of place to discuss it here (an attractor hasn't been reconstructed yet!), W_T is a time-parameter just like the sample-lag and it will be used for the estimation of N_e that follows. The minimal principle for choosing W_T is that the Theiler window should be at least as long as the autocorrelation decay, $W_T \geq (\lambda_n + 1)$ or perhaps larger by another factor. The more complicated bounds and estimates for W_T (such as [182, Provenzale *et al.* (1992)][108, NTSA (1997)]) can be forsaken here because the ECG is a cyclostationary signal. Here, it simply suffices to use the length of the ECG waves to eliminate those samples that appear close in time. Essentially, no points within the same wave should be counted as neighbours. This can be achieved by picking

$$\widehat{W_T} \approx \frac{1}{2} \langle \Delta_R(n) \rangle \tag{6.47}$$

In the case of a synthetic fECG signal used as an example through this section, this amounts to $\widehat{W_T} \approx 100$, as shown in Fig. 6.4. For better generalization to multiple sampling frequencies,

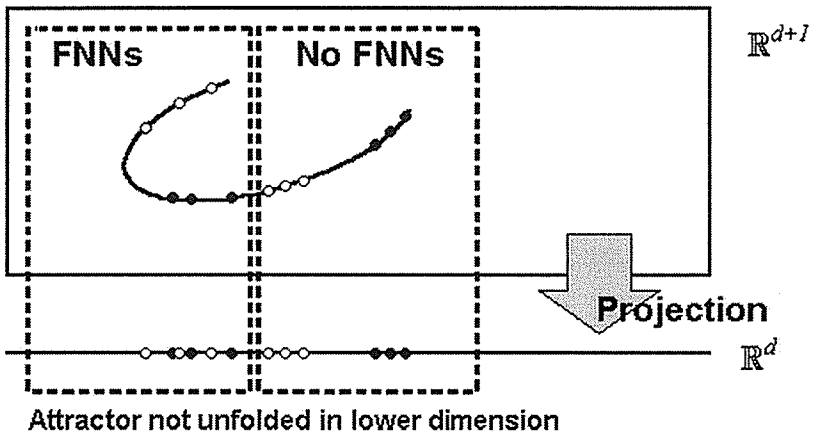


Fig. 6.5 The projection of higher-dimensional points onto subspaces introduces the false neighbour effect.

this is written in Theiler's product form as

$$\widehat{W_T} \stackrel{\text{def}}{=} \widetilde{\lambda}(\lambda_n + 1) \tag{6.48}$$

which is invariant to sampling frequency. In this work, the stretch-factor $\widetilde{\lambda}$ in (6.48) is set to 20 to be consistent with the previous ECG-based estimate (6.47).

6.4.3 Embedding Dimension

The estimation of the embedding dimension, N_e, for attractor reconstruction can be approached in multiple ways. The *false nearest neighbour* (FNN) technique [114, Kennel *et al.* (1992)][84][113, (2002)][108, NTSA, Ch. 9] is a common approach which identifies the significance of embedding dimension to dynamical variables. Essentially, if N_e is too small, the observed reconstruction is a projection of the attractor into a smaller embedding space and therefore parts of the attractor that are actually distant appear overlapping, as shown in Fig. 6.5. This can introduce error in statistics and other dynamical characterizations of the attractor. The weakness in the FNN methodology, however, is the sensitivity to its parameters

(*e.g.*, thresholds) and a demonstrated insensitivity to surrogate data randomization [84, Hegger and Kantz (1999)]. A more primitive approach to estimating the embedding dimension examines a dimension estimate (such as D_2) for saturation as the embedding dimension N_e is increased [73, Grassberger *et al.* (1983)][84]. Here, as with FNN, projections introduce measurable artifacts in the dimension, which disappear once the attractor is sufficiently embedded. On the other hand, noise signals (and randomized surrogates in particular) should be distinguishable by their ability to fill the embedding space, and so have a linear relationship between $D_2(\mathbf{y} \in \mathbb{R}^{N_e})$ and N_e (with no saturation). Here, since the feature of interest from the attractor is an MFA, this approach to estimating embedding dimension is also more significant to the end goal of the feature estimation.

Using the convergence of a fractal dimension for the estimation N_e, however, requires either (i) many repeated calculations of the dimension for different N_e, or (ii) a parallelization to calculate them all at once. Since the underlying complexity of a dimension estimate (from a correlation partition) is $O(N^2)$, repeated calculations are prohibitive (since the pairwise comparisons are repeated each time). Hence, parallel methods to estimate the CP are preferred. Furthermore, since only one dimension is required for convergence analysis over N_e, the fastest of all dimensions, D_2, can be considered. This means that only the correlation sum Z_C, (4.169), is required and not a full CP, (4.160). To streamline this calculation even further, Hegger, Kantz, and Schreiber [83, (1999)] use a random sampling approach[4] to reduce the number of comparisons to achieve a given accuracy. This takes advantage of the property that larger scales do not require as many comparisons to achieve reasonable statistics as the smaller scales. Of course, "large and small scales" vary somewhat with changing embedding dimension (since the mean distance between points increases with N_e). The TISEAN

[4]Random sampling for correlation-sums has been considered since 1992 [112, Kember *et al.*].

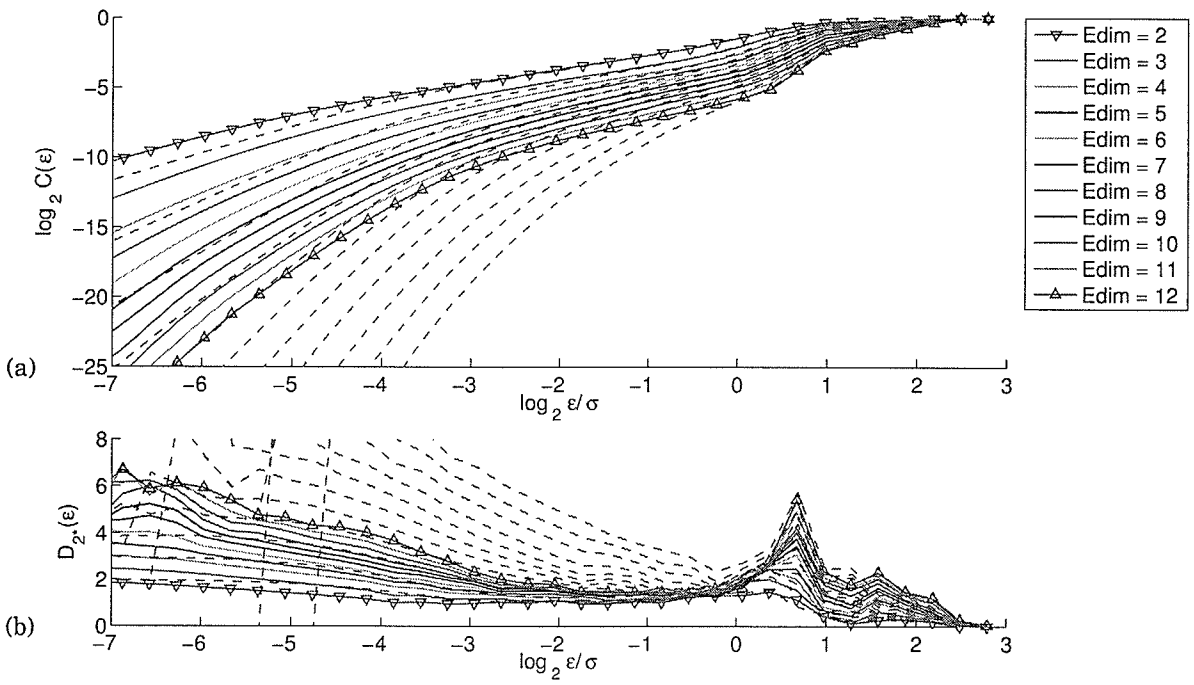


Fig. 6.6 Correlation sum scaling (a) and local gradient (b) for a synthetic fECG signal (dark solid) with increasing embedding dimension. (The CI curves are essentially monotonic from top to bottom in panel (a) (bottom to top in panel (b)) with increasing N_e.) The CI curves for a representative T21-surrogate is also shown (grey dashed), identifying the plateau at the right of panel (a) (spike in panel (b)) as a non-scaling artifact.

package [83] has an algorithmic optimization which (i) uses only a "sufficient" number of points to achieve a Z_C calculation accuracy at the given scale $\epsilon(i)$ and embedding dimension N_e, (ii) processes the embedding dimensions in parallel, and (iii) uses a box-assist technique to find neighbours with greater efficiency. Note that, while useful in this context for Z_C, this randomization technique should not be considered for a general MFA (such as the one used in this work; *c.f.*, Table. 4.1) because small q increases the sensitivity of the entropy to the small statistics. Thus, in order to achieve accuracy for $q < 2$, all possible comparisons should be considered.

The TISEAN estimate is implemented in C as `d2.c` to which the author has added a Matlab

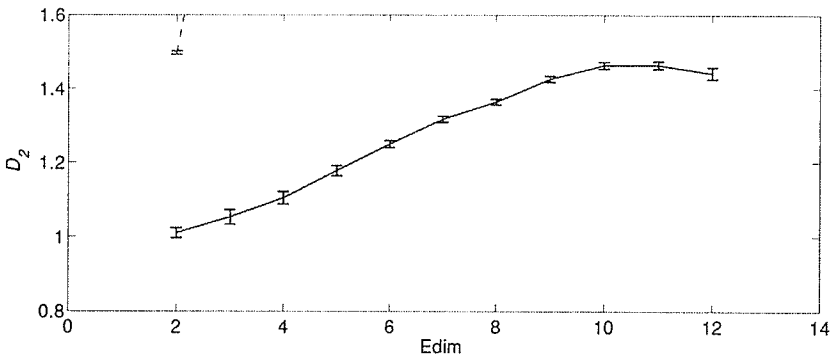


Fig. 6.7 Correlation dimension $D_2(N_e)$ as a function of embedding dimension N_e for a synthetic fECG signal (solid). Saturation occurs for the fECG signal at $N_e = 10$. The representative T21 surrogate (dashed) from Fig. 6.6 exceeds the scale of the graph.

interface as `d2.m`. The inputs to the algorithm include the time series itself, the sample-lag, the Theiler window, and the maximum embedding dimension to be considered. The output from the algorithm is (i) an array of values consisting of the embedding dimension $N_e(j)$, (ii) the scales $\epsilon(i)$, and (iii) the correlation sum, $Z_C(\epsilon(i), N_e(j))$, all of which are stored for further analysis. In particular, this output can be presented graphically via the function `ViewEdimDims.m`, as in Fig. 6.6, for a selection of scaling interval. Scaling intervals are evident by a clustering of horizontal lines in the lower panel (*e.g.*, Fig. 6.6(b) at -1 on the horizontal axis), which then translates into a region of linear behaviour in the upper panel (*e.g.*, Fig. 6.6(a)).

Once this interval is selected by visual inspection, it can be added to the function call to yield a presentation of the linear fit estimate $\widehat{D_2}(N_e)$ for that interval. If a saturation can be determined by visual inspection of the $\widehat{D_2}(N_e)$ over N_e plot, Fig. 6.7, then the embedding dimension for MFA can be set to that value using the `AssignEdim.m` function. The synthetic fECG signal represented in Fig. 6.7 has an embedding dimension of $N_e = 10$.

Having completely determined Δn and N_e, the finalized attractor is defined by (6.38) to

yield the set

$$\mathcal{A} \stackrel{\text{def}}{=} \left\{ \boldsymbol{y}^{[\Delta n, N_e]}(n) \mid n = 1, \ldots, N_o \right\} \tag{6.49}$$

which is stored as an array of size $N_o \times N_e$. The total number of points in the attractor, N_o, is given by

$$N_o \stackrel{\text{def}}{=} N - (N_e - 1)\Delta n \tag{6.50}$$

Remark 6.2. Since the embedding dimension requires user intervention to

(a) determine by visual inspection, and then specify, the scaling interval for $\widehat{D_2}$ analysis; and

(b) determine by visual inspection, and then specify, the embedding dimension;

this feature extraction must be separated from other automated elements in the experimental procedure.

6.4.4 Generating Surrogates

As mentioned in Sec. 5.5, only the T21 and PPS models are applied to generate surrogate data from the time series.

The T21 surrogates are generated from the algorithm of Ch. 5 in a `Matlab` implementation by Leontitsis[5] and packaged into the `surrogate.m` function. Since convergence of the T21 algorithm is very slow under ECG input, the number of allowed iterations was increased to 60 000, though typical runs would terminate around 2300–8000 iterations.

While a T21 surrogate requires no extra parameters apart from the time series itself, PPS requires embedding parameters and a PPS noise radius. The embedding parameters are taken

[5] With some additional cosmetic changes by the author.

in common from the previous analysis of the observed time series. The noise radius, however, is estimated internally by the `findrho.m` utility according to the optimal model suggested by Small (*c.f.*, Ch. 5).

The user can specify the number of surrogates to be generated. Since the purpose of the surrogates here does not include a rigorous statistical analysis, a sample of four surrogates is considered sufficient to identify a surrogate "representative" for comparison to the original without exhausting memory resources. Experiments indicate that this is sufficient to validate the selection of scaling intervals.

All representatives of the T21 and PPS surrogates to a time series $x(n)$ are generated by a single call to the `ThesisATTsurr2.m` function.

6.4.5 Correlation Partition

The calculation of the correlation partition is the most computationally intensive portion of the MFA routine. A CP is distinct from the more basic correlation sum because the reduction to a running sum is not possible: the individual contributions from each ball centre $y(n) \in \mathcal{A}$ must be kept distinct for calculation with the multiple values of q, (*c.f.*, Table 4.1). Though the algorithm for the calculation of the CP, `CIHptn.m`, is the subject in depth of App. C, its use and form are outlined here.

The fundamental arguments passed to the algorithm are the embedded data $\mathcal{A}$, the Theiler window, and a vector of scales

$$\epsilon = [\epsilon(1), \ldots, \epsilon(N_\epsilon)]^{\mathrm{T}} \tag{6.51}$$

which is not the scaling interval precisely, but the scales of calculation (which should certainly include the anticipated scaling interval). The vector ϵ must be provided by user input from the context of the $\widehat{D_2}$ scaling interval identified in Sec. 6.4.3.

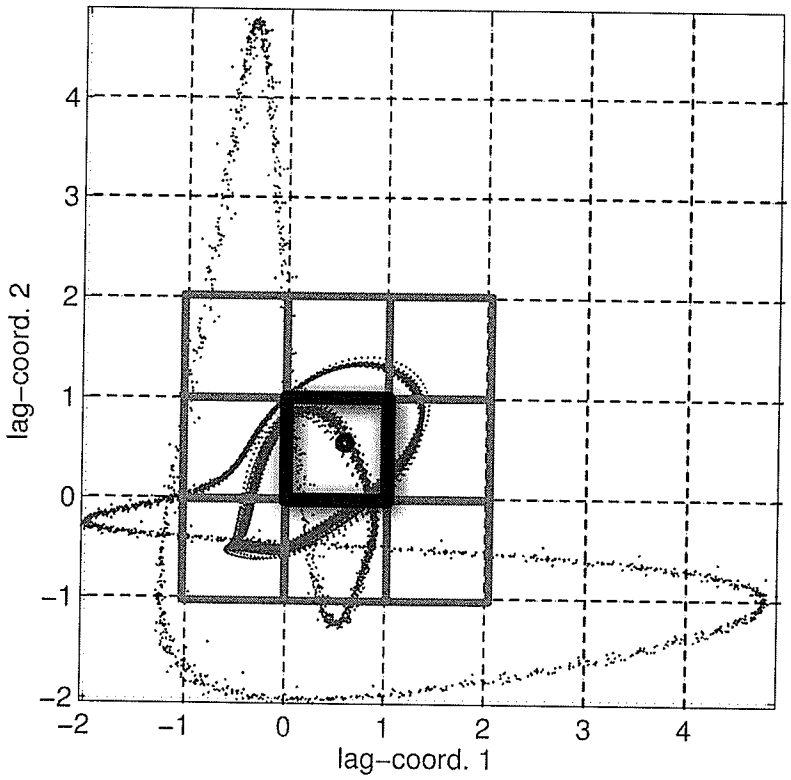


Fig. 6.8 CP box-assist: A 2D grid of size $\max_i \epsilon(i)$ is applied to the first two lag-coordinates of the attractor data. Neighbours on the attractor with distance less than $\epsilon(i)$ from the point $\boldsymbol{y}(n) \in \mathcal{A}$ (dark circle) must therefore be indexed within the centre box (dark) or one of its eight neighbours (grey).

Internally, the pairwise comparisons of the CP use a box-assist method on the first two lag-coordinates, $\left\{ \left(y_1^{[\Delta n, N_e]}(j), y_2^{[\Delta n, N_e]}(j) \right) \mid j = 1, \ldots, N_o \right\}$ and therefore, some overhead is spent in identifying the membership of each box. By using the Chebyshev norm, $\|\cdot\|_\infty$, distance relationships across increasing embedding dimensions are kept simple. By choosing the box width to be $\epsilon_g = \max_i \epsilon(i)$, all points $\boldsymbol{y}(j) \in \mathcal{A}$ within $\epsilon(i)$ of $\boldsymbol{y}(n)$ will be found in the neighbouring boxes to that of $\left(y_1^{[\Delta n, N_e]}(n), y_2^{[\Delta n, N_e]}(n) \right)$. Thus only the subset of points indexed in the box itself and its 8 neighbours need to be considered for pairwise analysis, as shown in Fig. 6.8. To preserve a benefit to the box-assist overhead, the scale vector $\epsilon(i)$ may be reduced by a factor

of 2 if the attractor is covered by fewer than 9 boxes.

The main loop of the CP algorithm proceeds by:

(a) Considering the point $\boldsymbol{y}(n) \in \mathcal{A}$ as a centre;

(b) Identifying by the box-assist indexing the possible neighbours of $\boldsymbol{y}(n)$ as a list;

(c) For those neighbours in the list satisfying the Theiler window criterion, calculating as a vector their distance from the centre point in each embedding dimension j (up to the maximum);

(d) Then, for each specified scale $\epsilon(i)$ and each embedding dimension j, the elements of the partition array (i, j) for the point $\boldsymbol{y}(n)$ are incremented if the interpoint distance is smaller than the specified scale $\epsilon(i)$;

(e) Updating to the next centre point, $n \leftarrow n + 1$, and repeat the loop until all points are exhausted.

After the main loop of the algorithm, the integer-valued CP is returned as a $N_o \times N_\epsilon \times N_e$ array (*i.e.*, the $N_i(\epsilon)$ of Table 4.1) and is normalized to make the true partition (*i.e.*, the $u_i(\epsilon)$ of Table 4.1). To preserve memory, only the full embedding is considered (and not all dimensions smaller than it) and therefore the final CP, $\mathcal{P}$, is an $N_o \times N_\epsilon$ array of floating point values. This calculation is fully automated in the function `CIHptn.m`, but because of long time series, it may take a substantial time to complete (*e.g.*, 30 min per time series of 30K points).

A visual representation of the regional properties of the CP is shown in Fig. 6.9. Here, the projection of the fECG attractor into 3D (shown in perspective) is coloured by the CP values. Note that the region of minimal CP probability is the R-wave loop while the region of maximal CP probability is the ST-interval (at least in this 3D projection from $\mathbb{R}^{10}$). Such an image is new to the MFA literature.

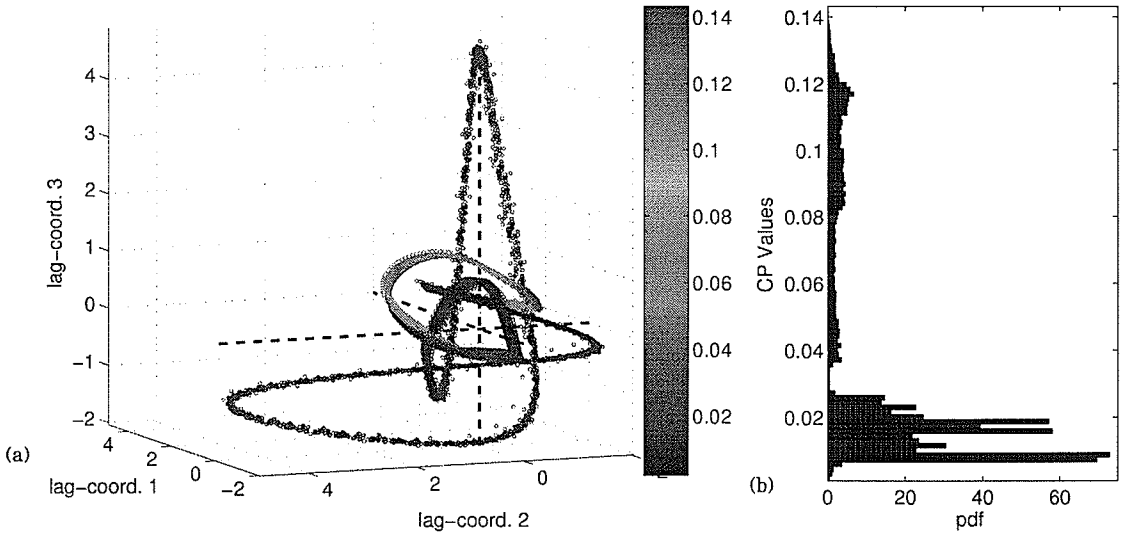


Fig. 6.9 ECG attractor CP density: (a) A 3D (perspective) scatterplot of the reconstructed fECG attractor coloured by the CP probability at scale $\epsilon = 2^{-0.1}$. Note that the R-wave loop is the region of minimal CP probability while the ST-interval is of maximal CP probability in this projection from $\mathbb{R}^{10}$. This image is a novel representation of an attractor. (b) The distribution of CP probability values at scale $\epsilon = 2^{-0.1}$.

6.4.6 Scaling Entropies and MFA Calculation

From the CP resulting from the preceding steps, the calculation of the Rényi, Hölder, and Mandelbrot entropies is then a straightforward application of the formulae from Table 4.1. The values of the exponent q are set by the user in a vector format. Note that to avoid numerical instability in the Rényi entropy, it is chosen that $q \neq 1$, and a higher density of q-values is recommended in the $[-1, 2]$ interval to capture the nonlinearity of the RS and the maximum of the MS.

Remark 6.3. In order to allow some user directed editing to q, the generalized scaling entropies themselves are not stored as arrays over $\epsilon(i)$ and $q(j)$. Instead, they can be recalculated from the correlation partition $\mathcal{P}$ which remains in memory. Since the CP is of the size $N_o \times N_\epsilon$, this is not an efficient use of memory, but the CP is more primitive, and recalculation of the

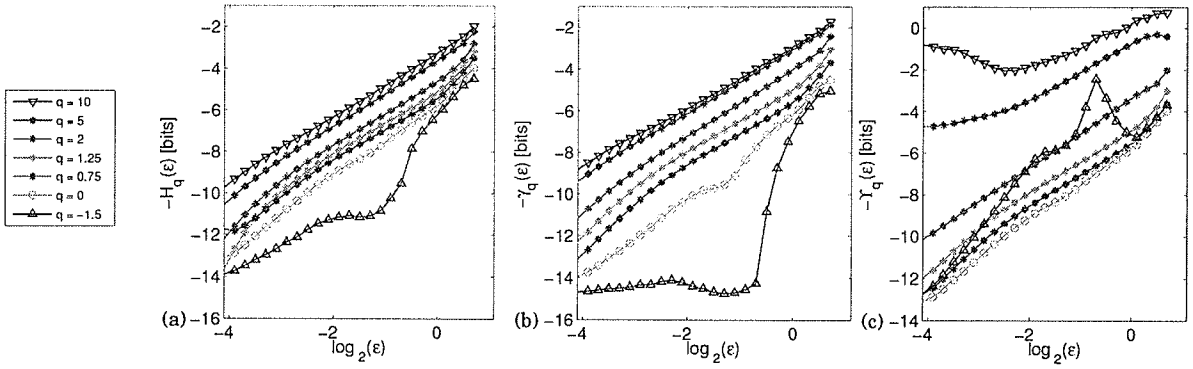


Fig. 6.10 Example of the MFA Scaling Entropies: Simultaneous scaling of (a) Rényi, (b) Hölder, and (c) Mandelbrot entropies of an fECG attractor (at 500 sps in $\mathbb{R}^{10}$ with $\Delta n = 17$). The legend at the left applies to all panels.

generalized scaling entropies with different q is much cheaper than a recalculation of the CP itself.

Using the scales vector ϵ, the CP, $\mathcal{P}$, and the q-vector as arguments, the `ViewEntropyScaling.m` function can plot the three scaling entropies as shown in Fig. 6.10. The analysis of these images are an important element of this work, and, therefore, some discussion of these figures will now be made. Note that the legend at the left applies to all panels.

As discussed in Ch. 4, it is expected that the finite size of the attractor introduce a saturation at large scales, while the finite sampling of the attractor or noise floor should induce a saturation at small scales. Here, panel (a) shows the scaling of the Rényi entropies from large q (top) to small (bottom). Note that saturation does not appear on the right in this image because $\max_i \epsilon(i)$ is sufficiently smaller than the size of the attractor (as required by the box-assist technique). At the bottom left, the finite sampling saturation affects the entropies differently depending on q. Specifically, $H_q(\epsilon)\big|_{q>1}$ diverges downwards as $\epsilon \to 0$, while $H_q(\epsilon)\big|_{q<1}$ appears concave-up in the same limit. Panel (b) shows the scaling of the Hölder entropies which are again ranked from large q (top) to small (bottom). Again, no saturation is

observed at large scales, but finite sampling saturation flattens out the curves at small scales. This effect is predominant in the small q entropies (*i.e.*, $q < 0$), which are more sensitive to the small probability sections of the attractor. For example, γ_q scaling for $q = 2$ persists along the interval $\log_2 \epsilon \in [-3, 0]$, while for $q = -2$, saturation effects dominate for $\log_2 \epsilon < -0.8$. Further discussion of the finite sampling effect is discussed later. Panel (c) shows the scaling of the Mandelbrot entropies. These are loosely ranked by $|q|$, so that $\min_q \Upsilon_q(\epsilon)$ is achieved at $q = 0$. The dense packing of curves at the bottom is expected by the higher density of q-values around zero. Furthermore, the large $q > 0$ curves can be distinguished from the $q < 0$, because they appear more smooth and do not approach the saturation asymptote at small scales. The significance of representing all entropies in the same figure is that the user can consider all entropies simultaneously while determining a scaling region. This step is discussed next.

6.4.6.1 Scaling Region

In keeping with the requirement that a scaling region be determined by inspection [108, NTSA (1997)], the user must personally analyze the generalized scaling entropies for a scaling region with images of the form of Fig. 6.10. Through this work, several useful elements have been identified to provide guidelines for specifying the scaling interval. These elements are specified in Table 6.1, but, generally speaking, the scaling interval is the largest interval with a strong linear approximation for $q \geq 0$ and as few nonlinear artifacts for $q < 0$ as considered reasonable. The application of these elements to scaling entropies is shown in Fig. 6.11 for an fECG attractor and in Fig. 6.12 for an fECG PPS surrogate.

Once a scaling interval is identified, it can be assigned to the CP using the `ViewEntropyScaling.m` command. Once assigned, these scaling intervals appear as vertical bars in the scaling entropy figures (*c.f.*, Fig. 6.13).

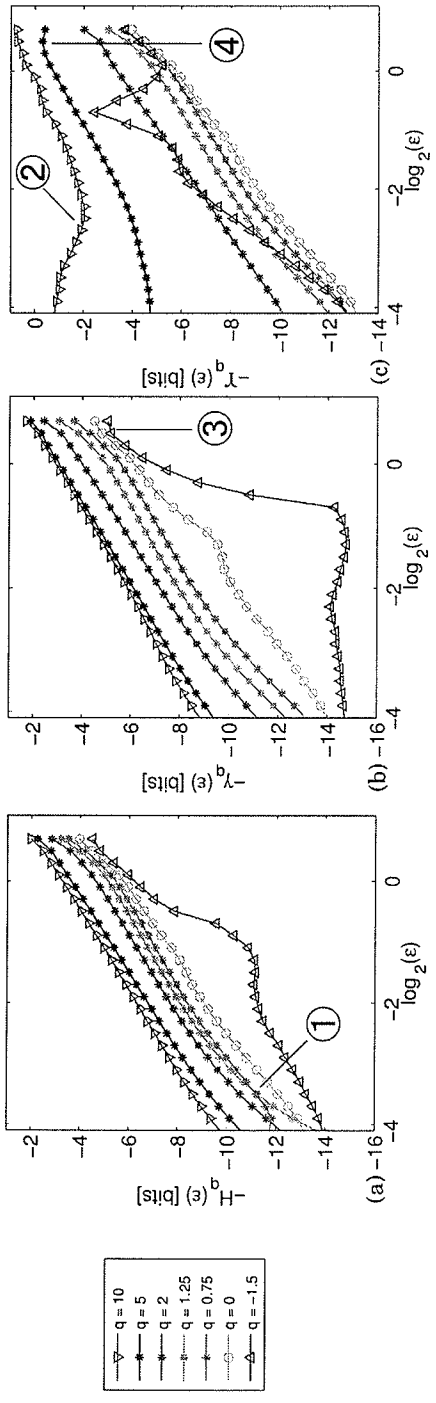


Fig. 6.11 Features of Scaling Entropies for Interval Selection: Simultaneous scaling of (a) Rényi, (b) Hölder, and (c) Mandelbrot entropies of an fECG attractor (at 500 sps in $\mathbb{R}^{10}$ with $\Delta n = 17$) with features from Table 6.1 as marked.

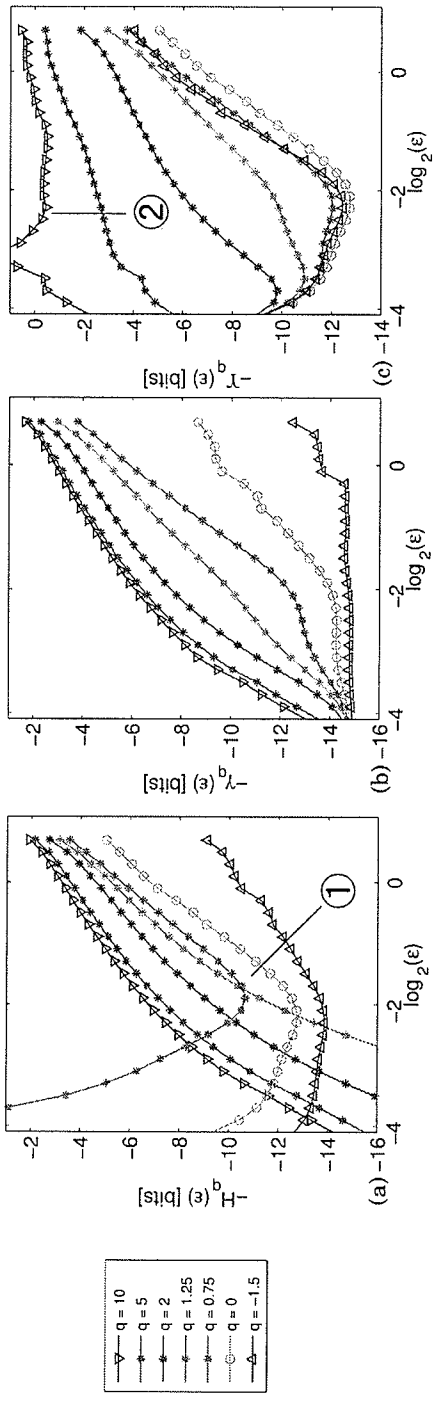


Fig. 6.12 Features of Scaling Entropies for Interval Selection: Simultaneous scaling of (a) Rényi, (b) Hölder, and (c) Mandelbrot entropies of an fECG PPS surrogate (of $\Delta n = 17$ in $\mathbb{R}^{10}$) with features from Table 6.1 as marked. Here, features ③ and ④ do not appear.

Table 6.1 Guidelines for Selection of Scaling Interval

Lower end of scaling region must be greater than the small ϵ crossover point of $H_q\big|_{\min q>1}$ and $H_q\big|_{\max q<1}$ (*c.f.*, ① in Fig. 6.11 and Fig. 6.12);

Lower end of scaling region should not be excessively smaller than any turning point of $\Upsilon_q\big|_{\max q}$ at small ϵ (*c.f.*, ② in Fig. 6.11 and Fig. 6.12);

Upper end of scaling region should not extend beyond any turning point of $\gamma_q\big|_{q<0}$ at large ϵ (*c.f.*, ③ in Fig. 6.11 and Fig. 6.12);

Upper end of scaling region should not be excessively larger than any turning point of $\Upsilon_q\big|_{\max q}$ at large ϵ (*c.f.*, ④ in Fig. 6.11 and Fig. 6.12);

Both ends should approximately balance any saturation curvature for H_q, γ_q, and Υ_q for $q \approx 2$, if possible.

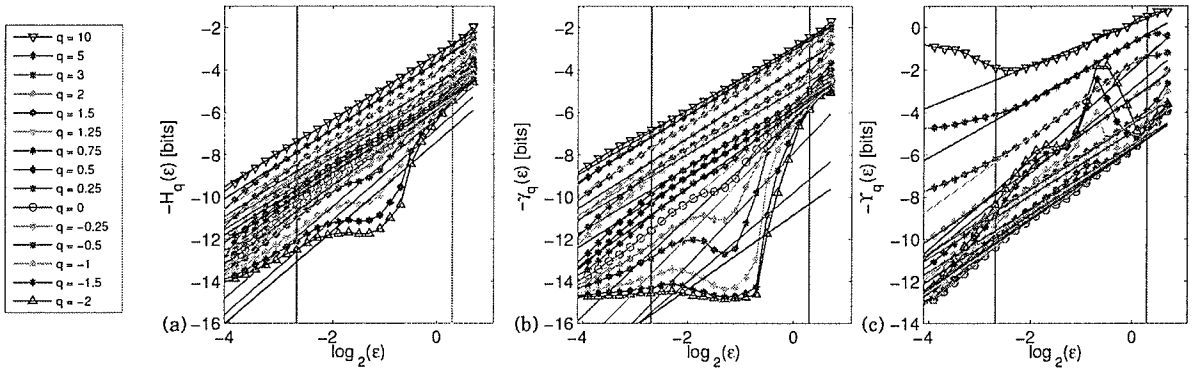


Fig. 6.13 MFA Scaling Entropies with Linear Fit over Scaling Interval: Simultaneous scaling of (a) Rényi, (b) Hölder, and (c) Mandelbrot entropies of an fECG attractor (at 500 sps in $\mathbb{R}^{10}$ with $\Delta n = 17$) with scaling interval and best linear fits as marked.

6.4.6.2 *Slope Estimates and Uncertainties*

Once the scaling interval for a CP has been established, the calculation of both the RS and MS is a straightforward application of the box-counting formalism to the formulae of Table 4.1. Implemented in the `calcMFA.m` function, a slope is fitted to the generalized entropies in the scaling region by minimum-absolute-deviation [180, Num. Rec. (1992), § 15.7]. That is, representing $\Gamma(i)$ as the numerator of the box-counting formalism (*e.g.*, $-H_q(\epsilon(i)) \equiv \Gamma(i)$ and similarly), the objective function

$$\sum_i \left| \Gamma(i) - \widehat{d}_1 \log \epsilon(i) - \widehat{d}_0 \right| \tag{6.52}$$

is minimized over the parameters $\widehat{d}_j$ for the samples i in the scaling interval. In particular, the slope estimate $\widehat{d}_1$ is returned as the generalized dimension of interest. This is the preferred fitting mechanism (over the more traditional minimum-squared-error) because it is more robust to outlier effects introduced by lacunarity or edge saturation. As previously discussed in Ch. 4, however, it is desired to have a measurement of uncertainty to the slope that provides context over the multiple value of q. For example, since saturation artifacts are still strong for $q = -2$ in the scaling region of Fig. 6.13, the "best linear fit" is not as reliable as for $q = 2$, and graphical representation of this imprecision is useful. Once the `calcMFA.m` function has been applied to calculate the MFA, a repeat call to `ViewEntropyScaling.m` includes in the plots the best straight line estimates, as shown in Fig. 6.13, and the variability in the precision of the fit can be discerned graphically. Quantitatively, relative imprecision in the fit is approached by estimating it as the slope uncertainty from a minimum-square-error model (*c.f.*, [180, Num. Rec. (1992), § 15.2]). This is somewhat mismatched from the estimate of the slope itself, but more convenient for implementation. The effective uncertainty in the estimated slope $\widehat{d}_1$ from the data $\Gamma(i)$ and $\log \epsilon(i)$ is of the form [180, Num. Rec. (1992), Eq. 15.2.20]

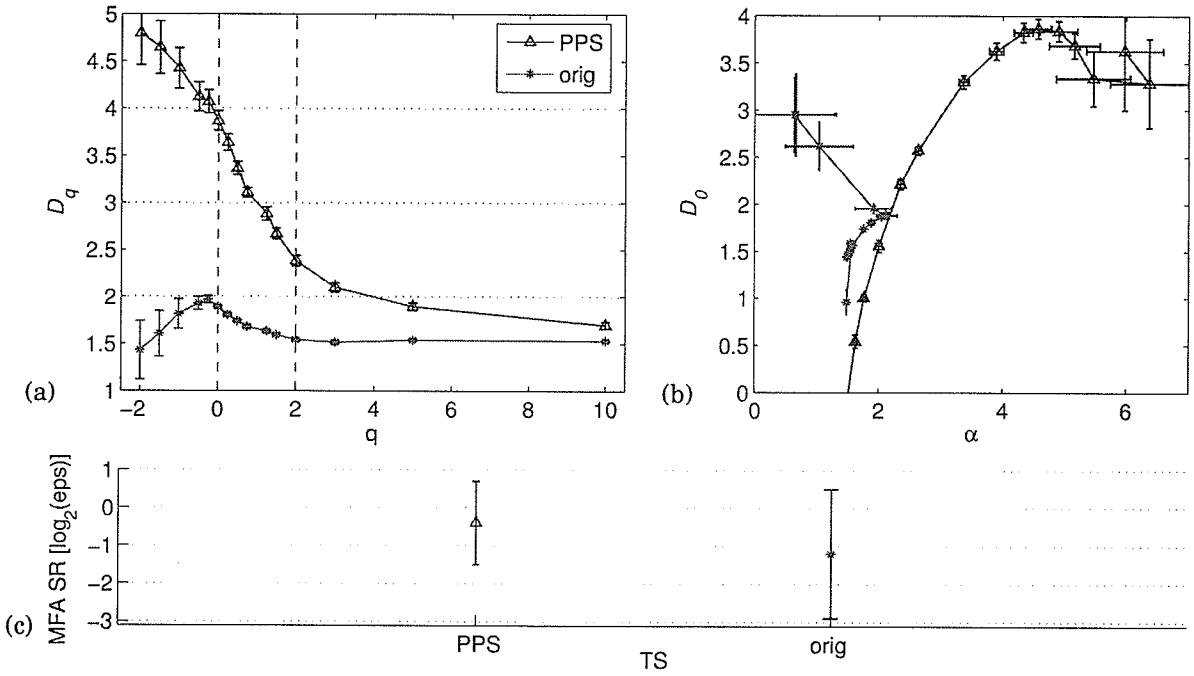


Fig. 6.14 Practical MFA spectra of fECG attractor via CP: (a) Rényi fractal dimension spectrum, and (b) Mandelbrot fractal singularity spectrum. Note that the uncertainty of the dimension estimates are nonuniform, and that panel (b) has uncertainties along both axes. Panel (c) also indicates the scaling regions

$$N_\epsilon{}' \stackrel{\text{def}}{=} \sum_i 1 \tag{6.53}$$

$$\ell(i) \stackrel{\text{def}}{=} \log \epsilon(i) - \frac{\sum_i \log \epsilon(i)}{N_\epsilon{}'} \tag{6.54}$$

$$\sigma_{\widehat{d}_1} \stackrel{\text{def}}{=} \left(\frac{1}{N_\epsilon{}' - 2} \frac{\sum_i \left(\Gamma(i) - \widehat{d}_1 \log \epsilon(i) - \widehat{d}_0 \right)^2}{\sum_i \ell^2(i)} \right)^{\frac{1}{2}} \tag{6.55}$$

where, again, the sum is made over all the $N_\epsilon{}'$ samples that are in the scaling interval. By this formalism, larger $\sigma_{\widehat{d}_1}$ signifies a decreased quality in the linear fit. The visualization of the MFA spectra, (4.165)-(4.167) including their uncertainty estimates and their supporting scaling interval, is provided by `ViewAllMFA.m` with results as shown in Fig. 6.14.

To conclude the discussion of feature extraction, a table summarizing the attractor and

Table 6.2 Summary of Practical Nonlinear Features

Feature	Symbol	Type	Defined
	Attractor		
Embedding Lag	Δn	auto	2nd zero of R, *c.f.* Sec. 6.4.1
Embedding Dimension	N_e	user	visual inspection, *c.f.* Sec. 6.4.3
	Parameters[‡]		
q-vector	$\boldsymbol{q}$	user	*c.f.* Sec. 6.4.6
ϵ-vector	$\boldsymbol{\epsilon}$	user	*c.f.* (6.51)
	Fractal		
Theiler Window	W_T	auto	(6.48)
Correlation Sum	Z_C	auto	(4.169) and Sec. 6.4.3
D_2 Scaling Interval	—	user	visual inspection, *c.f.* Sec. 6.4.3
Correlation-Partition	$\mathcal{P}$	auto	(4.160) and *c.f.* Sec. 6.4.5
MFA Entropies	$H_q(\epsilon)$, $\gamma_q(\epsilon)$, $\Upsilon_q(\epsilon)$	auto	(4.162)–(4.164)
MFA Scaling Interval	—	user	visual inspection, *c.f.* Sec. 6.4.6.1 and Table 6.1
MFA Spectra	D_q, $D_0(\alpha)$	auto	(4.165)–(4.167) and *c.f.* Sec. 6.4.6.2
MFA Uncertainties	—	auto	(6.55) and *c.f.* Sec. 6.4.6.2

[‡] These are not truly nonlinear features, but parameters required for the fractal features.

MFA features discussed in this section is displayed in Table 6.2.

6.5 SQM Convergence: ICA Error Space

As described in Chs. 1 and 2, the focus of this thesis is to evaluate feature convergence as a measure of practical ICA demixing performance. To accomplish this, the features described in the previous sections are used to construct feature-based SQMs to complement the usual variety of SQMs described in Sec. 3.5.7.5 and the ICA cost functions of Sec. 6.2.2. Specifically using the parameterization of the Stiefel manifold presented in Sec. 6.2.1, these SQMs, $\aleph_i$, create an *ICA error space*

$$[\aleph_1(\theta), \ldots, \aleph_{N_\aleph}(\theta)]^{\mathrm{T}} \tag{6.56}$$

in which the convergence relationships of the contributing SQMs can be analyzed.

The calculation of these SQMS will now be described fully.

6.5.1 B-Class SQMs: ICA Cost Functions

B-Class SQMs are those performance measures that can be calculated without using *a priori* information. They are represented by the ICA cost functions. In practice, however, these cost curves are not standardized in any way, making comparison difficult. (That is, they have both an arbitrary origin and arbitrary scale.) To ease comparison between them, the minima are translated to 0 and the scale of each SQM is normalized to produce an observed unit mean. Therefore, all the SQMs defined in (6.12), (6.16), (6.26), and (6.31) are updated according to

$$\aleph^I(\theta_j) \leftarrow \aleph^I(\theta_j) - \min_j \aleph^I(\theta_j) \tag{6.57}$$

$$\aleph^I(\theta_j) \leftarrow \aleph^I(\theta_j)\left(\frac{1}{N_\theta}\sum_{j=1}^{N_\theta}\aleph^I(\theta_j)\right)^{-1} \tag{6.58}$$

to be normalized for analysis.

6.5.2 A-Class SQMs

A-Class SQMs are those performance measures that use *a priori* information. They are categorized into four sub-classes: matrix SQMs, SOS-SQMs, HOS-SQMs, and nonlinear SQMs.

6.5.2.1 Matrix Sub-Class

The matrix sub-class of the A-Class SQMs are those performance measures that are explicitly based on the matrix parameters of the fetal ECG simulation: *i.e.*, the demixing matrix, $\mathbf{W}$; the sphering matrix, $\mathbf{W_0}$; and the mixing matrix $\mathbf{M}$. As presented in Sec. 3.5.7.5, the Amari performance index (3.86), the matrix Frobenius error MFE (3.103), and the vector norm error

VNE (3.101) are considered in this work; *i.e.*

$$\aleph^{A_\epsilon}(\theta) \stackrel{\text{def}}{=} A_\epsilon(\mathbf{W}(\theta)\mathbf{W_0}\mathbf{M}) \tag{6.59}$$

$$\aleph^{\text{MFE}}(\theta) \stackrel{\text{def}}{=} \left\|\mathbf{W}(\theta)\mathbf{W_0} - \mathbf{M}^{-1}\right\|_F \tag{6.60}$$

$$\aleph^{\text{VNE}}(\theta) \stackrel{\text{def}}{=} \left\|\mathbf{e}_2^{\text{T}}(\mathbf{W}(\theta)\mathbf{W_0} - \mathbf{M}^{-1})\right\| \tag{6.61}$$

where the VNE is specific to the fetal ECG estimate (consistently preserved in row 2.) Each of these SQMs are optimized at a minimum of zero, and they may be considered in their absolute units. It is also useful, however, to consider them in relative terms by normalizing them against their sample mean

$$\aleph^{A_\epsilon}(\theta_j) \leftarrow \aleph^{A_\epsilon}(\theta)\left(\frac{1}{N_\theta}\sum_{j=1}^{N_\theta}\aleph^{A_\epsilon}(\theta_j)\right)^{-1} \tag{6.62}$$

$$\aleph^{\text{MFE}}(\theta_j) \leftarrow \aleph^{\text{MFE}}(\theta)\left(\frac{1}{N_\theta}\sum_{j=1}^{N_\theta}\aleph^{\text{MFE}}(\theta_j)\right)^{-1} \tag{6.63}$$

$$\aleph^{\text{VNE}}(\theta_j) \leftarrow \aleph^{\text{VNE}}(\theta)\left(\frac{1}{N_\theta}\sum_{j=1}^{N_\theta}\aleph^{\text{VNE}}(\theta_j)\right)^{-1} \tag{6.64}$$

6.5.2.2 SOS Sub-Class

The SOS sub-class of the A-Class SQMs are those performance measures that explicitly compare the original s and estimated $\widehat{s}(\theta)$ signals using second-order statistics. As presented in Sec. 3.5.7.5, the standardized signal-to-noise ratio, SNR*, (6.36), and the cross-correlation index, CCI, (6.35), are considered in this work; *i.e.*

$$\aleph^{\text{SNR}^*}(\theta) \stackrel{\text{def}}{=} \text{SNR}^*(\widehat{s}(\theta), s) \tag{6.65}$$

$$\aleph^{\text{CCI}}(\theta) \stackrel{\text{def}}{=} \text{CCI}(\widehat{s}(\theta), s) \tag{6.66}$$

where the calculations of SNR* in (6.36) and CCI in (6.35) use the properties that both original fetal ECG time series s and its demixed estimate $\widehat{s}(\theta)$ are standardized to unit variance. Note

that the SNR diverges to ∞ as $\widehat{s}(\theta) \rightarrow s$, while the CCI increases to 1.

6.5.2.3 HOS Sub-Class

The HOS sub-class of the A-Class SQMs are those performance measures that explicitly compare the original s and estimated $\widehat{s}(\theta)$ signals using higher-order statistics. Three such measures are considered: (i) the *cross-mutual information* (xMI) of the source and the estimate, $\mathcal{M}(\widehat{s}(\theta), s)$, (ii) the *absolute deviation in the kurtosis* (ADK), $\left|\widehat{k}(\widehat{s}) - \widehat{k}(s)\right|$, and (iii) the *absolute deviation in the entropy* (ADE), $\left|\widehat{H}(\widehat{s}) - \widehat{H}(s)\right|$. These are, therefore,

$$\aleph^{\text{xMI}}(\theta) \stackrel{\text{def}}{=} \widehat{\mathcal{M}}(\widehat{s}(\theta), s) \tag{6.67}$$

$$\aleph^{\text{ADK}}(\theta) \stackrel{\text{def}}{=} \left|\widehat{k}(\widehat{s}(\theta)) - \widehat{k}(s)\right| \tag{6.68}$$

$$\aleph^{\text{ADE}}(\theta) \stackrel{\text{def}}{=} \left|\widehat{H}(\widehat{s}(\theta)) - \widehat{H}(s)\right| \tag{6.69}$$

where the calculations of $\widehat{\mathcal{M}}$, $\widehat{k}$, and $\widehat{H}$ are from (6.29), (6.37), and (6.27) respectively. Note that, in theory, the xMI diverges to ∞ as $\widehat{s}(\theta) \rightarrow s$, while the ADK and the ADE decrease to 0 in the same limit. Here, it is useful to normalize the absolute deviations by the values of the original signal's features, as

$$\aleph^{\text{ADK}}(\theta) \leftarrow \aleph^{\text{ADK}}(\theta)\left(\left|\widehat{k}(s)\right|\right)^{-1} \tag{6.70}$$

$$\aleph^{\text{ADE}}(\theta) \leftarrow \aleph^{\text{ADK}}(\theta)\left(\left|\widehat{H}(s)\right|\right)^{-1} \tag{6.71}$$

in order to express the significance of the error in relative terms.

6.5.2.4 Nonlinear Sub-Class

The nonlinear sub-class of the A-Class SQMs are those performance measures that explicitly compare the original nonlinear features of s to the estimated signals $\widehat{s}(\theta)$. Five elements from the MFA features of Ch. 4 are considered here: the embedding lag Δn, the embedding

dimension N_e, the MFA scaling interval, and the two MFA spectra. Since the embedding parameters are integer-valued, their "convergence" is considered differently from the previous features. No SQM-type scalar metric is formalized, but clearly *consistency* in the values is desired. The MFA scaling interval is also a difficult feature for which to specify a meaningful scalar SQM-type error. Considering this difficulty and the essentially supportive role the scaling interval has with the MFA spectra, no formal SQM will be applied to the MFA scaling interval either, but a consideration of its consistency will be made.

The comparison of MFA spectra was addressed in Sec. 4.8 by considering L^p-style norms on the parametric forms of the spectra over q (*c.f.*, (4.144) and (4.147)). Two such norms are implemented: specifically, a Euclidean-style quadratic norm and a Chebyshev-style (maximum) norm (*c.f.*, $\overline{\Lambda_2}[\cdot]$ and $\overline{\Lambda_\infty}[\cdot]$ from Ch. 4). These are referred to collectively as the "MFA error-norms", and individually as the MSE error-norm and the CHE error-norm respectively.

Since the estimation of the MFA fractal dimensions include a quantitative measure of uncertainty calculated from (6.55) (*e.g.*, $\sigma_{\widehat{D_q}}$, $\sigma_{\widehat{\alpha}}$, and $\sigma_{\widehat{D_0}}$), weighted norms are used to incorporate this uncertainty in the measurement of MFA matching. Thus the four MFA error-norms

are defined practically as

$$\aleph^{\text{MFA-MSE-RS}}(\theta) \stackrel{\text{def}}{=} \frac{1}{N_q}\sum_{q}\frac{\left|\widehat{D_q}(\widehat{s}) - \widehat{D_q}(s)\right|^2}{\sigma^2_{\widehat{D_q}}(\widehat{s}) + \sigma^2_{\widehat{D_q}}(s)} \tag{6.72}$$

$$\aleph^{\text{MFA-CHE-RS}}(\theta) \stackrel{\text{def}}{=} \max_{q}\frac{\left|\widehat{D_q}(\widehat{s}) - \widehat{D_q}(s)\right|}{\left(\sigma^2_{\widehat{D_q}}(\widehat{s}) + \sigma^2_{\widehat{D_q}}(s)\right)^{1/2}} \tag{6.73}$$

$$\aleph^{\text{MFA-MSE-MS}}(\theta) \stackrel{\text{def}}{=} \frac{1}{2}\left(\frac{1}{N_q}\sum_{q}\frac{\left|\widehat{\alpha}(q,\widehat{s}) - \widehat{\alpha}(q,s)\right|^2}{\sigma^2_{\widehat{\alpha}}(q,\widehat{s}) + \sigma^2_{\widehat{\alpha}}(q,s)}\right) + \frac{1}{2}\left(\frac{1}{N_q}\sum_{q}\frac{\left|\widehat{D_0}(q,\widehat{s}) - \widehat{D_0}(q,s)\right|^2}{\sigma^2_{\widehat{D_0}}(q,\widehat{s}) + \sigma^2_{\widehat{D_0}}(q,s)}\right) \tag{6.74}$$

$$\aleph^{\text{MFA-CHE-MS}}(\theta) \stackrel{\text{def}}{=} \frac{1}{2}\left(\max_{q}\frac{\left|\widehat{\alpha}(q,\widehat{s}) - \widehat{\alpha}(q,s)\right|^2}{\sigma^2_{\widehat{\alpha}}(q,\widehat{s}) + \sigma^2_{\widehat{\alpha}}(q,s)}\right)^{\frac{1}{2}} + \frac{1}{2}\left(\max_{q}\frac{\left|\widehat{D_0}(q,\widehat{s}) - \widehat{D_0}(q,s)\right|^2}{\sigma^2_{\widehat{D_0}}(q,\widehat{s}) + \sigma^2_{\widehat{D_0}}(q,s)}\right)^{\frac{1}{2}} \tag{6.75}$$

The implementation of these MFA feature-based norms is provided by a `Matlab` function `MFAdiff.m`.

6.6 Signal Synthesis

In order to satisfy the practical analysis of ICA in fetal ECG separation, the development of significantly novel ECG synthesis methods is required. These synthesized signals $s(n)$ should represent the relevant features discussed in Sec. 2.5. As described in Sec. 2.6.2.2, this has been accomplished by the generalization of the McSharry *et al.* ECGsyn model [140]. The morphological properties of the ECG are modelled dynamically by bell-shaped features off a baseline. The multiple amplitudes and widths of these features correspond to the P-, Q-, R-, S-, and T-waves. Keeping this model of the ECG morphology essentially intact (with only minor modifications, *c.f.*, Sec. A.3.2), the dynamical generation approach has been modified

to fit the beat annotations of an ECG template in the form of a sequence of RR-, PR-, and RT-intervals. Thus the successful generation of a noise-free ECG is characterized by

(a) a beat annotation template of finitely many event-intervals (suitable for maternal or fetal synthesis);

(b) a sampling frequency; and

(c) morphological modelling parameters (representing wave amplitude, width, and phase position).

The details of this work's contribution to the ECG synthesis model, particularly the interesting developments that identify a proper mapping from discrete event-intervals to a continuous heart rate, is included in App. A. Moreover, the morphological ECG parameters which hereafter are simply used with default values, are described there. Here, the context of the ECGfm algorithm is assumed, and the discussion will consist of the first two elements mentioned above, as follows.

6.6.1 fECG Event-Intervals

Event-interval data for maternal ECG signals is freely available online from the Physionet Physiobank Archive [1] but must undergo a selection process before being used for synthesis as shown in Fig. 6.15. The data considered in this study was drawn from the QT database, and therefore comes in annotated form. A subset of the QT database is, itself, drawn from the Physiobank Normal Sinus Rhythm database and does not exhibit any significant pathological rhythm. From this subset of files, records by females at childbearing ages were considered for analysis. From these recordings, the beat annotations were analyzed by the `structanno.m` function for long contiguous segments of clean normal beats (*i.e.*, annotation 'N'). The longest

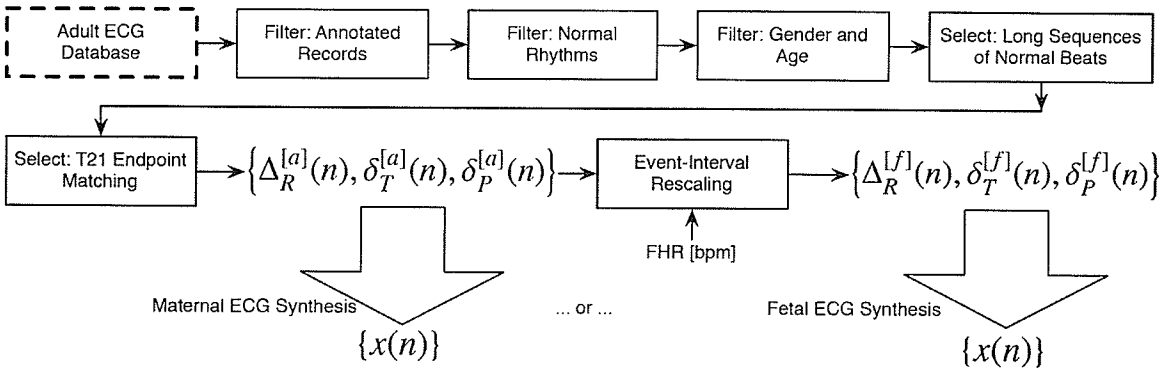


Fig. 6.15 Event-Interval Selection Process: The process for creating synthetic ECG begins with extracting event-intervals from a subset of the Physiobank QT database.

of these segments were then analyzed for suitable reduction of periodic artifacts according to the metrics of Ch. 5 implemented by the `SurrCycleTest.m` function. These refined sequences of event-intervals were then extracted into a Matlab structure suitable for reconstruction by the ECGfm algorithm. From these event-interval structures, then, maternal ECG time series were constructed with the normal parameters at multiple sampling frequencies.

This (already lengthy) process is not, however, directly capable of providing fetal event-intervals for fECG synthesis. Since fetal ECG and its properties are effectively absent in the literature, this work has developed an *ad hoc* strategy which is intended only to be sufficient to provide a simulated ECG dynamics that is statistically independent (in time-integrated distribution) from a (more rigorous) maternal surrogate ECG. Specifically, the fetal ECG time series should maintain ECG morphology, RR-variability, and representative frequencies and nonlinearities of a true dynamical ECG. To this end, this work made due with what was at hand and mapped the event-intervals from adult ECG recordings into a suitable fetal ECG template (*c.f.*, Fig. 6.16). Reference [196, Nakao *et al.* (2007)] provides a plot of an averaged fetal ECG beat extracted from a scalp electrode which can be taken as a target template. This target would indicate that at a heart rate of about 130 bpm (TT-interval of 450 ms), the

PR-interval is approximately 90 ms, and RT-interval is approximately 140 ms, while the QRS width is approximately 50 ms. Furthermore, though the different data samples are imprecise, the results of Nakao *et al.* suggest that the fetal PR-interval increases with gestational age, and consequently is under 100 ms for much of fetal observation. In order to maintain HRV in the fetal ECG, however, a mapping from adult intervals to levels matching the target template is required. Implemented in the `FetalIntConv.m` function, this considers an adult ECG template consisting of the event intervals $\{\Delta_R{}^{[m]}(n), \delta_T^{[m]}(n), \delta_P^{[m]}(n)\}$ for all beats n and maps them into a set of intervals $\{\Delta_R{}^{[f]}(n), \delta_T^{[f]}(n), \delta_P^{[f]}(n)\}$ consistent with a fetal ECG at a mean heart rate $\rho_\heartsuit^{[f]}$ in bpm. Multiple QT-correction formulae were evaluated in [184, (1990)], and some of these transformations were considered as prototypes for the conversion of adult intervals to fetal intervals. With trial and error, it was eventually determined that the mapping

$$\Delta_R{}^{[f]} = \Delta_R{}^{[m]} \frac{\rho_\heartsuit^{[m]}}{\rho_\heartsuit^{[f]}} \tag{6.76}$$

$$\delta_i^{[f]} = \delta_i^{[m]} \frac{\rho_\heartsuit^{[m]}}{\rho_\heartsuit^{[f]}} \frac{1}{\beta_i} \tag{6.77}$$

where the event-intervals δ_i for $i \in \{P, T\}$ are defined as in (A.21), and

$$\beta_i \stackrel{\text{def}}{=} \begin{cases} \beta_i \equiv \beta_T = 0.92 & i = T \\ \beta_i \equiv \beta_P = 0.80 & i = P \end{cases} \tag{6.78}$$

and

$$\rho_\heartsuit^{[m]} = 60 / \left\langle \Delta_R{}^{[m]}(n) \right\rangle \tag{6.79}$$

would be sufficient.[6] The selection parameters for the fetal event-interval templates is shown in Table 6.3.

[6] Discussion with medical doctors at Winnipeg's Women's Hospital Fetal Assessment Unit neither confirmed nor denied this claim, since fetal ECG analysis is not clinically implemented (*c.f.*, Ch. 2).

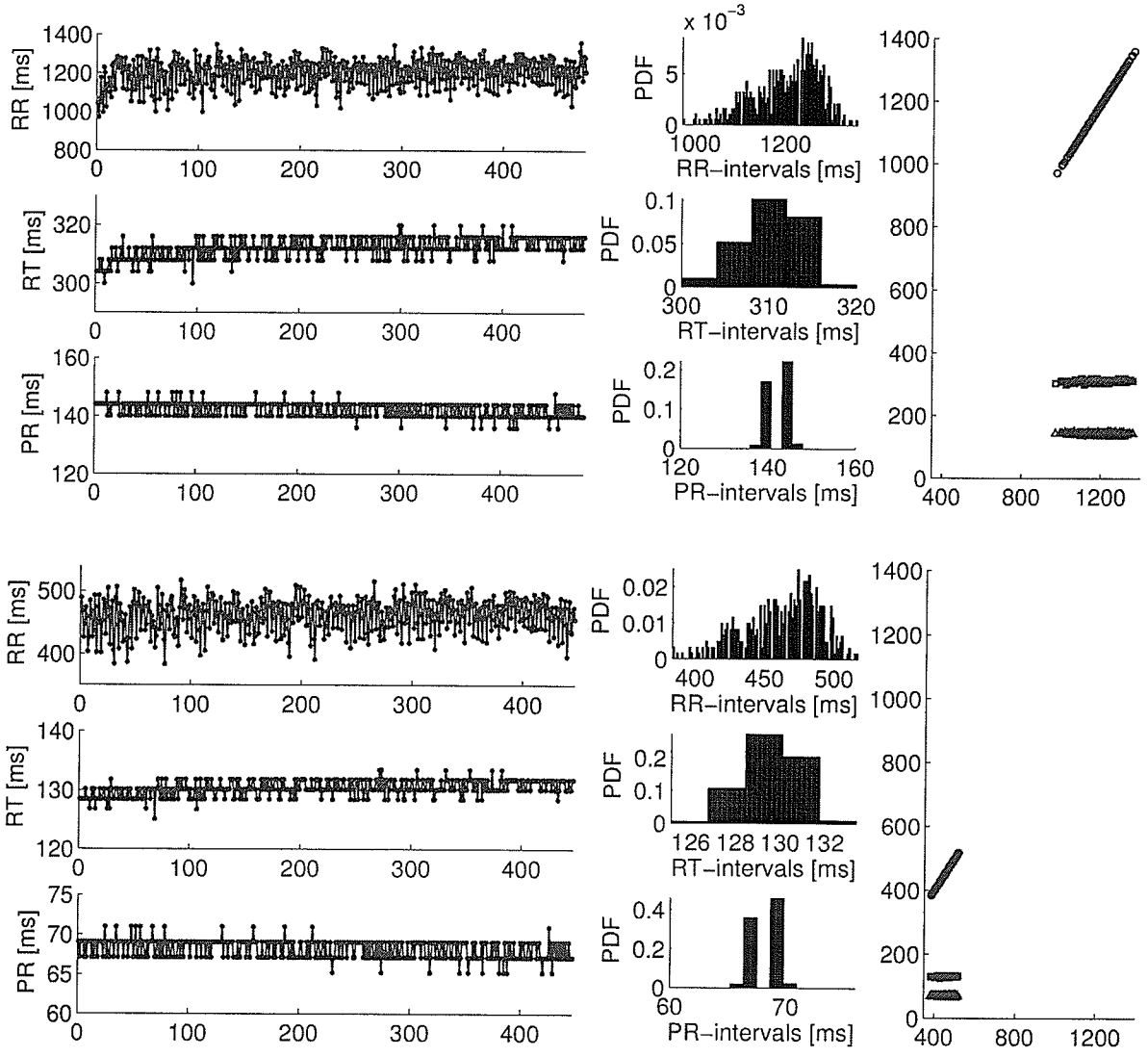


Fig. 6.16 FECG Event-Interval Methodology: Template of adult ECG event-intervals (top) from a selection of QT database record 16795.pu0 is mapped by a linear formulae to some representative fetal ECG event-intervals at 130 bpm (below).

Table 6.3 Origin of Representative Maternal and Fetal Event-Intervals

Output Template	Source File	Beat Sequence ID	Index Subset‡	Fetal HRV [bpm]
		Maternal		
`MWaves1.mat`	16272.pu0	16272_N140	(1:140)	—
`MWaves2.mat`	16795.pu0	16795ns_N250	(2:94)	—
`MWaves3.mat`	16795.pu0	16795ns_N250	(118:205)	—
		Fetal — Long		
`FWaves1_L.mat`	16795.pu0	16795s_N400	(28:474)	130
`FWaves2_L.mat`	16273.pu0	16273_N450	(5:450)	130
		Fetal — Short		
`FWaves1.mat`	16795.pu0	16795s_N400	(28:189)	130
`FWaves2.mat`	16795.pu0	16795s_N400	(231:373)	130
`FWaves3.mat`	16795.pu0	16795s_N400	(349:474)	130
`FWaves4.mat`	16273.pu0	16273_N450	(44:187)	130
`FWaves5.mat`	16273.pu0	16273_N450	(203:338)	130

‡ For the avoidance of periodic artifacts under T21 surrogates.

Remark 6.4. Since the fetal heart rate is significantly higher than an adult, the fetal ECG has a shorter beat length and, consequently, long annotation sequences are required to fill any given time window. As an example, the graphical output of the `ViewECGIntervals.m` function applied to the original adult template selected from the `16795.pu0` dataset and the finalized fetal representation at 130 bpm is shown in Fig. 6.16. Note that the fetal signal of 447 beats only measures to 206 seconds of ECG signal.

6.6.2 Sampling Frequency

With the event-interval templates finalized as described, the last requirement in this work for signal synthesis is the sampling frequency for the ECGfm algorithm. Since the objective here is to enable a bridge between "theory" and "practice", the sampling frequency should be chosen in accordance with practical guidelines. Specifically, what sampling frequency would be used in a device for fetal ECG separation? From the discussion of ECG sampling in Ch. 2, it was identified that a minimum of 500 sps sampling frequency is used in pediatric situations.

Consequently, the minimum sampling frequency to be considered here is 500 sps. Is there benefit, however, to using a higher sampling frequency? While that may be the case in empirical systems (*c.f.*, [148, Mochimaru *et al.* (2004)] and [90, Huang (2004)]), it is unlikely to matter in this work since the source of the noise-free surrogate ECG is a Runge-Kutta solution to a dynamical system (*i.e.*, the `Matlab ode45.m` algorithm). The Runge-Kutta solver is a variable step-size algorithm, (controlled by error tolerances), and not a fixed step algorithm (controlled by integration step size) and therefore it is not clear that a higher sampling frequency provides more information. In this context, a higher sampling frequency actually reflects only a more frequent interpolation, and not a finer (or better) calculation. The higher sampling frequency, of course, extends the Nyquist interval, and broadens the dynamic range of the spectrum. Spectral features, however, are linear and are not the target for the SQM analysis presented here. The most obvious outcome of a higher sampling frequency on an MFA, is the increase in the number of samples and the increased complication of the $O(N^2)$ MFA calculation it instigates. To verify that a higher sampling frequency is unnecessary for this work, the scaling intervals of the correlation sum implemented by the fast TISEAN randomization algorithm described in Sec. 6.4.3 were examined for any different behaviour as the sampling frequency was increased. The correlation sum scaling is shown in Fig. 6.17 for the same representative fetal event-interval template but at a sampling frequency of 500 sps (top) and 1000 sps (bottom). Little difference is observed between them, and a further analysis of some surrogates (dashed) did not demonstrate any other changes either. From this it is concluded that a sampling frequency of 500 sps is sufficient for the generation of surrogate ECG time series from the ECGfm algorithm.

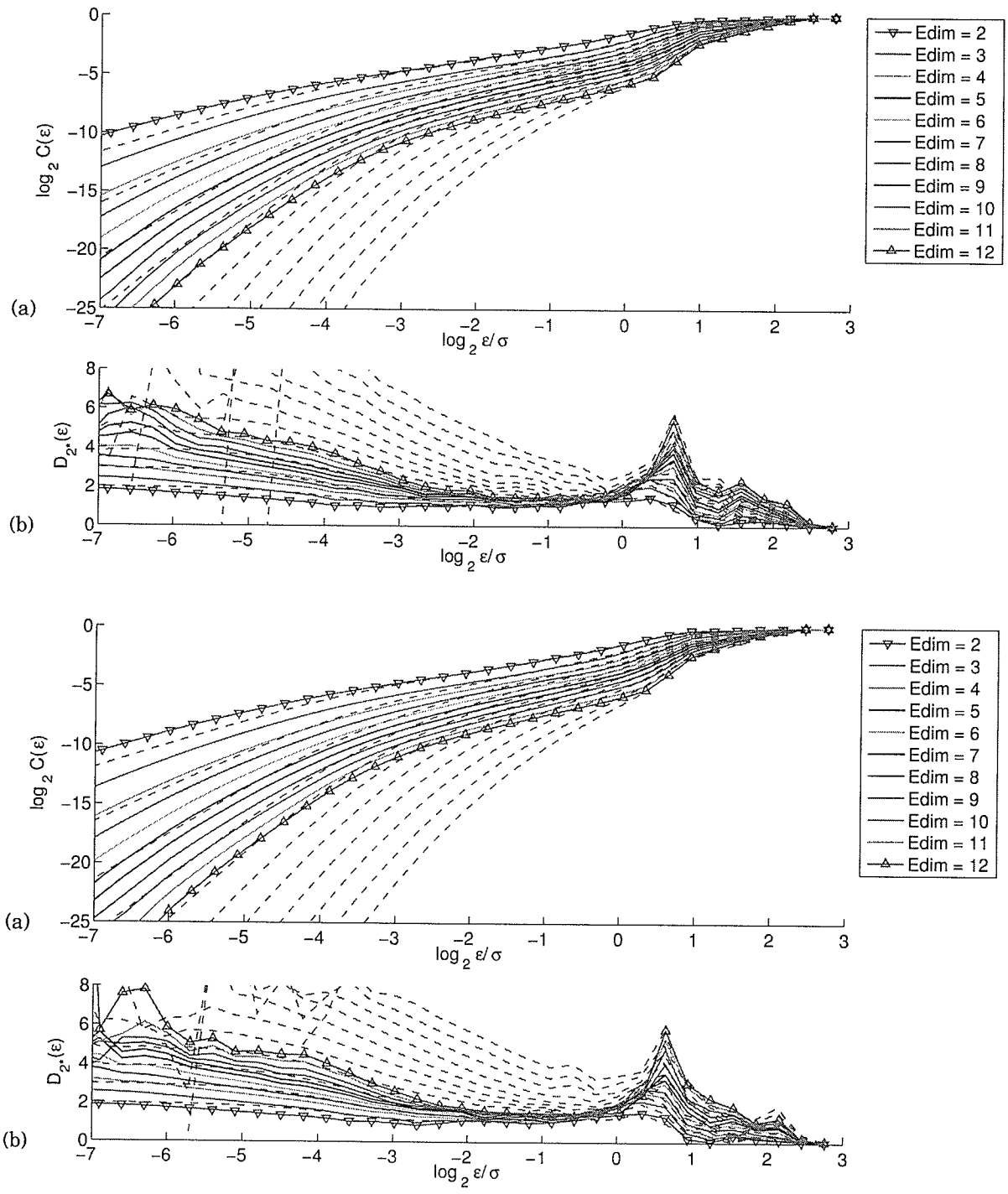


Fig. 6.17 Effect of sampling frequency on Z_C scaling interval: correlation sum scaling of fetal ECG attractor (event-intervals `FWaves1.mat`) constructed from time series sampled at 500 sps (top) and 1 ksps (bottom). Scaling of PPS surrogates also shown (dashed).

6.6.3 Record Length

Originally, the intent of this work was to use time series lengths that would be as long as feasibly possible and still within clinical bounds for fetal assessment (*e.g.*, 10 min). Due to the restricted number of beats available in the representative fetal event-interval templates, however, only signal lengths of approximately 3 mins were achievable. Even so, it was found that these recording lengths at 500 sps, when compounded with the multiple levels of ICA mixing (each yielding a potential fECG) and the necessity of multiple surrogate data signals of equal length, exceeded the memory capacity of the `Matlab` environment used to run the body of the experiments (*i.e.*, single core Mac G5 at 1.6 GHz with 3GB ram). To eliminate this complexity, the source signals were considered at a reduced length of about 1 min.

What are the effects of using a shorter time series? At the sampling rate considered, it is unlikely that ICA is greatly affected by the reduced number of samples. As shown in Table 6.4, the kurtosis of the two time series are equivalent. Furthermore, many of the other preliminary characteristics for an MFA are equal as well. It is of concern, however, that the number of beats (or cycles around the ECG attractor) has dropped from 400 to 130. The effect on the MFA scaling entropies is shown in Fig. 6.18. From this figure, it can be observed that the fewer samples has emphasized the small scale saturation in the $q < 0$ entropies, while the large-q entropies appear largely unchanged. This is consistent with the interpretation of the fECG attractor given by Fig. 6.9 and the properties of q. By reducing the amount of cycles around the attractor, the low-probability regions of the attractor become more coarsely represented by the limited sampling, and, hence, the small-q statistics become more statistically sensitive. In practical terms, the two scaling regions have coincident lower bounds, even when analyzed independently (since the main determining factor here is the large-q turning point of Υ_q, ② in Fig. 6.11 and Table 6.1).

Table 6.4 Significant Features of Synthetic ECG Time Series by Length

Feature	**Long**	**Short**
Num. Samples [K=1024]	100.4	30.7
Total Time [s]	205	63
Number of Beats	446	134
Kurtosis	5.2	5.3
Est. Δn	17	17
D_2 Sc. Reg.	$2^{[-3,\ 0.25]}$	$2^{[-3,\ 0.25]}$
D_2 Calc. Time [s]	128.3	17.6
Est. N_e	10	10
MFA Entropies	see text	
MFA Scaling Region	$2^{[-2.4,\ 0.5]}$	$2^{[-2.4,\ 0.5]}$
MFA Calc. Time [min]	452	51.4

Other non-essential characteristics change, (*e.g.*, the H_q crossover point ① occurs at larger ϵ), but the properties of the shorter time series' scaling interval are essentially the same. The *quality* of the linear fit within the scaling interval, however, is markedly different for small q. The γ_q saturation point for $q = -2$ happens about 1 octave higher than for the longer time series, increasing the corruption of the scaling within the marked interval. Consequently the slope estimate is subject to a larger bias and wider uncertainty (error bar) for the shorter time series. Since, however, (i) the properties of large q are well preserved, and (ii) the error bar serves to identify the poor fit for small q relative to large q, a reasonable interpretation of the MFA can still be estimated from the shorter time series. Simply, the MFA-based SQM should apply more weight to the positive q elements of the MFA as a result of this limitation in the experimental platform.

This concludes all discussion on the synthesis of both the maternal and fetal noise-free ECG. A representative segment of the final time series is shown in Fig. 6.19.

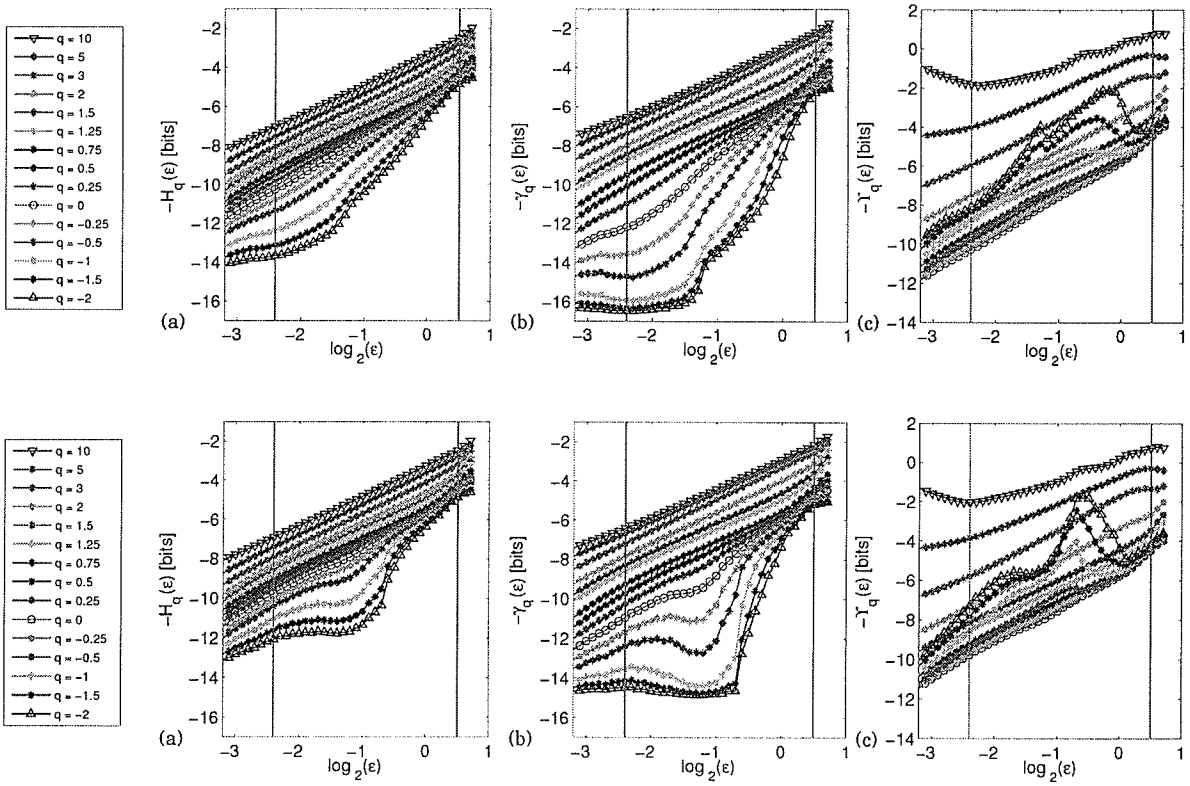


Fig. 6.18 Effects to MFA entropy scaling by signal length: The MFA scaling of a synthetic fetal ECG attractor reconstructed from 3 min (top) and 1 min (bottom) of signal at 500 sps (from the event-interval template `FWaves1.mat`). Scaling interval selection is the same, but quality of the fit for small q degrades for the shorter signal.

6.7 Overall Procedure of Experiments

Now that the design of the significant elements in the experimental procedure have been established, the final structure that binds them together will be presented. In consideration of the different levels of user intervention required for the calculation of the different features, the experimental procedure is broken up into a sequence of stages. Consequently, the general design of the thesis presented in Fig. 1.5 is implemented in a hierarchy as shown in Fig. 6.20. In particular, this figure demonstrates the flow of some of the main features used for SQM analysis, and roughly identifies the offline components in grey.

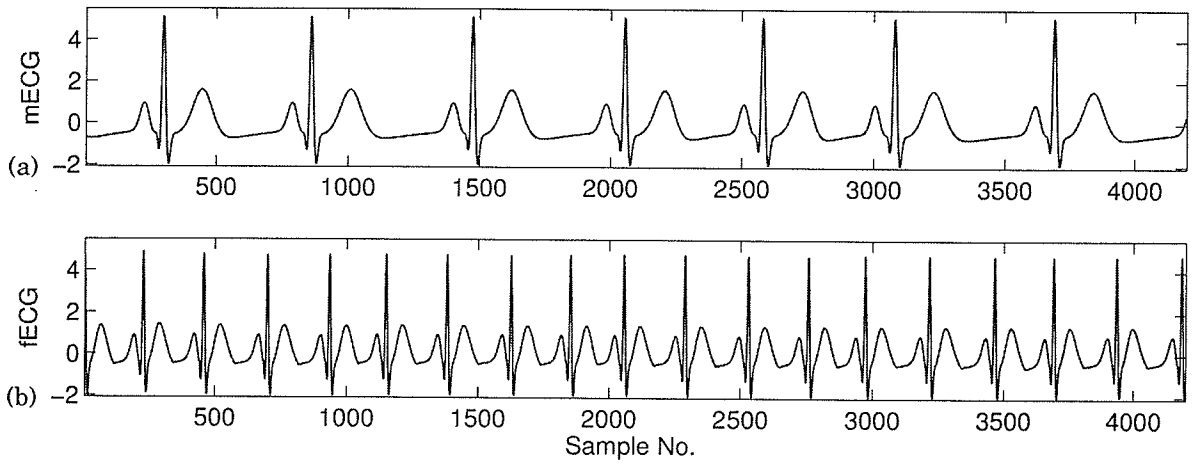


Fig. 6.19 Source signals for ICA Analysis: Synthetic ECG representing (a) maternal ECG (derived from an early segment of event-intervals from QT record `16795.pu0`) and (b) fetal ECG (derived from the event-interval template `FWaves1.mat`). The full 63 s of time series has $31K$ points sampled at 500 sps.

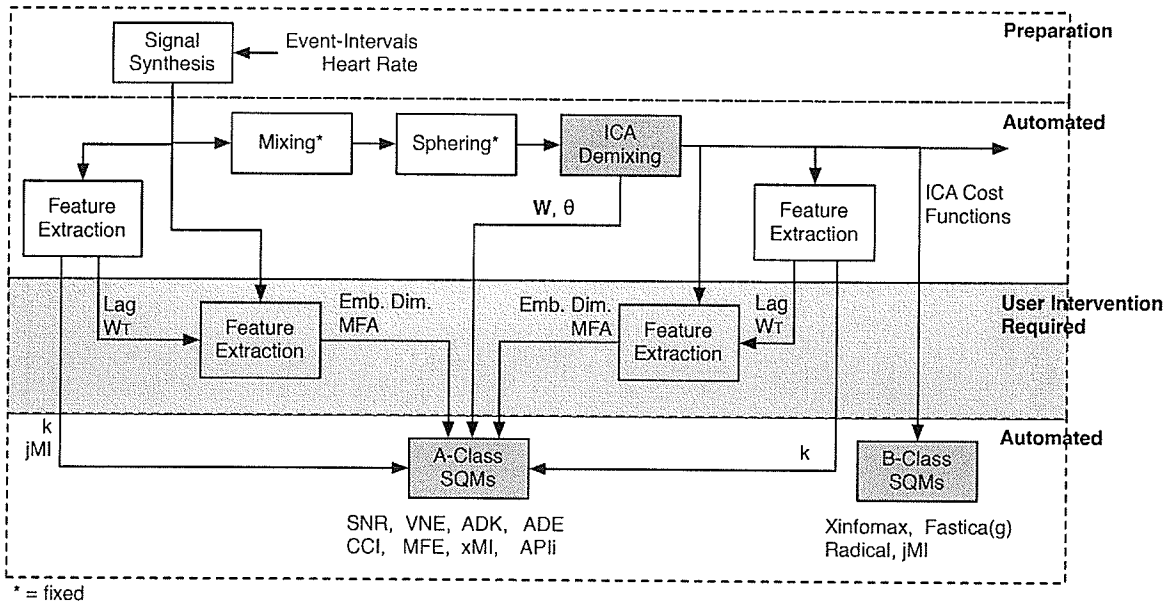


Fig. 6.20 The experimental method to assess ICA performance by SQM convergence.

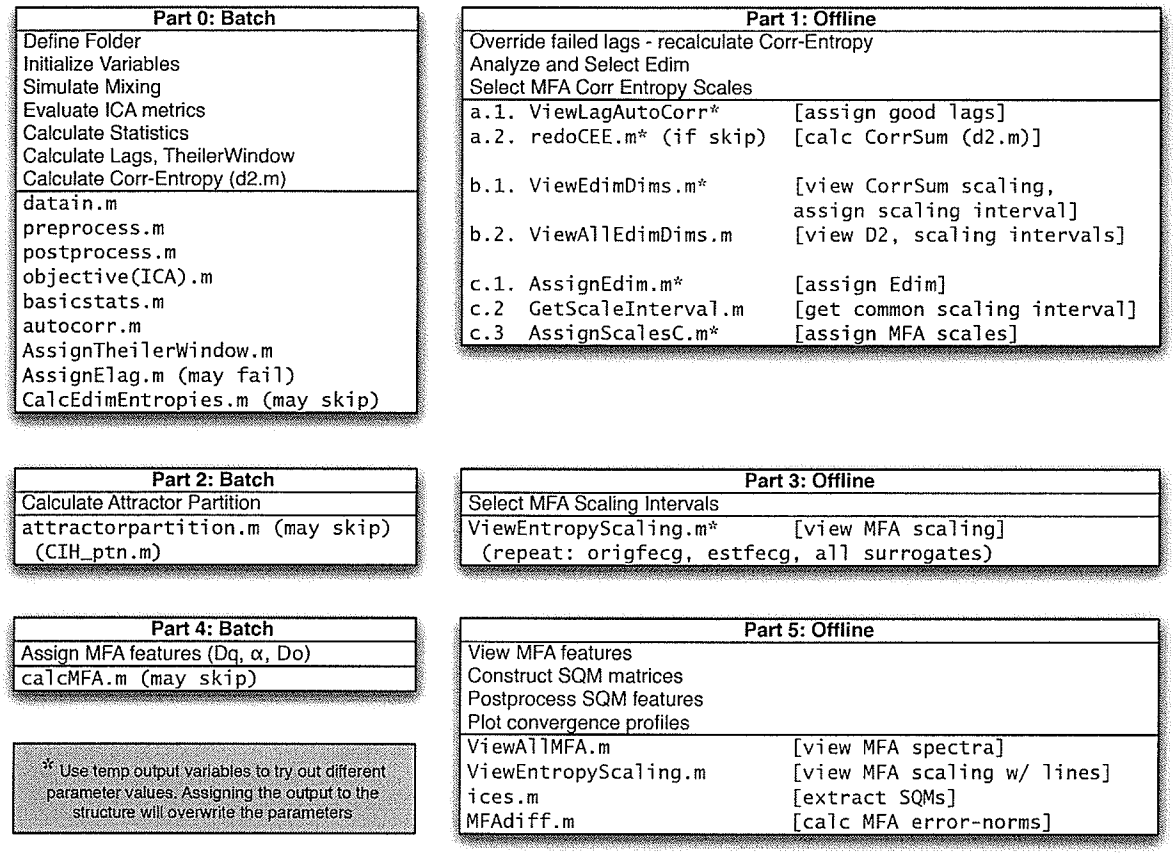


Fig. 6.21 The experimental procedure in its sequential form.

6.7.1 Organization of Experiments and Data

Implementation of the experimental procedure is organized through a `Matlab` function that invokes the tasks of each automated (online) stage of the experiment. The basic operation of this `thesisbody.m` function is summarized in Fig. 6.21 along with the offline components roughly identified in grey in Fig. 6.20. Offline components are required because of the user intervention necessary for MFA analysis. Data is passed between the stages of the experiment by saving the workspace data to a `Matlab .mat` file. Within the workspace, data is arranged in a detailed hierarchy of structures initialized by `thesisbody.m` in the 0th stage. Raw features

of the source fetal ECG and its demixing estimates are stored in the structures `origfecg` and `estfecg` respectively. The latter is a vector `struct` of the same length as the θ vector assigned at the primary input. The other significant structures generated by `thesisbody.m` include a vector `struct` of SQMs, and surrogates for the source fECG (`origsurr`) and demixing estimates (`estsurr`).

Experiment workspaces are saved in a directory created at the initialization of the 0th stage. At the beginning of each stage, the workspace of the previous stage is loaded and saved into a new file indexed by the stage number. A diary of the `Matlab` console is preserved during every stage of the procedure and saved in the experiment directory.

The main procedure `thesisbody.m` and all its invoked functions and scripts were first designed on a student version of `Matlab` (v.7r.14), then distributed for computing on a single core Mac G5 at 1.6 GHz with 3GB ram (running `Matlab` r.2007(b)).

6.7.2 Necessary Input for Stage 0

When `thesisbody.m` is invoked at stage 0 to initialize a new experiment, it is given an initialization file specifying the specific parameters for the experiment. This source file includes,

(a) A pointer to a source file containing a saved workspace of the noise-free fetal and maternal ECG;

(b) A vector of q values for MFA, (this is duplicated in each feature struct and can be modified later by user intervention);

(c) The value of the W_T stretchfactor $\widetilde{\lambda}$ (set at 20 for the experiments, here);

(d) A vector of θ values (in degrees) that parameterize the Stiefel manifold (roughly in the $[0, 45]$ interval);

(e) Maximum values to consider for the embedding lag Δn (set at 40) and the embedding dimension N_e (set at 12 for the experiments here);

(f) The number of specimens of each surrogate data type to generate (set at 4 for the experiments here).

Remark 6.5 (Selection of default parameters). Most of the parameters here were selected after some preliminary experimentation. With many of these variables, it is important to provide sufficient coverage with as few elements as possible in order to limit experimental burden. For this reason, the values of θ are not chosen uniformly, but concentrate in the region of convergence to the actual solution. To accentuate this in particular, some small values of $\theta < 0$ were also included in some experiments (*e.g.*, $[-5°,\ 30°]$). Similarly, the vector of q values is non-uniform, since large positive values become redundant and large negative values become dominated by saturation artifacts.

6.7.3 Source Input

As described previously, the synthesis of the source time series is generated offline and stored as a `Matlab .mat` workspace prior to the experiment initialization. A list of the available datasets created by the `makesources.m` script and their event-interval origins is shown in Table 6.3. In each file, the time series are in the form of a $2 \times N$ matrix,

$$s(n) = \begin{bmatrix} s_1(n) \\ s_2(n) \end{bmatrix} \quad \text{for } n = 1, \ldots, N \tag{6.80}$$

with the maternal ECG in the row 1 and the fetal ECG in row 2. As previously discussed in Prop. 3.26, it is desired that the source signals are standardized to zero-mean and unit variance (specifically, the fetal ECG time series). Since these properties are not constrained by the

ECGfm synthesis algorithm, this is enforced by a simple update with the `standardized.m` function in the `thesibody.m` procedure after being loaded into the experiment workspace.

6.7.4 Signal Mixing

Once loaded and standardized into the experiment workspace, the `thesisbody.m` function multiplies the sources with a mixing matrix. Specifically, the fixed matrix

$$\mathbf{M} = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix} \tag{6.81}$$

is used to cause an appropriate disparity in power in the "observations"

$$\boldsymbol{x}(n) = \mathbf{M}\boldsymbol{s}(n) \quad \text{for } n = 1, \ldots, N \tag{6.82}$$

Here the mECG is given an amplitude 5 times larger than the fECG, consistent with experimental observation (but not critically important because of the sphering process to follow). Note that since there are no nonzero cross terms in the mixing, no active decorrelation is required from a sphering stage. Furthermore, since sphering normalizes all channels to unit variance, it follows that effectively $\mathbf{W}_0\mathbf{M} = \mathbf{I}$ to numerical accuracy. This is designed to make the true post-sphering demixing matrix $\mathbf{W}$ known to be the identity. If $\mathbf{M}$ did have cross terms, then the sphering process can introduce a factor that necessitates a non-zero rotation in $\mathbf{W}$. From the consideration of an ICA profile, this is the most convenient form of mixing when using non-uniformly distributed sampling of the Stiefel manifold. Notwithstanding this design, it is still considered important to practically mix and sphere the data to simulate the same level of numerical inaccuracy that appears in real systems.

6.7.5 Preprocessing

As it would in a practical ICA system, the data $\boldsymbol{x}$ is then preprocessed for ICA. The preprocessing action is threefold: (i) standardization to zero mean, (ii) standardization to unit

variance, and (iii) sphering to identity covariance. Even though the sources $\boldsymbol{s}$ were standardized to fulfill (i) and (ii), the `preprocess.m` algorithm includes these actions for portability.

The sphering process is performed by calculating the covariance matrix as

$$\boldsymbol{\Sigma}_x = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}(n)\boldsymbol{x}^{\mathrm{T}}(n) \tag{6.83}$$

using division by N since this is not an estimate of a Gaussian covariance. This is then factored into an eigenvalue decomposition of a unitary matrix $\mathbf{W}_e$ and a positive diagonal matrix $\mathbf{D}$,

$$(\mathbf{W}_e, \mathbf{D}) = \mathrm{eig}(\boldsymbol{\Sigma}_x) \tag{6.84}$$

using the `Matlab` implementation of the LAPACK `DSYEV` algorithm in the function `eig.m` [220]. Since $\boldsymbol{\Sigma}_x$ is symmetric and positive definite, this is very efficient. Consequently the 2×2 covariance matrix is of the form

$$\boldsymbol{\Sigma}_x = \mathbf{W}_e \mathbf{D} \mathbf{W}_e^{-1} \tag{6.85}$$

The sphering matrix is then calculated as

$$\mathbf{W}_0 = (\sqrt{\mathbf{D}})^{-1} \mathbf{W}_e^{-1} \tag{6.86}$$

where the square root is performed element-wise in the diagonal matrix $\mathbf{D}$ and the matrix inversions are performed by the `Matlab` function `inv.m`. The data is then sphered by the transformation

$$\boldsymbol{x}(n) \leftarrow \mathbf{W}_0 \boldsymbol{x}(n) \tag{6.87}$$

and updated to the same variable. As mentioned previously, the choice of mixing matrix makes the sphered system effectively independent so the practical demixing matrix would be the identity.

6.7.6 Parameterized ICA (Stage 0 Loop Begins)

As described in Sec. 6.2.1, the "demixing" portion of the experiment uses a parameterization of the Stiefel manifold. As such, an algorithmic loop is created here on the values of θ specified in the initialization file. The first action in the loop is to construct the demixing matrix as

$$\mathbf{W} \leftarrow \mathbf{G}_2(\theta) = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \tag{6.88}$$

and produce the "demixed estimate"

$$z(n) \leftarrow \mathbf{W}\boldsymbol{x}(n) \tag{6.89}$$

By design of the mixing and preprocessing system, $\mathbf{W}|_{\theta=0}$ should be close to optimal. The next action in the loop is to postprocess the data as discussed next.

6.7.7 Postprocessing (Within Stage 0 Loop)

The goal of the postprocessing routine is to standardize the demixing estimates according to the items in Prop. 3.26. This includes the standardization of the channel mean and variance as well as standardization of the ICA indeterminacies by setting the orientation that correctly identifies the "best estimate of the fetal ECG". This proceeds by

(a) removing the mean (if any);

(b) removing the permutation and sign invariance by the CCI;

(c) removing the invariance to scaling by standardizing the variance.

All three actions are packaged into the function `postprocess.m` which is called in the form

$$(z, \mathbf{W}) \leftarrow \text{postprocess}(s, z, \mathbf{W}) \tag{6.90}$$

The first action is accomplished by the simple update

$$z_0 \leftarrow \frac{1}{N}\sum_{j=1}^{N} z(n) \tag{6.91}$$

$$z(n) \leftarrow z(n) - z_0 \tag{6.92}$$

The most difficult procedure is the second, which matches the fECG estimate with the original in terms of permutation and sign. As described in Prop. 3.26, this is achieved through cross-correlation analysis. The $N_c \times N_c$ correlation matrix[7]

$$\mathbf{C} = \sum_{n=1}^{N} z(n)\boldsymbol{s}^{\mathrm{T}}(n) = [C_{ij}] \tag{6.93}$$

is then analyzed to identify the maximal absolute elements per column

$$i'(j) \stackrel{\text{def}}{=} \arg\max_i \left|C_{ij}\right| = \arg\max \left|\mathbf{C}\mathbf{e}_j\right| \tag{6.94}$$

This can be achieved in vector form using the `Matlab` call

$$(\boldsymbol{m}, \boldsymbol{i}') = \max \mathrm{abs}(\mathbf{C}) \tag{6.95}$$

although the values of $\boldsymbol{m}$ are not used. Then a matrix is defined

$$\mathbf{U} \stackrel{\text{def}}{=} [U_{ij}] \tag{6.96}$$

where

$$U_{ij} \stackrel{\text{def}}{=} \begin{cases} \operatorname{sign} C_{ij} & : \ i = i'(j) \\ 0 & : \ \text{otherwise} \end{cases} \tag{6.97}$$

which has a single nonzero element per column. This signifies that each source s_j in $\boldsymbol{s}$ (*i.e.*, column j of $\boldsymbol{s}^{\mathrm{T}}$) should be dominated by a single "best estimate" (*i.e.*, row of $\boldsymbol{z}$) which is

[7] Although $N_c = 2$ by design in this work, this action is derived in its generality for portability.

identified by $i'(j)$. In a proper demixing of independent sources, each source should identify a *single and unique* estimate and therefore each row of $\mathbf{U}$ should have a single non-zero entry. This can be confirmed experimentally by checking the condition

$$\sum_{j=1}^{N_c} |U_{ij}| = 1 \quad \forall i \tag{6.98}$$

which, if true, signifies that $\mathbf{U}$ is a unitary matrix (and, specifically, $|\mathbf{U}|$ is a permutation matrix). It follows that if $\mathbf{U}$ satisfies the condition of (6.98), then the update

$$z \leftarrow \mathbf{U}^{\mathrm{T}} z \tag{6.99}$$

matches each estimate z_j with its dominating source. For this new system,

$$\mathbf{C}' = \sum_{n=1}^{N} \mathbf{U}^{\mathrm{T}} z(n) s^{\mathrm{T}}(n) \tag{6.100}$$

$$= \mathbf{U}^{\mathrm{T}} \sum_{n=1}^{N} z(n) s^{\mathrm{T}}(n) \tag{6.101}$$

$$= \mathbf{U}^{\mathrm{T}} \mathbf{C} = [C'_{ij}] \tag{6.102}$$

which implies that the cross correlation is maximal and positive along the diagonal.[8] This update makes the estimate z_j the standardized estimate for s_j according to the principles of Prop. 3.26. Note that if the condition (6.98) fails, then the demixing solution is so poor that a "best estimate" cannot be determined. In that case, a warning is issued, and the update (6.99) is left trivial (*i.e.*, $\mathbf{U} = \mathbf{I}$). Finally, note that in order to maintain the relationship

$$z = \mathbf{W}x \tag{6.105}$$

[8] This can be observed by the relationship

$$C'_{jj} = \mathbf{e}_j^{\mathrm{T}} \mathbf{U}^{\mathrm{T}} \mathbf{C} \mathbf{e}_j = \left(\operatorname{sign} C_{i'j}\right) \mathbf{e}_{i'}^{\mathrm{T}} \mathbf{C} \mathbf{e}_j \tag{6.103}$$

$$= \left(\operatorname{sign} C_{i'j}\right) C_{i'j} = |C_{i'j}| \tag{6.104}$$

which by construction was greater than all other $|C_{ij}|$.

the update (6.99) must be supplemented with the update

$$\mathbf{W} \leftarrow \mathbf{U}^{\mathrm{T}}\mathbf{W} \tag{6.106}$$

The final action of the postprocessing module is to standardize the variance of the source estimates according to Prop. 3.26. (Even though this is unnecessary by account of all the previous processing, this is included for portability and generality.) This normalization of amplitude is done by defining the matrix

$$\mathbf{D} \stackrel{\text{def}}{=} [D_{ij}] \tag{6.107}$$

where

$$D_{ij} \stackrel{\text{def}}{=} \begin{cases} 1/\sqrt{\widehat{\mu_2}(z_j)} & : \ i = j \\ 0 & : \ \text{otherwise} \end{cases} \tag{6.108}$$

and performing the updates

$$z \leftarrow \mathbf{D}z \tag{6.109}$$

$$\mathbf{W} \leftarrow \mathbf{D}\mathbf{W} \tag{6.110}$$

These three postprocessing actions ensure that the estimates z_j, and specifically the fECG estimate z_2, satisfy the standardization specified in Prop. 3.26. Verification of proper postprocessing during experimentation was ensured by analysis of the time series z_2 and scatterplots of z using the `presentTSpair.m` function.

6.7.8 Automated Feature Extraction and SQMs (End of Stage 0 Loop)

By this point in the stage 0 loop, both the original fECG source time series $s_2(n)$ and its best estimate $z_2(n)$ under the demixing of $\mathbf{W}(\theta)$ are well defined. Here the automated statistical features for SQM analysis are extracted. This includes

(a) the basic statistics of the demixed system z: the scalar $\widehat{\mathcal{M}}(z)$, the $\widehat{\mu_1}(z)$ and $\widehat{k}(z)$ vectors, and the $\widehat{\Sigma}_z$ matrix;

(b) the absolute difference in kurtosis, $\Delta k = |k(z) - k(s)|$,

(c) the absolute difference in entropy, $\Delta H = |H(z_2) - H(s_2)|$,

(d) the cross-MI, $\widehat{\mathcal{M}}(z_2, s_2)$;

(e) all ICA cost functions, $\mathcal{I}$, Sec. 6.2.2;

(f) the cross-correlation, $\mathrm{CCI}(z_2, s_2)$;

(g) the API, $A_\epsilon(\mathbf{W}(\theta)\mathbf{W_0}\mathbf{M})$;

(h) the standardized signal-to-noise ratio, $\mathrm{SNR}^*(z_2, s_2)$;

(i) the matrix Frobenius-error, $\mathrm{MFE}(\mathbf{W}(\theta)\mathbf{W_0} - \mathbf{M}^{-1})$; and

(j) the vector norm-error, $\mathrm{VNE}(\mathbf{W}(\theta)\mathbf{W_0} - \mathbf{M}^{-1})$;

where each item is calculated according to their methods described in Sec. 6.5.

This nearly completes the main algorithmic loop in stage 0.

6.7.9 Nonlinear Feature Extraction (Stages 0–4)

6.7.9.1 Attractor Preliminaries (Stage 0 — Online)

Prior to the end of stage 0 of `thesisbody.m`, the automated features for nonlinear time series analysis are calculated. This includes

(a) the autocorrelation time, λ_n;

(b) the embedding lag, Δn;

(c) the Theiler window, W_T; and

(d) the correlation sum, Z_C, for embedding dimensions up to the maximal value prescribed in the initialization over an automated set of scales;

where each item is calculated according to their methods described in Sec. 6.4. This completes stage 0 of an experiment and all variables in the `Matlab` workspace are saved to the stage 0 storage file.

6.7.9.2 Attractor Characterization (Stage 1 — Offline)

As described in Fig. 6.20, and Table 6.2, and Sec. 6.4, the nonlinear features require user intervention for the selection of

(a) the correlation sum scaling interval;

(b) the embedding dimension;

(c) the scales for MFA calculation; and

(d) the scaling interval for the MFA entropies.

Stage 1 of the experimental procedure occurs offline during which the user specifies the first three items from the list above. To proceed, first the stage 0 storage file is loaded. Then, using the `ViewEdimDims.m`, `VEDtag.m`, and `ViewAllEdimDims.m` tools, the user analyses the correlation sum of the original and each estimated fECG time series for scaling intervals. Once a suitable scaling interval is identified for a time series, the user can analyze the convergence of the correlation dimension $\widehat{D_2}$ and assign an embedding dimension for that attractor reconstruction using `AssignEdim.m`.

Once each time series has been assigned a value for N_e, the user must determine a sequence of scales for MFA calculation. Typically, this is chosen uniformly in a base 2 exponent, as in

$\epsilon(i) = \epsilon_0 2^{\delta_\epsilon i}$ for $i = 1, \ldots, N_\epsilon$. The rule of thumb used for the selection of the increment δ_ϵ is that $N_\epsilon \approx 30$ over the union of all the observed $\widehat{D_2}$ scaling intervals. The choice of ϵ_0 should position the selection of scales over the full extent (and a small extra margin) of the union of all the observed $\widehat{D_2}$ scaling intervals. Note that oversampling of the scaling interval drastically increases the memory required for data results since the CP calculated in the next stage is an $N_o \times N_\epsilon$ array. With the vector of scales ϵ finally determined, it must be assigned to the original fECG and each fECG estimate using the `AssignScalesC.m` function.

This completes the actions required in stage 1 of an experiment and all (updated) variables in the `Matlab` workspace are saved to the stage 1 storage file.[9]

6.7.9.3 Correlation Partition and Surrogates (Stage 2 — Online)

Stage 2 of the experimental procedure returns to the online calculation of nonlinear features. It accomplishes the actions of

(a) calculating the correlation partition, $\mathcal{P}$, for all the reconstructed attractors of the original and estimated fetal ECG time series;

(b) generating T21 and PPS surrogates, if required; and

(c) calculating the correlation partition, $\mathcal{P}$, for all surrogates, if required.

Since these algorithms are the most time-intensive, this stage may take a substantial time for completion. The stage begins by (i) trying to load any stage 2 storage file that may have been created (*e.g.*, if the processing requires a restart), and upon failure, (ii) trying to load any stage 1 storage file (*e.g.*, for a fresh start).

[9] It is also recommended that the user save the workspace periodically throughout the stage.

At the end of (and throughout) the stage, all (updated) variables in the `Matlab` workspace are saved to the stage 2 storage file. Note that the size of the stage 2 storage file is much larger than previous stages because of (i) the large memory required for the CP, $\mathcal{P}$, and (ii) the extra time series and CPs introduced by the surrogates.

6.7.9.4 MFA Entropies and Scaling Interval (Stage 3 — Offline)

The final stage of nonlinear feature extraction requiring user intervention consists of selecting the scaling intervals for the MFA entropies. This must be processed offline for every time series; including the original fECG time series, its demixing estimates, and all surrogates. This is approached in stage 3 by loading the results from the stage 2 storage file and sequentially identifying the scaling intervals from the output of the `ViewEntropyScaling.m` function as in Fig. 6.11. Guidelines for the selection of the scaling interval has been described in Table 6.1. The value for the scaling interval is also assigned through the `ViewEntropyScaling.m` function.

Remark 6.6 (Calculation of MFA Entropies). An in depth discussion of the calculation of MFA entropies from a CP is presented in Sec. C.4 of App. C. One significant fact to be considered is that a large CP (of size $N_o \times N_\epsilon$) generates three large arrays (of size $N_q \times N_o \times N_\epsilon$) used in the simultaneous calculation of the three MFA entropies (of size $N_q \times N_\epsilon$). Consequently, this calculation can be a significant, but temporary, drain on computing memory resources.

Remark 6.7 (Evaluation of Surrogates). The purpose behind generating surrogate data is to provide a validation of the MFA scaling entropies' scaling interval. Once the scaling of the surrogates has been evaluated and compared to the actual scaling (*c.f.*, Figs. 6.11 and 6.12), further processing is not required.

Once all MFA scaling intervals have been processed, all the actions required in stage 3 of

an experiment are complete and all (updated) variables in the `Matlab` workspace are saved to the stage 3 storage file.[10]

6.7.9.5 MFA Calculation (Stage 4 — Online)

The MFA features of the original and estimated fetal ECG time series, as well as their surrogates, are finalized in stage 4 of the `thesisbody.m` procedure. Here, the MFA scaling intervals are applied to the MFA scaling entropies for linear fitting and slope extraction as described in Sec. 6.4.6.2. Specifically, the `calcMFA.m` function provides for each scaling entropy and for each value of q:

(a) a slope for the linear fit on the scaling region representing the fractal dimension of that entropy according to the box-counting formalism;

(b) an error for that slope; and

(c) an intercept for the linear fit on the scaling interval.

The intercept is only used for representing the best fit lines in the figures of scaling entropies using the `ViewEntropyScaling.m` function, as shown in Fig. 6.13. At the end of (and throughout) the stage, all (updated) variables in the `Matlab` workspace are saved to the stage 4 storage file.

6.7.10 SQM Analysis

The final analysis of experiments is performed offline by assembling the features of the individual demixing estimates into a matrix form for numerical and graphical analysis. Since these features are stored in various complicated positions within the structures that organize the experimental workspace, an extraction tool `ices.m` is used to simplify this process.

[10] It is also recommended that the user save the workspace periodically throughout the stage.

The actual feature extraction, SQM processing, and generation of SQM convergence plots is organized through a series of Matlab scripts calling the `ices.m`, `ViewLagAutoCorr.m`, `ViewAllEdimDims.m`, `ViewEntropyScaling.m`, and `ViewAllMFA.m` tools. Specifically they analyze the SQMs by sub-class:

(a) ICA cost functions and the Matrix SQM sub-class, (*c.f.*, `CF_MAT_convergence.m`);

(b) Statistical A-Class SQMs (including SOS and HOS forms: *c.f.*, `ABS_convergence.m`);

(c) Nonlinear A-Class SQMs (*c.f.*, `MFA_convergence.m`)

and plot the convergence behaviour of multiple experiments concurrently. These scripts proceed linearly by loading one experiment's workspace, processing the data fully into the convergence figures, and then repeating the procedure with the data from another experiment.

A study of tolerances to the different features extracted is also organized in the `SourcePhase.m` script.

6.8 Summary

This chapter has presented the technical and organizational aspects of the experiment design for the evaluation of the convergence of feature-based SQMs in a simulation of the fetal ECG separation problem. Specifically, the motivation, design, and implementation of convergence profiles has been finalized. The procedure for the calculation of these profiles from basic feature extraction, processing, to the SQM error-space vector has been explained with sufficient depth to facilitate the reproducibility of the work.

A discussion of the ECG synthesis methods — including the limitations and compromises in the selection of ECG event-intervals, sampling frequency, and record length — has been

provided. The nonlinear dynamical properties of the synthesized fetal ECG (including attractor reconstruction, correlation partition, and MFA features) has been demonstrated. In particular, the implementation of scalar MFA error-norms for the evaluation of MFA convergence has contributed to the ICA error space.

In total, the SQM error space is composed of 5 classes of SQMs:

(a) B-Class: ICA cost functions (FastICA, Xinfomax, RADICAL, and MILCA (also referred to as jMI));

(b) A-Class: Matrix SQMs (API, MFE, and VNE);

(c) A-Class: SOS-feature-based SQMs (CCI, SNR*);

(d) A-Class: HOS-feature-based SQMs (ADK, ADE, xMI);

(e) A-Class: Nonlinear dynamical feature-based SQMs (Δn, N_e, MFA scaling interval, and the MSE and CHE MFA error-norms),

whose relationships are analyzed in the next chapter.

Chapter VII

Results and Discussion

7.1 Overview

The experiments described in the previous chapter have been executed and analyzed fully for two source datasets. This chapter presents these results and discusses their interpretation and significance.

Before the in-depth analysis of the fetal ECG simulation convergence profiles, some natural uncertainties to the simulation are analyzed in Sec. 7.2 to provide context (*i.e.*, tolerances) for the convergence analysis of SQMs. In particular, uncertainties due to

(a) the random phase of the recorded signals; and

(b) the selection of the MFA scaling entropy scaling interval;

are considered.

Thereafter, the collection of SQMs are analyzed in four parts. The first of these, Sec. 7.3, analyzes the convergence of the fundamental elements of the ICA system: (i) the ICA cost functions, which represent the B-Class SQMs; and (ii) the matrix sub-class of A-Class SQMs. Second, the convergence of the second-order measures are analyzed in Sec. 7.4. Third, the higher-order feature-based SQMs are analyzed in Sec. 7.5. This proceeds by first considering the basic statistical SQMs — the ADK and ADE — and also the cross-MI. Last, Sec. 7.6 presents the study of the convergence of nonlinear dynamical features. This section is considered in two parts. The first, Sec. 7.6.1, studies the effect of mixing on the underlying

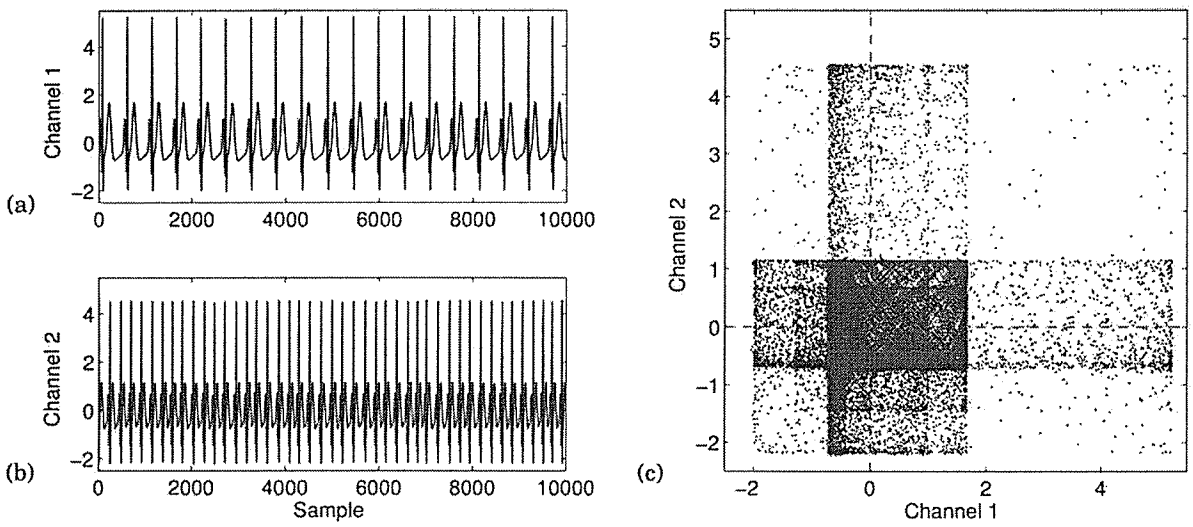


Fig. 7.1 Visualization of joint ECG sources for dataset 1: (a) maternal ECG, s_1; (b) fetal ECG, s_2; (c) scatterplot visualization of joint distribution.

embedding parameters for attractor definition. Thereafter, the convergence of the fetal ECG MFA spectra is presented in Sec. 7.6.2.

7.2 Natural Uncertainties

Prior to an analysis of the convergence of fetal ECG features, it is useful to establish benchmarks for interpretation. The following subsections answer specifically:

(a) In the fetal ECG scenario, how accurate is the assumption of independent sources? How sensitive is it to the (random) relative phase of the ECG oscillators?

(b) What precision can be assigned to the HOS estimates of the synthetic fetal ECG?

(c) What values of MFA error-norm can be considered a match? How sensitive are these values to the selection of the scaling interval?

Table 7.1 Practical Source Independence of Analyzed Datasets

Dataset	$\widehat{\mathrm{CCI}}(s_1, s_2)$	$\widehat{\mathcal{M}}(s_1, s_2)$ [nats]
Dataset 1 (`MWaves1.mat`, `FWaves1.mat`)	0.014	0.003
Dataset 2 (`MWaves1.mat`, `FWaves2.mat`)	0.001	0.011

7.2.1 Source Independence

As discussed in Sec. 3.5.6, source signals originating from independent systems are not necessarily independent in the *time-integrated* sense, which is the basis for ICA. Since the fetal and maternal ECGs have different mean heart rates and both exhibit HRV, the cyclostationary nature of ECG restricts the amount of random dependence that can appear in the joint time-integrated system. Nonetheless, the following analyzes the second-order and higher-order dependence that can occur randomly in their joint system, which will assist in the interpretation of convergence profiles.

Figure 7.1 shows a visualization of the joint system for dataset 1. As described by the standardization procedures of the previous chapter, both sources are standardized to zero mean and unit variance, with the fetal ECG in the second channel. The joint distribution portrayed in panel (c) appears at right angles and is well approximated as independent. The dependence of the joint source system is considered in Table 7.1 by measurements at the second-order, $\widehat{\mathrm{CCI}}(s_1, s_2)$, and at the higher-order via the MI, $\widehat{\mathcal{M}}(s_1, s_2)$. All values are small, but dataset 2 has weaker linear correlation (*i.e.*, smaller CCI), but larger higher-order dependence as measured by Grassberger's MI estimator. Taking into account that these oscillators are joined at an arbitrary relative phase, a context for the values of Table 7.1 can be determined by observing the variability of both measures to a change in phase. To accomplish this, the fetal ECG time series has been shortened by 700 points ($\widetilde{s_2}(n) \leftarrow s_2(n)$), and joined to the maternal ECG at different phase offsets n_o, similar to a sliding window analysis. The values of

$\mathrm{CCI}(s_1(n+n_o), \widetilde{s_2}(n))$ and $\mathcal{M}(s_1(n+n_o), \widetilde{s_2}(n))$ of these joint systems from equivalent sources are plotted in Fig. 7.2. The phase offset length of 700 samples is equivalent to more than double the average fetal ECG beat.

The CCI results in Fig. 7.2(a) for the two datasets demonstrate a variability within $\approx \pm 0.04$. The MI estimates in Fig. 7.2(b) exhibit variability from -0.01 to 0.025 [nats]. (Note that the Grassberger estimator can produce small negative values even though the MI is a nonzero quantity by definition.) The actual estimates for the datasets from Table 7.1 appear in the figures as horizontal lines. This indicates that the estimates of the CCI and MI are subject to a natural variability of approximately 0.04 for the CCI and 0.02 for the MI. The significance of these numerical uncertainties under visual inspection of the distribution scatterplot, however, is relatively small. Note that dataset 1 (solid) in Fig. 7.2(a) has a maximal correlation and anticorrelation identified by stars around a phase offset of $n_o \approx 400$. The joint distributions of these systems are plotted in Fig. 7.3(b) and (c) next to the actual system in panel (a). No perceptible rotation is observed between them, but panel (c) has fewer samples appearing in the first quadrant and more samples appearing in the fourth quadrant. This would seem to indicate that the CCI variability is influenced by the placement of a relatively small subset of points at large amplitude.

7.2.2 Fetal ECG HOS

Recall that the higher-order statistics of the fetal ECG have been identified as important features for convergence consideration. To estimate the natural variability anticipated in the estimation of HOS features from the fetal ECG time series, a similar study to the one presented in the previous subsection has been performed. Here, a large sliding window (width = $N - 700$) has been applied to the fetal ECG. The kurtosis and the entropy of the windowed time series has been estimated under an increasing phase offset (from 1 to 700 samples). Shown in

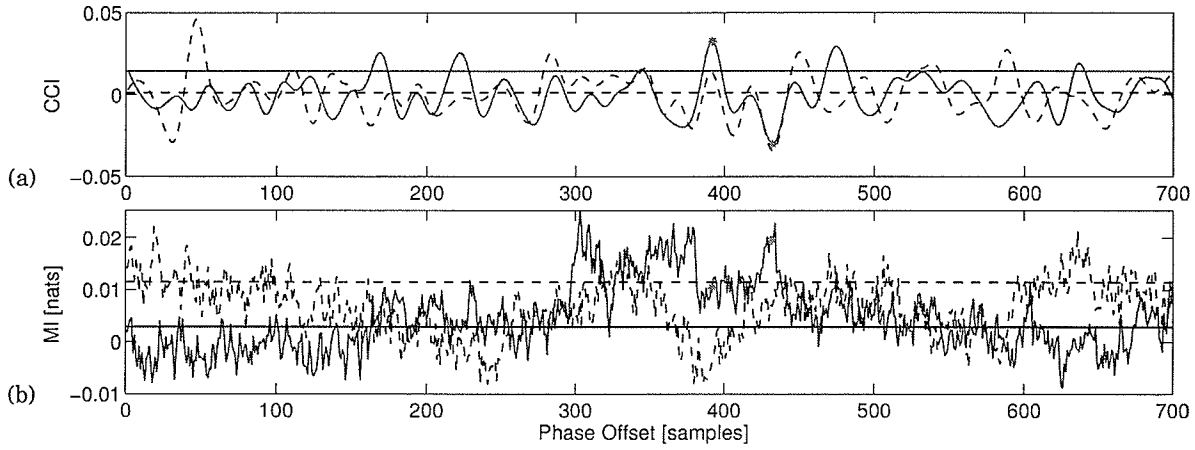


Fig. 7.2 Natural variability in joint statistics due to phase offset of ECG sources: (a) linear correlation $\mathrm{CCI}(s_1(n+o), \widetilde{s_2}(n))$ of the joint ECG sources of dataset 1 (solid) and dataset 2 (dashed) under a phase translation; (b) Estimated dependence $\widehat{\mathcal{M}}(s_1(n+o), \widetilde{s_2}(n))$ of the joint ECG sources of dataset 1 (solid) and dataset 2 (dashed) under a phase translation. Actual CCI and MI values for the datasets (*c.f.*, Table 7.1) are shown as horizontal lines.

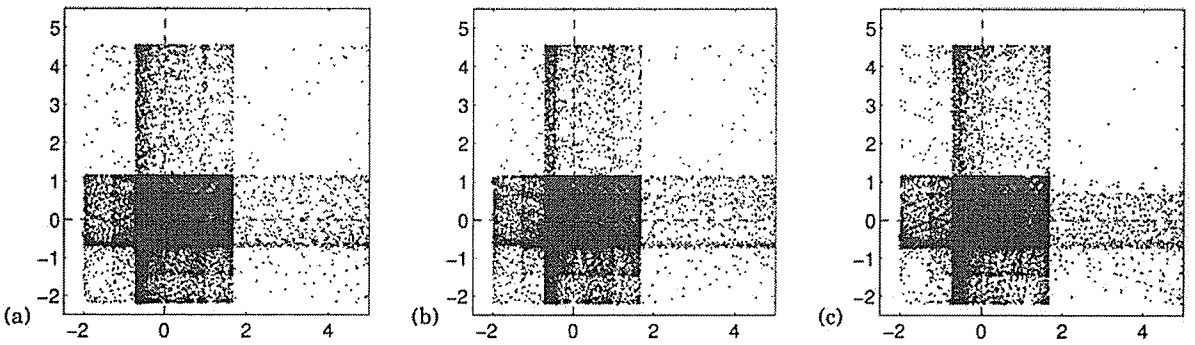


Fig. 7.3 Scatterplot visualization of dataset 1 joint ECG sources at different phase offset correlations: joint system with (a) no offset, (b) offset producing maximal correlation, and (c) offset producing maximal anticorrelation. Offsets defining panels (b) and (c) correspond to the stars in Fig. 7.2(a). Overall patterns seem similar except for the right hand side of panel (c), which contributes to the anticorrelation.

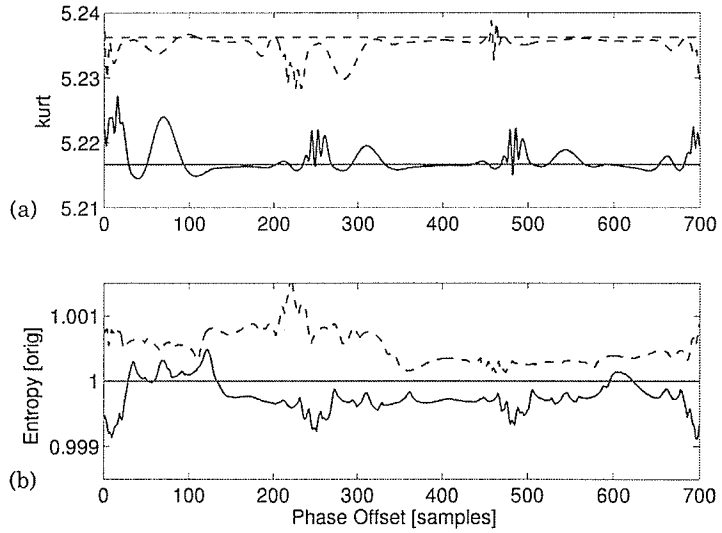


Fig. 7.4 Natural variability in fetal ECG HOS estimates expected from the phase of a random window: (a) kurtosis $k(\widetilde{s_2})$ of the fetal ECG signal on a very long sliding window for the two datasets. Variability is about 0.2%.(b) entropy estimate in units of actual estimate, $\widehat{H}(\widetilde{s_2})/\widehat{H}(s_2)$, of the two fetal ECG datasets have variability between 0.1–0.2%. Both dataset 1 (solid) and dataset 2 (dashed) are represented and the actual values are represented as horizontal lines.

Fig. 7.4, the results indicate that a natural variability of 0.2% for the k and 0.1–0.2% for the entropy is expected for this length of fECG time series. The kurtosis is estimated in its natural unitless form, while the entropy estimate (calculated by the modified Vasicek estimator by Miller (6.21)) is normalized against the entropy estimate of the original; *i.e.*, as $\widehat{H}(\widetilde{s_2})/\widehat{H}(s_2)$. The fetal ECG of dataset 2 has a slightly larger kurtosis, but both are highly super-Gaussian. Since the statistical variabilities are less than a percent, it is likely that demixing error will overwhelm any natural variability from the phase of the recording itself.

7.2.3 MFA Error-Norms

The previous quantities have been examined for statistical deviation under the nuisance parameter of random phase which exists in the acquisition of any random signal. It is not

feasible to consider the natural uncertainty in the MFA of the reconstructed attractor from phase because of the time required for analysis. In the analysis of the MFA, however, there are other sources of variability. Specifically, the selection of a scaling interval for the MFA scaling entropies by visual inspection is a source of uncertainty. To determine the effect of the scaling interval variability on the scalar MFA error-norms defined in Sec. 6.5.2.4, the MFA of the actual fetal ECG source signals have been modified by slight changes to the chosen scaling interval. Since the correlation partition remains unchanged — only the entropies and their fitting need be reevaluated — this is a reasonable analysis. The results that are shown in Fig. 7.5, demonstrate that the MSE and CHE error-norms behave similarly, but with the MSE being generally larger. Furthermore, the MSE error-norm returned similar values from both the Rényi spectrum or the Mandelbrot spectrum, whereas the CHE error-norm applied to the MS appears as the most sensitive. In response to the changing scaling interval shown in panel (c), changes to the MFA spectra by the error-norms are measurable in the range of 2–4. From this, it seems reasonable to allow an MFA error-norm tolerance of 4–6 as a match to account for variability introduced by visual inspection.

Now that tolerances have been specified by identifying the natural uncertainties in the HOS and MFA characterizations of the data, the discussion now proceeds to the convergence of the fundamental ICA SQMs: namely, the ICA cost functions and the matrix-based SQMs.

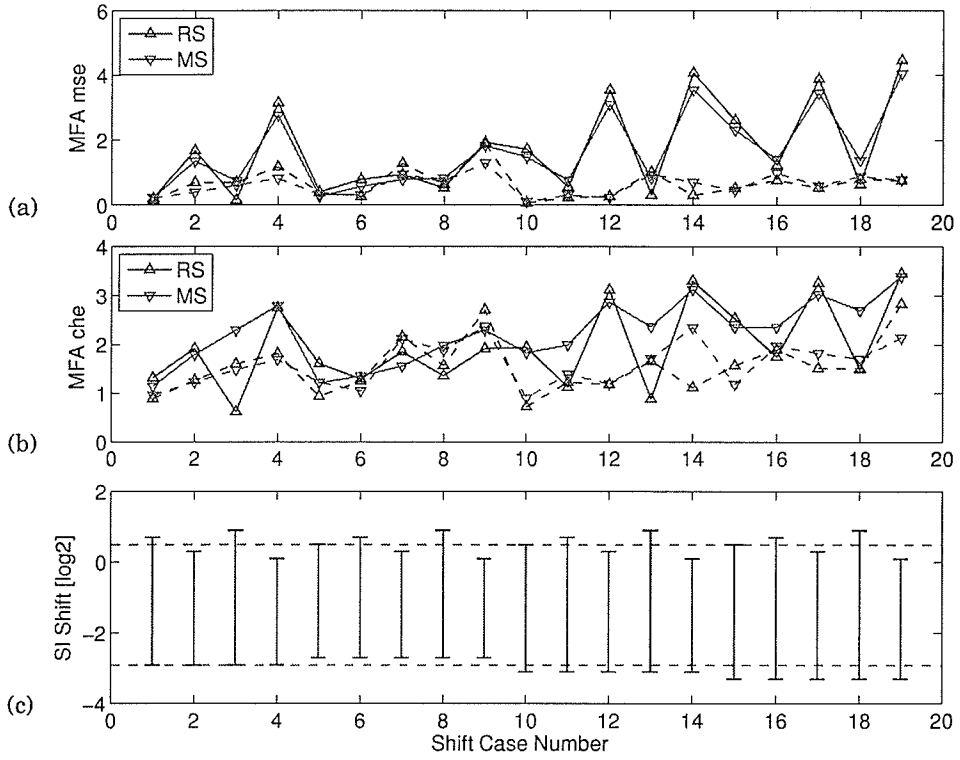


Fig. 7.5 Variability in MFA error-norms under small changes to scaling interval: (a) MSE-based MFA error-norm, and (b) CHE-based MFA error-norm as a function of (c) adjusted choice of scaling interval for the original source fetal ECG attractors of dataset 1 (solid) and dataset 2 (dashed). Both RS and MS spectra are analyzed as marked. Dataset 1 is most sensitive to a shortening of the scaling region, while dataset 2 is more robust.

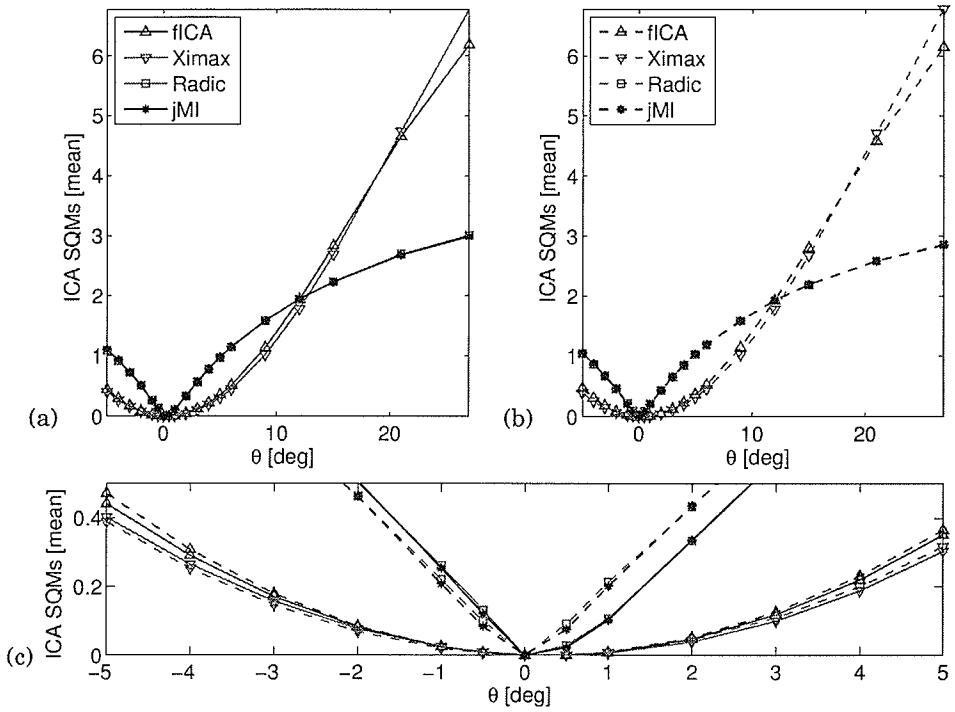


Fig. 7.6 Convergence of ICA cost functions for (a) dataset 1 (solid), and (b) dataset 2 (dashed). FastICA and Xinfomax have similar superlinear convergence, while RADICAL and MILCA have similar sublinear convergence. Panel (c) shows both datasets in the convergence limit. RADICAL and MILCA have better discrimination in this region. Note that an interpolated minimum for dataset 1 should appear for non-zero demixing, $\theta \approx 0.25°$. (Note also that all cost functions are normalized to zero minimum and unit mean; *c.f.*, Sec. 6.5.1.)

7.3 Convergence of ICA Cost Functions and Matrix SQMs

7.3.1 ICA Cost Functions (B-Class)

The normalized convergence profiles of the FastICA, Xinfomax, RADICAL, and MILCA SQMs on the two datasets are displayed in Fig. 7.6. From these relationships, it is clear that FastICA and Xinfomax have a *superlinear* convergence (*i.e.*, power-law with exponent greater than 1), while RADICAL and MILCA have an overall *sublinear* convergence (*i.e.*, exponent less than 1). For convenience, the former will be identified as the superlinear group, and the latter as the sublinear group. Examining the profiles in the convergence limit (panel (c)),

the optimal demixing identified by the minima of the cost functions is not necessarily zero as expected. The superlinear group appears symmetric around $\theta \approx 0.25°$ for both datasets, but discrimination is difficult because of the small derivatives in this region. The approach of the sublinear group is effectively linear, however, and consequently a distinct difference in optimum between the two datasets is perceptible. Dataset 2 (dashed) converges to a demixing $\theta \approx 0°$, whereas dataset 1 (solid) converges to positive angle, $\theta \approx 0.25°$. Notwithstanding that both of the cost functions in the sublinear group are directly derived from MI formulae, the consistency of the two different estimates is striking.

The two classes of profiles shown here in Fig. 7.6 will be an important reference for the remainder of the analysis since it captures the sensitivity of ICA to the independent variable of these experiments, θ. The significance of θ from a joint distribution perspective is visualized in the scatterplots of Figs. 7.7 and 7.8. In particular, panels (c) and (d) from Fig. 7.7 at $\theta = 0°$ and $\theta = 1°$, respectively, both appear to give reasonable, but imperfect, independence overall. Explicitly, the *horizontal* edges of Fig. 7.7(c) do not appear parallel to the frame, while the *vertical* edges of Fig. 7.7(d) are slightly askew. This can be interpreted as an optimal mixing for the maternal ECG in the former, and an optimal demixing for the fetal ECG in the latter. The plots of Fig. 7.8, however, display without confusion that $\theta = 0$ is optimal. Specifically, even the demixing of 1° is perceptibly rotated.

7.3.2 Matrix SQMs (A-Class)

The matrix SQMs are fundamental metrics of ICA convergence because they characterize the convergence of the weight matrix $\mathbf{W}$ within the optimization space. The normalized API, MFE, and VNE (channel 2, fetal ECG) are shown in Fig. 7.9 for dataset 1 (solid) and dataset 2 (dashed). The MFE and API both effectively have linear convergence to the optimum. The VNE, on the other hand, is superlinear when distant from the optimum. In the convergence

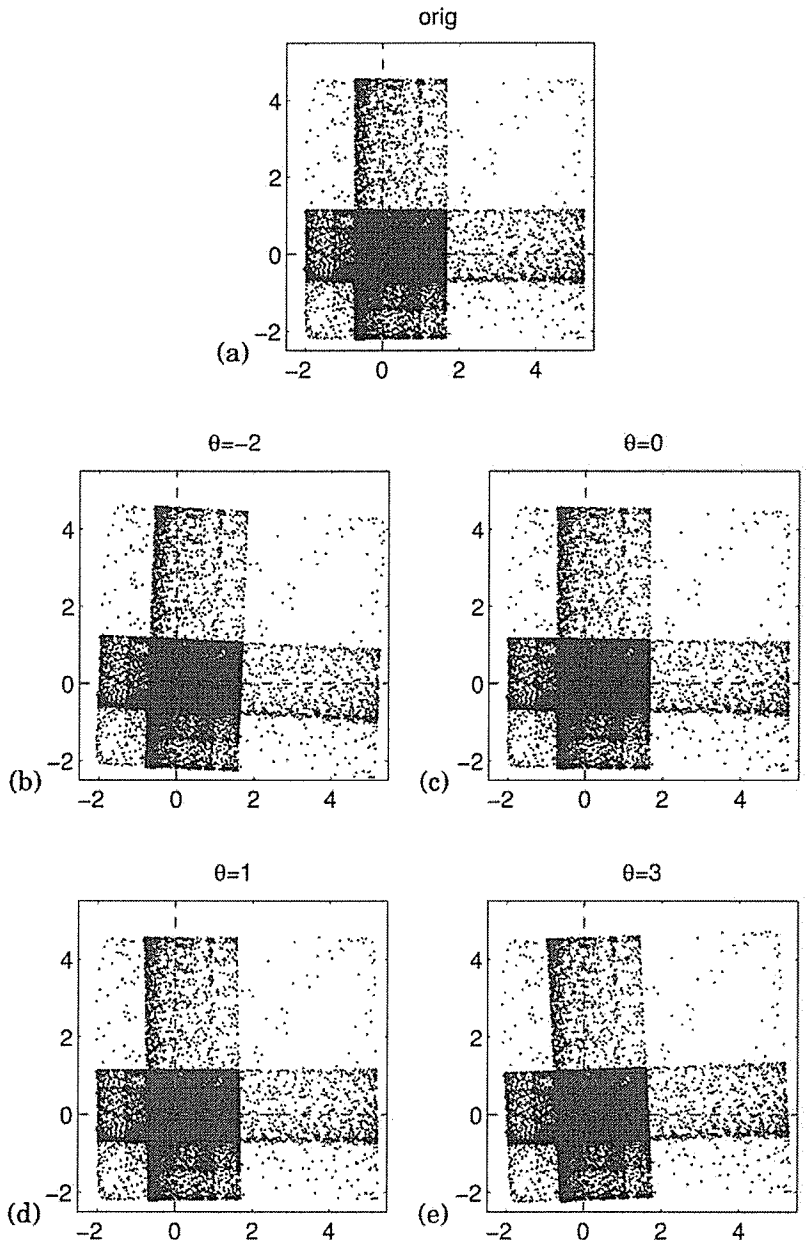


Fig. 7.7 Scatterplot visualization of the dataset 1 joint ICA system estimates at significant demixing levels: (a) actual sources; (b) demixing estimate at $\theta = -2°$; (c) demixing estimate at $\theta = 0°$; (d) demixing estimate at $\theta = 1°$; (e) demixing estimate at $\theta = 3°$. Panels (b) and (e) are observably rotated from an optimal. Panels (c) and (d) seem to compromise on the optimization: the former optimizing the vertical edges, and the latter optimizing the horizontal edges.

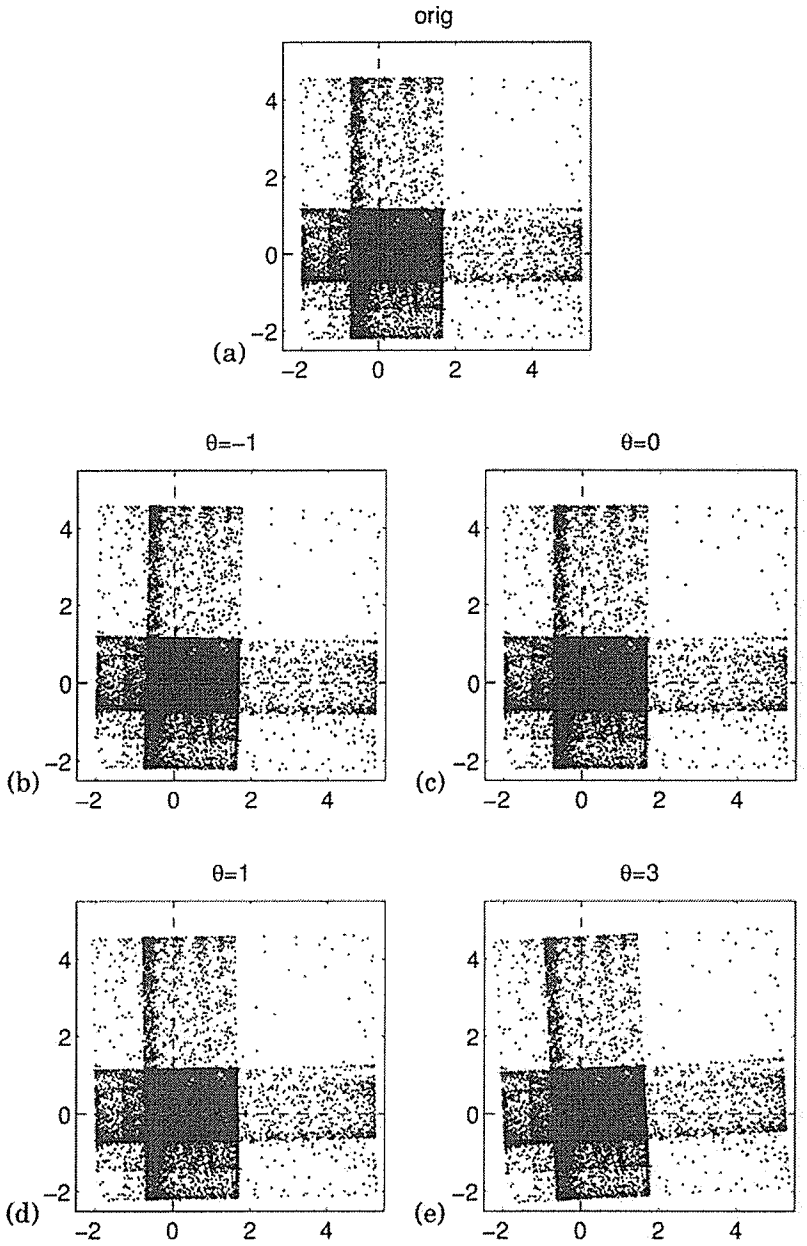


Fig. 7.8 Scatterplot visualization of the dataset 2 joint ICA system estimates at significant demixing levels: (a) joint system of dataset 2 sources; (b) demixing estimate at $\theta = -1°$; (c) demixing estimate at $\theta = 0°$; (d) demixing estimate at $\theta = 1°$; (e) demixing estimate at $\theta = 3°$. Panels (b) and (d) seem mildly rotated, while panel (e) is observably rotated from an optimal.

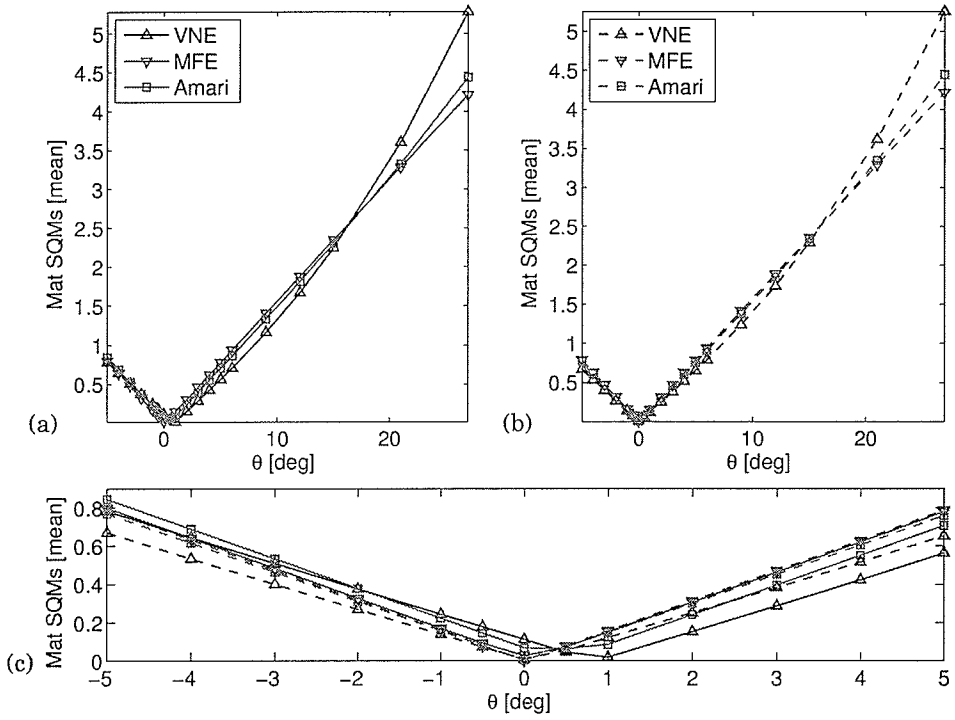


Fig. 7.9 Convergence of matrix SQMs for (a) dataset 1 (solid), and (b) dataset 2 (dashed). MFE (∇) and API ($\square$) essentially converge linearly in θ, while the VNE (Δ) is superlinear. Panel (c) shows both datasets in the convergence limit. All SQMs optimize at zero angle for dataset 2. For dataset 1, VNE optimum for occurs for $\theta = 1°$, while MFE still remains at zero angle, and the API flattens on the [0,1] interval making an optimum more difficult to interpret.

limit shown in panel (c), all metrics for dataset 2 (dashed) converge to $\theta = 0$. For dataset 1 (solid), however, the SQMs disagree about the optimum. The $\aleph^{[\mathrm{MFE}]}$ (∇) converges to $\theta = 0$, but the $\aleph^{[\mathrm{VNE}]}$ ($\triangle$) converges to $\theta = 1°$. The $\aleph^{[A_\epsilon]}$ in this case ($\square$) actually flattens over the [0, 1°] interval and does not reach zero. The minimal values for the MFE and VNE are also both positive, but smaller than the minimum API. Recall that the VNE is the only direct measure of the fetal demixing, while the other SQMs measure the joint system. It follows, therefore, that the disagreement among the metrics over the dataset 1 optimum is consistent with the observation of the scatterplots in Fig. 7.7 that $\theta = 0$ seemed to optimize the maternal ECG and $\theta = 1°$ optimized the fetal ECG.

Comparing the convergence profiles of the matrix SQMs in Fig. 7.9 with those of the ICA cost functions in Fig. 7.6, it is interesting to consider the effect of the original cross-correlation that existed in the source datasets. Dataset 1 had a correlation of less than 2%, which is typical of the up to 4% tolerance that can be introduced by phase alone, but 10 times larger than that of dataset 2. These experiments appear to suggest that this difference in random correlation can complicate the proper convergence of ICA. The ICA estimates are assumed in sphered form, and, consequently, the demixing is unable to optimize both channels simultaneously if the random CCI is at the higher end of the tolerance identified in Sec. 7.2. The sensitivity to this effect is not uniform, however. The MFE as an A-Class SQM and the superlinear group of B-Class ICA cost functions seem the least sensitive to this effect. The sublinear group of ICA cost functions and the API are more sensitive, whereas the VNE is clearly the most affected, since it measures only the convergence of the fetal ECG channel.

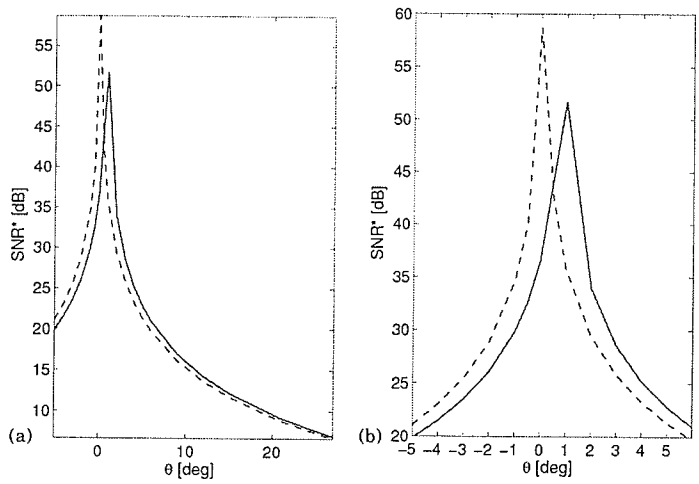


Fig. 7.10 Convergence of SNR for dataset 1 (solid) and dataset 2 (dashed): (a) overall profile,; (b) in the convergence limit. Performance of greater than 20 dB occurs for demixing angle within ±5° of optimum. Optimum for dataset 1 approaches $\theta \approx 1°$ while the other approaches $\theta = 0$. Recall that dataset 1 started with larger native CCI, while the estimates are constrained to have none.

7.4 Convergence of Second-Order A-Class Features

7.4.1 Standardized SNR

The convergence profile of the standardized SNR SQM, $\aleph^{[\mathrm{SNR}^*]}$, is shown in Fig. 7.10. Continuing the trend of previous SQMs which distinguish between the two datasets, dataset 1 is optimized at $\theta = 1°$ while dataset 2 is optimized at $\theta = 0$. Note that dataset 2 improves over the quality of dataset 1 by more than 5 dB in the convergence limit, but both exceed 20 dB performance with demixing angles within ±5° of optimum. According to the statement of Principe *et al.* [88, (2001)] quoted on p. 136, the threshold for "good separation", therefore, should be considered to be ±5°. A more stringent requirement of 30 dB performance might require that the demixing angle be roughly within 2° of optimum. In a more realistic fetal ECG scenario, of course, the added corruption introduced by additional sources and noise would lower the SNR values observed here, and serve to shrink the interval allowed demixing values.

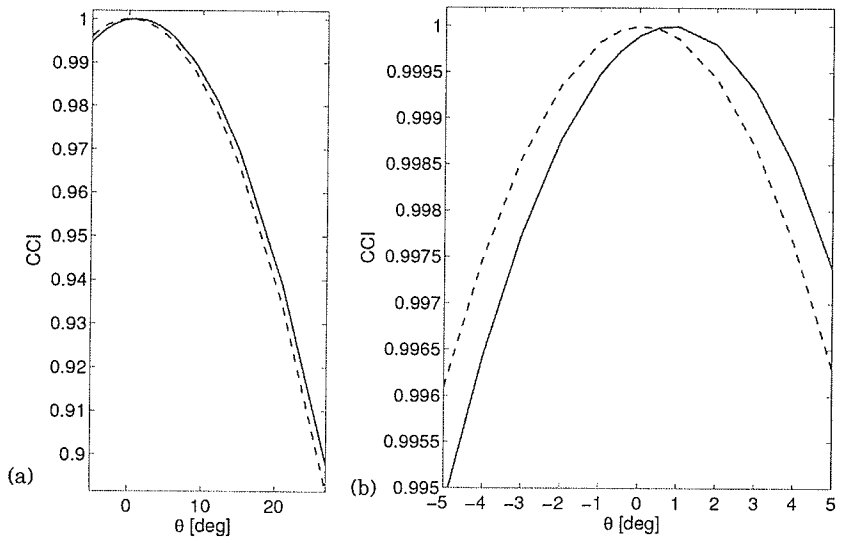


Fig. 7.11 Convergence of CCI for dataset 1 (solid) and dataset 2 (dashed): (a) overall profile; (b) in the convergence limit. Convergence is superlinear but highly insensitive. Discrimination occurs realistically in the third decimal. Optimum for dataset 1 is $\theta \approx 1°$ while the other is zero.

7.4.2 CCI

The convergence profile of the correlation SQM, $\aleph^{[\mathrm{CCI}]}$, is shown in Fig. 7.11. As with the SNR, dataset 1 is optimized at $\theta = 1°$ while dataset 2 is optimized at $\theta = 0$. In contrast to the SNR, however, the profiles appear of the same quality. Overall, the CCI is a very insensitive SQM, since separation in the $\pm 5°$ interval is discriminated only within the third decimal (*i.e.*, CCI > 0.995). Although, as with the SNR, the characteristic CCI values would likely drop in situations with more corruptible influences.

In general, the second-order absolute SQMs are sensitive to the initial correlation that can randomly appear in the joint fetal-maternal system. In this they are best compared to the VNE in the matrix sub-class. Furthermore, when considered against the two classes of convergence observed in the ICA cost function SQMs, it is expected that an ICA estimate positioned at the minimum of the ICA cost functions would be $\pm 1°$ from a SOS A-Class optimum for the

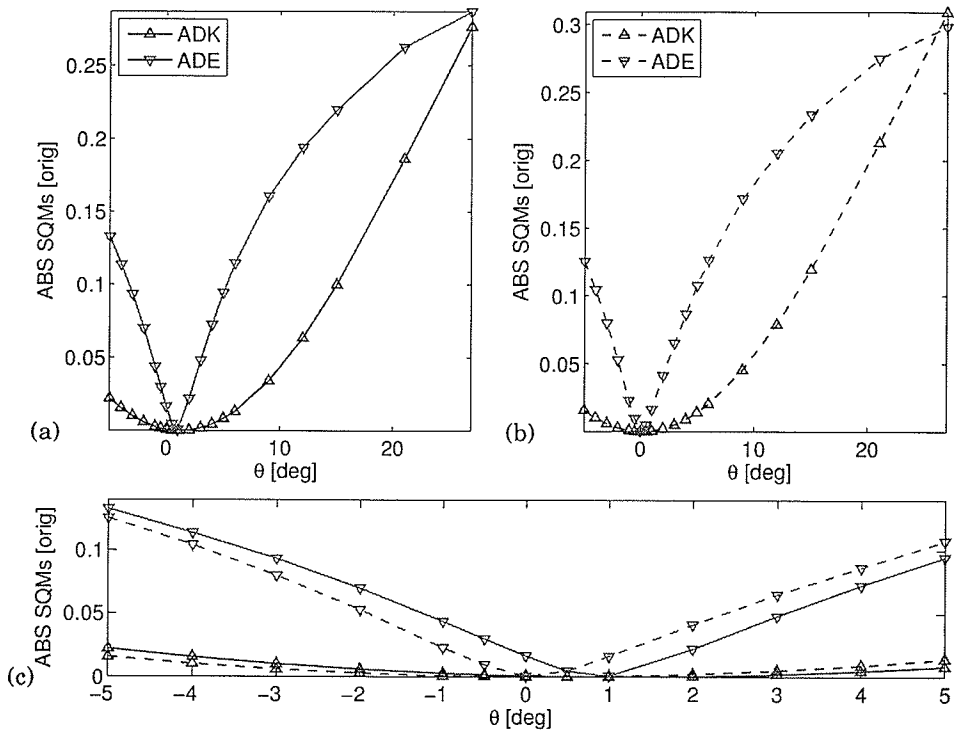


Fig. 7.12 Convergence of absolute HOS SQMs for (a) dataset 1 (solid), and (b) dataset 2 (dashed). Panel (c) shows both datasets in the convergence limit. The kurtosis ($\triangle$) converges much faster than the Shannon entropy ($\triangledown$), (superlinearly to sublinearly respectively). Optimum for dataset 1 is $\theta \approx 1°$ while the other is zero. This is similar to the second-order A-Class SQMs and demonstrates a sensitivity to initial correlation in the joint fetal-maternal system.

sublinear group and about $\pm 1.5°$ for the superlinear group. Convergence to that level would amount to an SNR* performance of greater then 30 dB.

7.5 Convergence of Higher-Order A-Class Features

7.5.1 Basic HOS A-Class (ADK and ADE)

The convergence of the higher-order statistics are an important point of interest of this work. The convergence profile of two of the proposed HOS-based SQMs are shown in Fig. 7.12; namely, the ADK ($\triangle$) and the ADE ($\triangledown$). As described in Sec. 6.5.2.3, these values are normalized against the actual values of the kurtosis and Shannon entropy of the source

fetal ECG. As with the ICA cost functions, there is a distinctive difference in convergence rate: superlinear convergence by the kurtosis, and a sublinear convergence by the entropy. In the convergence limit shown in in panel (c), it is clear that the ADE is optimized by $\theta = 1°$ for dataset 1, as was found for the SOS-based SQMs. A stronger magnification can also reveal that the ADK of the same dataset is optimized for $\theta = 1°$ as well. Note, however, that the slow convergence of the ADE implies that for θ within $\pm 3°$ from optimum is only within 10% accuracy, while the kurtosis is less than 0.6%. Furthermore, this implies by the observations of the previous sections, that an SNR^* of 30 dB only assures an entropy estimate of 5% accuracy, which is substantially larger than the natural variability in the estimator (which was less than 1%, *c.f.*, Sec. 7.2). It follows, therefore, that demixing error is the main contributor to the estimation of entropy. Considering, also, that the superlinear ICA group (*e.g.*, FastICA and Xinfomax) has weak sensitivity around the optimum, demixing processes from this class should have increased variance in entropy estimation, (*e.g.*, 10% of actual). The members of the sublinear class, however, have stronger sensitivity here, and consequently both kurtosis and entropy estimation should be reasonably accurate (*e.g.*, $< 1\%$ and 2% of actual, respectively).

To compare the superlinear behaviours of the ICA cost functions to that of the kurtosis, Fig. 7.13 expresses their joint behaviour parameterized by θ. The relationship is mildly sublinear. By examining the convergence behaviour in panel (c), it can be seen that the spreading effect introduced by the source correlation of dataset 1 subjects less than 0.5% error to the kurtosis estimate. Since the CCI observed in dataset 1 is typical according to Sec. 7.2, this would seem to indicate that the kurtosis would be accurate to within 1% subject to random correlation in the sources from phase matching.

A similar comparison of the sublinear ICA class to the entropy is shown in Fig. 7.14. This relationship is essentially linear (and not surprising given the relationship of the MI to marginal

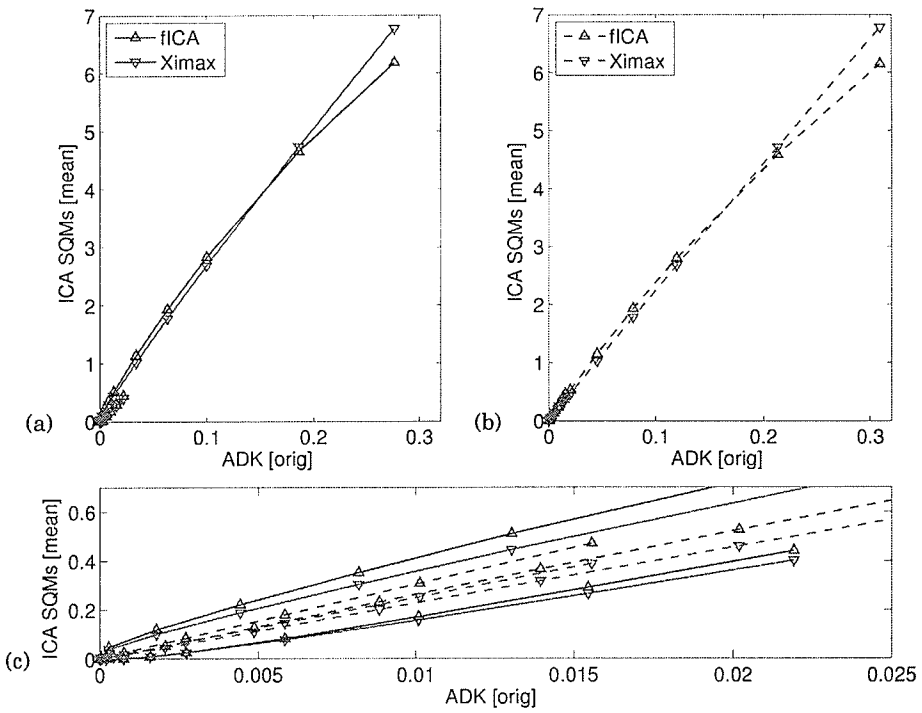


Fig. 7.13 ADK relationship to FastICA and Xinfomax for (a) dataset 1 (solid), and (b) dataset 2 (dashed) is mildly sublinear. Panel (c) shows both datasets in the convergence limit. Dataset 1 spreads out wider from the convergence point.

entropy). The effect of the source correlation, however, is enhanced here. Specifically, the optimal ICA solution only makes the kurtosis estimate accurate to just under 2%. This would appear to be the lower bound on entropy accuracy due to the random correlation in the sources.

7.5.2 Cross-HOS A-Class (xMI)

The unconventional use of cross-MI for an SQM gives the results shown in Fig. 7.15. Its form is surprisingly similar to that of SNR*, but with a shallower peak. This is consistent with the fact that both the SNR and the cross-MI diverge to ∞ if the two signals are identical. Further study is required to identify whether any further relationship between the second-order and higher-order SQMs is significant.

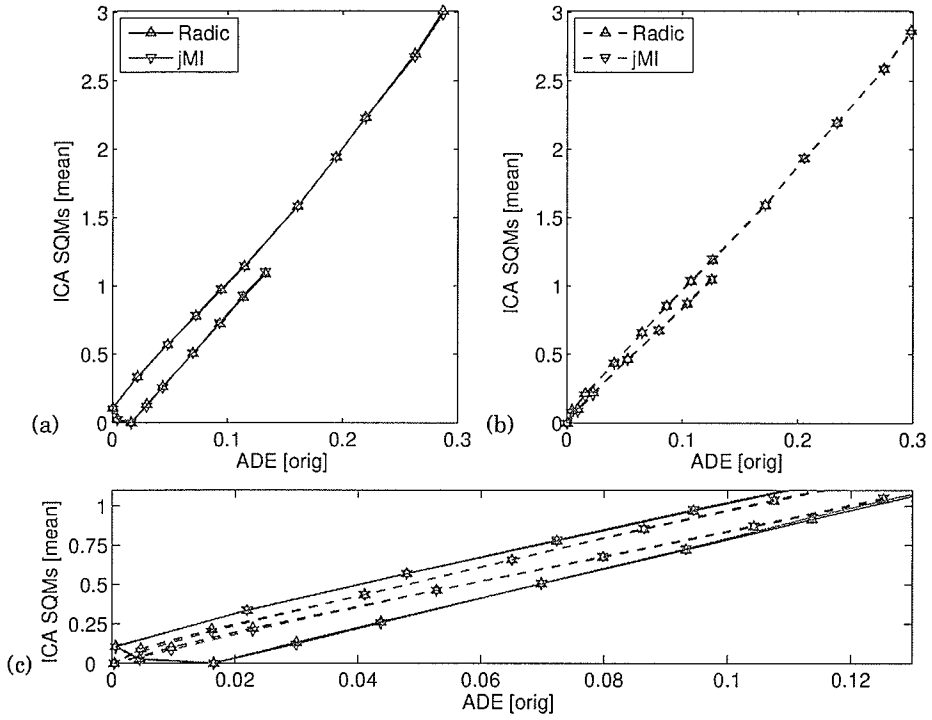


Fig. 7.14 ADE relationship to RADICAL and MILCA for (a) dataset 1 (solid), and (b) dataset 2 (dashed) is roughly linear. Panel (c) shows both datasets in the convergence limit. There is an offset in their optimal point from the original correlation in the sources of dataset 1.

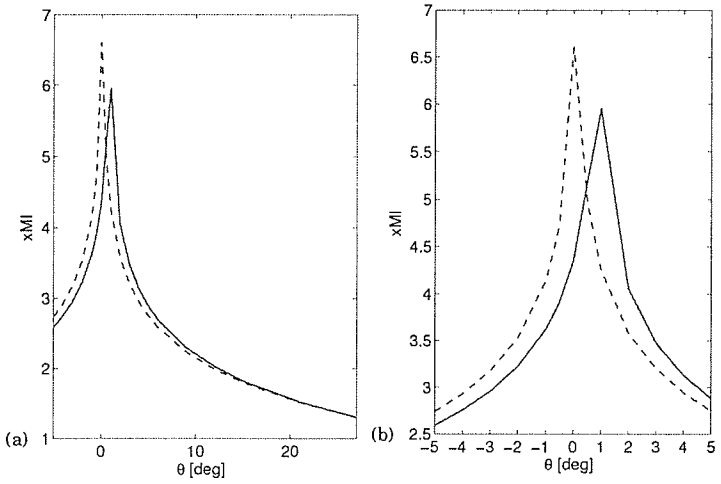


Fig. 7.15 Convergence of xMI for dataset 1 (solid) and dataset 2 (dashed): (a) overall profile,; (b) in the convergence limit. Optimum for dataset 1 approaches $\theta \approx 1°$ while the other approaches $\theta = 0$. At its best representation, dataset 1 is more independent than dataset 2.

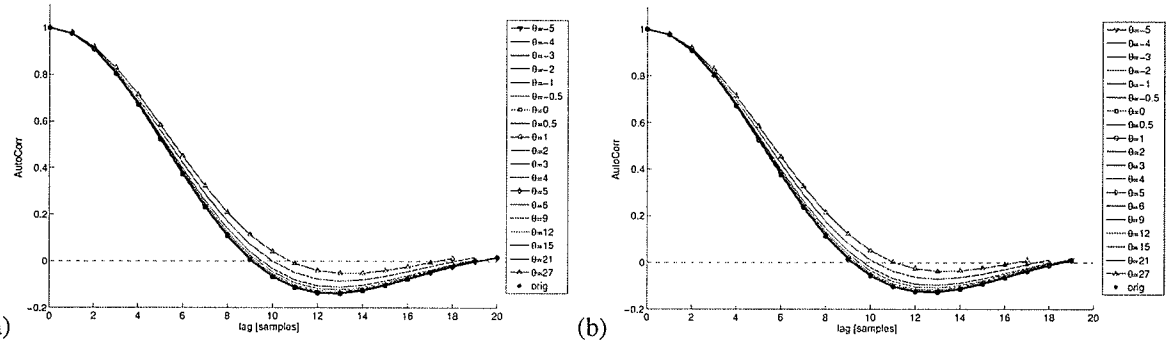


Fig. 7.16 Estimation of embedding lag for (a) dataset 1 and (b) dataset 2. Lag is insensitive to moderate mixing, with $\widehat{\Delta n} \approx 18$.

7.6 Convergence of Nonlinear Dynamical A-Class Features

7.6.1 Preamble: Attractor Parameters

As described in Sec. 6.4, the embedding lag Δn and the embedding dimension N_e are prerequisites for the MFA of a fetal ECG time series or its demixing estimate. The automated results that determine the embedding lag for the two datasets are shown in Fig. 7.16. The second zero of the autocorrelation was successfully defined for all mixing angles considered, and is rather insensitive to moderate levels of mixing.

The embedding dimension is a more complicated feature: requiring user interpretation of (i) a D_2 scaling interval, and (ii) a D_2 saturation point over increasing embedding dimension. Both of these elements are subject to changes under mixing, as is demonstrated in Fig. 7.17 and Fig. 7.18. Here, Fig. 7.17(a) shows the D_2 saturation curves for the different estimates of dataset 1, as well as the original fetal ECG source (bold). Note that the saturation points vanish under moderate mixing ($\theta \geq 15°$), but otherwise are approximately determined as 9 or 10. The scaling intervals of panel (b) also demonstrate that mixing reduces the scaling region in a rather continuous way. At heavier levels of mixing, however, a true scaling interval becomes too short for proper identification, and is hence interpreted as a lacunarity effect within the

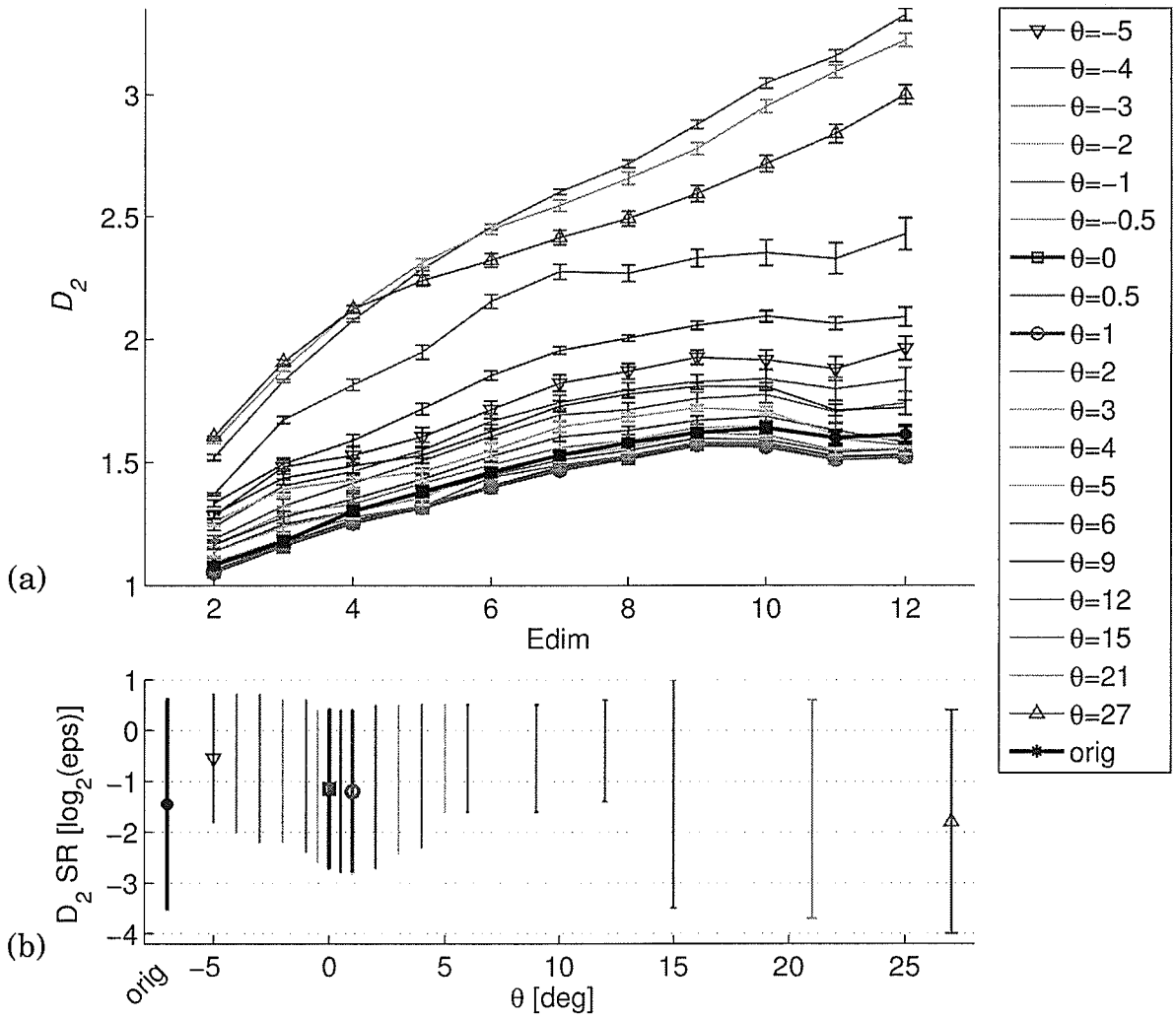


Fig. 7.17 Estimation of embedding dimension for dataset 1: (a) D_2 saturation curves, and (b) their supporting scaling region. At moderate mixing no saturation occurs, and therefore no dimension is assigned. Otherwise, saturation of estimates is around 9 or 10. Note also that the scaling interval tends to maximize around 0.

larger behaviour of the correlation sum. Thus the scaling intervals at moderate mixing are identified as very large, but no saturation is found. The features of the embedding dimension analysis of dataset 2 (displayed in Fig. 7.18) are similar, except some saturation points are identified at $N_e = 8$. The estimated embedding dimension for all demixing estimates of both datasets are compiled in Table 7.2.

Table 7.2 Determined Embedding Dimensions for Fetal ECG Estimates

Dataset	$\widehat{N_e(\widehat{s_2})}$																		
θ [°]	-5	-4	-3	-2	-1	-0.5	0	0.5	1	2	3	4	5	6	9	12	15	21	27
Dataset 1	9	9	10	10	9	9	9	9	9	9	9	9	9	9	10	7	—	—	—
Dataset 2	10	10	8	8	8	10	10	10	10	10	10	10	8	11	11	9	—	—	—

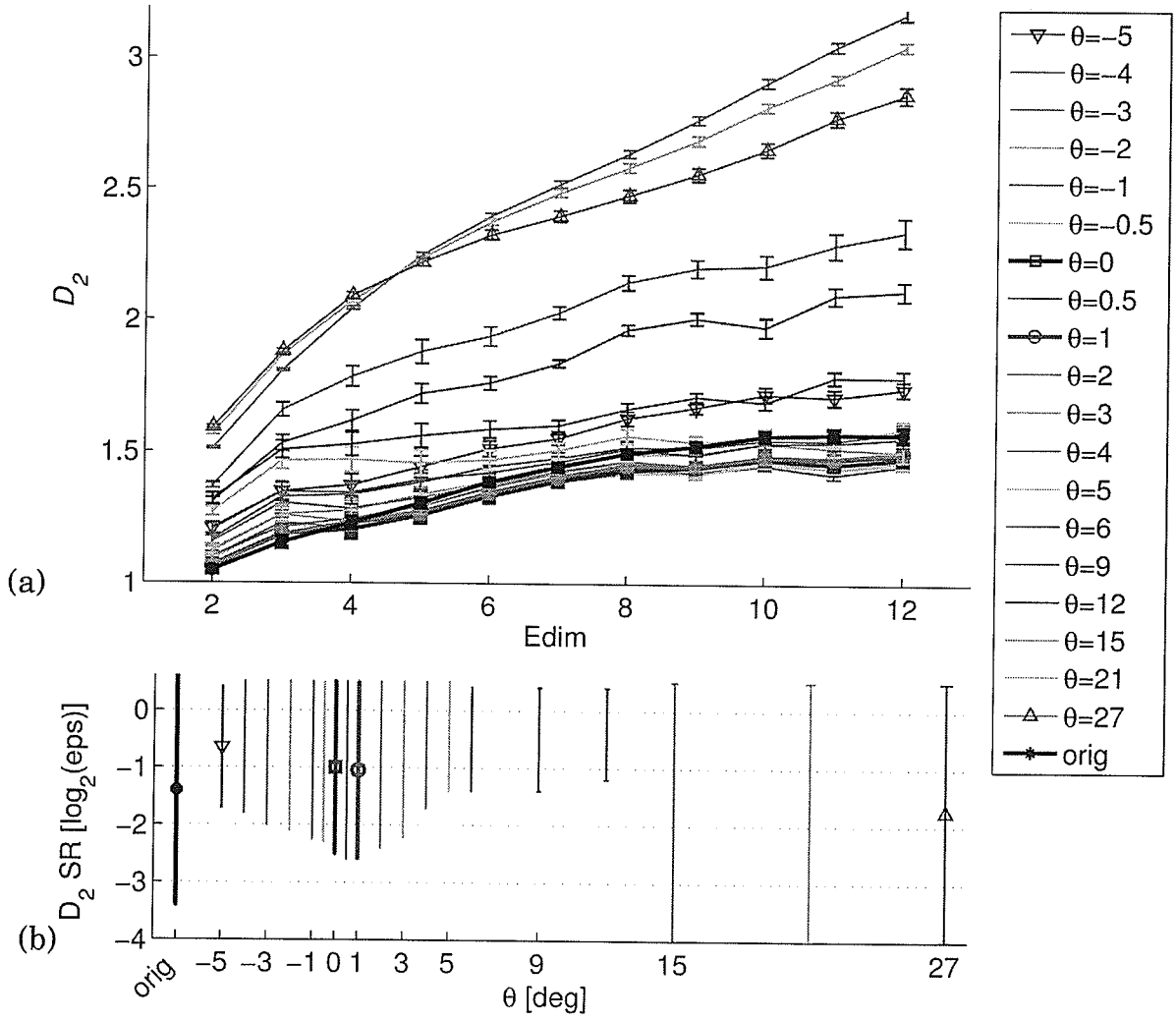


Fig. 7.18 Estimation of embedding dimension for dataset 2: (a) D_2 saturation curves, and (b) their supporting scaling region. At moderate mixing, no saturation occurs, and therefore no dimension is assigned. Otherwise, saturation of estimates is around 8–10. Note also that the scaling interval tends to maximize around 0.5° – 1°.

7.6.2 MFA Convergence

At long last, the thesis question can now be examined. Does the MFA of the ICA estimated signal converge to that of the original? Figure 7.19 presents the convergence profile of the MFA error-norms. As with the previous A-Class SQMs, the two datasets behave differently. With the thresholds determined from Sec. 7.2.3 included in panels (c) and (d) as horizontal lines, it appears that dataset 2 (dashed) successfully converges around $\theta = 0$ in keeping with the other features previously analyzed. Further details will also be discussed later. In contrast, however, the profile of dataset 1 (solid) does not cross below the threshold. The origins of this behaviour will be discussed further in the next subsections by evaluating the MFA and scaling entropies directly. Discussion will focus first on dataset 2, because it is the simpler case.

7.6.3 Dataset 2: MFA Details at Significant θ

The interpretation of dataset 2 is somewhat straightforward. According to the convergence profile of Fig. 7.19, the demixing estimates of $\theta = 0$ and 1° should be matching the MFA spectra of the original. These MFA are shown in Fig. 7.20. In fact, inspection of these curves demonstrates that the $\theta = 0$ estimate is almost a point-by-point reproduction of the original. The $\theta = 1^\circ$ estimate is slightly offset from the original outside the $q \in [0, 2]$ interval, but is within the numerical uncertainties. The estimate at $\theta = 3^\circ$, meanwhile, only matches the original within the $q \in [0, 2]$, and is defined on a shorter scaling region than the original, and so is correctly identified as a non-match.

A significant element of dataset 2's convergence profile, however, is the drastic difference in MFA error-norm between $\theta = -1^\circ$ and $\theta = -0.5^\circ$. This high measure of disparity is consistent with the contrasting MFA spectra of Fig. 7.20 they measure. Since the scaling intervals for the MFA estimation are similar, this is not the source of disparity. Inspection

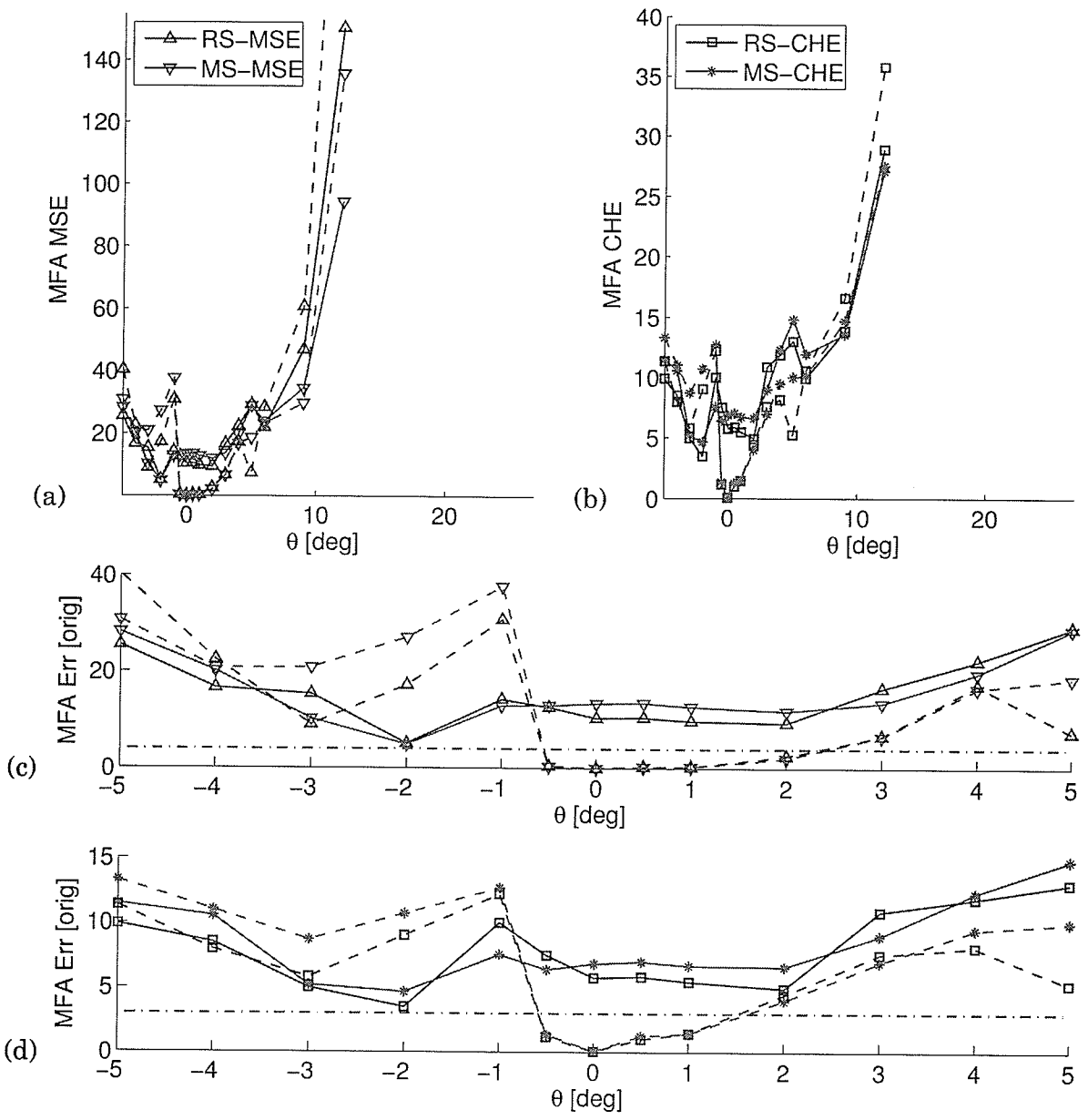


Fig. 7.19 Convergence of MFA error SQMs for dataset 1 (solid) and dataset 2 (dashed): (a) MSE-based MFA error-norm applied to both the RS and MS representations (as marked); (b) as in panel (a), but using the CHE-based MFA error-norm; Panels (c) and (d) amplify the convergence limit for the MSE-norm SQM and CHE-norm SQM respectively. Tolerance levels described in Sec. 7.2.3 are plotted as horizontal lines. Dataset 2 converges with a tolerance of $\Delta\theta < 1°$, while dataset 1 does not converge. Estimated embedding dimensions are in Table 7.2.

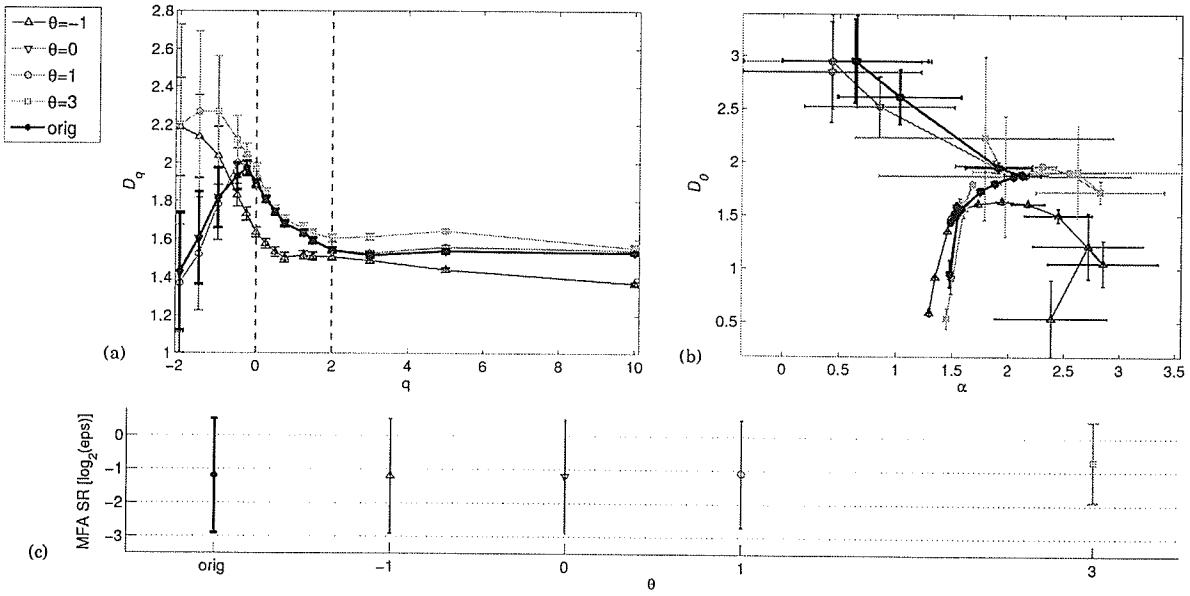


Fig. 7.20 MFA spectra for dataset 2 at significant demixing angles θ: (a) Rényi spectrum; (b) Mandelbrot spectrum; and (c) scaling interval for the demixing estimates as marked. Actual MFA of original source fetal ECG is also shown (heavy solid). Scaling intervals match for all except $\theta = 3°$. Spectra of $\theta = 0, 1°$ match the original, however, the $\theta = -1°$ curve is quite different. Visual inspection of the curves corresponds to the error-norm estimates.

of the MFA scaling entropies (Figs. 7.21 and 7.22) confirms that, indeed, the slopes of the entropies are different (*c.f.*, 7.22(a) and (c)) and the scaling in each is well-defined. In fact, the source of this MFA disparity is not in the MFA analysis itself, but a sensitivity to an underlying change in assigned *embedding dimension*. The N_e values included in Table 7.2 change from 8 at $\theta = -1°$ to 10 at $\theta = -0.5°$. In fact, the section of MFA error-norm over $\theta \in [-3°, -1°]$ which is intuitively going the wrong way (*i.e.*, increasing instead of decreasing), all have been embedded in 8 dimensions instead of 10, and, therefore, are not consistent with the original. This would seem to indicate that embedding dimension has an important effect on the MFA estimates, even though the feasibility of the embedding dimension cannot be determined from the scaling entropies alone.

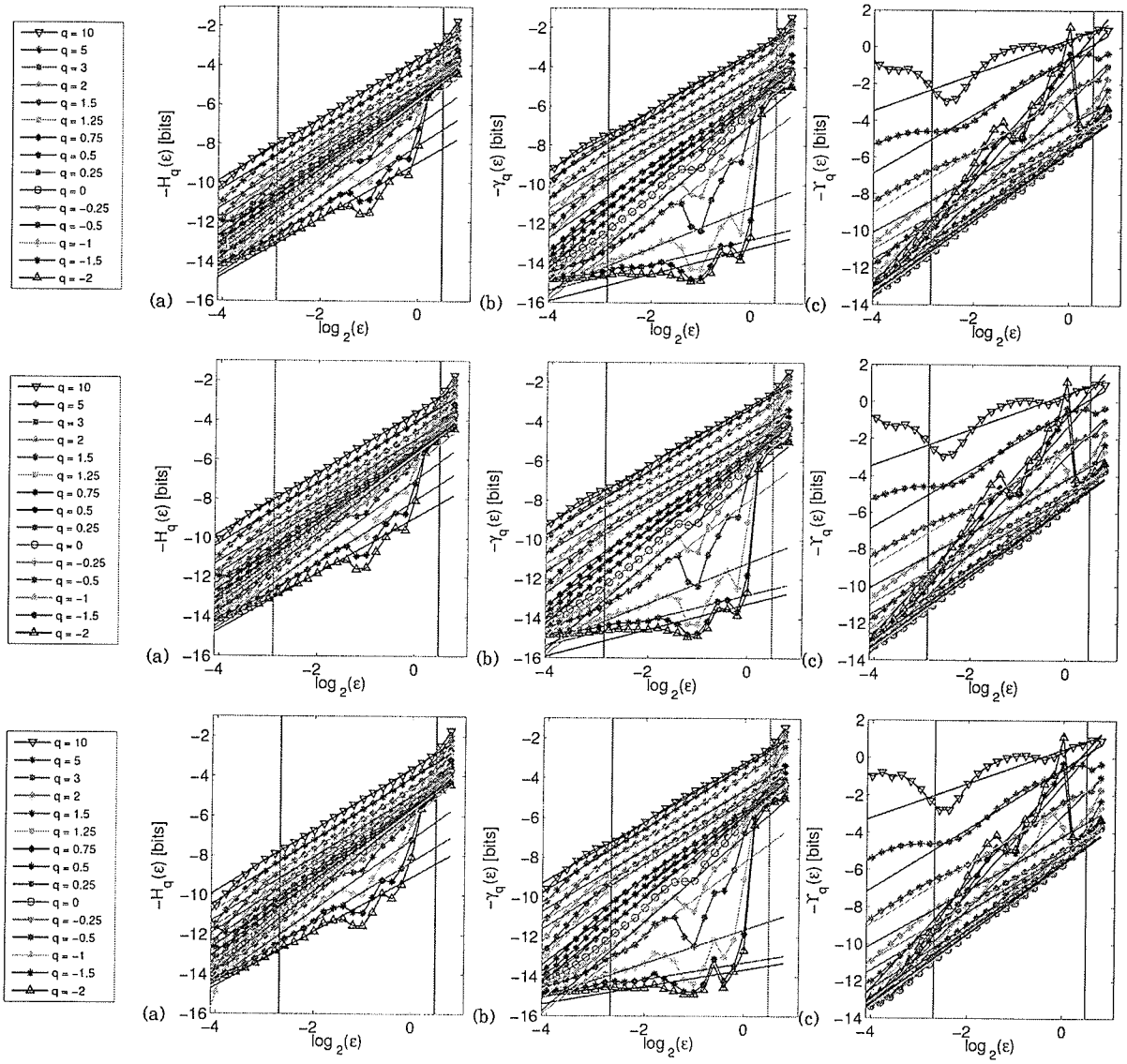


Fig. 7.21 Scaling entropies of dataset 2 at significant demixing levels: Simultaneous scaling of (a) Rényi, (b) Hölder, and (c) Mandelbrot entropies of the original dataset 1 fECG attractor (top row), and its reconstruction at a mixing of $\theta = 0°$ (middle row) and optimal $\theta = 1°$ (bottom row).

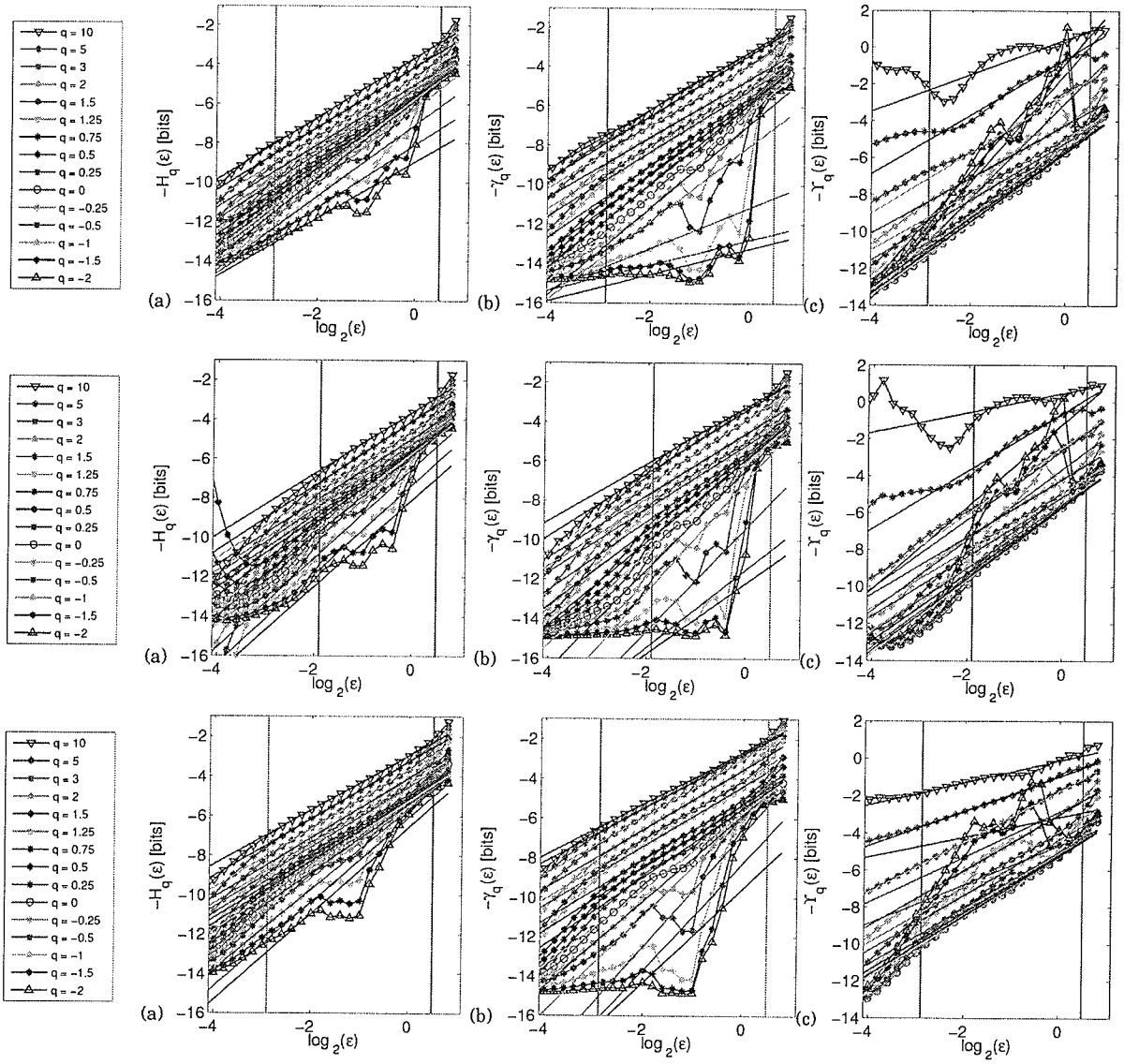


Fig. 7.22 Scaling entropies of dataset 2 at significant demixing levels: Simultaneous scaling of (a) Rényi, (b) Hölder, and (c) Mandelbrot entropies of the original dataset 1 fECG attractor (top row), and its reconstruction at a mixing of $\theta = 3°$ (middle row) and $\theta = -1°$ (bottom row).

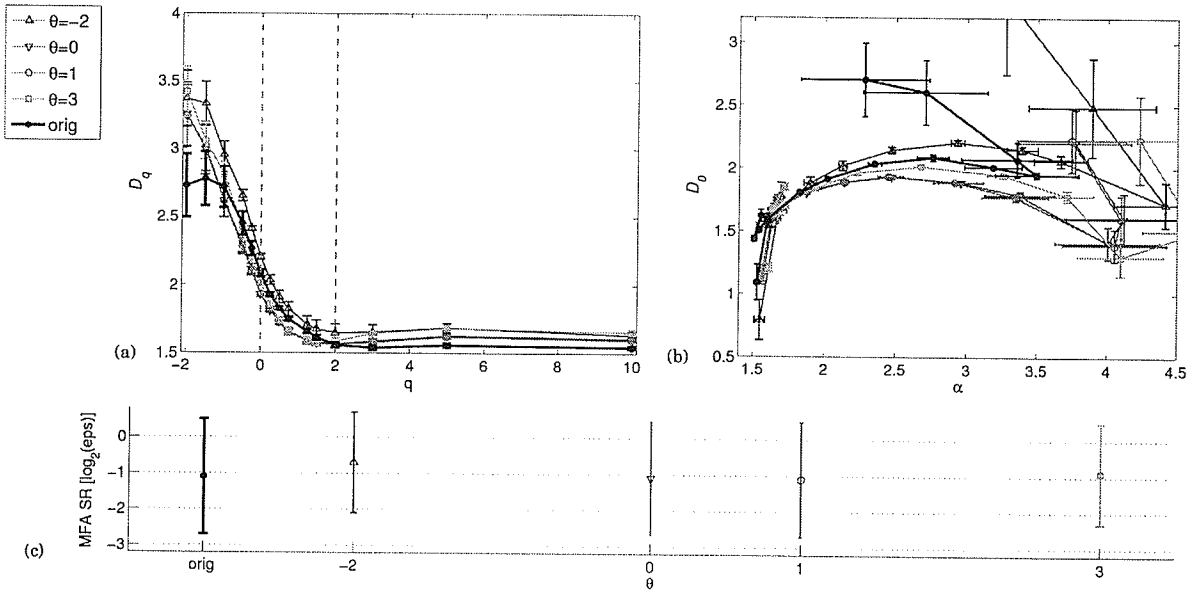


Fig. 7.23 MFA spectra for dataset 1 at significant demixing angles θ: (a) Rényi spectrum; (b) Mandelbrot spectrum; and (c) scaling interval for the demixing estimates as marked. Actual MFA of original source fetal ECG is also shown (heavy solid). Even though scaling intervals of $\theta = 0, 1°$ correspond to the original scaling region, the MFA curves do not follow the original. The estimate at $\theta = -2°$ applies to a shorter scaling region and does not match the original curve but has consistently larger uncertainty that contributes to a smaller error-norm.

7.6.4 Dataset 1: MFA Details at Significant θ

Several features of the convergence of dataset 1 should be addressed. These include (i) an interpretation for the constant MFA error-norm on the $[-1°,\ 2°]$ interval, (ii) an interpretation for the minimal MFA error-norm at $\theta = -2$, and (iii) an explanation for the lack of convergence. Figure 7.23 shows the MFA spectra from the original source fetal ECG as well as the demixing estimates at $\theta = -2, 0, 1$ and $3°$ (as marked). By inspection of panels (a) and (b), it is clear that none of the MFA spectra derived from the demixing estimates actually agree with original. The $\theta = 0$ and $\theta = 1°$ MFA spectra, however, are overlapping throughout both the RS and MS. They also have matching scaling regions, as shown in Fig. 7.23(c). It is reasonable, therefore, to conclude that the MFA estimates *do* converge to *some limit MFA*: the issue is this

limit MFA is not a match to the original fetal ECG MFA.

Moreover, according to Fig. 7.19, the minimal MFA error is measured at $\theta = -2°$. Looking at Fig. 7.23, however, this is not a match to the original MFA either, and is also not fitted over the same scaling region. Consequently, it should not be considered a "best estimate" of the original MFA; notwithstanding its preferred status as the demixing value with minimum MFA error-norm.

Further analysis of the scaling entropies that defined these spectra, shown in Figs. 7.25 and 7.26, does not demonstrate anything else remarkable. In particular, the quality of the entropy scaling of the original, and the $\theta = 0$ and $\theta = 1°$ demixing estimates seem equivalent. The quality of the fit on the other demixing estimates is reduced, but is not particularly out of the ordinary for a short scaling region.

In fact, there is one explanation for all three elements observed in the convergence profile of dataset 1: mismatch in embedding dimension. All the demixed estimates in the convergence limit had an N_e evaluated as 9, whereas the original was determined at 10. Note also, that the segment of minimal MFA error-norm occurs on $\theta \in [-3°, -2°]$ which corresponds to the two demixing estimates assigned $N_e = 10$ — the same as the original. To verify that this was the cause of the observed convergence, a second analysis was performed whereby the embedding dimension of the original fetal ECG source was (artificially) set to 9 to be consistent with the demixing estimates. (This involved only a recalculation of one correlation partition.) The convergence profile comparing the MFA to this modified reference is shown (dashed) against the original dataset 1 profile (solid) in Fig. 7.24. The modification to the embedding dimension solves the lack of convergence and makes its profile qualitatively equivalent to the profile of dataset 2.

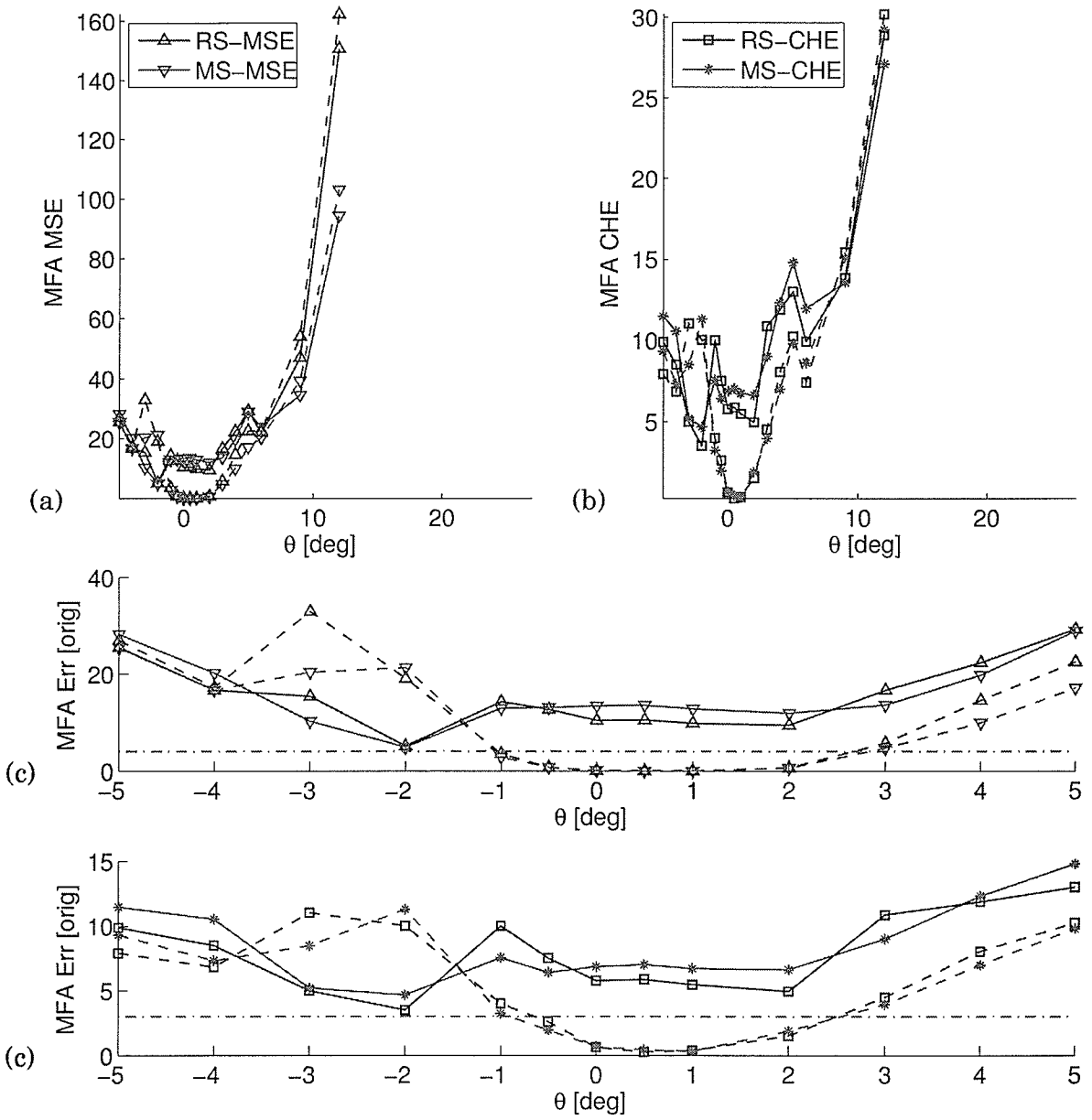


Fig. 7.24 Convergence of MFA error SQMs for dataset 1 against the original source reference (solid) and the modified reference with $N_e = 9$ (dashed): (a) MSE-based MFA error-norm applied to both the RS and MS representations (as marked); (b) as in panel (a), but using the CHE-based MFA error-norm; Panels (c) and (d) amplify the convergence limit for the MSE-norm SQM and CHE-norm SQM respectively. Tolerance levels described in Sec. 7.2.3 are plotted as horizontal lines. The modified dataset converges with a tolerance of $\Delta\theta < 1.5°$.

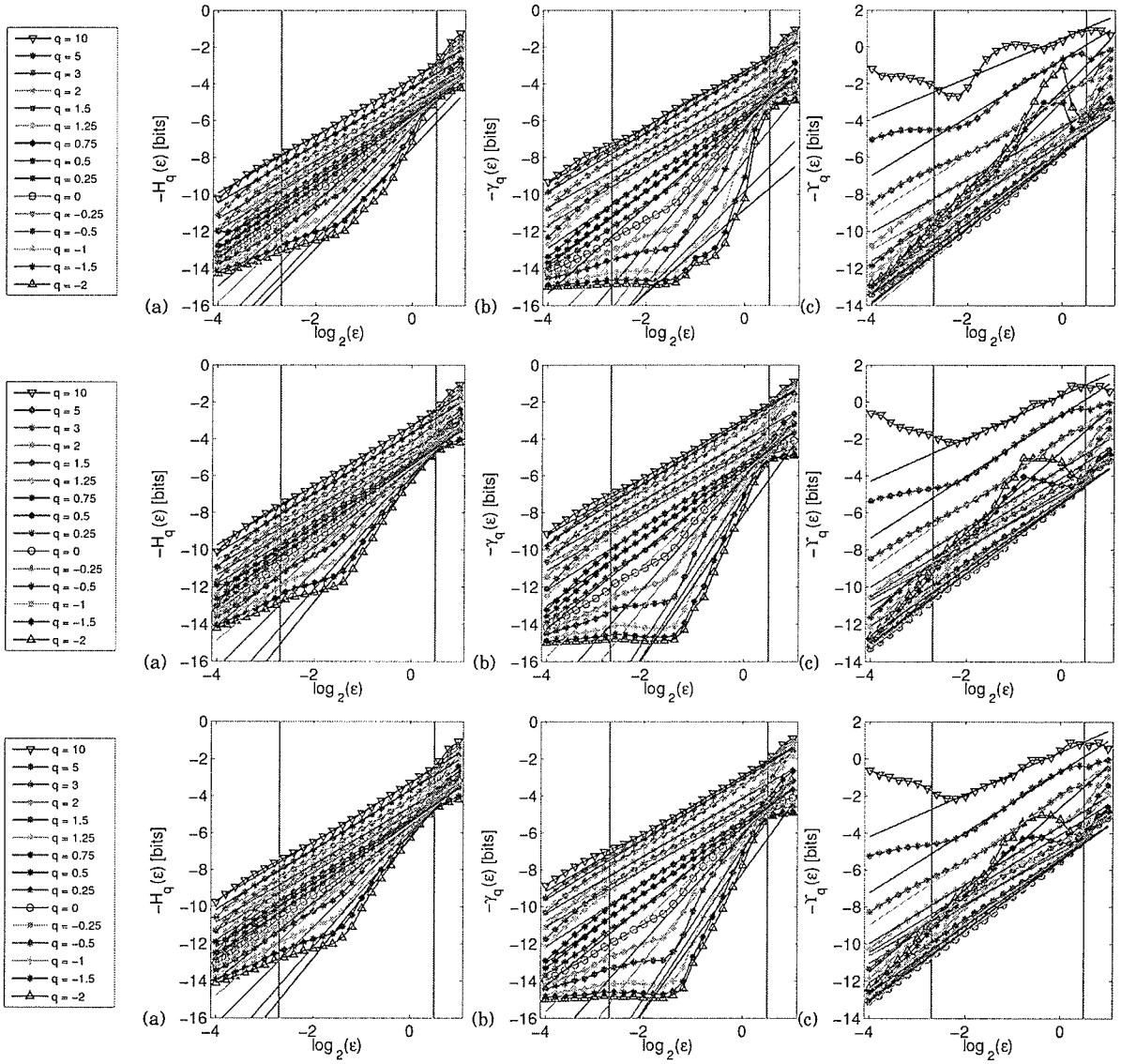


Fig. 7.25 Scaling entropies of dataset 1 at significant demixing levels: Simultaneous scaling of (a) Rényi, (b) Hölder, and (c) Mandelbrot entropies of the original dataset 1 fECG attractor (top row), and its reconstruction at a mixing of $\theta = 0°$ (middle row) and optimal $\theta = 1°$ (bottom row).

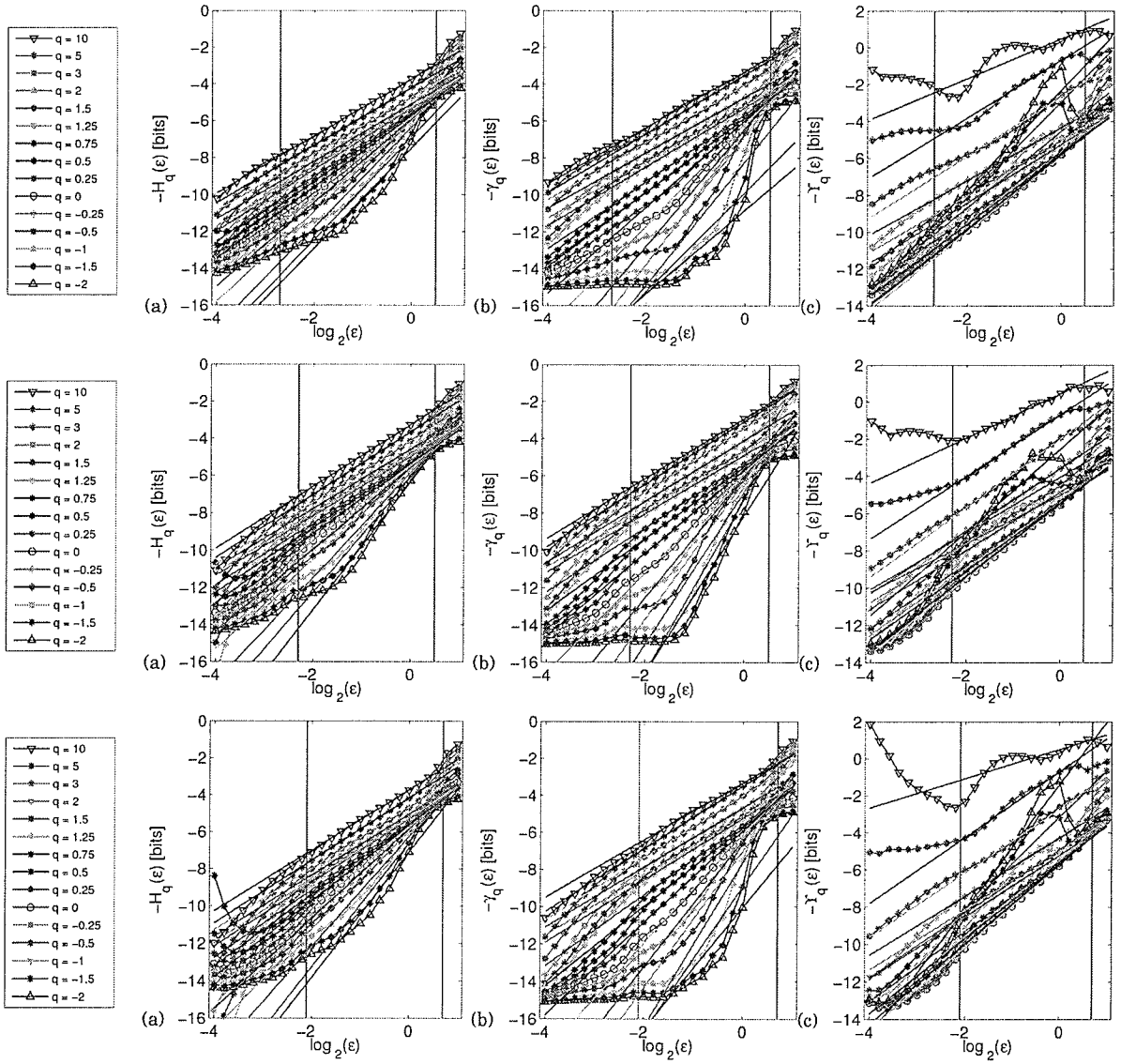


Fig. 7.26 Scaling entropies of dataset 1 at significant demixing levels: Simultaneous scaling of (a) Rényi, (b) Hölder, and (c) Mandelbrot entropies of the original dataset 1 fECG attractor (top row), and its reconstruction at a mixing of $\theta = 3°$ (middle row) and $\theta = -2°$ (bottom row).

7.7 Discussion

7.7.1 MFA Feature Convergence

MFA convergence requires a more complicated discussion than the other A-Class SQMs because it represents a vector feature and not a scalar (as CCI, SNR, kurtosis, and entropy are). As such, it is more difficult to express the accuracy of the feature in concrete numerical terms (*e.g.*, as a percent-error). The MFA error-norms that were proposed, however, did seem successful at guiding insight into the convergence of MFA. Furthermore, the estimates coming from the RS and the MS spectra are consistent. It is easier, although, to asses MFA convergence graphically using the Rényi spectrum because (i) the support is common among the estimates, and (ii) uncertainties are applied to only one axis.

These experiments helped identify a relationship of MFA convergence to the other features of interest. Specifically, MFA convergence can be interpreted as a heightened form of entropy convergence. MFA is a scaling analysis of the entropies of the attractor distribution. This work has demonstrated that as the *entropy of the time-integrated distribution* of the fetal ECG times series converges under demixing (*i.e.*, the ADE), the *distribution on the fetal ECG attractor* also converges under demixing (*i.e.*, the MFA spectra). A precise comparison of the convergence dependencies profiles is hampered by the intermediary step of the (somewhat arbitrary) MFA error-norms. The basic interpretation seems to apply, however, that the MFA and entropy estimation share the properties that: (i) the optimum demixing was sensitive to the initial source correlation and the effect of sphering; and (ii) a tight interval of convergence around the optimum demixing angle is required for accurate estimation. Item (i) indicates that ICA convergence from the *joint system perspective* alone may contribute to bias in the feature estimates from demixed signals. Item (ii), furthermore, indicates that estimation of MFA or entropy from a superlinear-class of ICA algorithm may also be subject to higher variance.

Several unique elements of MFA convergence over basic entropy convergence can also be observed. It is possible that the *scaling nature* of MFA on the attractor entropy adds extra sensitivity under moderate mixing because the MFA error-norms grow very quickly outside the convergence limit. The convergence of MFA overall, however, is distinguished by the sensitivity of the *embedding dimension* to mixing and its consequent effects in the MFA estimate. That is, the estimation of the attractor parameters themselves — which define the attractor support and precede the entropies of the attractor — are the significant contributor of nonlinear effects in the feature convergence of MFA.

7.7.2 Limitations and Generalization

In order for conclusions from this work to generalize to the broader context of ICA application to fetal ECG separation, several limitations of this work must be addressed. These can be considered in two different categories as discussed below.

7.7.2.1 Fetal ECG Simulation

First, the fetal ECG simulation in this work has identified that nonzero source cross-correlation can contribute to bias in the estimation of fetal ECG features. Furthermore, this cross-correlation in the sources can originate purely by the random relative phase of the QRS-complexes of the two ECG time series. A specific study of feature convergence with fixed time series (and, consequently, fixed features) at different relative phase should be considered. In particular, the effect of source correlation on the misspecification of embedding dimension is of interest.

Second, the fetal ECG simulation in this work used a single ECG morphology. A repeat study using other ECG morphologies can identify whether the SQM convergence profiles observed here are sensitive to, or consistent with, these changing conditions.

Third, the MFA of the fetal ECG was based on attractor reconstruction with traditional equispaced lag-embedding. Experiments using more advanced embedding approaches should be considered that better represent more efficiently the nonlinear and scaling aspects of the fetal ECG dynamics.

7.7.2.2 General ICA Topics

First, the SQM approach used in this work was simplified by considering only two sources, and consequently a Stiefel manifold parameterized by a single Givens angle. Studies with more sources should be designed to observe whether the SQMs become anisotropic as the Stiefel manifold becomes more complex, or whether scalar relationships can generalize to higher dimensional systems. (*e.g.*, Is it sufficient to consider convergence against a scalar measure such as $\text{MFE}(\mathbf{W}(\theta_1, \ldots, \theta_j))$, or must the entire vector $[\theta_1, \ldots, \theta_j]^{\text{T}}$ be considered?)

Second, the SQM approach in this work identified that a kurtosis estimate is reliable under the "20 dB SNR" rule of thumb (*c.f.*, [88, (2001)] quoted on p. 136), but the entropy and the MFA were not. Further experimentation should be applied to determine if the demixing relationship of SNR changes with increasing complexity of the Stiefel manifold, as above. Specifically, does the convergence of the SNR^* feature degrade or perhaps just drop in absolute value so that the demixing interval of 20 dB shrinks? How does the "20 dB SNR" rule of thumb apply to the more complicated situations?

Third, the SQM approach used in this work has been made in application to fetal ECG separation. Are the relationships observed here (in particular the classification of a sub- and superlinear ICA cost functions) sensitive to the source kurtosis, just their sign, or is it consistent for other statistics?

Fourth, the SQM approach here has considered feature convergence only in noise-free situations. Experiments to analyze the effect of noise on feature convergence is required.

7.8 Summary

This chapter has presented the analysis of feature convergence for two runs in a fetal ECG separation simulation. Using a Stiefel manifold sampling approach to define convergence profiles, two types of ICA cost function convergence (sublinear and superlinear) have been identified. Furthermore, the same two types of convergence for HOS and nonlinear dynamical features were identified. These convergence profiles indicate that kurtosis can be reliably estimated from demixed signals produced by either ICA cost function class, but the entropy and MFA may require the sublinear (MI-based) ICA cost functions in order to achieve reliable estimation. Some features, such as cross-correlation or embedding lag, are affected only mildly by mixing, while others, such as embedding dimension, can behave erratically.

Furthermore, differences in the initial source correlation for the experiments indicate that minor levels of random source correlation can contribute to bias in the feature estimates. the most significant bias is in the entropy, SNR, and MFA (through the embedding dimension), though all feature-based SQMs are affected to some degree. MFE is the most insensitive SQM to the effect of random source correlation as it measures the entire demixing system, and not the features of a single channel.

The MFA error-norms defined in Ch. 4 and applied here also successfully provide insight into MFA convergence. The nonlinear nature of MFA affects convergence through (i) its narrow convergence region and (ii) a sensitivity to embedding dimension. Provided attractor estimates remain embedded in the same embedding space, convergence of MFA is related strongly to the convergence of time-integrated entropy. Sensitivity of the MFA to embedding dimension establishes that separation performance of ICA is subject to some complications when using nonlinear features.

Chapter VIII

CONCLUSIONS

8.1 Conclusions

This work has presented a study of feature convergence for simulated fetal electrocardiograms (ECG) under the action of independent component analysis (ICA) processing. Specifically, the convergence of multifractal spectra characterizing the attractors reconstructed from the fetal ECG estimates was compared to the convergence of second and higher-order statistics using a Stiefel manifold sampling approach to define convergence profiles. Results from the study indicate that

(a) ICA cost function converge in at least two classes: superlinear (Xinfomax and FastICA) and sublinear (RADICAL and MILCA);

(b) higher-order statistics (HOS) converge in at least two classes; kurtosis converges superlinearly, while entropy converges sublinearly;

(c) the convergence of the multifractal analysis (MFA) spectra can be intuitively estimated by MFA error-norms;

(d) the rate of MFA convergence is closer to entropy than kurtosis;

(e) the estimation of the attractor embedding dimension is subject to nonlinear effects and subjective measurement: namely, the estimate of N_e in the limit of weak mixing is not necessarily the same as the original source;

(f) the fetal ECG MFA spectra (and, consequently, MFA convergence) are sensitive to the embedding dimension.

These results confirm the original thesis statement that (i) not all ICA cost functions converge at the same rate, and (ii) not all higher-order statistics converge at the same rate. Furthermore, the rates of the two ICA classes effectively match the two classes of HOS convergence. Consequently, the variance of the sublinear HOS (namely entropy) is expected to be higher when demixing is provided by the superlinear ICA cost functions as its convergence lags behind. Kurtosis estimates should be reliable under the convergence of either cost function class.

These results also confirm the original thesis statement that (iii) nonlinear dynamical features extracted from ICA estimates may not converge to the features of the original independent sources because of nonlinear sensitivities; and (iv) traditional bounds using second-order performance metrics fail to account for sensitivities in some higher-order and dynamical feature extraction. The embedding dimension was found to be a significant nonlinear parameter that may be consistently estimated in the demixing convergence limit, yet still be inaccurate to the value estimated for the original source. Since the MFA spectra are sensitive to the embedding dimension, then, they may converge on demixing to improper spectra. In particular, this demonstrates that high values of SNR do not necessarily indicate that the MFA of the estimate is an accurate representation of the original. Consequently, any SOS-based performance rule (*e.g.*, "20 dB is satisfactory") is subject to limited interpretation.

This work has also demonstrated preliminary evidence that random phase correlations in the sources of the fetal ECG separation problem may affect the performance of ICA estimation. Specifically, cross-correlation between the sources may be significant enough to prevent a *sphered* ICA estimate (*i.e.*, without correlations) to properly estimate both sources simultaneously. Experiments indicate that this can introduce bias into features extracted from the

optimized demixing estimates. This bias was most clearly identified in the entropy, but may also play a role in the changed estimation of the embedding dimension.

Moreover, the novel simulation of the fetal ECG separation problem developed here represents a sound beginning for the analysis of ICA separation in controlled experiments that bridge the divide between "theory" and "practice". Noise-free ECG signals were synthesized with morphology, heart rate variability, and dynamical features representative of natural signals. Thus the generic features considered for analysis are suitable representatives for the features in practical situations, unlike the traditional disconnect contained in ICA studies in the literature.

8.2 Contributions

This work represents the first *feature-based* analysis of ICA performance (*i.e.*, where signal "quality" is measured generically in a sense beyond simple matrix or power measures). This novelty is tandem with the introduction of the new ICA error-space approach that bridges "theory" and "practice". This work also clearly demonstrates the complexity involved in maintaining an experimental paradigm suitable to connect the two "ends" of engineering. This work also presents the analyzes of multifractal (dynamical scaling) feature convergence as a new paradigm for ICA performance measurement.

Four contributions to science and technology are made with this work. First, the convergence profile methodology for ICA analysis is proposed and demonstrated. Since ICA performance analysis has been restricted in the literature to either (i) power-based, or matrix-based methods in toy examples, or (ii) subjective analysis in blind situations, this new approach provokes a new discussion of what *separation quality* means in theory and practice. In particular, this work has raised concerns over a non-uniform bias in feature estimation occurring from

random phase correlations in ECG signals.

Second, the comparative analysis of ICA SQM-convergence establishes the first relationship of multifractal feature convergence to other ICA convergence measures. Consequently, this work has identified specific concerns regarding the variance in the estimation of entropy or MFA from systems separated by FastICA or Xinfomax.

Third, the direct $f(\alpha)$ correlation-integral technique is defined, providing an extension to the canonical theory of multifractal analysis. In particular, the utility of having three simultaneous entropies for the selection of an MFA scaling interval has been demonstrated.

Last, the ECGfm surrogate ECG method for ECG synthesis improves the dynamical fidelity of the state of the art Oxford ECG dynamical model.

This work also contributes several resources to the research community by

(a) publishing code for the generation of surrogate ECG;

(b) publishing code for the calculation of an attractor MFA from a correlation partition;

(c) synthesizing a survey on the fetal ECG separation problem;

(d) synthesizing a historical and geometrical background on ICA; and

(e) synthesizing a background on multifractal analysis from a measure-theoretic perspective (that applies to attractor or wavelet methodologies), and the significance of surrogate data for the validation of multifractal analysis.

8.3 Limitations and Future Work

This work has restricted its focus to an analysis of feature convergence in a two-source simulation of the fetal ECG separation problem. In order for conclusions from this work to

generalize to the broader context of ICA application to fetal ECG separation, several limitations of this work must be addressed. These can be considered in two different categories as discussed below.

8.3.1 Fetal ECG Simulation

First, the fetal ECG simulation in this work has identified that nonzero source cross-correlation can contribute to bias in the estimation of fetal ECG features. Furthermore, this cross-correlation in the sources can originate purely by the random relative phase of the QRS-complexes of the two ECG time series. A specific study of feature convergence with fixed time series (and, consequently, fixed features) at different relative phase should be considered. In particular, the effect of source correlation on the misspecification of embedding dimension is of interest.

Second, the fetal ECG simulation in this work used a single ECG morphology. A repeat study using other ECG morphologies can identify whether the SQM convergence profiles observed here are sensitive to, or consistent with, these changing conditions.

Third, the MFA of the fetal ECG was based on attractor reconstruction with traditional equispaced lag-embedding. Experiments using more advanced embedding approaches should be considered that represent the nonlinear and scaling aspects of the fetal ECG dynamics more efficiently. In particular, the effect of the limits on data sampling and record length on practical signals is of interest. Advanced embedding methods should be considered in the hope that they may alleviate some of these restrictions and make the MFA of ECG more practical.

Fourth, it has been conjectured that entropy estimates of fetal ECG signals should have a higher variance under superlinear ICA demixing (*i.e.*, Xinfomax and FastICA) than sublinear (*i.e.*, RADICAL and MILCA). This behaviour should be confirmed by a study with the practical ICA optimizations (*i.e.*, complete algorithms and not simply convergence profiles).

Fifth, the fetal ECG event-intervals used in this work were created by an *ad hoc* process. It is recommended that a better source of fetal ECG event-intervals may improve on the practical MFA modelling of the fetal ECG attractors. Other recommendations for an improved ECG model are included in App. A.

8.3.2 General ICA Topics

First, the SQM approach used in this work was simplified by considering only two sources, and consequently a Stiefel manifold parameterized by a single Givens angle. Studies with more sources should be designed to observe whether the SQMs become anisotropic as the Stiefel manifold becomes more complex, or whether scalar relationships can generalize to higher dimensional systems. (*e.g.*, Is it sufficient to consider convergence against a scalar measure such as $\mathrm{MFE}(\mathbf{W}(\theta_1, \ldots, \theta_j))$, or must the entire vector $[\theta_1, \ldots, \theta_j]^{\mathrm{T}}$ be considered?)

Second, the SQM approach in this work identified that a kurtosis estimate is reliable under the "20 dB SNR" rule of thumb (*c.f.*, Principe *et al.* [88, (2001)] quoted on p. 136), but the entropy and the MFA were not. Further experimentation should be applied to determine if the demixing relationship of SNR changes with increasing complexity of the Stiefel manifold, as above. Specifically, does the convergence of the SNR^* feature degrade or perhaps just drop in absolute value so that the demixing interval of 20 dB shrinks? How does the "20 dB SNR" rule of thumb apply to the more complicated situations?

Third, the SQM approach used in this work has been made in application to fetal ECG separation. Are the relationships observed here (in particular the classification of a sub- and superlinear ICA cost functions) sensitive to the source kurtosis, just their sign, or is it consistent for other statistics?

Fourth, the SQM approach here has considered feature convergence only in noise-free situations. Experiments to analyze the effect of noise on feature convergence is required.

Last, the SQM approach here used the ICA cost functions as B-Class SQMs, but were evaluated under offline conditions so normalization could be applied. Can other B-Class be derived (*e.g.*, from the derivatives of ICA cost functions) which measure convergence blindly but also locally (*i.e.*, well-defined in an online optimization)?

REFERENCES

[1] (2004, May) Physiobank: Physiologic signal archives for biomedical research. [Online]. Available as of Jan. 15, 2006: www.physionet.org/physiobank

[2] S. S. Abeysekera, U. R. Abeyratne, and S. B. Goh, "Spectral information changes in obtaining heart rate variability from tachometer R-R interval signals," *Critical Reviews in Biomed Eng.*, vol. 28, no. 1/2, pp. 149–155, 2000.

[3] P. S. Addison, *Fractals and Chaos.* Philadelphia, PA, USA: Institute of Physics, 1997, 256 pp.

[4] R. Almeida, S. Gouveia, A. P. Rocha, E. Pueyo, J. P. Martinez, and P. Laguna, "QT variability and HRV interactions in ECG: Quantification and reliability," *IEEE Trans Biomed Eng*, vol. 53, no. 7, pp. 1317–1329, July 2006.

[5] S.-I. Amari, "Natural gradient works efficiently in learning," *Neural Computation*, vol. 10, no. 2, pp. 251 – 276, Feb. 1998.

[6] S.-I. Amari and A. Cichocki, "Adaptive blind signal processing: Neural network approaches," *Proceedings IEEE*, vol. 86, no. 10, pp. 2026–2048, Oct. 1998.

[7] S.-I. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," in *Advances in Neural Information Processing Systems*, D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, Eds., vol. 8. Cambridge, MA, USA: The MIT Press, 1996, pp. 757–763.

[8] S.-I. Amari and S. Douglas, "Why natural gradient?" in *Proc. 1998 IEEE Int. Conf. Acoustics Speech Signal Proc, ICASSP98*, vol. 2. (May 12-15; Seattle, WA, USA): IEEE Press, 1998, pp. 1213–1216, (ISBN: 0-7803-4428-6/98).

[9] K. V. K. Ananthanag and J. S. Sahambi, "Investigation of blind source separation methods for extraction of fetal ECG," in *Proc. 2003 Cdn. Conf. on Elec. Comp. Eng, CCECE03.* (May, 2003; Montreal, PQ, Canada): IEEE, 2003.

[10] A. Arneodo, B. Audit, E. Bacry, S. Manneville, J.-F. Muzy, and S. Roux, "Thermodynamics of fractal signals based on wavelet analysis: application to fully developed turbulence data and dna sequences," *Physica A*, vol. 254, pp. 24–45, 1998.

[11] K. T. Assaleh and H. A. Al-Nashash, "Non-iterative method for two-lead fetal ECG extraction," in *Proc. 2004 IEEE Int. Conf. EMBS, EMBC04*, vol. 2, 2004, pp. 929–932.

[12] K. Assaleh, "Extraction of fetal electrocardiogram using adaptive neuro-fuzzy inference systems," *IEEE Trans Biomed Eng*, vol. 54, no. 1, pp. 59–68, Jan. 2007.

[13] K. Assaleh and H. Al-Nashash, "A novel technique for the extraction of fetal ECG using polynomial networks," *IEEE Trans Biomed Eng*, vol. 52, no. 6, pp. 1148–1152, June 2005.

[14] M. Babaie-Zadeh, C. Jutten, and K. Nayebi, "Differential of the mutual information," *IEEE Signal Processing Letters*, vol. 11, no. 1, pp. 48–51, Jan. 2004.

[15] F. R. Bach and M. I. Jordan, "Kernel independent component analysis," *J of Machine Learning Research*, vol. 3, pp. 1–48, July 2002.

[16] E. Bacry, J.-F. Muzy, and A. Arneodo, "Singularity spectrum of fractal signals: Exact results," *J Stat Physics*, vol. 70, no. 3/4, pp. 635–674, 1993.

[17] H. Barlow, "Possible principles underlying the transformation of sensory messages," *Sensory Communication*, pp. 217–234, 1961.

[18] A. K. Barros, R. Vigário, V. Jousmäki, and N. Ohnishi, "Extraction of event-related signals from multichannel bioelectrical measurements," *IEEE Trans Biomedical Engineering*, vol. 47, no. 5, pp. 583–588, May 2000.

[19] V. N. Batchvarov, A. Ghuran, P. Smetana, K. Hnatkova, M. Harries, P. Dilaveris, A. J. Camm, and M. Malik, "QT-RR relationship in healthy subjects exhibits substantial intersubject variability and high intrasubject stability," *Am J Physiol Heart Circ Physiol*, vol. 282, pp. 2356–2363, 2002.

[20] J. Beirlant, L. Györfi, and E. van der Meulen, "Nonparametric entropy estimation: An overview," *Int J Math Stat Sci*, vol. 6, no. 1, pp. 17–39, 1997.

[21] A. J. Bell and T. J. Sejnowski, "An information-maximisation approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129–1159, 1995.

[22] J. J. Benedetto and M. W. Frazier, Eds., *Wavelets: mathematics and applications*. Boca Raton, Florida: CRC Press, 1994, 575 pp.

[23] J.-F. Bercher and C. Vignat, "Estimating the entropy of a signal with applications," *IEEE Trans. Sig. Proc.*, vol. 48, no. 6, pp. 1687–1694, June 2000.

[24] R. D. Berger, S. Akselrod, D. Gordon, and R. J. Cohen, "An efficient algorithm for spectral analysis of heart rate variability," *IEEE Trans Biomed Eng*, vol. 33, no. 9, pp. 900–904, Sept. 1986.

[25] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, UK: Oxford University Press, 1995, 500 pp., (ISBN: 0-19-853864-2).

[26] S. A. Briller, D. B. Geselowitz, S. D. Arlinger, G. K. Danielson, D. Jaron, and C. R. Joyner, "The electrical interaction between artificial pacemakers and patients, with applications to electrocardiography," *American Heart Journal*, vol. 71, no. 5, pp. 656–665, May 1966.

[27] D. H. Brooks and R. S. MacLeod, "Electrical imaging of the heart," *IEEE Signal Processing Mag.*, vol. 14, no. 1, pp. 24–42, Jan. 1997.

[28] J. Butcher, "Runge-Kutta methods," *Scholarpedia*, vol. 2, no. 9, 2007, 3147. [Online]. Available: http://www.scholarpedia.org/article/Runge-Kutta_methods

[29] V. D. Calhoun, T. Adali, L. K. Hansen, J. Larsen, and J. J. Pekar, "ICA of functional MRI data: An overview," in *Proc. 4th Int. Symp. on Independent Component Analysis and Blind Signal Separation (ICA2003)*, (April 2003, Nara, Japan), 2003, pp. 281–288.

[30] D. Callaerts, W. Sansen, J. Vandewalle, G. Vantrappen, and J. Janssens, "Description of a real-time system to extract the fetal electrocardiogram," *Clin. Phys. Physiol. Meas.*, vol. 10, no. Suppl. B, pp. 7–10, Nov. 1989. [Online]. Available as of Jan. 15, 2006: http://www.iop.org/EJ/abstract/-search=11346796.1/0143-0815/10/4B/001

[31] M. Capinski and E. Kopp, *Measure, Integral and Probability*, ser. Springer Undergraduate Mathematics Series. New York, NY, USA: Springer-Verlag, 1999, 227 pp., (ISBN: 3-540-76260-4).

[32] J.-F. Cardoso and B. Laheld, "Equivariant adaptive source separation," *IEEE Trans Sig Proc*, vol. 44, no. 12, pp. 3017–3030, Dec. 1996.

[33] J.-F. Cardoso, "Performance and implementation of invariant source separation algorithms," in *Proc. 1996 IEEE Int Symp Circuits and Systems, ISCAS '96*, vol. 2. (May 12-15, 1996): IEEE Press, 1996, pp. 85–88.

[34] J.-F. Cardoso, "Infomax and maximum likelihood for blind source separation," *IEEE Signal Processing Letters*, vol. 4, no. 4, pp. 112–114, Apr. 1997.

[35] J.-F. Cardoso, "Blind signal separation: Statistical principles," *Proceedings IEEE*, vol. 86, no. 10, pp. 2009–2025, Oct. 1998.

[36] J.-F. Cardoso, "Learning in manifolds: The case of source separation," in *Proc. Ninth IEEE SP Workshop on Statistical Signal and Array Processing*. (Sep. 14-16, 1998): IEEE Press, 1998, pp. 136–139.

[37] J.-F. Cardoso and P. Comon, "Independent component analysis, a survey of some algebraic methods," in *Proc. 1996 IEEE Int Symp Circuits and Systems, ISCAS '96*, vol. 2. (May 12-15, 1996): IEEE Press, 1996, pp. 93–96.

[38] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals," *IEE Proceedings-F (Radar and Signal Processing)*, vol. 140, no. 6, pp. 362–370, Dec. 1993.

[39] A. B. Chhabra and R. V. Jensen, "Direct determination of the $f(\alpha)$ singularity spectrum," *Physical Review Letters*, vol. 62, no. 12, pp. 1327–1330, Mar. 1989.

[40] A. Cichocki and S.-I. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. West Sussex, UK: John Wiley and Sons, 2002, 554 pp., (ISBN: 0-471-60791-6), with CD-rom.

[41] G. D. Clifford. ecgsyn24.m. [Online]. Available as of July 10, 2007: http://www.mit.edu/~gari/CODE/ECGSYN/MATLAB/ecgsyn24/

[42] G. D. Clifford and P. E. McSharry, "A comparison of nonlinear noise reduction and independent component analysis using a realstic dynamical model of the electrocardiogram," *Proc of SPIE*, vol. 5467, pp. 78–88, 2004.

[43] G. D. Clifford and P. E. McSharry, "A realistic coupled nonlinear artificial ECG, BP and respiratory signal generator for assessing noise performance of biomedical signal processing algorithms," *Proc of SPIE*, vol. 5467, pp. 290–301, 2004.

[44] G. D. Clifford and L. Tarassenko, "Quantifying errors in spectral estimates of HRV due to beat replacement and resampling," *IEEE Trans Biomed Eng*, vol. 52, no. 4, pp. 630–638, Apr. 2005.

[45] P. Comon, "Independent component analysis, A new concept?" *Signal Processing (Elsevier)*, vol. 36, no. 3, pp. 287–314, 1994, (Special Issue on Higher-Order Statistics).

[46] J. A. Copel, C. S. Kleinman, and A. H. Friedman, *Fetal Medicine: Basic Science and Clinical Practice*. London, UK: Churchill Livingstone, 1999, ch. Ch. 51: Diagnosis and Management of Fetal Heart Disease, pp. 641–650, (ISBN: 0-4430-5357X).

[47] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 3rd ed., ser. Wiley Series in Telecommunications. New York: John Wiley and Sons, 2001, no. 1.

[48] R. Dansereau and W. Kinsner, "New relative multifractal dimension measures," in *Proc. 2001 IEEE Int. Conf. Acoustics, Speech, and Sig. Proc. ICASSP'01*, vol. 3. (Salt Lake City, Utah, May 7-11, 2001): IEEE Press, 2001, pp. 1741–1744.

[49] G. A. Darbellay and I. Vajda, "Estimation of the information by an adaptive partitioning of the observation space," *IEEE Trans. Inf. Theory*, vol. 45, no. 4, pp. 1315–1320, May 1999.

[50] P. Davey, "A new physiological method for heart rate correction of the QT interval," *Heart*, vol. 82, pp. 183–186, 1999.

[51] L. De Lathauwer, B. De Moor, and J. Vandewalle, "Fetal electrocardiogram extraction by blind source subspace separation," *IEEE Trans Biomed Eng*, vol. 47, no. 5, pp. 567–572, May 2000.

[52] R. W. DeBoer, J. M. Karemaker, and J. Strackee, "Comparing spectra of a series of point events particularly for heart rate variability data," *IEEE Trans Biomed Eng*, vol. 31, no. 4, pp. 384–387, Apr. 1986.

[53] D. L. Donoho and Y. Tsaig, "Recent advances in sparsity-driven signal recovery," in *Proc. 2005 IEEE Int. Conf. Acoustics, Speech and Signal Processing, ICASSP'05*. IEEE Press, 2005, pp. (V–713)–(V–716), (ISBN: 0-7803-8874-7).

[54] S. Douglas, "On the design of gradient algorithms employing orthogonal matrix constraints," in *Proc. 2007 IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP07*, vol. 4, (Honolulu, HW; April 15-20,2007), 2007, pp. IV–1401–IV–1404.

[55] D. Erdogmus, K. E. Hild II, and J. C. Principe, "Blind source separation using Rényi's -marginal entropies," *Neurocomputing*, vol. 49, pp. 25–38, 2002.

[56] K. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*. Chichester, NY, USA: Wiley, 1990, 288 pp., (ISBN: 0471922870).

[57] D. Farina, M.-F. Lucas, and C. Doncarli, "Optimized wavelets for blind separation of nonstationary surface myoelectric signals," *IEEE Trans Biomed Eng*, vol. 55, no. 1, pp. 78–86, Jan. 2008.

[58] E. R. Ferrara and B. Widrow, "Fetal electrocardiogram enhancement by time-sequenced adaptive filtering," *IEEE Trans Biomed Eng*, vol. 29, no. 6, pp. 458–460, June 1982.

[59] S. Fiori, A. Uncini, and F. Piazza, "Neural learning and weight flow on Stiefel manifold," 1998.

[60] F. L. Foresta, N. Mammone, and F. Morabito, "Independent component and wavelet analysis for fECG extraction: The ST waveform evaluation," in *Proc. 2005 IEEE Int. Conf. Comp. Int. for Measure. Sys. and Apps., CIMSA05*, (Giardini Naxos, Italy; July 20–22, 2005), 2005, pp. 86–90.

[61] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Physical Review A*, vol. 33, no. 2, pp. 1134–1140, Feb. 1986.

[62] F. N. Fritsch and R. E. Carlson, "Monotone piecewise cubic interpolation," *SIAM J Numerical Analysis*, vol. 17, no. 2, pp. 238–246, Apr. 1980.

[63] N. Gadhok, "A study of outliers for robust independent component analysis," Master's thesis, University of Manitoba, 2007.

[64] U. Gembruch, "Fetal tachyarrythmia," in *Fetal Cardiology: embryology, genetics, physiology, echocardiographic evaluation, diagnosis and perinatal management of cardiac diseases*, S. Yagel, N. H. Silverman, and U. Gembruch, Eds. London, UK: Martin Dunitz, 2003, ch. 30, pp. 355–372, (ISBN: 1-85317-905-1).

[65] M. A. Girolami and J. Taylor, Eds., *Advances in Independent Component Analysis*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2000, (ISBN: 1852332638).

[66] L. Glass, M. R. Guevara, A. Shrier, and R. Perez, "Bifurcation and chaos in a periodically stimulated cardiac oscillator," *Physica D*, vol. 7, pp. 89–101, 1983.

[67] L. Glass and M. C. Mackey, *From clocks to chaos: the rhythms of life*. Princeton, N.J.: Princeton University Press, 1988, 248 pp., (ISBN: 0691084963).

[68] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, June13 2000, Circulation Electronic Pages: http://circ.ahajournals.org/cgi/content/full/101/23/e215. [Online]. Available: http://www.physionet.org/

[69] A. L. Goldberger, *Clinical Electrocardiography: A Simplified Approach*, 6th ed. St. Louis, MI, USA: Mosby Inc., 1999, 292 pp., (ISBN: 0-323-00252-8).

[70] J.-F. Gouyet, *Physics and Fractal Structures*. New York, NY, USA: Springer-Verlag, 1996, 234 pp., (ISBN: 0-387-94153-3).

[71] P. Grassberger, "Finite sample corrections to entropy and dimension estimates," *Physics Letters A*, vol. 128, no. 6-7, pp. 369–373, Apr.11 1988. [Online]. Available: http://www.sciencedirect.com/science/article/B6TVM-46MDDJ8-1J1/2/80dd9197071751d17d315ae72c82441e

[72] P. Grassberger and I. Procaccia, "Characterization of strange attractors," *Physical Review Letters*, vol. 50, no. 5, pp. 346–349, Jan. 1983.

[73] P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors," *Physica D*, vol. 9, pp. 189–208, 1983.

[74] H. S. Greenside, A. Wolf, J. Swift, and T. Pignataro, "Impracticality of a box-counting algorithm for calculating the dimensionality of strange attractors," *Physical Review A*, vol. 25, no. 6, pp. 3453–3456, June 1982.

[75] T. C. Halsey, M. H. Jensen, L. P. Kadanoff, I. Procaccia, and B. I. Shraiman, "Fractal measures and their singularities: The characterization of strange sets," *Physical Review A*, vol. 33, no. 2, pp. 1141–1151, Feb. 1986.

[76] K. Han, J. H. Nagel, and N. Schneiderman, "A continuous representation of heart rate," in *Proc. 14th IEEE Int. Conf. EMBS, EMBC92*, vol. 2, (Oct. 29-Nov. 1, 1992), 1992, pp. 785–786.

[77] S. Harmeling, F. Meineck, and K.-R. Muller, "Analysing ICA components by injecting noise," in *Proc. 4th Int. Symp. on Independent Component Analysis and Blind Signal Separation ICA2003*, (April 2003; Nara, Japan), 2003, pp. 149–154.

[78] S. Haykin, Ed., *Unsupervised Adaptive Filtering*. New York, NY, USA: Wiley-Interscience, 2000, vol. 1: Blind Source Separation, 446 pp., (ISBN: 0-471-29412-8).

[79] S. Haykin and Z. Chen, "The cocktail party problem," *Neural Computation*, vol. 17, no. 9, pp. 1875–1902, Sept. 2005.

[80] S. Haykin and B. Kosko, Eds., *Intelligent Signal Processing*. Piscataway, NJ, USA: Wiley-IEEE Press, 2001, 608 pp., (ISBN: 0-7803-6010-9).

[81] S. Haykin and B. Van Veen, *Signals and Systems*, 2nd ed. J. Wiley and Sons, Oct. 2002, 820 pp., (ISBN: 0-471-16474-7).

[82] T. He, G. D. Clifford, and L. Tarassenko, "Application of ICA in removing artefacts from the ECG," *Neural Comput. & Applic.*, vol. 15, no. 2, pp. 105–116, 2006.

[83] R. Hegger, H. Kantz, and T. Schreiber, "Practical implementation of nonlinear time series methods: The TISEAN package," *CHAOS*, vol. 9, p. 413, 1999. [Online]. Available: http://www.mpipks-dresden.mpg.de/~tisean

[84] R. Hegger and H. Kantz, "Improved false nearest neighbor method to detect determinism in time series data," *Physical Review E*, vol. 60, no. 4, Oct. 1999.

[85] H. Hentschel and I. Procaccia, "The infinite number of generalized dimensions of fractals and strange attractors," *Physica D*, vol. 8, no. 3, pp. 435–444, 1983.

[86] J. Herault and C. Jutten, "Space or time adaptive signal processing by neural network models," in *Workshop on Neural Networks for Computing*, (Snowbird, Utah; April 13-16, 1986), 1986.

[87] J. Herrmann and F. Theis, "Statistical analysis of sample-size effects in ICA," in *Intelligent Data Engineering and Automated Learning - IDEAL 2007*, H. Y. et. al., Ed., 2007, pp. 416–425. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-77226-2_43

[88] K. E. Hild II, D. Erdogmus, and J. Príncipe, "Blind source separation using Rényi's mutual information," *IEEE Signal Processing Letters*, vol. 8, no. 6, pp. 174–176, June 2001.

[89] J. Himberg and A. Hyvärinen, "Icasso: software for investigating the reliability of ICA estimates by clustering and visualization," in *Proc IEEE 13th Wkshop on Neur Nets for Sig Proc, NNSP03*, (Sept. 17-19, 2003), 2003, pp. 259–268.

[90] B. Huang, "Compression of ECG signals with feature extraction, classification, and browsability," Ph.D. dissertation, Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB, Canada, May 2004.

[91] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans Neural Networks*, vol. 10, no. 3, pp. 626–634, May 1999.

[92] A. Hyvärinen, "Survey on independent component analysis," *Neural Computing Surveys*, vol. 2, pp. 94–128, 1999. [Online]. Available as of Dec. 12, 2005: http://www.cs.helsinki.fi/u/ahyvarin/papers/review.shtml

[93] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York, NY, USA: Wiley-Interscience, 2001, 504 pp.

[94] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, vol. 9, no. 7, pp. 1483–1492, 1997.

[95] IBM. (2001, Oct.) Autonomic computing: IBM's perspective on the state of information technology. [Online]. Available as of Sep. 10, 2005: http://www-03.ibm.com/autonomic/library.shtml

[96] H. Ito and L. Glass, "Theory of reentrant excitation in a ring of cardiac tissue," *Physica D: Nonlinear Phenomena*, vol. 56, no. 1, pp. 84–106, Apr. 1992. [Online]. Available: http://www.sciencedirect.com/science/article/B6TVK-46MM90Y-61/2/2869fd73ed484b7c7e38be7786b5a05b

[97] P. C. Ivanov, L. A. N. Amaral, A. L. Goldberger, S. Havlin, M. G. Rosenblum, Z. R. Struzik, and H. E. Stanley, "Multifractality in human heartbeat dynamics," *Nature*, vol. 399, pp. 461–465, June3 1999.

[98] P. C. Ivanov, M. G. Rosenblum, C.-K. Peng, J. Mietus, S. Havlin, H. E. Stanley, and A. L. Goldberger, "Scaling behaviour of heartbeat intervals obtained by wavelet-based time-series analysis," *Nature*, vol. 383, pp. 323–327, Sept.26 1996.

[99] M. G. Jafari and J. A. Chambers, "Fetal electrocardiogram extraction by sequential source separation in the wavelet domain," *IEEE Trans Biomed Eng*, vol. 52, no. 3, pp. 390–400, Mar. 2005.

[100] C. J. James and C. W. Hesse, "On the use of spectrally constrained ICA applied to single-channel ictal EEG recordings within a dynamical embedding framework," in *Proc. 2005 Int Conf EMBS, EMBS05.* (Sep. 1-4, 2005; Shanghai, China): IEEE, 2005.

[101] C. J. James, K. Kobayashi, and D. Lowe, "Isolating epileptiform discharges in the unaveraged EEG using independent component analysis," in *Proc. IEE Colloq. Medical Appl. Sig. Proc.*, vol. 2. (Oct. 6, 1999): IEE, 1999, pp. 1–6.

[102] K. Judd, "An improved estimator of dimension and some comments on providing confidence intervals," *Physica D: Nonlinear Phenomena*, vol. 56, no. 2-3, pp. 216–228, May 1992. [Online]. Available: http://www.sciencedirect.com/science/article/B6TVK-46MM91H-6C/2/21df09abfc17ad015820d05fc612c85a

[103] K. Judd and A. Mees, "Embedding as a modeling problem," *Physica D: Nonlinear Phenomena*, vol. 120, no. 3-4, pp. 273–286, Sept.15 1998.

[104] T.-P. Jung, S. Makeig, T.-W. Lee, M. J. McKeown, G. Brown, A. J. Bell, and T. J. Sejnowski, "Independent component analysis of biomedical signals," in *2nd Int. Workshop on Independent Component Analysis and Signal Separation*, (Helsinki, Finland, June 19-22, 2000), 2000, pp. 633–644.

[105] C. Jutten and J. Herault, "Blind separation of sources, part I: An adaptive algorithm based on neuromimatic architecture," *Signal Processing (Elsevier)*, vol. 24, no. 1, pp. 1–10, July 1991.

[106] P. P. Kanjilal, S. Palit, and G. Saha, "Fetal ECG extraction from single-channel maternal ECG using singular value decomposition," *IEEE Trans Biomed Eng*, vol. 44, no. 1, pp. 51–59, Jan. 1997.

[107] H. Kantz and T. Schreiber, "Human ECG: Nonlinear deterministic versus stochastic aspects," *IEE Proc.-Sci. Meas. Techno*, vol. 145, no. 6, pp. 279–284, Nov. 1998.

[108] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, ser. Cambridge Nonlinear Science Series. Cambridge, UK: Cambridge University Press, 1997, no. 7, 304 pp., (ISBN: 0-521-55144-7).

[109] D. T. Kaplan, "Simultaneous QRS detection and feature extraction using simple matched filter basis functions," in *Proc. 1990 Computers in Cardiology Conf.* (Chicago, IL; Sep. 23-26, 1990): IEEE, 1990, pp. 503–506, (ISBN: 0-8186-2225-3).

[110] D. T. Kaplan, "The analysis of variability," *J Cardiovascular Electrophysiology*, vol. 5, no. 1, pp. 16–19, Jan. 1994.

[111] J. Karhunen, A. Hyvärinen, R. Vigário, J. Hurri, and E. Oja, "Applications of neural blind separation to signal and image processing," in *Proc. 1997 IEEE Int. Conf. Acoustics Speech Sig. Proc. ICASSP97.* IEEE Press, 1997, pp. 131–134.

[112] G. Kember and A. Fowler, "Random sampling and the Grassberger-Procaccia algorithm," *Physics Letters A*, vol. 161, pp. 429–432, Jan. 13 1992.

[113] M. B. Kennel and H. D. I. Abarbanel, "False neighbors and false strands: A reliable minimum embedding dimension algorithm," *Phys. Rev. E*, vol. 66, no. 2, p. 026209, Aug 2002.

[114] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Phys. Rev. A*, vol. 45, no. 6, pp. 3403–3411, Mar 1992.

[115] A. Khamene and S. Negahdaripour, "A new method for the extraction of fetal ECG from the composite abdominal signal," *IEEE Trans Biomed Eng*, vol. 47, no. 4, pp. 507–516, Apr. 2000.

[116] W. Kinsner, "Entropy-based fractal dimensions: Probability and pair-correlation algorithms for E-dimensional images and strange attractors," University of Manitoba, Winnipeg, MB, Canada R3T 5V6, Tech. Rep. 94-5, June 1994, 44 pp.

[117] W. Kinsner, "Fractal dimensions: Morphological, entropy, spectrum, and variance classes," University of Manitoba, Winnipeg, MB, Canada R3T 5V6, Tech. Rep. 94-4, May 1994, 146 pp.

[118] W. Kinsner, "Fractional Brownian noise and the variance fractal dimension," University of Manitoba, Winnipeg, MB, Canada R3T 5V6, Tech. Rep. 94-3, Apr. 1994, 86 pp.

[119] W. Kinsner, "Characterizing chaos through Lyapunov metrics," in *Proc. Second IEEE International Conference on Cognitive Informatics ICCI'03*, 2003.

[120] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Physical Review E*, vol. 69, no. 6, pp. 066 138–1–066 138–16, 2004.

[121] D. Kugiumtzis, "Test your surrogate data before you test for nonlinearity," *Physical Review E*, vol. 60, no. 3, pp. 2808–2816, Sept. 1999.

[122] P. Laguna, R. G. Mark, A. Goldberger, and G. B. Moody, "A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG," in *Proc. 1997 Computers in Cardiology Conf.*, vol. 24. (Lund; Sep. 7-10, 1997): IEEE Computer Society Press, 1997, pp. 673–676.

[123] E. G. Learned-Miller and J. W. Fisher III, "ICA using spacings estimates of entropy," *J of Machine Learning Research*, vol. 4, pp. 1271–1295, Dec. 2003.

[124] T.-W. Lee, *Independent Component Analysis, Theory and Applications*, 1st ed. Boston, Mass USA: Kluwer Academic Publishers, 1998, (ISBN: 0-7923-8261-7).

[125] T.-W. Lee, M. Girolami, and T. J. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources," *Neural Computation*, vol. 11, no. 2, pp. 417–441, Feb. 1999.

[126] T.-W. Lee and M. S. Lewicki, "Unsupervised image classification, segmentation, and enhancement using ICA mixture models," *IEEE Trans Image Processing*, vol. 11, no. 3, pp. 270–279, Mar. 2002.

[127] R. Linsker, "Local synaptic learning rules suffice to maximize mutual information in a linear network," *Neural Computation*, vol. 4, pp. 691–702, 1992.

[128] R. Liston, J. Crane, and S. W. Group, "SOGC clinical practice guidelines - fetal health surveillance in labour - part 1," *J Ob Gyn Canada*, vol. 112, pp. 1–13, Mar. 2002. [Online]. Available: http://www.sogc.org/guidelines/index_e.asp#Obstetrics

[129] S. Mallat, *A Wavelet Tour of Signal Processing*. San Diego, CA, USA: Academic Press, 1998, 577 pp., (ISBN: 0124666051).

[130] S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 617–643, Mar. 1992.

[131] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 7, pp. 710–732, July 1992.

[132] B. Mandelbrot, *Multifractals and 1/f Noise.* New York, NY, USA: Springer-Verlag, 1999, 442 pp., (ISBN: 0-387-98539-5).

[133] B. B. Mandelbrot, *Les Objets Fractals: Forme, Hasard et Dimension.* Paris: Flammarion, 1975.

[134] B. B. Mandelbrot, C. J. Evertsz, and Y. Hayakawa, "Exactly self-similar left-sided multifractal measures," *Physical Review A*, vol. 42, no. 8, pp. 4528–4536, Oct. 1990.

[135] A. Mansour and C. Jutten, "What should we say about the kurtosis?" *Signal Processing Letters, IEEE*, vol. 6, no. 12, pp. 321–322, Dec. 1999.

[136] D. E. Marossero, D. Erdogmus, N. Euliano, J. C. Principe, and K. E. Hild II, "Independent component analysis for fetal electrocradiogram extraction: A case for the data efficient Mermaid algorithm," in *Proc. 2003 IEEE XIII Wkshp Neur. Net. Sig. Proc., NNSP03*, (Toulouse, France; Sept. 7-19, 2003), 2003, pp. 399–408.

[137] J. P. Martinez, R. Almeida, S. Olmos, A. P. Rocha, and P. Laguna, "A wavelet-based ECG delineator: Evaluation on standard databases," *IEEE Trans Biomed Eng*, vol. 51, no. 4, pp. 570–581, Apr. 2004.

[138] J. Mateo and P. Laguna, "Improved heart rate variability signal analysis from the beat occurrence times according to the IPFM model," *IEEE Trans Biomed Eng*, vol. 47, no. 8, pp. 985–996, Aug. 2000.

[139] J. Mateo and P. Laguna, "Analysis of heart rate variability in the presence of ectopic beats using the heart timing signal," *IEEE Trans Biomed Eng*, vol. 50, no. 3, pp. 334–343, Mar. 2003.

[140] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith, "A dynamical model for generating synthetic electrocardiogram signals," *IEEE Trans Biomed Eng*, vol. 50, no. 3, pp. 289–294, Mar. 2003.

[141] F. Meinecke, A. Ziehe, M. Kawanabe, and K.-R. Muller, "A resampling approach to estimate the stability of one-dimensional or multidimensional independent components," *IEEE Trans Biomedical Engineering*, vol. 49, no. 12, pp. 1514–1525, Dec. 2002.

[142] C. A. Menihan, *Electronic Fetal Monitoring: Concepts and Applications.* Philadelphia, USA: Lippincott, 2001, 267 pp., (ISBN: 0781726263).

[143] M. Merri, D. C. Farden, J. G. Mottley, and E. L. Titlebaum, "Sampling frequency of the electrocardiogram for spectral analysis of the heart rate variability," *IEEE Trans Biomed Eng*, vol. 37, no. 1, pp. 99–106, Jan. 1990.

[144] A. Meyer-Baese, A. Wismueller, and O. Lange, "Comparison of two exploratory data analysis methods for fMRI: Unsupervised clustering versus independent component analysis," *IEEE Trans Information Technology in Biomedicine*, vol. 8, no. 3, pp. 387–398, Sept. 2004.

[145] M. Mihoko and S. Eguchi, "Robust blind source separation by beta divergence," *Neural Computation*, vol. 14, no. 8, pp. 1859–1886, 2002.

[146] E. G. Miller, "A new class of entropy estimators for multi-dimensional densities," in *Proc. 2003 Int Conf Acoustics Speech Sig Proc ICASSP03*. IEEE, 2003.

[147] F. Mochimaru, Y. Fujimoto, and Y. Ishikawa, "Detecting the fetal electrocardiogram by wavelet theory-based methods," *Progress in Biomed Res*, vol. 7, no. 3, pp. 185–193, Sept. 2002.

[148] F. Mochimaru, Y. Fujimoto, and Y. Ishikawa, "The fetal electrocardiogram by independent component analysis and wavelets," *Japan J Physiology*, vol. 54, no. 5, pp. 457–463, 2004.

[149] R. Moddemeijer, "On estimation of entropy and mutual information of continuous distributions," *Signal Processing*, vol. 16, no. 3, pp. 233–246, 1989.

[150] G. B. Moody. (2005, Mar.) RR intervals, heart rate, and HRV howto. [Online]. Available as of Feb.28, 2006: http://www.physionet.org/tutorials/hrv/

[151] G. E. Morgan, Jr., M. S. Mikhail, M. J. Murray, and C. P. Larson, Jr., Eds., *Clinical Anesthesiology*, 3rd ed. McGraw-Hill, 2002, (ISBN: ISBN: 0-8385-1553-3).

[152] Z. Moussavi, "Introduction to biomedical engineering: Course notes," Dec. 2003, University of Manitoba.

[153] T. Musha, "1/f fluctuations in biological systems (review)," in *6th International Conference on Noise in Physical Systems*. Elsevier, 1981.

[154] J.-F. Muzy, E. Bacry, and A. Arneodo, "Wavelets and multifractal formalism for singular signals: Application to turbulence data," *Physical Review Letters*, vol. 67, no. 25, pp. 3515–3518, Dec. 1991.

[155] J.-P. Nadal and N. Parga, "Non linear neurons in the low noise limit: a factorial code maximizes information transfer," *Network*, vol. 5, pp. 565–581, 1994.

[156] J.-P. Nadal and N. Parga, "Redundancy reduction and independent component analysis: Conditions on cumulants and adaptive approaches," *Neural Computation*, vol. 9, no. 7, pp. 1421–1456, 1997.

[157] M. Nakao, M. Norimatsu, Y. Mizutani, and M. Yamamoto, "Spectral distortion properties of the integral pulse frequency modulation model," *IEEE Trans Biomed Eng*, vol. 44, no. 5, pp. 419–426, May 1997.

[158] NASDAQ. (2005, Dec.) 2005 NASDAQ 100 (SPX) index. [Online]. Available as of Dec.2, 2005: http://dynamic.nasdaq.com/dynamic/nasdaq100_activity.stm

[159] National Library of Medicine. (2003) The Visible Human Project. [Online]. Available as of Sept. 16, 2005: http://www.nlm.nih.gov/research/visible/

[160] D. Obradovic and G. Deco, "Information maximization and independent component analysis: Is there a difference?" *Neural Computation*, vol. 10, no. 8, pp. 2085–2101, 1998.

[161] E. Oja, "Simplified neuron model as a principal component analyzer," *Journal of Mathematical Biology*, vol. 15, no. 3, pp. 267–273, Nov. 1982. [Online]. Available: http://www.springerlink.com/content/u9u6120r003825u1

[162] E. Oja, "Neural networks, principal components, and subspaces," *International Journal of Neural Systems (IJNS)*, vol. 1, no. 1, pp. 61–68, 1989.

[163] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, ser. Prentice-Hall Signal Processing Series. Englewood Cliffs, NJ, USA: Prentice-Hall, 1989, 879 pp., (ISBN: 0-13-216292-X).

[164] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans Biomed Eng*, vol. 32, no. 3, pp. 220–236, 1985.

[165] A. Papoulis and S. U. Pillai., *Probability, Random Variables, and Stochastic Processes*, 4th ed., ser. McGraw-Hill Series in Electrical and Computer Engineering. Boston, USA: McGraw-Hill, 2002, 862 pp., (ISBN: 0-07-112256-7).

[166] S. Parmar, H. Patel, and J. Sahambi, "Separation performance of ICA algorithms on FECG and MECG signals contaminated by noise," in *Applied Computing: Proc. 2nd Asian Applied Computing Conference, AACC 2004*. (Kathmandu, Nepal: Oct. 29-31, 2004): Springer Berlin / Heidelberg, 2004, pp. 184–190, (ISBN: 978-3-540-23659-7). [Online]. Available: www.springerlink.com/content/1tewwcrpfu1v10jp

[167] K. Pawelzik and H. G. Schuster, "Generalized dimensions and entropies from a measured time series," *Physical Review A*, vol. 35, no. 1, pp. 481–484, Jan. 1987.

[168] D. T. Pham, P. Garat, and C. Jutten, "Separation of mixture of independent sources through a maximum likelihood approach," in *Proc. European Signal Processing Conference EUSIPCO '92*, ser. Signal Processing VI, J. Vandewalle, R. Boite,

M. Moonen, and A. Oosterlinck, Eds. Bruxelles, Belgium: Elsevier, 1992, pp. 771–774.

[169] D.-T. Pham and J.-F. Cardoso, "Blind separation of instantaneous mixtures of nonstationary sources," *IEEE Trans Signal Processing*, vol. 49, no. 9, pp. 1837–1848, Sept. 2001.

[170] R. Plonsey, *Bioelectric Phenomena*, ser. McGraw-Hill Series in Bioengineering. New York, NY, USA: McGraw-Hill, 1969, 380 pp.

[171] M. Plumbley, "Geometry and manifolds for independent component analysis," in *Proc. 2007 IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP07*, vol. 4, (Honolulu, HW; April 15-20,2007), 2007, pp. IV–1397–IV–1400.

[172] C.-S. Poon and C. K. Merrill, "Decrease of cardiac chaos in congestive heart failure," *Nature*, vol. 389, pp. 492–495, Oct.2 1997.

[173] M. Potter, N. Gadhok, and W. Kinsner, "Separation performance of ICA on simulated EEG and ECG signals contaminated by noise," *Canadian Journal of Electrical and Computer Engineering*, vol. 27, no. 3, pp. 123–127, July 2002.

[174] M. Potter and W. Kinsner, "Competing ICA techniques in biomedical signal analysis," in *IEEE Canadian Conference on Electrical and Computer Engineering CCECE'01*, vol. II, (Toronto, ON, Canada; May 13–16), 2001, pp. 987–992, (ISBN 0-7803-6715-4).

[175] M. Potter and W. Kinsner, "Surrogate ECG signals for the null-hypothesis analysis of heart monitoring," in *Proc. 2005 IEEE Canadian Conference Electrical and Computer Engineering CCECE'05*. (Saskatoon, SK, Canada; May 1–4, 2005): IEEE, 2005, pp. 1668–1672.

[176] M. Potter and W. Kinsner, "A biologically motivated phase-offset to the piecewise-linear model of instantaneous heart-rate interpolation," in *Proc. 5th IEEE Int. Conf. on Cognitive Informatics, ICCI06*, vol. 1. (Beijing, China; July 17–19, 2006): IEEE, 2006, pp. 218–225, (ISBN: 1-4244-0475-4).

[177] M. Potter and W. Kinsner, "Direct calculation of the $f(\alpha)$ fractal dimension spectrum from high-dimensional correlation-integral partitions," in *Proc. 2007 IEEE Int. Conf. Acoustics Speech Signal Proc, ICASSP07*, vol. III. (Honolulu, USA; Apr. 15–20, 2007): IEEE, 2007, pp. 989–992, (ISBN: 1-4244-0728-1).

[178] M. Potter and W. Kinsner, "Instantaneous heart rate: Should RR-intervals be resampled?" in *Proc. 30th IEEE Int. Conf. EMBS, EMBC08*, (Vancouver, BC, Canada; Aug. 20–24, 2008), (accepted June 16, 2008).

[179] M. Potter and W. Kinsner, "Improved event-interval reconstruction in synthetic electrocardiograms," *IEEE Trans Biomed Eng*, (accepted May 21, 2008).

[180] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C : The Art of Scientific Computing*, 2nd ed. Cambridge, UK: Cambridge University Press, Oct.30 1992, 1020 pp., (ISBN: 0521431085).

[181] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications*, 3rd ed. Upper Saddle River, NJ, USA: Prentice Hall, 1996, 1016 pp., (ISBN: 0-13-373762-4).

[182] A. Provenzale, L. A. Smith, R. Vio, and G. Murante, "Distinguishing between low-dimensional dynamics and randomness in measured time series," *Physica D: Nonlinear Phenomena*, vol. 58, no. 1-4, pp. 31–49, Sept.15 1992. [Online]. Available: http://www.sciencedirect.com/science/article/B6TVK-46FYTH6-18/2/85f222429a0c3282f14f7b4227c0d6cd

[183] R. M. Rangayyan, *Biomedical Signal Analysis: A Case Study Approach*, 1st ed., ser. IEEE Press Series on Biomedical Engineering. New York, NY: Wiley-Interscience, Dec. 2001, 552 pp., (ISBN: 0-471-20811-6).

[184] P. M. Rautaharju, J. W. Warren, and H. P. Calhoun, "Estimation of QT prolongation: A persistent, avoidable error in computer electrocardiography," in *J Electrocardiography*, vol. 23, no. Supplement 1, (15th ISCE Conf on Comp. Appl. in Card., April 22-27, 1990, Virginia Beach, VA, USA), 1990, pp. 111–117. [Online]. Available: http://www.sciencedirect.com/science/article/B6WJ4-4C4FH56-30/2/23f446b9d653d85a4cdde307d1981c67

[185] M. Richter, T. Schreiber, and D. T. Kaplan, "Fetal ECG extraction with nonlinear state-space projections," *IEEE Trans Biomed Eng*, vol. 45, no. 1, pp. 133–137, Jan. 1998.

[186] P. R. Rijnbeek, J. A. Kors, and M. Witsenburg, "Minimum bandwidth requirements for recording of pediatric electrocardiograms," *Circulation*, vol. 104, pp. 3087–3090, Dec.18/25 2001.

[187] S. Roberts and R. Everson, Eds., *Independent Component Analysis: Principles and Practice*. New York, NY, USA: Cambridge University Press, 2001, (ISBN: 0521792983).

[188] C. H. Rodeck and M. J. Whittle, Eds., *Fetal Medicine: Basic Science and Clinical Practice*. London, UK: Churchill Livingstone, 1999, 1162 pp., (ISBN: 0-4430-5357X).

[189] A. M. Rudolph, "Fetal circulation," in *Fetal Cardiology: embryology, genetics, physiology, echocardiographic evaluation, diagnosis and perinatal management of cardiac diseases*, S. Yagel, N. H. Silverman, and U. Gembruch, Eds. London, UK: Martin Dunitz, 2003, ch. 9, pp. 107–120, (ISBN: 1-85317-905-1).

[190] F. B. Sachse, C. D. Werner, K. Meyer-Waarden, and O. Dössel, "Applications of the Visible Man dataset in electrocardiology: Calculation and visualization of body surface potential maps of a complete heart cycle," in *The Second Visible Human Project Conference Proceedings*. Bethesda, Maryland USA: NIH, Oct.1-2 1998, pp. (html–1). [Online]. Available as of Jan. 15, 2006: http://www.nlm.nih.gov/research/visible/vhpconf98/AUTHORS/SACHSE2/sachse2.htm

[191] R. Sameni, M. Shamsollahi, C. Jutten, and G. Clifford, "A nonlinear Bayesian filtering framework for ECG denoising," *IEEE Trans Biomed Eng*, vol. 54, no. 12, pp. 2172–2185, Dec. 2007.

[192] R. Sameni, G. D. Clifford, C. Jutten, and M. B. Shamsollahi, "Multichannel ECG and noise modeling: Application to maternal and fetal ECG signals," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, pp. Article ID 43 407, 14 pp., 2007, doi:10.1155/2007/43407.

[193] R. Sameni, C. Jutten, and M. B. Shamsollahi, "What ICA provides for ECG processing: Application to noninvasive fetal ECG extraction," in *Proc. 2006 IEEE Int Symp on Signal Processing and Information Technology*, (Vancouver, BC, Canada; Aug. 2006), 2006, pp. 656–661.

[194] V. D. Sánchez A, "Frontiers of research in BSS/ICA," *Neurocomputing*, vol. 49, pp. 7–23, 2002.

[195] M. Santos, J. V. Fonseca, and A. K. Barros, "New method to heart instantaneous frequency estimation," in *Proc. 2004 IEEE Workshop Machine Learning for Signal Processing*, 2004, pp. 473–479.

[196] M. Sato, Y. Kimura, S. Chida, T. Ito, N. Katayama, K. Okamura, and M. Nakao, "A novel extraction method of fetal electrocardiogram from the composite abdominal signal," *IEEE Trans Biomed Eng*, vol. 54, no. 1, pp. 49–58, 2007.

[197] T. Sauer, J. A. Yorke, and M. Casdagli, "Embedology," *J. Statistical Physics*, vol. 65, no. 3/4, pp. 579–616, Nov. 1991.

[198] L. L. Scharf, *Statistical Signal Processing: Detection, estimation, and time series analysis*. Reading, Mass.: Addison-Wesley Pub. Co., 1991, 524 pp., (ISBN: 0-201-19038-9).

[199] K. G. Schmidt, "Fetal bradyarrythmia," in *Fetal Cardiology: embryology, genetics, physiology, echocardiographic evaluation, diagnosis and perinatal management of cardiac diseases*, S. Yagel, N. H. Silverman, and U. Gembruch, Eds. London, UK: Martin Dunitz, 2003, ch. 29, pp. 345–354, (ISBN: 1-85317-905-1).

[200] T. Schreiber, "Constrained randomization of time series data," *Physical Review Letters*, vol. 80, no. 10, pp. 2105–2108, Mar. 9 1998.

[201] T. Schreiber and D. T. Kaplan, "Signal separation by nonlinear projections: The fetal electrocardiogram," *Physical Review E*, vol. 53, no. 5, pp. R4326–R4329, May 1996.

[202] T. Schreiber and A. Schmitz, "Improved surrogate data for nonlinearity tests," *Physical Review Letters*, vol. 77, no. 4, pp. 635–638, July 22 1996.

[203] T. Schreiber and A. Schmitz, "Discrimination power of measures for nonlinearity in a time series," *Physical Review E*, vol. 55, no. 5, pp. 5443–5447, May 1997.

[204] T. Schreiber and A. Schmitz, "Surrogate time series," *Physica D*, vol. 142, pp. 346–382, 2000.

[205] S. R. Seydnejad and R. I. Kitney, "Real-time heart rate variability extraction using the Kaiser window," *IEEE Trans Biomed Eng*, vol. 44, no. 10, pp. 990–1005, Oct. 1997.

[206] M. Small and C. Tse, "Detecting determinism in time series: the method of surrogate data," *IEEE Trans on Circuits and Systems I: Fundamental Theory and Applications*, vol. 50, no. 5, pp. 663–672, May 2003.

[207] M. Small, D. J. Yu, J. Simonotto, and R. G. Harrison, "Nonlinear analysis of human ECG during sinus rhythm and arrhythmia," in *Proc. 2000 Computers in Cardiology Conf.*, (Cambridge, MA; Sept. 24-27, 2000), 2000, pp. 147–150.

[208] M. Small and C. K. Tse, "Applying the method of surrogate data to cyclic time series," *Physica D: Nonlinear Phenomena*, vol. 164, no. 3-4, pp. 187–201, Apr. 15 2002.

[209] M. Small and C. K. Tse, "Optimal embedding parameters: a modelling paradigm," *Physica D: Nonlinear Phenomena*, vol. 194, no. 3-4, pp. 283–296, July 15 2004.

[210] M. Small, D. Yu, and R. G. Harrison, "Surrogate test for pseudoperiodic time series data," *Physical Review Letters*, vol. 87, no. 18, p. 188101, Oct. 29 2001.

[211] M. Small, D. Yu, J. Simonotto, R. G. Harrison, N. Grubb, and K. Fox, "Uncovering non-linear structure in human ECG recordings," *Chaos, Solitons, and Fractals*, vol. 13, pp. 1755–1762, 2002.

[212] J. C. Sprott. (1995, Mar.) arfusien.gif. [Online]. Available as of Dec. 2, 2005: http://sprott.physics.wisc.edu/fractals/arfusien.gif

[213] J. C. Sprott. (1995, Feb.) ilurcego.gif. [Online]. Available as of Dec. 2, 2005: http://sprott.physics.wisc.edu/fractals/ilurcego.gif

[214] J. C. Sprott. (1997, Dec.) trees.jpg. [Online]. Available as of Dec. 2, 2005: http://sprott.physics.wisc.edu/fractals/natural/trees.jpg

[215] H. Stögbauer, A. Kraskov, S. A. Astakhov, and P. Grassberger, "Least-dependent-component analysis based on mutual information," *Physical Review E*, vol. 70, pp. 066 123–1–066 123–17, 2004.

[216] J. V. Stone, *Independent Component Analysis: A Tutorial Introduction.* Cambridge, MA, USA: MIT Press, 2004, (ISBN: 0262693151).

[217] Z. R. Struzik, "Determining local singularity strengths and their spectra with the wavelet transform," *Fractals*, vol. 8, no. 2, pp. 163–179, June 2000.

[218] A. Stuart and J. K. Ord, *Kendall's Advanced Theory of Statistics*, 6th ed. Toronto: Halsted Press (John Wiley and Sons), 1994, vol. 1, (ISBN: 0-340-61430-7).

[219] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980*, ser. Lecture Notes in Mathematics, D. A. Rand and L. S. Young, Eds. New York, NY: Springer-Verlag, 1981, vol. 898, pp. 366–381, (ISBN: 3540111719).

[220] The Mathworks. Documentation (matlab function reference): eig.m. [Online]. Available as of Jun 23, 2008: http://www.mathworks.com/access/helpdesk/help/techdoc/ref/eig.html

[221] J. Theiler, "Spurious dimension from correlation algorithms applied to limited time-series data," *Physical Review A*, vol. 34, pp. 2427–2432, Sept. 1986.

[222] J. Theiler, "Estimating fractal dimension," *J. Optical Society of America A*, vol. 7, no. 6, pp. 1055–1073, June 1990.

[223] J. Theiler, "On the evidence for low-dimensional chaos in an epileptic electroencephalogram," *Physics Letters A*, vol. 196, no. 5-6, pp. 335–341, Jan. 1995. [Online]. Available: http://www.sciencedirect.com/science/article/B6TVM-3YB9Y2G-8W/2/a22428a60f404624d3da1ba776d1c98d

[224] J. Theiler and S. Eubank, "Don't bleach chaotic data," *Chaos*, vol. 3, pp. 771–782, 1993. [Online]. Available as of Jan. 15, 2006: http://nis-www.lanl.gov/~jt/Papers/bleach.ps.gz

[225] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, "Testing for nonlinearity in time series: the method of surrogate data," *Physica D*, vol. 58, no. 1-4, pp. 77–94, Sept. 1992.

[226] J. Theiler and D. Prichard, "Constrained-realization Monte-Carlo method for hypothesis testing," *Physica D*, vol. 94, no. 4, pp. 221–235, July 1996.

[227] K. L. Thornburg, *Fetal Medicine: Basic Science and Clinical Practice.* London, UK: Churchill Livingstone, 1999, ch. 13: Development of the Cardiovascular System, pp. 141–154, (ISBN: 0-4430-5357X).

[228] J. Vanderschoot, D. Callaerts, W. Sansen, J. Vandewalle, G. Vantrappen, and J. Janssens, "Two methods for optimal MECG elimination and FECG detection from skin electrode signals," *IEEE Trans Biomed Eng*, vol. 34, no. 3, pp. 233–243, Mar. 1987.

[229] R. Vigário, J. Särelä, V. Jousmäki, M. Hämäläinen, and E. Oja, "Independent component approach to the analysis of EEG and MEG recordings," *IEEE Trans Biomedical Engineering*, vol. 47, no. 5, pp. 589–593, May 2000.

[230] V. Vigneron, A. Paraschiv-Ionescu, A. Azancot, O. Sibony, and C. Jutten, "Fetal electrocardiogram extraction based on non-stationary ICA and wavelet denoising," in *Proc. 7th Int. Symp. on Sig. Proc. and Appl.*, vol. 2, (July 1-4, 2003), 2003, pp. 69–72.

[231] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Trans Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1462–1469, July 2006.

[232] F. Vrins, J. Lee, M. Verleysen, V. Vigneron, and C. Jutten, "Improving independent component analysis performances by variable selection," in *Proc. IEEE 13th Workshop on Neural Networks for Signal Processing, NNSP'03*, (Sep. 17-19, 2003), 2003, pp. 359–368, (ISBN: 1089-3555).

[233] C. D. Werner, F. B. Sachse, and O. Dössel, "Applications of the Visible Human Male data set in electrocardiology: Simulation of the electrical excitation propagation," in *Second Visible Human Project Conference.* (National Institute of Health, Bethesda, Maryland USA, October 12, 1998): NIH, 1998, Electronic.

[234] B. Widrow, J. R. Glover, J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, E. Dong Jr., and R. C. Goodlin, "Adaptive noise cancelling: Principles and applications," *Proceedings IEEE*, vol. 63, no. 12, pp. 1692–1716, Dec. 1975.

[235] Wikipedia. Matrix norm. [Online]. Available as of May 1, 2008: http://en.wikipedia.org/w/index.php?title=Matrix_norm&oldid=208598828

[236] Wikipedia. (2005, Nov.16) Central limit theorem. [Online]. Available as of Nov.20, 2005: http://en.wikipedia.org/wiki/Central_limit_theorem

[237] Wikipedia. (2005, Aug.) Heart rate. [Online]. Available as of Aug. 31, 2005: http://en.wikipedia.org/wiki/Heart_rate

[238] Wikipedia. (2005, Sept.) Lists of indices of refraction. [Online]. Available as of Sept 13, 2005: http://en.wikipedia.org/wiki/List_of_indices_of_refraction

[239] Wikipedia. (2005, Oct.) Partition function (statistical mechanics). [Online]. Available as of Nov.11, 2005: http://en.wikipedia.org/wiki/Partition_function_(statistical_mechanics)

[240] D. C. Wood and J. C. Huhta, *Fetal Medicine: Basic Science and Clinical Practice.* London, UK: Churchill Livingstone, 1999, chapter 61: Fetal Cardiac Arrhythmias, pp. 817–828, (ISBN: 0-4430-5357X).

[241] M. S. Woolfson, W. Peasgood, S. Sahota, and J. A. Crowe, "Signal processing of the fetal electrocardiogram," in *J Electrocardiography*, vol. 23, no. Supplement 1, (15th ISCE Conf. on Comp. Appl. in Card., April 22-27, 1990, Virginia Beach, VA, USA), 1990, pp. 51–57.

[242] D. J. Yu, M. Small, R. G. Harrison, C. Robertson, G. Clegg, M. Holzer, and F. Sterz, "Complexity measurements for analysis and diagnosis of early ventricular fibrillation," in *Proc. 1999 Computers in Cardiology Conf.*, (Hannover; Sept. 26-29, 1999), 1999, pp. 21–24.

[243] V. Zarzoso, J. Millet-Roig, and A. Nandi, "Fetal ECG extraction from maternal skin electrodes using blind source separation and adaptive noise cancellation technique," in *Proc. 2000 Computers in Cardiology Conf.*, vol. 27. IEEE Press, 2000, pp. 431–434.

[244] V. Zarzoso and A. K. Nandi, "Noninvasive fetal electrocardiogram extraction: Blind separation versus adaptive noise cancellation," *IEEE Trans Biomed Eng*, vol. 48, no. 1, pp. 12–18, Jan. 2001.

Appendix A

Details of the ECGfm Surrogate ECG Algorithm

A.1 Overview

In fulfillment of the fetal ECG simulation requirement of noise-free uncontaminated ECG, this thesis makes an important contribution to the synthesis of ECG signals with natural HRV. Given, however, the thesis focus on ICA and its performance measurement (and not the ECG itself), the details of the novel synthesis technique have been separated from the body of the thesis. Now, this appendix presents the details of the ECGfm synthesis algorithm for surrogate ECG.

The significance of the ECGfm algorithm is it provides a mechanism for converting a record of ECG RR-intervals into a waveform with ECG properties. In this way, an ECG time series can exhibit either natural or unnatural HRV by design. Since this technique can synthesize a noise-free representation of any recorded ECG, the author calls this contribution to the field of ECG synthesis a *surrogate ECG*, in analogy to the formalisms described in Ch. 5 and its application to the field.

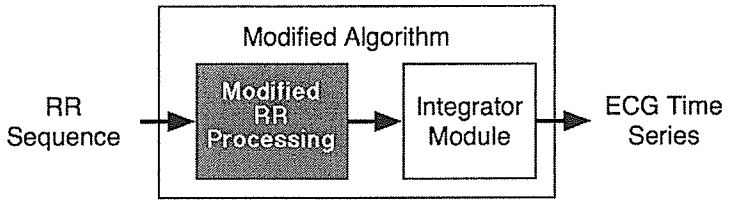


Fig. A.1 The surrogate ECG model converts an event series (*e.g.*, tachogram) into an ECG time series via two modules: an IHR-processing module, and an integrator module.

The ECGfm system is based on two parts, as shown in Fig. A.1. First, a complete description of the cyclostationary system will be discussed in Sec. A.2. In a top down manner, the entire system and the interface of the modules is overviewed in Sec. A.2.1. The integrator module is then discussed in Sec. A.2.2 while the IHR-module is described in Sec. A.2.3. An interpretation on the significance of IHR to the ECG attractor reconstruction is then presented in Sec. A.2.4. Specifically, it is discussed how the significant variables in the ECGfm synthesis process affect the reconstructed attractor, and also how they might contribute to an MFA.

Thereafter, in Sec. A.3, the ECGfm design is discussed in comparison to the ECGsyn algorithm [140]. Specifically, a set of experiments for ECG synthesis comparison is established, and the practical limitations of the original integrator and IHR-processing modules of the ECGsyn algorithm are discussed separately in Sec. A.3.2 and Sec. A.3.3 respectively. It is also demonstrated how the ECGfm overcomes these response limitations. Last, the most significant code of the ECGfm algorithm is presented in Sec. A.4.

A.2 Understanding the ECGfm Model for Surrogate ECG

A.2.1 Driven Periodic Systems for Cyclostationarity

The surrogate ECG model is based on the separation of cyclostationarity into conjoined periodic $z(\theta)$ and aperiodic $\theta(t)$ elements. The continuous periodic form $z(\theta)$ contains the characteristic information of the cyclic regularity, including the amplitude and duration of the different ECG waves. The aperiodic increasing variable, $\theta(t)$, will be responsible for representing cyclic variability and, specifically, HRV.

To visualize the behaviour of the joint system $z(t) = z \circ \theta(t)$, (where $\circ$ is the functional composition operator), consider Fig. A.2. Here, a three-dimensional system of differential

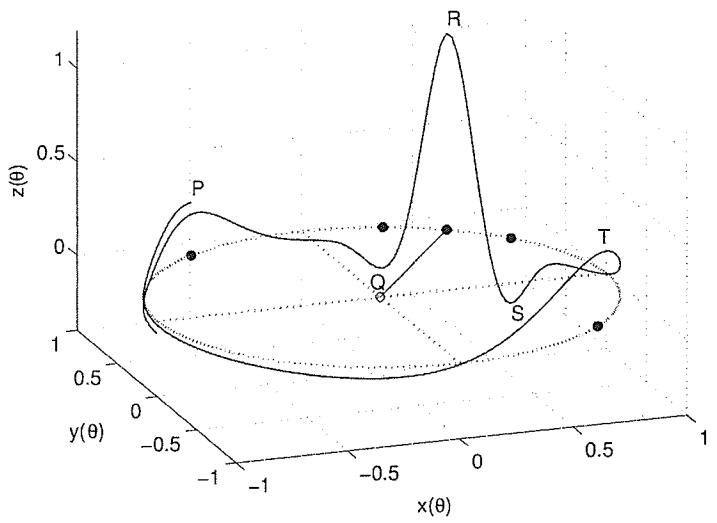


Fig. A.2 The ECGfm/ECGsyn three-dimensional dynamical space. [After [140].] The variables x and y are dummy variables used to define $\theta = \theta(x, y)$. However, different degrees of variability away from strict periodicity in $z(\theta)$ can be designed by increasing the degrees of freedom in this three-dimensional motion, $(x(\theta), y(\theta), z(\theta))$. Dots identify the parametric angles θ_i at which the P-,R-, and T-waves occur (as shown).

equations in a cylindrical coordinate frame can represent the ECG signal. The regular periodic features can then be mapped into the 3D space of the model. That is, by introducing asymmetrically repulsive points on the baseline plane, morphological features like the the P-, QRS-, and T-waves can be modelled in the $z(t)$ time series. The angular position, θ_i, of these repulsive points defines the location in phase of the fiducial extrema of the ECG waves. Here, θ_i is best indexed by the wave identifiers themselves: namely $i \in \{P, Q, R, S, T\}$. Thus, as the system cycles around the origin with positive angular velocity, $\omega(t)$, the repulsive points (in a proper positioning) sequentially deflect the system trajectory from the baseline and influence the morphology to correctly represent sinus rhythm. This is a rather straightforward approach for the modelling of generic cyclostationary signals as a driven periodic system.

As a parametric oscillator, this ECG model can be tuned for the multiple morphologies that

can appear in various ECG leads (or other cyclostationary biological signals [43]) by adapting the parameters of the repulsive points. More significant to the current discussion, however, this approach to ECG synthesis needs a well-defined angular velocity (as a time series). This angular velocity time series is equivalent philosophically to IHR [140][176]. The purpose of the IHR-processing module is to define this IHR signal from an input representing the *event-template* of a recorded ECG. In the case of a tachogram input, this defines an IHR signal consistent with natural HRV. With the addition of the shorter ECG-intervals (*e.g.*, PR- and RT-intervals) at the input, the IHR signal can also synthesize these elements.

Now that the decomposition of the cyclostationary ECG model has been identified, the details of the two modules describing the cyclic $z(\theta)$ and aperiodic $\theta(t)$ components will be presented. First, the cyclic $z(\theta)$ system, equivalent to the integrator module of Fig. A.1, will be described, under the assumption that $\theta(t)$ is already well-defined. Afterwards, the proper definition of $\theta(t)$ from an event-interval template is discussed, and is presented in the form of the *frequency-modulated phase interpolation* (FMPI) technique. This technique will be contrasted to spectral or resampling techniques for IHR definition.

A.2.2 Integrator Module

The basic form of the ECGfm integrator module is inspired by, and adapted from, the McSharry *et al.* model [140]. The final form of the adaptations will be presented here, while a comparison to the original ECGsyn system will be made in Sec. A.3.2.

The simplest interpretation to the integrator module would be to consider only one degree of freedom, whereby the amplitude $z(\theta)$ in Fig. A.2 is defined as a fixed periodic function of one angular variable, θ. A slightly more general model could define z as a static surface over two variables (x, y). Under this generalization the system is allowed to leave the unit circle trajectory (*i.e.*, the "the average" behaviour) and explore the surface around it as well. More

Table A.1 Default Values for ECGfm/ECGsyn Repellor Properties

Parameter	**Value** $i = P$	**Value** $i = Q$	**Value** $i = R$	**Value** $i = S$	**Value** $i = T$
θ_i at 60 bpm	$-\pi/3$	$-\pi/12$	0	$\pi/12$	$\pi/2$
a_i	1.2	-5.0	30.0	-7.5	0.75
b_i	0.25	0.10	0.10	0.10	0.4

variability in the output $z(t)$ morphology could be observed here.

The surrogate ECG model goes further, in fact, by going beyond an explicitly defined z surface. Instead, the model uses a true three-dimensional deterministic system $(x(t), y(t), z(t))$ with a periodic attractor given by (x, y) in the unit circle and z on the curve of minimum potential. Thus the "average ECG trajectory" at the output is equivalent to the restricted 1-dimensional model, but at instantaneous times, the model presents more degrees of freedom. Thus the fundamental motion of the system can be considered as a simple 1-dimensional attractive manifold, but the system is free to indulge in mild dynamical effects, thus adding morphological variability to the ECG waveforms.

As previously mentioned, the ECG waves are controlled by asymmetric repellors positioned in the 3D space of Fig. A.2. These repellors for $i \in \{P, Q, R, S, T\}$ are positioned on the unit circle in the $z = 0$ plane and are described by the parameters

(a) angular position, θ_i;

(b) signed amplitude, a_i (*i.e.*, magnitude and direction of deflection); and

(c) width parameter, b_i.

Default values for these parameters are shown in Table A.1. Note however, that due to the dependence of the ECG event-intervals to mean heart rate, a correction is made to applied values of θ_i. That is, to compensate for the nonlinear shortening of the event-intervals with

increased heart rate, a Bazett-style correction

$$\begin{cases} \theta_i \leftarrow \left(\frac{\overline{\rho_\heartsuit}}{60}\right)^{\frac{1}{4}} \theta_i & ;\ i \in \{P, T\} \\ \theta_R \leftarrow \theta_R \\ \theta_i \leftarrow \left(\frac{\overline{\rho_\heartsuit}}{60}\right)^{\frac{1}{2}} \theta_i & ;\ i \in \{Q, S\} \end{cases} \tag{A.1}$$

is applied before the initialization of the integrator module, where the mean heart rate $\overline{\rho_\heartsuit}$ is measured in bpm.

With these parameters and an IHR time series, $\omega(t)$, the system of ordinary differential equations defining the integrator module,

$$\alpha(x, y) = 1 - \sqrt{x^2 + y^2} \tag{A.2}$$

$$\theta(x, y) = \text{atan2}(yx) \tag{A.3}$$

$$\frac{dx}{dt} = \alpha\, x - \omega(t) y \tag{A.4}$$

$$\frac{dy}{dt} = \alpha\, y + \omega(t) x \tag{A.5}$$

$$\frac{dz}{dt} = -\omega(t) \sum_i a_i\, (\theta \ominus \theta_i)\, \exp\left(-\frac{1}{2}\left(\frac{\theta \ominus \theta_i}{bi}\right)^2\right) - (z - z_0(t)) \tag{A.6}$$

are well-defined. Here, the $\ominus$ subtraction operator is a circular phase subtraction defined as $\theta \ominus \theta_i \stackrel{\text{def}}{=} (\theta - \theta_i) \mod 2\pi$. Note that the form of (A.6) accounts for (i) a baseline behaviour defined by the $z_0(t)$ function and (ii) events with a Gaussian profile at the θ_i. For the synthetic ECG generated in this work, the default values of Table A.1 and $z_0 = 0$ were used.

The system of equations above are then integrated from an initial condition using the Dormand-Prince explicit (4,5)-Runge-Kutta solver [28]. This is a variable step-size algorithm, whose output is then interpolated onto a uniformly sampled time vector with sampling frequency ν_{si}. As with any dynamical system, there is a transient from the initial condition, hard-set to $[1, 0, 0.04]$, until when the trajectory settles onto the attractor (here, limit cycle)

of the system. To eliminate this transient from the synthetic ECG output, a transient control protocol is enabled, whereby the IHR signal of the first ECG beat drives the system for 20 consecutive repetitions to allow for convergence. The maximum $z_m = \max z(t)$ over the final (repeated) beat is then taken for the effective initial condition

$$(x, y, z)|_{t=0} = (1, 0, z_m) \tag{A.7}$$

which is consistent with beginning the ECG surrogate at the R-wave apex, since $\theta(0) = 0$.

One legacy feature of the ECGfm algorithm is the ability to downsample the returned time series from the internal sampling frequency ν_{si}. The default implementation here is that the internal sampling will be twice that of the returned time series, $\nu_{si} = 2\nu_s$. At high sampling frequencies, however, there is limited benefit of an upsampled internal frequency since it amounts only to more interpolation points being applied to the irregularly sampled Runge-Kutta solution.

One significant feature of the original ECGsyn algorithm is a self-annotating mechanism based on the 3D xyz system of Fig. A.2. This procedure continues to be implemented in the ECGfm algorithm. Essentially, the xy coordinates are used to calculate $\theta(t)$, whereby the closest sample to θ_i is assigned as the fiducial extrema of that wave $i \in \{P, Q, R, S, T\}$. This is implemented by assigning a numerical "peak identifier" to each sample of the ECG, 0 for no peak, and 1 though 5 mapping to P through T.

Once the peak identifier has established the fiducial markers of the ECG waves, the xy coordinates of the 3D system are discarded and only the vector of z-values is used for the ECG time series itself. The integrator module is thus responsible to return (i) the surrogate ECG amplitudes (from z), (ii) the time vector t_n of the time series, and (iii) the peak marker vector.

Attention is now turned to the modelling of the IHR time series $\omega(t)$ used to drive the

integrator module.

A.2.3 IHR-Processing and FMPI

The fundamental purpose of the integrator module just described was to synthesize the regular features of an ECG with some morphological allowance of dynamics. The significance of the IHR-processing module, however, is that it controls the HRV of the synthesized ECG, since the angular rate $\omega(t)$ in Fig. A.2 is equivalent to IHR. Were a constant angular velocity $\omega(t) = \omega_0$ to drive the integrator module, a perfectly periodic ECG would result. Thus by properly introducing nonlinearity into the monotonic continuous function $\theta(t)$, realistic cyclostationarity variability can be expressed in $z(t) = z \circ \theta(t)$.

The action of the integrator module is thus a kind of *frequency-modulation* (FM), since the nonlinear components of $\theta(t)$ from natural HRV are mild. The novel aspect of the ECGfm surrogate ECG algorithm is that the event-intervals, and specifically the RR-intervals that define HRV, are faithfully reconstructed in the time domain. Given an event-interval template, such as a tachogram, the FM-variability of the ECG is represented in discrete form. For use in the integrator module, however, an IHR time series, $\omega(t)$, should have specificity on time intervals smaller than the original template. (As an example, the IHR derived from a tachogram $\Delta_R(n)$, must be specified at time scales under the average beat.) At these time scales, the interpolation from discrete information to continuous is philosophically ill-posed. This level of specificity is often not required in clinical or spectral analysis, which focus on low-frequency behaviour and windowed-averages over multiple beats. Here, in the modelling of synthetic ECG, however, it is *necessary* that $\omega(t)$ be practically defined on such short time scales. Thus, some model for IHR must be assumed, but the choice of model should be justified as much as possible.

To preserve realistic cyclostationarity variability in the continuous time series, a careful

analysis of IHR modelling is required. Some novel possibilities will now be discussed and evaluated on the simplest form of HRV constraint that can be applied: the conversion of a tachogram into IHR. First, the context of RR-processing will be demonstrated to be an interpolation problem, then several solutions of increasing difficulty will be presented: (i) basic, (ii) offset piecewise, (iii) Hermite. This analysis will demonstrate that the Hermite solution, and its generalized form of higher-order *frequency-modulated phase-interpolation* (FMPI), are the most justified modelling approach to IHR processing. After the implementation of the Hermite-FMPI model for use with ECG event-interval templates is described, a theoretical comparison of FMPI to other methods will be presented. Later in this chapter, the superior performance of the FMPI-based system at preserving HRV will be demonstrated in experimental comparison to other methods.

A.2.3.1 RR-Processing is Interpolation

Consider an ECG signal with R-wave occurrence times t_n, $n \geq 0$. As described in Ch. 2, HRV can be represented as a sequence of RR-intervals, $\Delta_R(n) = t_n - t_{n-1}$ with $n \geq 1$. Recall, the tachogram is a sequence of discrete values, $\Delta_R(n)$ vs. n, well-ordered by time, but not uniformly sampled as a time series usually is. For the remainder of this chapter, tachograms are best considered graphically as a stem series $\Delta_R(n)$ vs. t_n, which directly uses the known timing of the events in a causal way.

Recall that the tachogram is a convenient feature of HRV even for relatively low quality ECG signals (either in signal to noise ratio, or sampling frequency), because RR-interval extraction is robust and well-automated [164][109][137], with a precision determined by the sampling frequency of the ECG signal [143]. From the chosen cyclostationary representation of an ECG, every event in z has a position jointly described by t and θ. The R-waves are identifiable in time as t_n under usual time-series analysis. They are also identifiable in angle

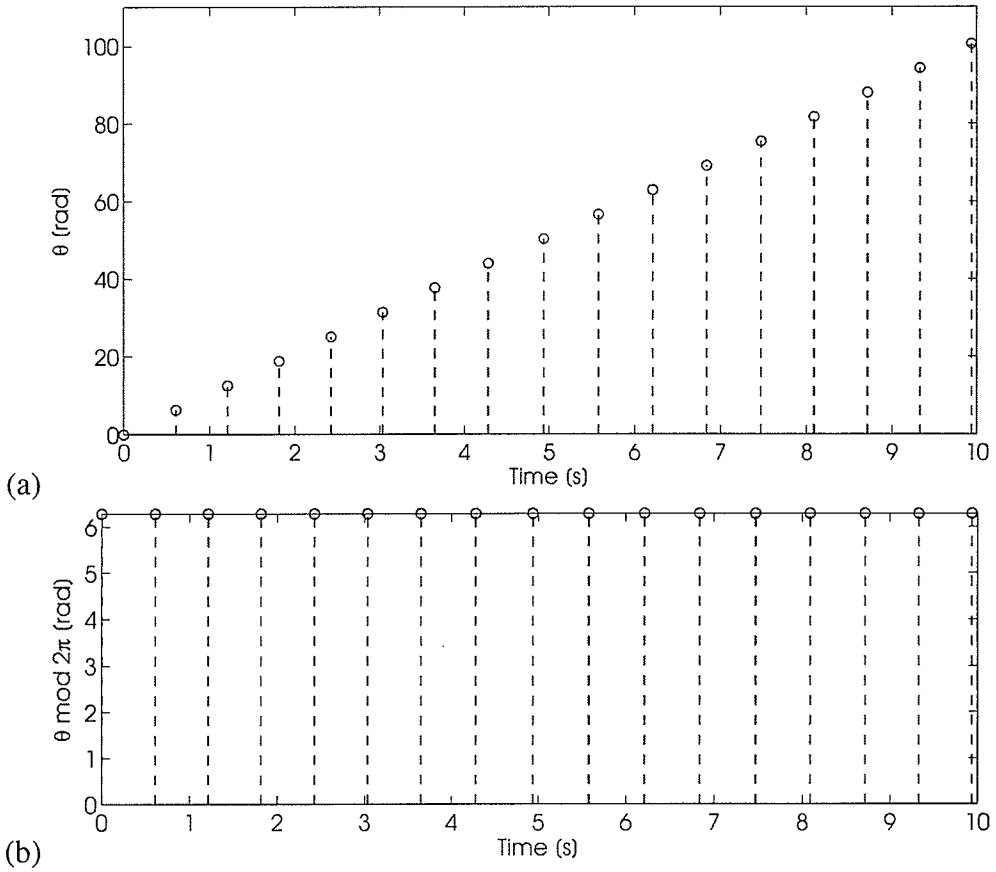


Fig. A.3 The procedure of modelling $\theta(t)$ is an interpolation problem with the known times of the R-waves acting as points each separated by 2π radians. (a) θ and t space; (b) the same space, now wrapped in periodic angle.

because they occur at the same angular point in the dynamical space of Fig. A.2. Including a measure of revolution, this would mean that the angles $\theta(t_n)$ are separated by 2π radians. Thus, observing a space of θ and t, there are a finite number of identified monotonic points from which to define $\theta(t)$, as shown in Fig. A.3(a). That is, the constraint

$$\theta(t_n) = 2\pi n + \theta_R \tag{A.8}$$

is given for some constant θ_R that represents the phase at which the R-wave occurs in the ECGfm/ECGsyn model, Fig. A.2.

The constraint on any model of $\theta(t)$ for the driven periodic oscillator is that it must be monotonic increasing and it must satisfy the known timings of the R-waves. This is essentially an interpolation problem [76]. The resulting interpolated $\theta(t)$ function should also be differentiable enough so that

$$\omega(t) = \frac{d\theta(t)}{dt} \tag{A.9}$$

is usefully defined. Recall that this interpolation is equivalent to defining the IHR through the relationship

$$\rho_{\heartsuit}(t) = \frac{60}{2\pi}\omega(t) \tag{A.10}$$

where *beats per minute* (bpm) are assumed as the units of IHR. Of course, it is more helpful to simply use $\omega(t)$ in rad/s or $\omega(t)/2\pi$ in Hz.

This is the fundamental framework of the *frequency-modulated phase-interpolation* (FMPI) model for the definition of IHR [178]. The event-interval data provides a physical constraint on an interpolation problem. The philosophical ill-posedness of the IHR at small time-scales appears as the flexibility of interpolation models that can be ascribed to the FMPI solution. Some examples of FMPI solutions with increasing difficulty will now be presented: (i) basic piecewise, (ii) offset piecewise, (iii) Hermite interpolation. The first is the simplest choice for interpolation: a piecewise-linear interpolation between the constraints of Fig. A.3(a). Adopting this model, however, necessarily introduces a singularity in $\omega(t)$ at the t_n. Instead, an offset can be introduced to relocate the discontinuity, as described in the second model. Last, when discontinuities improperly describe the continuous process to be modelled, Hermite interpolation can ensure a continuous IHR. A visualization of the three different FMPI solutions to a fixed dataset is shown in Fig. A.4.

Before going through these solutions in detail, the *de facto* signal processing model for IHR is defined [150].

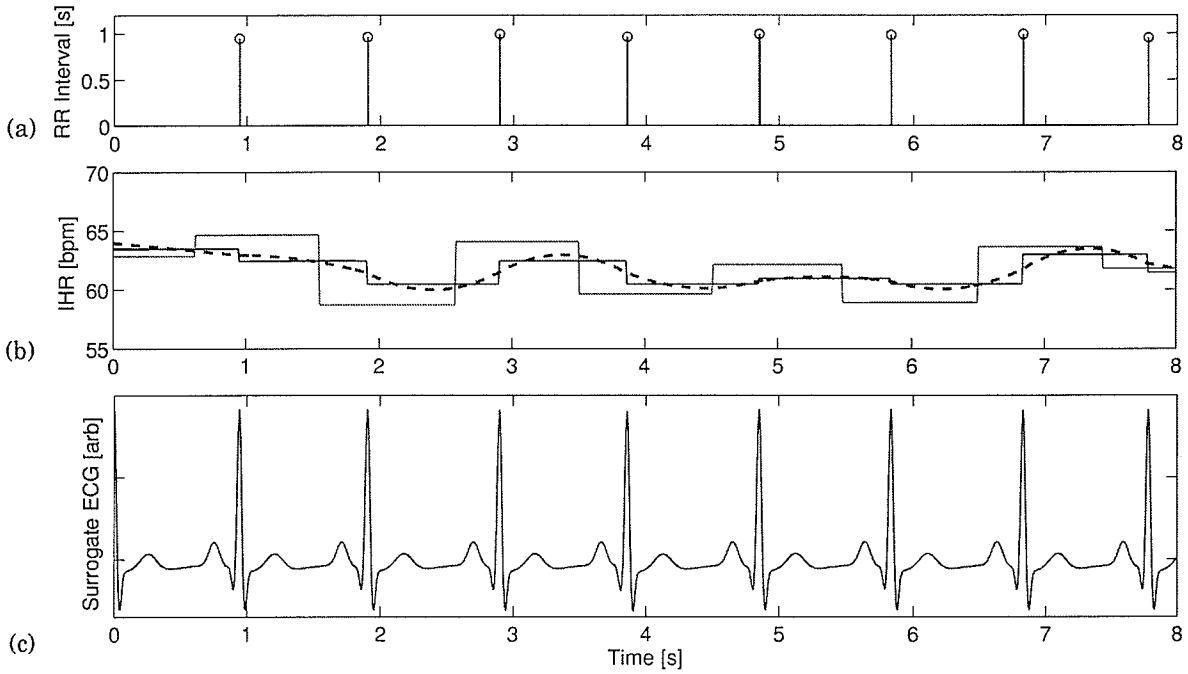


Fig. A.4 FMPI solutions to the surrogate ECG problem: (a) Natural RR-interval sequence from the Physionet NSRD (record 16272); (b) Traditional piecewise-linear interpolation gives constant IHR on the RR-interval (black), while a generalized phase-offset piecewise-linear interpolation can be used to create constant IHR throughout a heart cycle (red). Monotonic Hermite-interpolation methods can provide smooth and continuous IHR (blue dashes); (c) Surrogate ECG signal using the ECGsyn dynamical system for $z(\theta)$ and the Hermite FMPI for IHR.

Definition A.1 (Semicontinuous Tachogram). A *semicontinuous tachogram* is a piecewise-constant extension of a tachogram onto the real line. For the dynamical model used here, the semicontinuous tachogram based an a R-sequence t_n is defined as

$$\omega_{\mathrm{SCT}}(t) = \frac{2\pi}{t_n - t_{n-1}} \quad : \quad t \in (t_{n-1}, t_n] \tag{A.11}$$

in units of rad/s (or simply $(t_{n+1} - t_n)^{-1}$ in units of Hz). This approach to an IHR signal dates back to the original spectral work of DeBoer [52] and Akselrod [24].

Note that, in graphical terms, the semicontinuous tachogram "fills in between" the stem

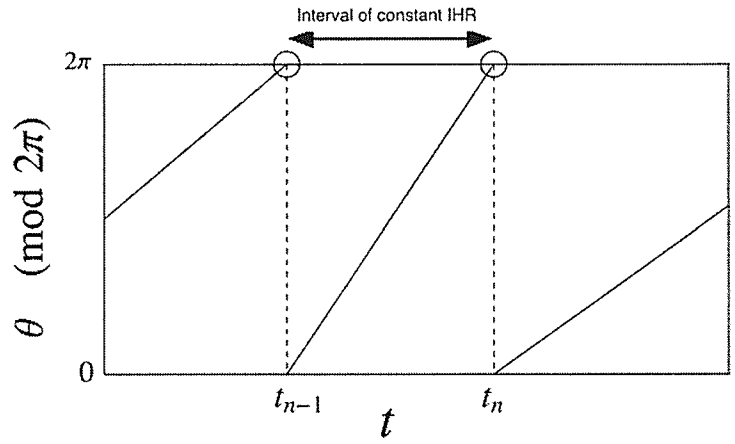


Fig. A.5 Schematic of $\theta(t)$ for the basic piecewise-linear model.

series created by a tachogram (but is not causal because the occurrence of the next R-wave must be known beforehand, $t < t_n$).

A.2.3.2 FMPI: Basic Piecewise Interpolation

The simplest approach to an FMPI solution is to simply apply a basic linear interpolation on the tachogram constraints

$$\theta(t_n) = \theta_n = \theta_R + 2\pi n \tag{A.12}$$

That is,

$$\theta(t) = \frac{t - t_{n-1}}{t_n - t_{n-1}}(\theta_n - \theta_{n-1}) + \theta_{n-1} \quad : \quad t \in (t_{n-1}, t_n] \tag{A.13}$$

$$= 2\pi\frac{t - t_{n-1}}{t_n - t_{n-1}} + 2\pi(n-1) + \theta_R \ : \ t \in (t_{n-1}, t_n] \tag{A.14}$$

and the schematic is shown in Fig. A.5. The linear interpolation assumes that the rate is constant in between the R-waves. Thus, the basic linear FMPI is equivalent to the semicontinuous tachogram of (A.11) through a simple derivative.

Apart from its simplicity, the basic linear FMPI has two other significant characteristics: (i) the average rate over the RR-interval is attained at all points in the interval, and (ii) a

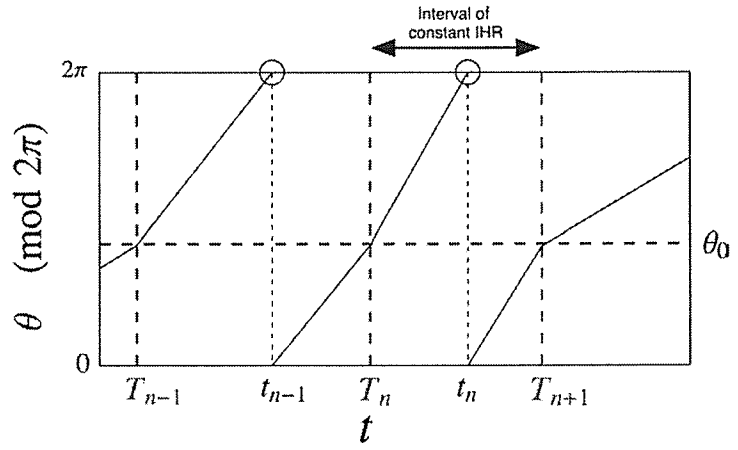


Fig. A.6 Schematic of $\theta(t)$ for the generalized θ_0 phase-offset piecewise-linear model.

discontinuity in $\omega(t)$ occurs at the t_n. Since this discontinuity is placed at the middle of the ECG's electrodynamical cycle described in Ch. 2, the physiological likelihood of this model is quite low. Instead, a small adaptation to the piecewise-linear model can be made to reposition the discontinuity, as described next.

A.2.3.3 FMPI: Offset Piecewise Interpolation

As mentioned, the basic piecewise-linear model has a non-physiological limitation in the placement of its discontinuity. One approach to address this limitation is to offset the discontinuity so it appears at the onset of the beat (*i.e.*, sinoatrial activation) and not in the middle. Thus, the solution is to introduce a phase offset θ_o with respect to θ_R where the discontinuity occurs. If $\theta_o \neq 0$, then the interval of linear interpolation is between new transition times T_n and T_{n+1}, where $T_n < t_n$, as shown in Fig. A.6.

The tachogram constraint on the RR-intervals now affects two values of unknown slope, making a recursive system of equations to be solved

$$\dot{\theta}_n^{-1} = \frac{1}{2\pi - \theta_o}\left(\Delta_R(n) - \theta_o \dot{\theta}_{n-1}^{-1}\right) \tag{A.15}$$

where the shorthand $\dot{\theta}_n^{-1}$ for $\left(\frac{d\theta}{dt}\right)^{-1}\Big|_n$ is used. This reduces correctly to the basic model as $\theta_o \to 0$. Note since the measured RR-interval sequence is being interpreted as a moving average of $\left(\frac{d\theta}{dt}\right)^{-1}$ values, (A.15), it is expected that these $\dot{\theta}_n^{-1}$ values have a greater oscillation than the RR-interval sequence itself. (That is, the RR-interval sequence is a smoothed version of the calculated $\dot{\theta}^{-1}$ values.)

The recursive system of (A.15) is unstable, however, for $\theta_o \geq \pi$. By expressing the system as a reverse recursion (from end to beginning), stability can be regained in the $\theta_o \in (\pi, 2\pi]$ region.

The final form of the phase-offset piecewise-linear IHR interpolation model is formalized as

$$\rho_{\heartsuit}(t) = \frac{60}{2\pi}\frac{d\theta}{dt} \tag{A.16}$$

$$\frac{d\theta(t)}{dt} = \frac{1}{\dot{\theta}^{-1}(t)} \tag{A.17}$$

$$\dot{\theta}^{-1}(t) = \dot{\theta}_n^{-1}, \quad \forall t \in (T_n, T_n + 1] \tag{A.18}$$

where

$$T_n = \begin{cases} t_{n-1} + \theta_o\,\dot{\theta}_{n-1}^{-1} & \forall n \quad : \quad \theta_o < \pi \\ t_n - (2\pi - \theta_o)\,\dot{\theta}_n^{-1} & \forall n \quad : \quad \theta_o > \pi \end{cases} \tag{A.19}$$

and, recursively,

$$\dot{\theta}_n^{-1} = \begin{cases} \frac{1}{2\pi-\theta_o}\left(\Delta_R(n) - \theta_o\dot{\theta}_{n-1}^{-1}\right) & : \quad \theta_o < \pi \\ \frac{1}{\theta_o}\left(\Delta_R(n+1) - (2\pi - \theta_o)\dot{\theta}_{n+1}^{-1}\right) & : \quad \theta_o > \pi \end{cases} \tag{A.20}$$

This model overcomes the direct issue of the discontinuity in the basic piecewise model, but there is no reason why a discontinuity in IHR is required at all. Thus the main approach

of FMPI for a surrogate ECG will look at smooth interpolation schemes, such as the one described in the next section.

A.2.3.4 FMPI: Monotonic Hermite Interpolation

To ensure a continuous instantaneous rate, a piecewise-cubic monotonic Hermite interpolation [62] can be used as a FMPI solution on (A.8). This approach can be simply evaluated in `Matlab` using the `pchip` mode of interpolation. As opposed to cubic spline interpolation, this approach is guaranteed to maintain monotonicity (and thus keep $\theta(t)$ well-defined). Upon taking the derivative, the IHR $\omega(t)$ is guaranteed to be continuous, but it may have singularities (*i.e.*, not be smooth) at the interval transitions.

This approach is the foundation of IHR-processing module of the ECGfm algorithm in order to preserve HRV in the time domain. In order to preserve the time domain reconstruction of shorter ECG event-intervals, such as the PR- and RT-intervals, extra constraints can be added to the FMPI interpolation problem, as described next.

A.2.3.5 FMPI: Supporting Shorter Event-Intervals

A true surrogate ECG should be capable of reconstructing all the event-intervals extractable from a recorded ECG. To this end, the ECGfm algorithm is able to support PR- and RT-intervals as optional arguments, in addition to the basic tachogram information. Specifically, the monotonic Hermite FMPI in the ECGfm algorithm is applied to the interpolation constraints

$$\theta(t_n + \delta_i(n)) = 2\pi n + \theta_i \tag{A.21}$$

where for $i \in \{P, R, T\}$

$$\begin{cases} \delta_P(n) < 0 & ; \text{ PR-interval of } n\text{th beat (offset to } t_n) \\ \delta_R(n) = 0 & ; \text{ no offset to } t_n \\ \delta_T(n) > 0 & ; \text{ RT-interval of } n\text{th beat (offset to } t_n) \end{cases} \tag{A.22}$$

are the time domain offsets provided by the ECG event-interval template, and the θ_i are the static $z(\theta)$-markers for the P-, R-, and T-waves from the ECGfm integrator module (presumed $\theta_P < 0$, $\theta_R = 0$, and $\theta_T > 0$). Clearly, (A.21) simplifies to (A.8) if only R-waves are considered: $i \in \{R\}$.

To accommodate the passing of this ECG template information multiple input options to the algorithm are allowed. Specifically, the template can be presented as a vector or a `Matlab` structure:

VECTOR INPUT— The minimum data requirement is a vector of R-times (in seconds), and the shorter intervals are not constrained. If the R-time vector is given, an extra calling option may be applied, with which a constant PR-interval constraint can be given and applied during FM-interpolation. This is satisfactory to at least model a constant PR-interval independent of HRV, which is a good first-order physiological approximation. The preferred form of use for the ECGfm algorithm, however, is a structure argument because it reduces ambiguity, as described next.

STRUCTURE INPUT— If a vector input is not given, then the ECGfm algorithm expects a `Matlab` structure with the subfields:

R: a vector of R-times;

RR: a vector of RR-intervals (only required if R subfield not available);

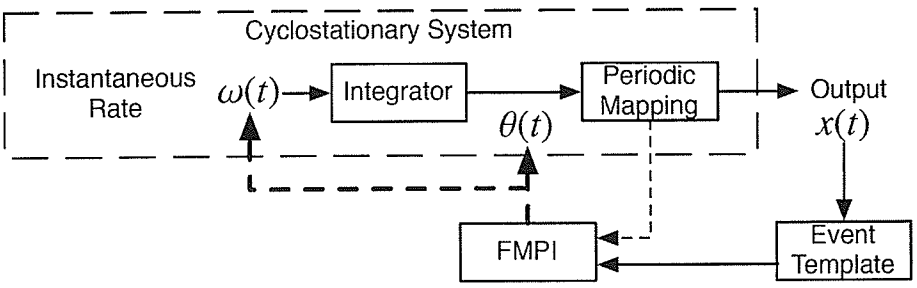


Fig. A.7 FMPI defines an instantaneous rate for a general cyclostationary model.

PR: a scalar value or vector of PR-intervals (optional);

RT: a scalar value or vector of RT-intervals (optional).

By convention, all event-intervals must be in seconds and positive. (Note the negative sign required for δ_P in (A.21) is internally applied.)

From the data in the input structure fields, the IHR-processing algorithm processes calculates all the individual $t_i(n)$ and $\theta_i(n)$ as necessary, then sorts them into aggregate monotonic $t_{Vi}(n)$ and $\theta_{Vi}(n)$ vectors. Then the θ vector is interpolated via the `pchip` interpolation onto a uniformly sampled time series over the $t_{Vi}(n)$ vector. As described in Sec. A.2.2, the ECG time series begins with an R-wave (*i.e.*, $\theta(t)$ starts at 0). As a consequence, the first PR interval value is applied at the first opportunity: namely on the *second* beat. Note, therefore, that to properly maintain the correspondence of a data input to a recorded ECG, some care is generally required when preparing a multi-structure argument.

A.2.3.6 Generalized FMPI and the IPFM Model

FMPI is a general approach to the definition of instantaneous rate for cyclostationary signals modelled by the system of Fig. A.7, and extends the techniques presented in [76],[2],

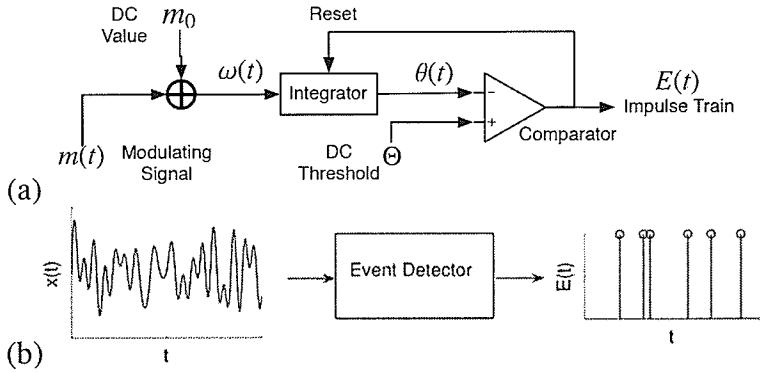


Fig. A.8 The IPFM model: (a) Design of an IPFM circuit; (b) IPFM system converts a continuous-time modulating signal into an impulse train (point process).

and [176]. As it has been described here, FMPI is related to the *integral-pulse frequency-modulation* (IPFM) model [76][157][138]. The IPFM model is often used as a "ramp-and-fire" model for biological processes, such as the depolarization of the sinoatrial node [157]. It is also used as a model for the extraction of a point process (event impulse series) from a continuous signal, *e.g.*, R-wave detection [2] as shown in Fig. A.8. The latter is essentially a conversion from continuous ECG (with a well-identified R-wave), into the sequence of R-wave times t_n.

The IPFM change of variable is often useful when applied to derive approximations in HRV spectral estimation [157]. The power of the IPFM model is it naturally describes the cyclic behaviour of the ECG by the relationship between time and an imaginary phase variable θ. It has been noted by the author [176], however, that this imaginary phase variable has important consequence in the context of the surrogate ECG/ECGsyn model: $\theta(t)$ is, in fact, the angular variable of Fig. A.2.

The goal of synthesizing an ECG with a given tachogram is, in one sense, the "inverse" of the IPFM point-process model shown in Fig. A.8(b). Here, a dynamical system as described in the earlier sections (Fig. A.2) is used to reconstruct the ECG morphology discarded by the

IPFM transformation. (Note this is independent of the biology of the sinoatrial node, and based solely on the cyclostationarity of the ECG waveform itself. The biological connection to the pacemaker activity is only necessary when conclusions are to be drawn specifically on the characteristic behaviour of autonomic influences on HRV.)

The idea of inverting the IPFM model is not a new idea (*i.e.*, IPFM demodulation). However, it is usually approached with some underlying assumptions in order to obtain a unique solution (*i.e.*, bandwidth assumptions that validate the spectral approximations [205]). This is, in practice, approximate linearized solutions to the nonlinear IPFM system.

By comparing Fig. A.8(a) to Fig. A.7, it is easily determined that IPFM can be considered a special case of the general cyclostationary system to which FMPI applies. While IPFM is limited to an impulse train output, however, the cyclic decomposition model for cyclostationary signals $x(t) = z \circ \theta(t)$ can be widely applied, as shown in Fig. A.9. For example, the FMPI approach to instantaneous rate also applies to

(a) *Non-uniform sampling* (Fig. A.9(b)), where sample values $x(t_n) = x_n$ are applied so that $z(\theta(t)) = \sum_n x_n \delta_1 \theta(t) - 2\pi n$;

(b) *Sinusoidal frequency modulation* (Fig. A.9(c)), such as $z(\theta(t)) = \cos(\theta(t) - \phi)$;

(c) *Pulse width modulation* (Fig. A.9(d)), such as $z(\theta) = \Xi((2\theta \mod 2\pi) - \pi/2)$, (using the Heaviside function Ξ);

(d) *Cyclostationary modelling* (Fig. A.9(e)), where $z(\theta(t)) = \int g(z, \dot{z}, \theta, \dot{\theta})dt$, as was described in the preceding sections for ECG or elsewhere [140] [43].

Note that the analysis that follows in Sec. A.3 is based on Fig. A.9(e), which is effectively using the cyclic dynamical system of Sec. A.2.2 as $z(\theta)$ to reconstruct the ECG morphology discarded in the reduction of an ECG to a tachogram. As was discussed in Sec. A.2.3, it is

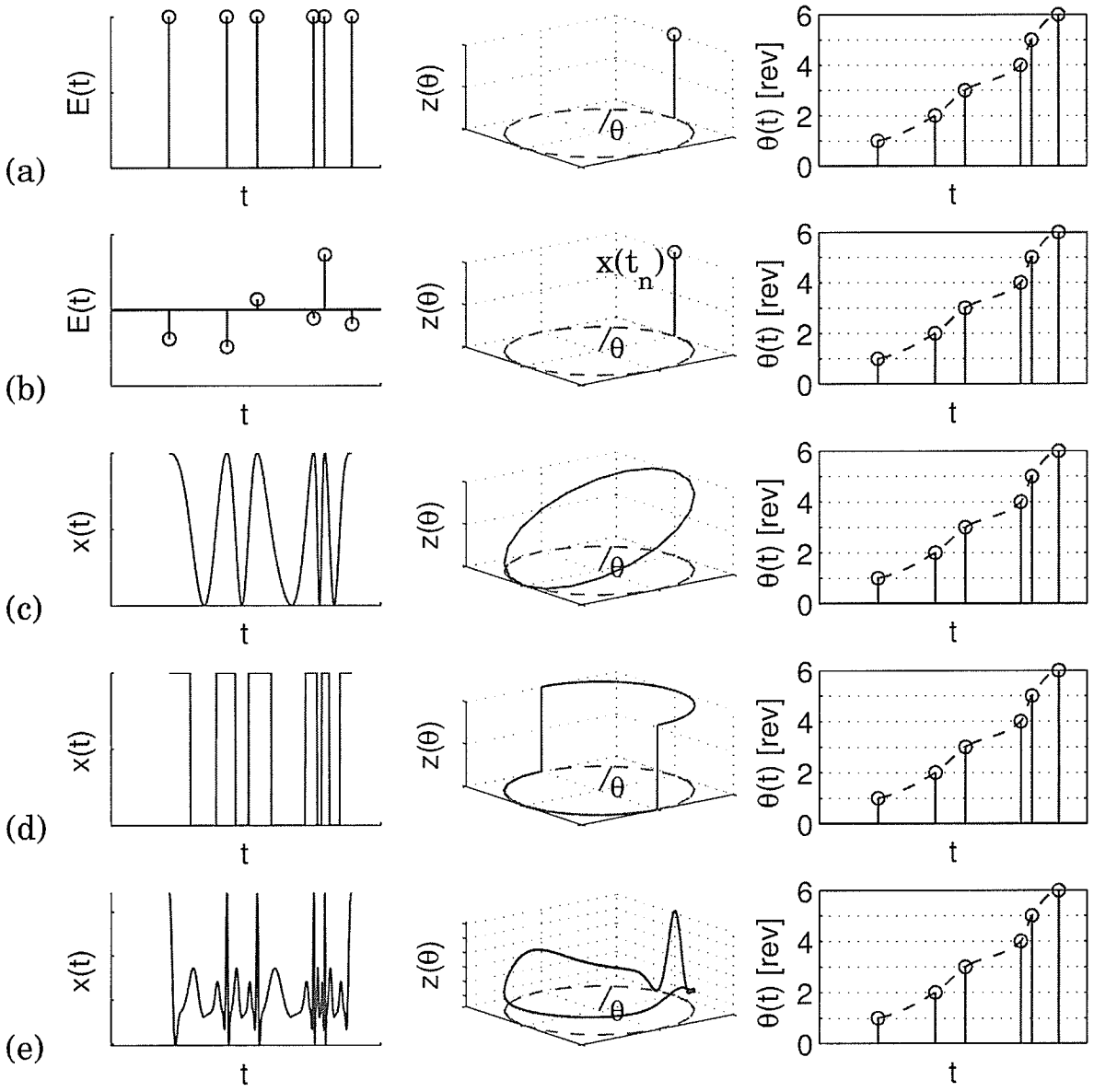


Fig. A.9 FMPI universality: FMPI applies to the instantaneous rate inverse problem of diverse signals including; (a) IPFM, (b) nonuniform sampling, (c) classical frequency modulation, (d) pulse width modulation, and (e) cyclostationary modelling, such as ECG synthesis. FMPI solves for a phase function under the common phase constraint that appears down the right hand column.

also particularly useful for this case that the FMPI model will accept multiple event-markers per cycle to preserve the short event-intervals.

By this analysis, therefore, FMPI can be considered a generalized extension to the IPFM demodulation problem. In contrast to the IPFM demodulation formulation, a numerical solution to FMPI requires a choice of interpolation scheme. As mentioned previously, this flexibility is consistent with the ill-posedness of the inverse problem on short-time scales. Since the defining FMPI constraint is preserved for all interpolation orders, this choice of interpolation order need only be justified based on the solution's utility and physiological credibility, instead of "correctness". That is, unique solutions are not available, but *useful* solutions can be described that provide mathematical insight. The preceding sections present an example of this, whereby new solutions were derived within the same IPFM context solely on the basis of the qualitative continuity of the final IHR signal.

It will be shown in the following that FMPI actually improves the time-domain fidelity of the synthesized ECG to the usual *tachogram resampling* techniques [150] for defining IHR. This is not totally surprising. Han identified as early as 1992 [76] that IHR based on a special case of FMPI (which interpolates than differentiates) yields different results than resampling methods (which numerically differentiates then interpolates). Furthermore, Mateo, Laguna, *et al.* [138] considered the spectral analysis of a potpourri of IHR techniques and determined that a so-called *heart timing signal* was the most unbiased. This signal, as it is defined with an IPFM model, amounts to a high-order spline interpolation of the *linear residual* of $\theta(t)$; *i.e.*, $m(t) = \theta(t) - \mu_1 \theta t$ [178]. The significance of this superiority against the other resampling interpolation methods, however, was not identified. From consideration of Han *et al.* and the arguments that follow in Sec. A.3.3 and in [178, Potter (2008)], it is clear that an IPFM interpolation on $\theta(t)$, or even its nonlinear residual $m(t)$, captures the constraint of the FMPI which

no interpolation on cumulative differences (*i.e.*, tachograms) ever will. This is significant because some authors will argue for direct spectral estimation of HRV from tachograms (*e.g.*, Lomb-Scargle estimation [44, Clifford (2005)]), but as will be shown in Sec. A.3.3, even this makes an invalid assumption regarding the "non-uniform sampling" of the tachogram.

A.2.4 System Attractor

Since the reconstruction of an ECG attractor from a time series is featured heavily in this work, it is of interest to consider the implications of the ECGfm model on this attractor. By design, ECGfm signals come from the attractor of a driven nonlinear oscillator, Fig. A.2. Note that, from the equations of motion, the morphological dynamics of surrogate ECG can be visualized in a three-dimensional system, but the variability dynamics contained in $\theta(t)$ are not represented here at all. IHR is essentially a conjugate variable to the ones depicted. Meanwhile, the reconstructed attractor is based on lag-embedding of the observed time series alone, as described in Chs. 4 and 6, and is to be analyzed as a low-dimensional manifold embedded in a high-dimensional space.

As a consequence of the surrogate ECG technique, the "ECG attractor" can also be considered at a conceptual midpoint between these two ends. Consider the images shown in Fig. A.10. Here a 2D lag-embedding of the ECGfm time series is enhanced by the IHR variable on the vertical axis. Images here were created by tachograms that increase monotonically across the normal heart rate range, while PR- and RT-intervals were constrained to constants. As compared to Fig. A.2, the manifold shown here reduces the morphological variables by one (and hence introduces false intersections) but realizes one of the conjugate variables instead; *e.g.*, IHR on the vertical axis. Thus, this representation captures a dynamical relationship between the angular rate and the ECG morphology as a geometry to the manifold.

By observation of these figures, it is observed that a larger spreading in IHR occurs in the

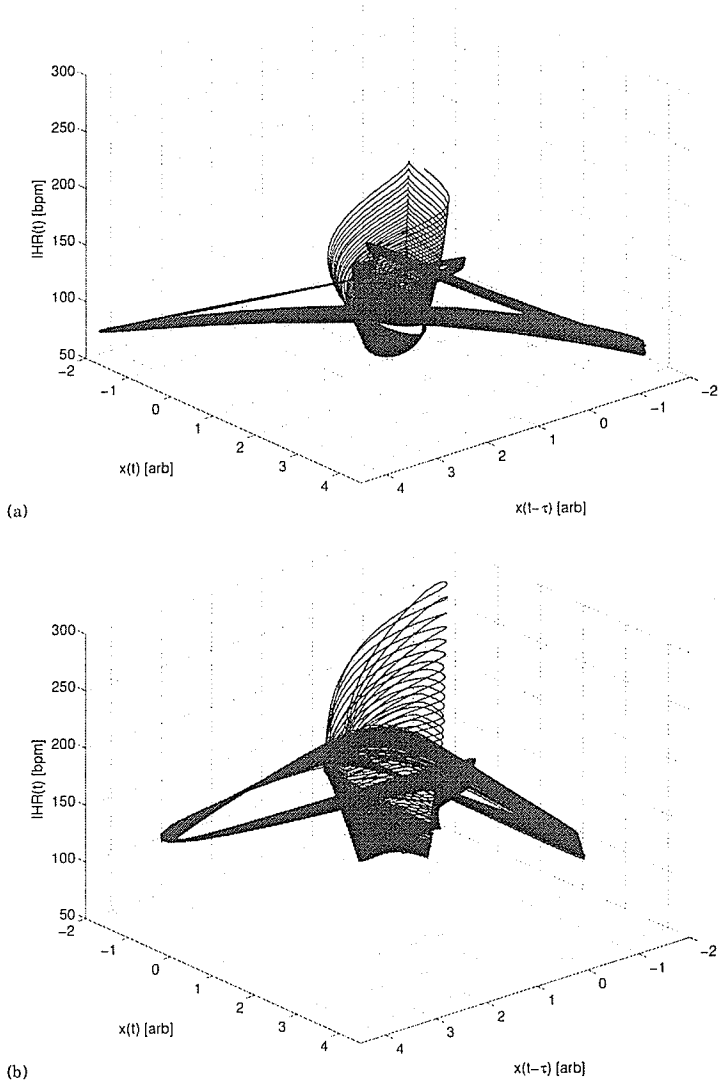


Fig. A.10 Visualization of surrogate ECG attractor support: The interactive dynamics of ECG morphology (base plane) and heart rate (vertical axis) for (a) adult parameters; and (b) fetal parameters constructs a manifold, observed here in 3D. Thus, the attractor of an observed ECG can be interpreted as a specific distribution on this manifold, depending on its IHR properties.

loop that represents the TP-interval as compared to the "ribboned" structure of the rest of the trajectory. These images can be interpreted as a projection in 3D of the support of the higher-dimensional attractor of the surrogate ECG model (with some assumed parameters). That is, the figures shown here do not capture the weighting or dynamics induced by the tachogram input, but only the manifold on which they act. A specific tachogram would constrain the ECG signal to traverse the shown patterns according to some specific density function. That is, the dynamics of a specific IHR time series leads to a density feature on the morphologically-based structure in Fig. A.10.

Therefore, an ECG attractor, once fully-reconstructed by lag-embedding, can be considered as a segment of the unraveled manifold of Fig. A.10, with a weighting according to its IHR dynamics. It is the multiscale behaviour of this distribution on this manifold that is being measured in multifractal analysis. This progression from ECG and IHR time series to a reconstructed attractor equivalent to Fig. A.10 is shown in Fig. A.11.

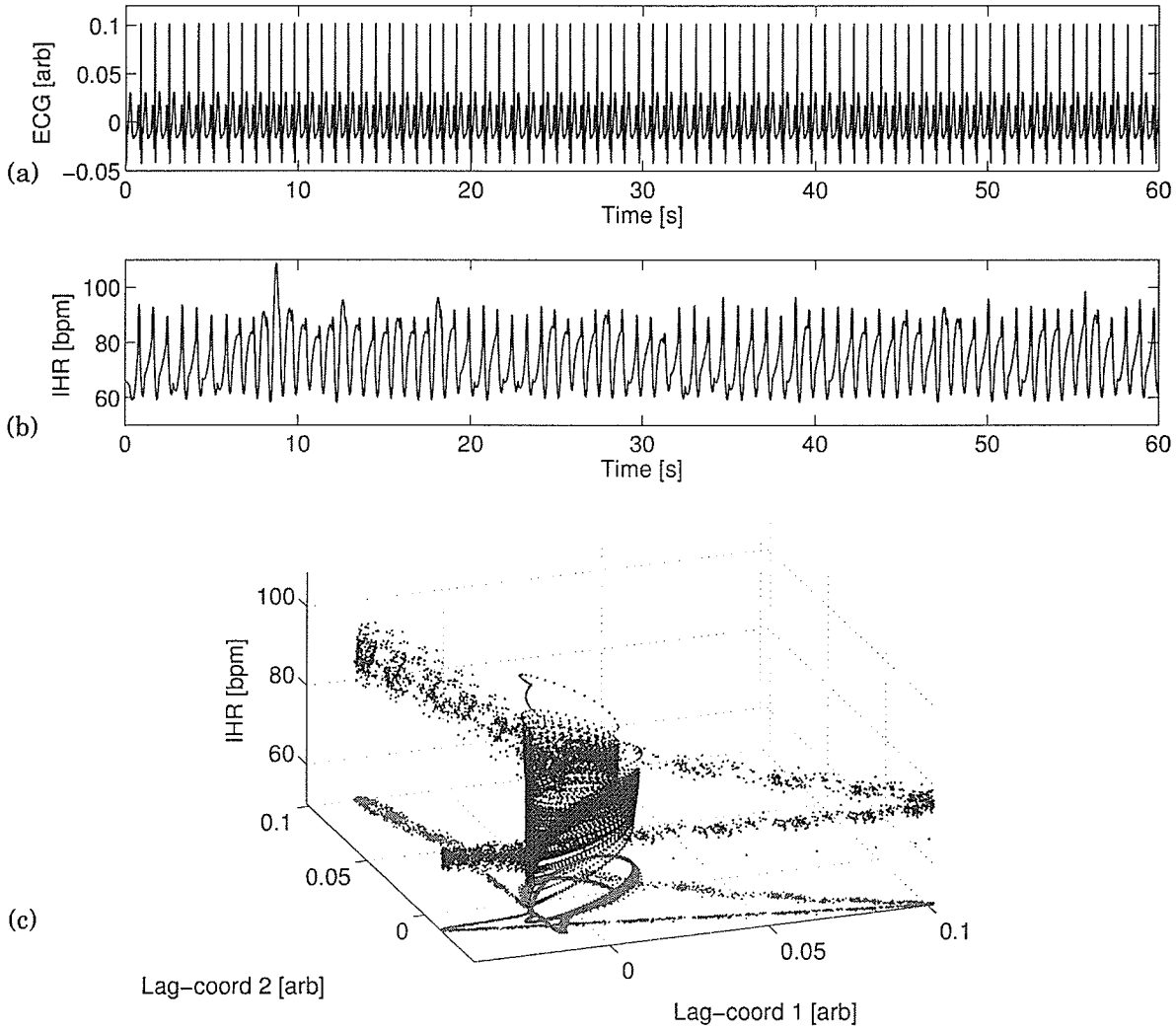


Fig. A.11 Visualization of specific surrogate ECG attractor: (a) synthesized time series which yields lag-coordinates; (b) its IHR time series; and (c) attractor specific to (a) and (b) reconstructed in 3D.

A.3 Changes to the ECGsyn Design

As mentioned in Sec. A.2, the ECGfm algorithm for surrogate ECG synthesis is based on the ECGsyn model by McSharry *et al.*. The following considers the differences that occur between the two algorithms when they are applied to the same input, and demonstrates the superior performance of ECGfm at preserving time domain fidelity.

A.3.1 Comparison of Synthesis Behaviours

Here the ECGfm algorithm presented in Sec. A.2 is compared to the `ecgsyn24.m` algorithm, which is the MIT-based (Clifford, 2005) extension to the original Oxford-MIT ECGsyn model. The distinction between the original ECGsyn and `ecgsyn24.m`, is that the former produced synthetic RR-intervals internally, whereas the latter could also accept RR-intervals as inputs. (Since the rest of the algorithms are equivalent, they will both still be identified as ECGsyn hereafter).

A simple example of the different response of the algorithms is shown in Fig. A.12. Here, a short piecewise-constant square-wave tachogram was presented to the algorithms. There is considerable amplitude distortion in the ECGsyn response, as well as a lack of time domain fidelity. The ECGfm approach, however, is more resistant to these artifacts. This is a result of a few minor changes to the integrator module, and the introduction of FMPI into the IHR-processing module. The use of FMPI specifically improves the RR-interval fidelity, as well as the shorter event-interval fidelity. In the following, these differences between the algorithms will be addressed in the context of experiments with natural intervals.

A.3.1.1 Experiments with natural intervals

To demonstrate the event-interval response of the `ecgsyn24.m` and surrogate ECG algorithms with natural ECG behaviour, contiguous segments of annotated normal ECG beats in

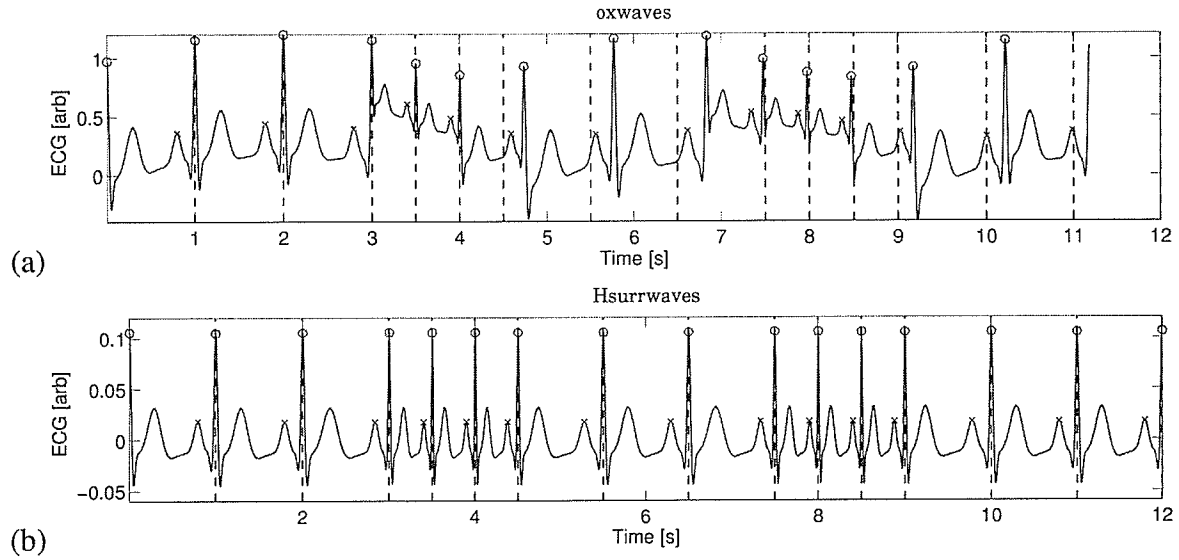


Fig. A.12 ECG synthesis results on square-wave tachogram: (a) ECGsyn algorithm displays artifacts in amplitude and time-domain fidelity; (b) ECGfm algorithm eliminates these artifacts. R-times t_n at the input are identified by dashed vertical lines.

sinus rhythm have been extracted from the Physionet QT database [122][68][1]. It is remarked that these selected recordings are not necessarily representative of universal HRV behaviour, but do indeed provide a documented dataset that can be used for the evaluation of ECG synthesis algorithms at the reconstruction of natural ECG behaviour. Figure A.13 shows a summary of the event-intervals for one of the datasets studied. Here, Fig. A.13(a) shows the tachogram of 100 consecutive beats drawn from the `sel16273.pu0` dataset. The PR- and RT-intervals for the same ECG is shown as sequenced by beat number in Fig. A.13(b) and (c) respectively. These event-interval features are also collected into a scatterplot in Fig. A.13(d) as a function of RR-interval length. Note that the PR- and RT-intervals do not follow a linear relationship with the RR-intervals. In fact, they exhibit very little variability (within the quantization produced by the 250 Hz sampling rate of the ECG). Thus, as is shown more explicitly by comparing the ECG traces of consecutive intervals as shown in Fig. A.14(a), it is observed that the

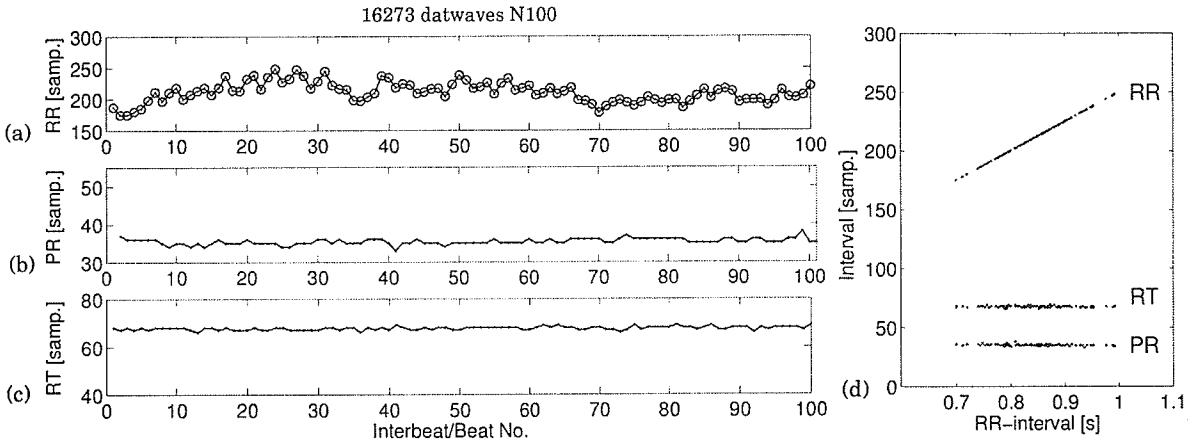


Fig. A.13 The event-interval behaviour of 100 beats from a natural sinus rhythm (Physiobank QT database record `sel16273.pu0`). (a) RR-interval sequence; (b) PR-interval sequence; (c) RT-interval sequence; (d) relative behaviour of intervals in samples in relationship to HRV (RR-intervals).

HRV is concentrated into the interbeat interval (*i.e.*, TP-interval). It is on this basis that HRV has been analyzed in the context of a stochastically driven oscillator [107].

Consecutive beats from the synthesized time series are presented in Fig. A.14 along with the beats from the original recording they represent. Throughout the remainder of the chapter, references will be made to the distinct panels in this figure. The impact of these distinctions best appears when in direct comparison, therefore, they are all presented here against the natural behaviour.

Now certain details in the ECGfm and ECgsyn integrator modules responsible for changes in surrogate ECG response will be demonstrated.

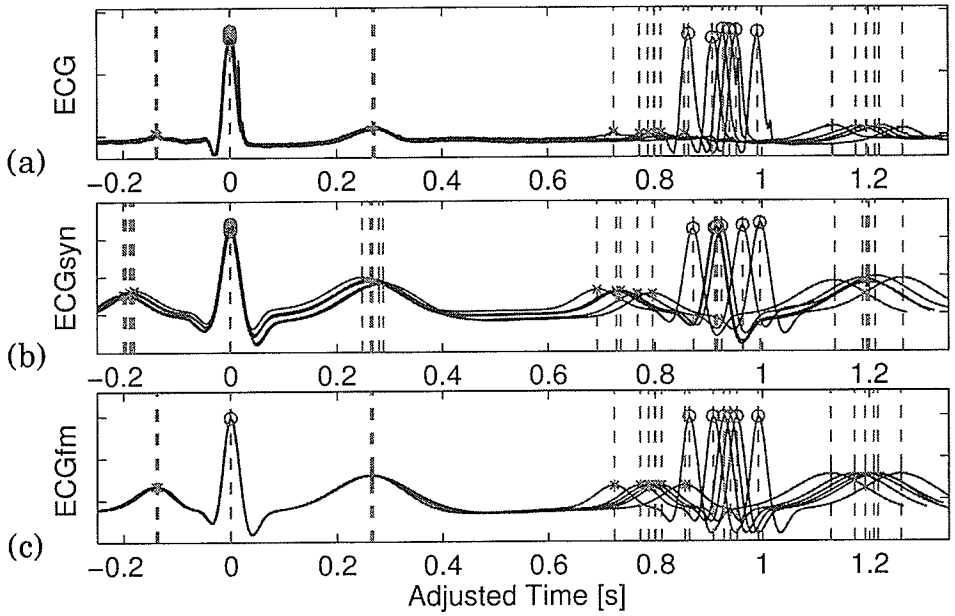


Fig. A.14 Evaluating time domain fidelity of IHR: The variability of 6 consecutive ECG beats from (a) a natural ECG recording (Physiobank QT database record `sel16273.pu0`) with HRV concentrated in the interbeat interval as a template, and equivalent (b) reconstructed ECG using a TR variant [41], and (c) reconstructed ECG using an IHR from cubic-Hermite FMPI [179]. Note the poor alignment of the R- (circles), P- (crosses), and T-waves (plusses) in panel (b) with the template. Alignment improves with the FMPI reconstruction of panel (c).(Differences in gross morphology are negligible since the morphology of the synthetic models was not tuned to match panel (a)).

A.3.2 Integrator Module

A.3.2.1 Low-Frequency Baseline

The ECG signals from the QT-database analyzed through this work did not exhibit strong low frequency baseline wander. On this basis, the $z_0(t)$ baseline wander term in Sec. A.2.2 defaults to zero. In ECGsyn, this term defaults to a non-zero sinusoid, which complicates the comparison of the techniques. Thus for all the experiments that follow, the ECGsyn low-frequency baseline was set to zero for consistency with ECGfm.

A.3.2.2 Amplitude Correction

As presented in Fig. 9 of the original ECGsyn article [140, (2003)], the QRS-amplitude in the ECGsyn model has a linear relationship to RR-interval. This strong linear coupling of the amplitude to the RR-interval may seem acceptable under normal HRV conditions (as [140] argues), but it is also responsible for the strange amplitude effect in the square wave response of Fig. A.12. The origins of this effect is that a lower angular rate $\omega(t)$ allows the nonlinear dynamical system of (A.6) more time to respond to the repellors, and thus leads to greater deviation from the baseline. The ECGfm model compensates for this behaviour by normalizing the repellor activation in (A.6) with the leading $\omega(t)$ factor, which is not present in the ECGsyn equations. This effectively makes the time integral (dt) of the repellor response into an angle-integral ($d\theta = \omega(t)dt$), and hence independent of instantaneous rate. To some degree, this is likely on over-compensation, as is shown in the amplitude response of natural ECG and the synthetic ECG methods of Fig. A.15-A.17.

Throughout these figures, panel (b) shows the scatterplot of the QRS-amplitude, defined here as the $(\max_{\delta t(R)} x(t)) - (\min_{\delta t(R)} x(t))$ where the small time window $\delta t(R)$ includes all the waves of the QRS-complex, in units of its mean as a function of the trailing RR-interval.

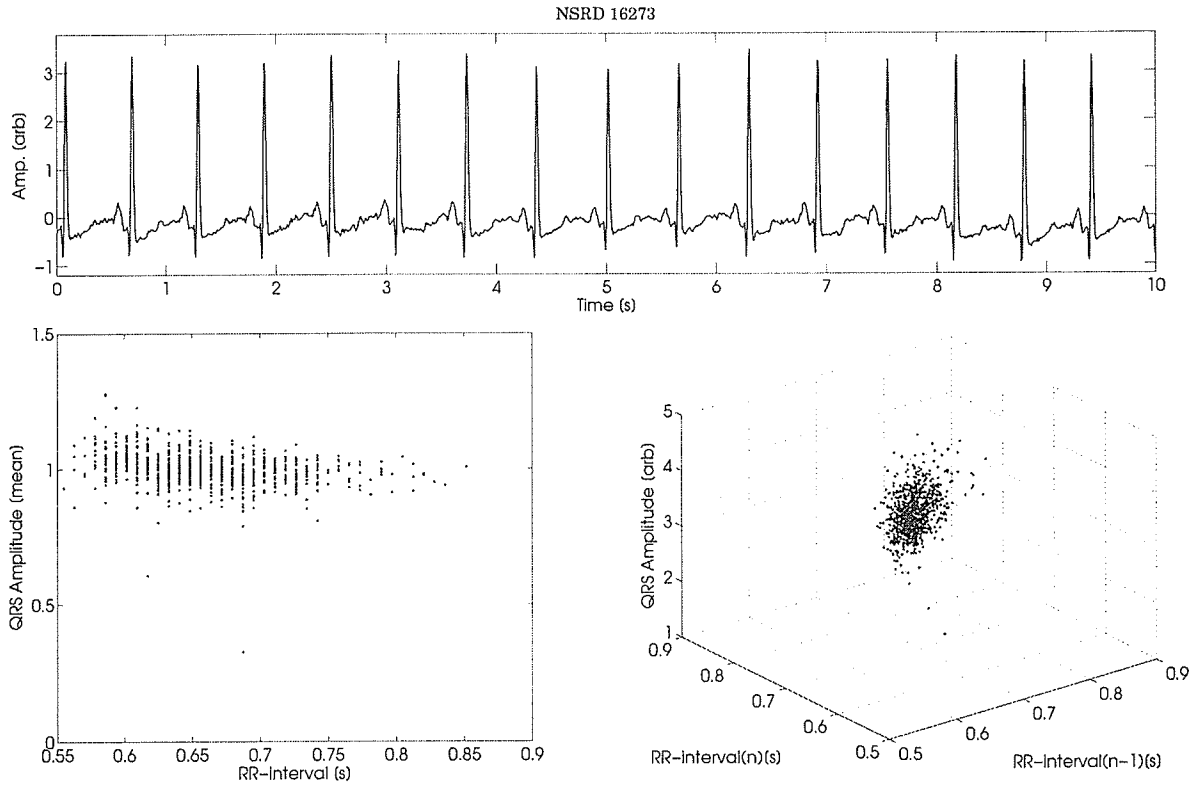


Fig. A.15 Natural QRS amplitude as function of heart rate: (a) Example of ECG time series; (b) QRS amplitude as a function of the trailing RR-interval (in units of its mean). Note the statistical spread, but no obvious trend; (c) Since QRS-complex occurs between RR-intervals, QRS amplitude as a function of preceding and trailing RR-intervals. Note that RR-intervals are correlated, but amplitude has no obvious trend.

The data for Fig. A.15 is a recorded natural female adult ECG, while A.16 is the ECGsyn-based surrogate ECG from the natural tachogram of the former figure. Likewise, A.17 is the ECGfm-based surrogate ECG from the same natural tachogram. There is significantly less variability in the synthesized signals, likely due to the elimination of baseline wander and noise. The ECGsyn relationship, however, is strongly linear and not consistent with the natural data. With the modification to the integrator module used in ECGfm, this linear relationship is eliminated, as shown in Fig. A.17.

Panel (c) of these figures also shows the relationship to the preceding RR-interval. No

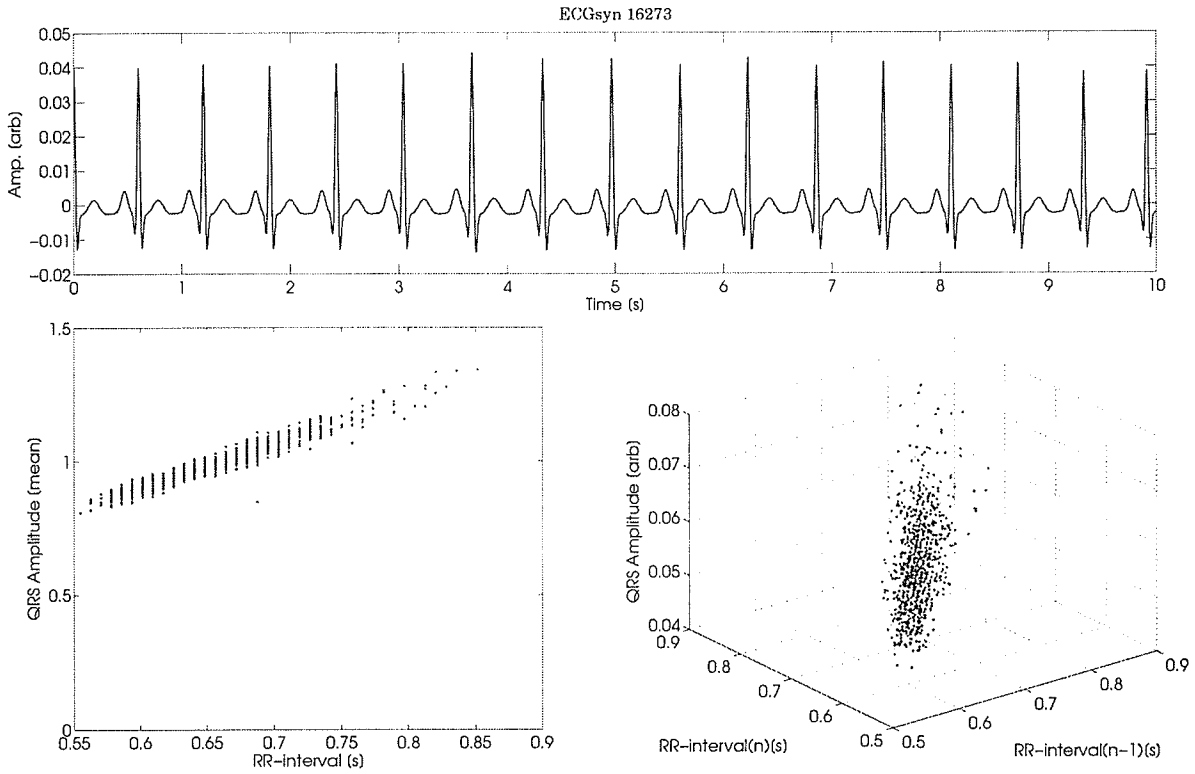


Fig. A.16 ECGsyn QRS amplitude as function of heart rate: (a) Example of time series; (b) QRS amplitude as a function of the trailing RR-interval (in units of its mean). Note there is significantly less statistical spread than Fig. A.15, but there is an obvious linear trend; (c) QRS amplitude as a function of preceding and trailing RR-intervals provides no extra information. Again, low-frequency baseline eliminated.

significant additional behaviour is observed, however, since the RR-intervals are already naturally correlated.

Last, note that by default in the ECGsyn algorithm, the entire time series output of the nonlinear dynamical system is rescaled just before it is returned by the algorithm. The goal of this rescaling is to match the amplitude values of the time series to millivolt values expected from a recorded ECG. In ECGfm this rescaling is turned off by default since the simulation of fetal ECG (i) would have different amplitudes altogether, and (ii) are indeterminate under ICA anyway. Since the simulation method of this work requires that the signals be preprocessed

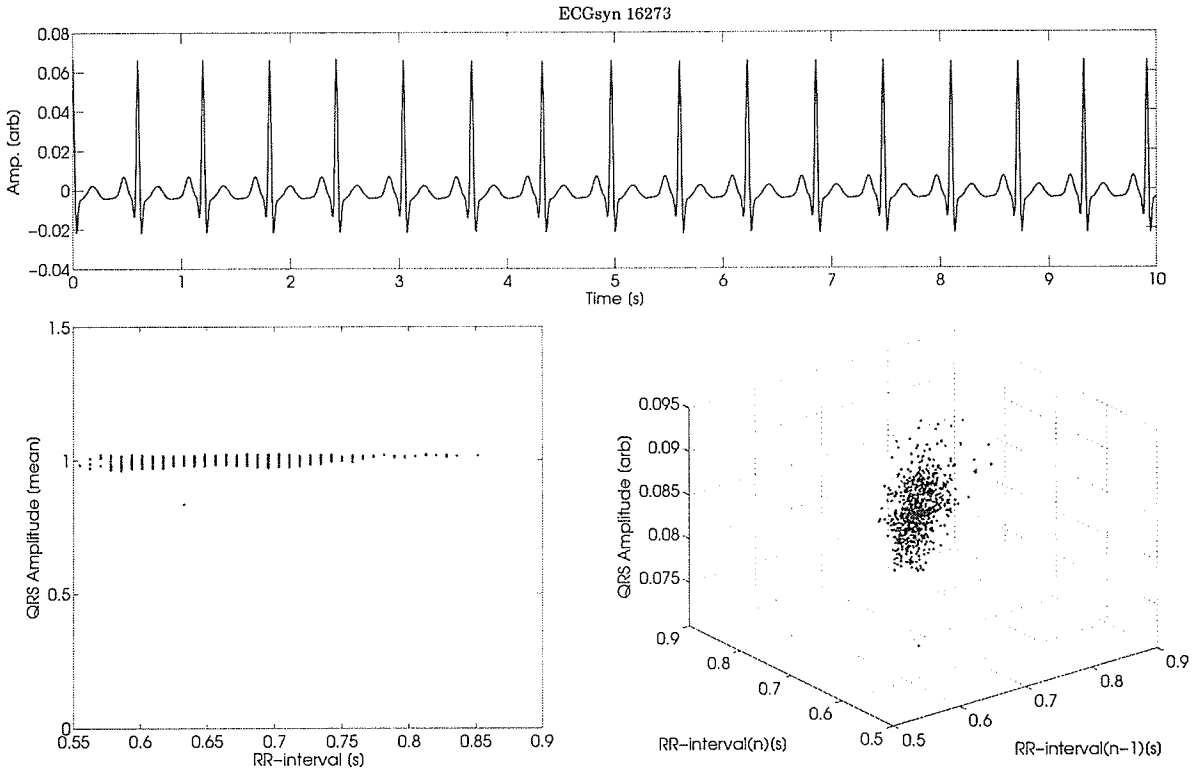


Fig. A.17 ECGfm QRS amplitude as function of heart rate: (a) Example of time series; (b) QRS amplitude as a function of the trailing RR-interval (in units of its mean). Note there is significantly less statistical spread than Fig. A.15, but the linear trend is removed; (c) QRS amplitude as a function of preceding and trailing RR-intervals provides no extra information. Again, low-frequency baseline eliminated.

and also postprocessed to a standard variance, the natural range of values produced by the dynamical system are not a concern.

A.3.2.3 Transient Correction

As noted in Sec. A.2.2 and Sec. A.2.4, the geometry of the nonlinear dynamical system limit cycle is dependent on the mean heart rate of the driving $\omega(t)$ signal. This makes any fixed initial condition to the system unable to be on the attractor in all conditions. Thus the ECGfm algorithm uses a "transient control" protocol, where a preliminary ECG transient is

synthesized with a constant tachogram (*i.e.*, a repeated single-beat IHR-profile) and whose sole purpose is to provide an estimate for an initial condition on the ECGfm attractor for the given heart rate.

This transient control protocol is novel to the ECGfm algorithm, and is not present in the ECGsyn algorithm integrator module. Since the ECGsyn algorithm used a fixed initial condition, it really only is suitable for ECG generation with mean heart rates near 60 bpm. Very obvious transient behaviour is observed when ECGsyn is applied to fetal rates.

A.3.3 IHR-processing and RR-Interval Fidelity

As mentioned in Sec. A.2.3, the fundamental purpose of the IHR-processing module is to generate a continuous IHR time series that can drive the integrator module in such a way as to be consistent with any discrete event-interval observations prescribed at the input. Thus a surrogate ECG can be synthesized to fit (i) natural HRV and (ii) natural event-intervals. To achieve this end, the FMPI technique is used to define $\omega(t)$.

This approach is distinct from the ECGsyn algorithm which is focused on preserving spectral, not time domain, HRV information. As a result, the properties of HRV in surrogate ECGs by ECGsyn and ECGfm can be drastically different. In [179, Potter (2008)] and [178, Potter (2008)], arguments are presented that identify the improved time domain fidelity that FMPI provides to the surrogate ECG process. The following sections present the same conclusions with mostly complementary material.

A.3.3.1 Tachogram Resampling and Time-Domain Fidelity of RR-Intervals

As in Sec. A.2.3, consider the set of R-wave timings t_n of an ECG recording, $n \geq 0$. Letting $\Delta_R(n) = t_n - t_{n-1}$, the ECG's tachogram is then the bounded sequence

$$(n, \Delta_R(n)) \; ; \; n \geq 1 \tag{A.23}$$

By convention [52][24][150], (A.23) is also extended to the similar expression

$$(t_n, \Delta_R(n)) \; ; \; n \geq 1 \tag{A.24}$$

which represents a non-uniformly sampled time series. *Tachogram resampling* (TR) is the technique of applying interpolation methods to (A.24) to get a signal $T(t)$, from which IHR is then defined as

$$\omega(t) = \frac{2\pi}{T(t)} \tag{A.25}$$

The earliest examples of TR consist of a piecewise-constant resampling [52][24], whereby the rate is assumed constant within the event-interval. This is identical to the SCT defined on page A-12. Linear and higher-order resampling methods are just as simple to apply, but their final form is less elegant analytically. Usually, the higher-order methods are preferred [150][138], since it is physiologically unlikely that $\omega(t)$ be discontinuous. The relative merits and limitations of the various interpolation approaches for HRV resampling have been addressed by many authors, but mostly in the context of power spectrum analysis. Here, however, the discussion is limited to the cubic spline interpolation that is used by default in the `ecgsyn24.m` algorithm [41] to be tested.

As mentioned earlier, tachogram resampling ignores the observation made by Han *et al.* [76] that differentiation and interpolation are not commutable operators . While this may be of smaller significance when limited to spectral analysis, it was observed in the test sequence of Fig. A.12 that time-domain fidelity was an issue, and will now be shown to have significant impact when using real ECG data.

By applying the `ecgsyn24.m` algorithm to the extracted real tachogram of Fig. A.13(a), a synthetic ECG is generated at 250 Hz for testing against the original ECG. (Note that the morphological elements of the model were not tuned during synthesis, and, therefore, differences in ECG morphology are considered inconsequential — only timings of the events are significant.) Event-intervals of the synthesized time series are then extracted from the (internal) peak markers identified by the annotating algorithm in Sec. A.2.2. (Note Since the synthesized ECG is annotated differently than the natural ECG, some nominal systematic mismatch must be tolerated.)

The original event-intervals of the natural ECG template are compared to those of the ECGsyn reconstruction in Fig. A.14(b) and Fig. A.18. The mismatch between the 6 R-waves

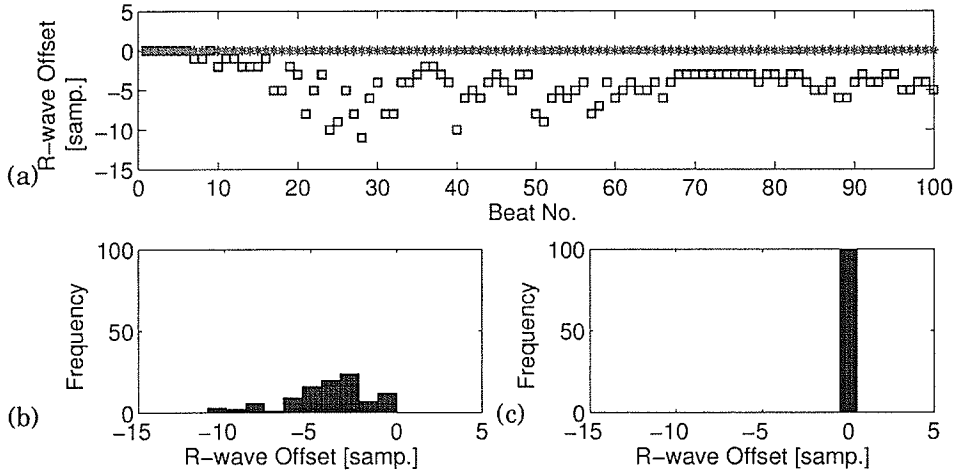


Fig. A.18 Errors in reconstructed R-wave positions for 100 natural beats (Physiobank QT database record `sel16273.pu0`): (a) As a function of beat number, the R-wave offset in the `ecgsyn24.m` reconstructed time series (squares) and ECGfm (stars); (b) Frequency of R-wave offsets for the `ecgsyn24.m` algorithm, and (c) the modified algorithm.

(circles) in the former with the original ECG in Fig. A.14(a) demonstrates the poor time domain fidelity of the `ecgsyn24.m` algorithm. A global analysis over the entire reconstruction of 100 beats is further illustrated in Fig. A.18(a) and (b). Here the offset of the reconstructed R-wave to the original is shown as a function of beat number (black squares). It is observed that the output R-waves are systematically resolved prematurely (*i.e.*, negative offset), and is not simply an accumulation of errors, since the worst offsets occur before beat 30 and then improves thereafter. In other experiments, R-waves were premature by up to 80 ms (`sel16795.pu0`, not shown). A common consequence of the poor time domain fidelity is that, for long input sequences, more beats are synthesized at the output than is actually prescribed at the input.

The magnitude of this systematic artifact, however, is not consistent and depends on the tachogram input. The practical effect of resampling on the tachogram data does not degrade time-domain fidelity uniformly. In the experiments shown in Fig. A.19 (using QT-database

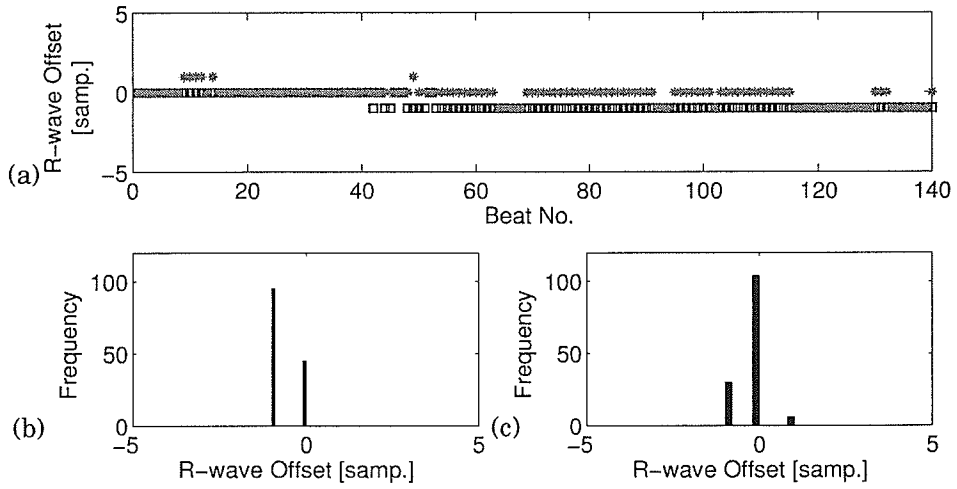


Fig. A.19 Errors in reconstructed R-wave positions for 100 natural beats (Physiobank QT database record `sel16272.pu0`): (a) As a function of beat number, the R-wave offset in the `ecgsyn24.m` reconstructed time series (squares) and ECGfm (stars); (b) Frequency of R-wave offsets for the `ecgsyn24.m` algorithm, and (c) the modified algorithm.

record `16272.pu0`), the temporal position of the R-waves were well preserved, in contrast to the previous figures (using QT-database record `16273.pu0`). For the data shown, all R-waves in the synthesized ECG were within one sample of the template ECG used for input.

It is true that the original ECGsyn algorithm was focused on preserving the spectral behaviour of a tachogram, and not its time domain features. These current observations have shown, in fact, that the extended ECGsyn algorithm (which accepts a tachogram input with well-defined time domain features) does not preserve the time domain information of the input tachogram at the output, even though its HRV may be spectrally equivalent. In fact, it is observed in the following that the RR-interval invariance of the `ecgsyn24.m` algorithm performed best when its resulting resampled $\omega(t)$ time series were close to the FMPI result.

Note that the `ecgsyn24.m` algorithm applies two processes to the tachogram input to define the IHR signal. First, cubic-spline TR is used, and, second, an SCT extension is resolved. Thus the final IHR time series is a step function, but its values are not determined via a process

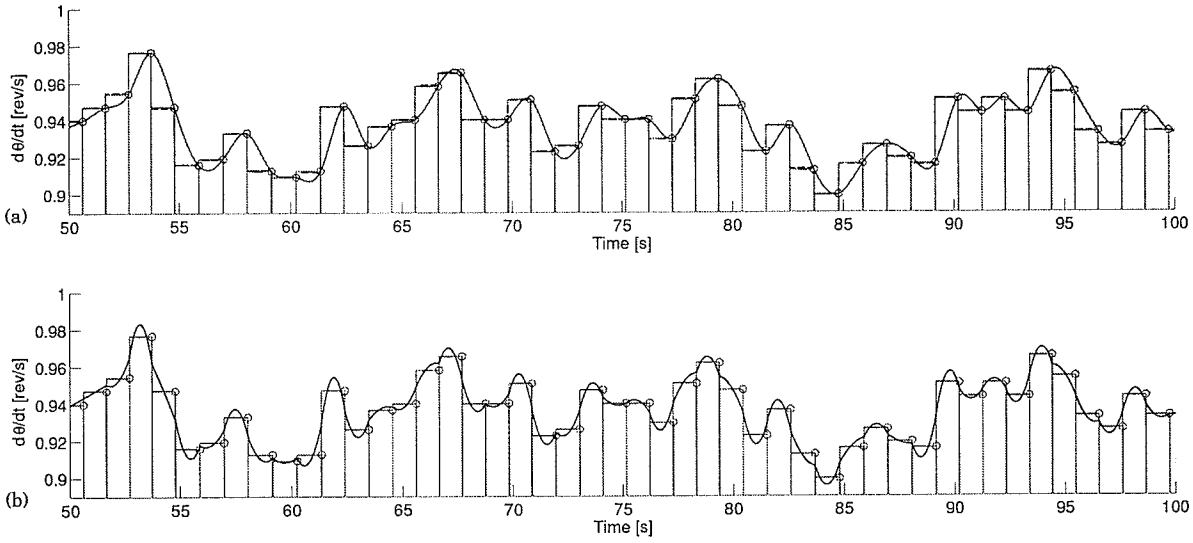


Fig. A.20 Details of IHR series: Example of good `ecgsyn24.m` behaviour; (a) Piecewise-constant IHR time series (dashed) derived from spline-resampled tachogram (solid curve). Tachogram is identified by stems at R-times t_n and amplitude $\Delta_R(n)$. Note how the piecewise-constant IHR series closely follows the SCT. This supports the global performance established in Fig. A.19; (b) Continuous IHR time series (solid) derived from Hermite FMPI of the tachogram. Note how it is not an interpolation of the tachogram itself. Tachogram and SCT are identified as in panel (a).

equivalent to FMPI. Note that many different resampling interpolations are made available in the responsible algorithm `interp_RR.m` of the supporting ECGsyn library, but cubic-spline interpolation is the default. Here, the spline interpolation between the tachogram points defines the "instantaneous RR" function, $T(t)$, which is uniformly sampled. From the resampled tachogram, an SCT is then reconstructed (*i.e.*, constant values from sample points are extended throughout the RR-interval). The transition points are determined by a rounding process and thus may be different from the points of the tachogram itself.

To examine the inner workings of the ECGsyn IHR-process, a visualization of the relevant IHR signals are presented in Fig. A.20. The discrete information of the tachogram appears there as a stem series at the times t_n, with amplitudes $2\pi/\Delta_R(n)$. The SCT defined from the

tachogram (A.11) appears as the step function (grey solid). The solid curve in Fig. A.20(a) interpolating through the tachogram stemplot is the cubic-spline TR. From this TR a second SCT (black dashed) is derived as the ECGsyn IHR function. Here, the dashed SCT effectively is indistinguishable from the first. The reader should recall that a non-resampled SCT is equivalent to a basic piecewise-linear FMPI and thus has good time domain fidelity. This correspondence between the FMPI-based SCT and the TR-based SCT for this tachogram identifies a reason for the RR-interval fidelity of Fig. A.19.

Panel (b) of Fig. A.20 shows how distinct the Hermite FMPI is by comparison. Note how on every SCT segment, the Hermite-based curve does *not* interpolate the tachogram, but instead crosses each segment. According to construction, the area under the smooth curve of each segment is equal to the area of the SCT rectangle. That is, the area in excess of the SCT is equal to the area in deficiency, maintaining a unit area constraint for every interval.

For comparison, the IHR functions for the dataset with poor fidelity, Fig. A.18, is now shown in Fig. A.21. Here, the tachogram stem series and SCT (grey solid) are presented in all panels, as before. Panel (a) shows a cubic-spline TR (black solid) and its secondary SCT (black dashed). As can be easily seen, the rounding introduced by the algorithm changes the values of the secondary SCT from the one consistent with the tachogram. In particular, the area under $\omega(t)$ is not preserved over each interval. Panel (b), which uses linear TR, provides no improvement. Panel (c) shows the cubic-Hermite FMPI, and as with Fig. A.20(b), the IHR time series crosses the SCT, but still preserves the area of every interval.

This conservation of area is critical, since the tachogram acts as an interpolation constraint on $\theta(t)$. Upon differentiation, this becomes an integral constraint on $\omega(t)$. Recall that the SCT assumes the "average rate" throughout the entire interval. According to the mean value theorem, the property of an FMPI solution is that this average rate must be attained *somewhere*

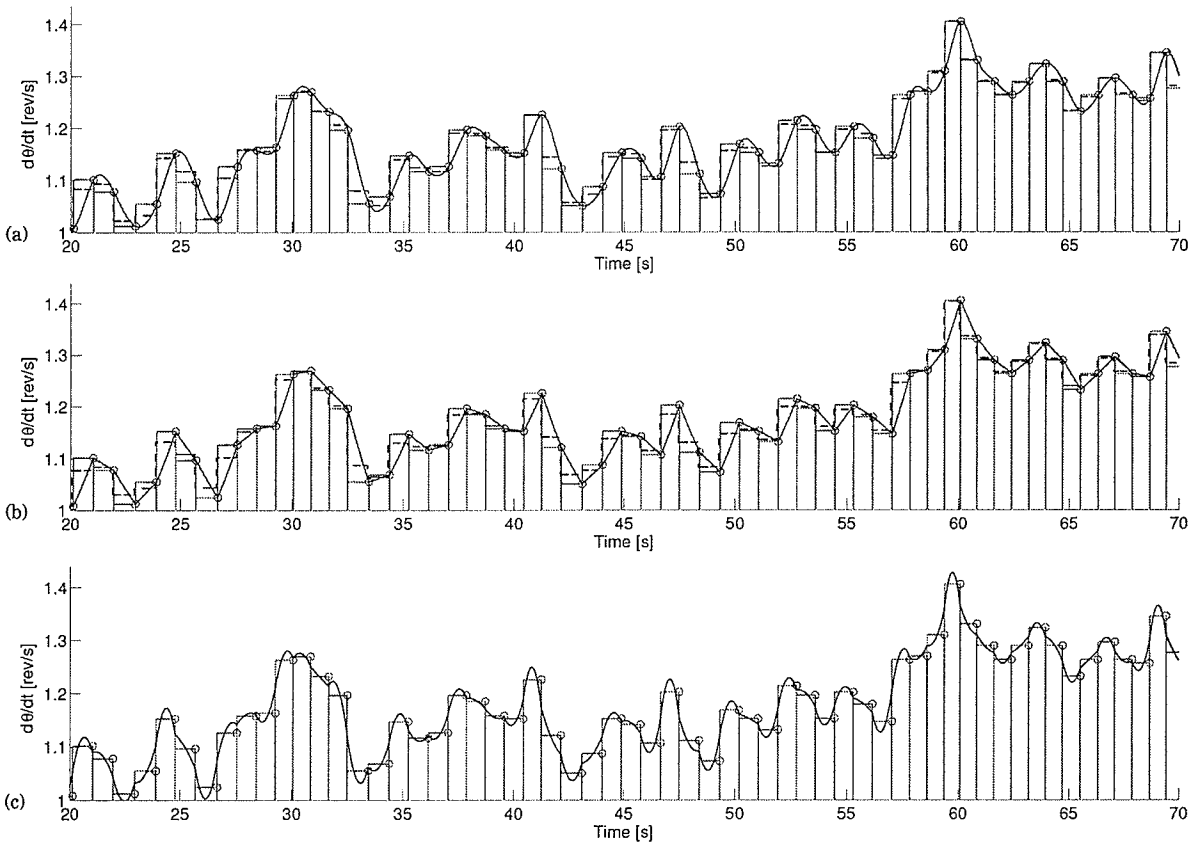


Fig. A.21 Details of IHR series: Example of poor `ecgsyn24.m` behaviour. Tachogram (identified by stems at R-times t_n and amplitude $\Delta_R(n)$) and SCT (grey solid) are the same in all panels; (a) Piecewise-constant IHR time series (dashed) derived from spline-resampled tachogram (solid curve). Note that the dashed IHR series deviates from the SCT. This identifies the source of the poor global performance established in Fig. A.18; (b) Piecewise-constant IHR time series (dashed) derived from linearly-resampled tachogram (solid curve). Note the deviation from the SCT persists. This identifies the limited effect of changing resampling algorithms. (c) Continuous IHR time series (solid) derived from Hermite FMPI of the tachogram. Note, as before, how it is not an interpolation of the tachogram itself. Evaluating this and previous figures, it is concluded that resampling is to be avoided. FMPI is the best approach to IHR definition.

within that interval. That is, an FMPI solution of any order must cross the SCT at least once per SCT interval. This is the important distinction between FMPI and TR. Tachogram resampling assumes this point is the *end point of the interval* in all cases, when in fact it might not be. Note that the SCT is both an FMPI and a TR because it applies the mean slope to *all points*, including the endpoints, of the interval. At higher orders, however, the endpoint assumption of the TR forces (i) the IHR signal to cross the SCT *only once*, and (ii) consequently, it violates the area constraint. Now, a TR could in fact match the SCT constraint if the points of "irregular sampling" were taken to be the crossing points of an FMPI with the SCT. These are the very points that satisfy the mean value theorem. These points, however, are not available *a priori*, and thus TR uses the endpoints and loses time domain fidelity.

Now the time domain benefit of FMPI over the original ECGsyn techniques is demonstrated for a tachogram constraint, and thus, HRV. In the following, the differences in time domain fidelity on the smaller event-intervals are analyzed.

A.3.4 IHR-processing and Other Event-Interval Fidelity

A.3.4.1 Linear Artifacts in Smaller Event-Intervals

The second artifact that can be improved by FMPI is the time-domain fidelity of the smaller ECG waves. In comparison to the natural events in Fig. A.14(a), it is quite clear from Fig. A.14(b) that ECGsyn synthesizes the P-waves (crosses) and T-waves (plusses) with a significantly larger spread.

Note that clinical analysis of ECG event-intervals will focus on elements such as the QT-interval and the PR-isoelectric segment [69]. For simplicity purposes, however the "return to baseline" event is forsaken for the maxima of the waves. By this, then, the PR-interval will mean the interval length between the fiducial maxima of the P-wave to the fiducial maxima of the R-wave, and similarly, for the RT-, RR-, or PP-intervals.

It is well-known that the time taken for ventricular repolarization, expressed by the QT-interval, is affected by the mean heart rate, and attempts to compensate for this behaviour using Bazett's formula or others [19] have been proposed. These approaches oversimplify the HRV connection to the QT-interval on a beat-by-beat basis, as commented in [19][50][4]. Considering this, the event-interval fidelity of synthetic ECGs should represent the variety of complex behaviours that have been observed in natural ECGs. This would include the behaviours in Fig. A.13.

Similar to that of Fig. A.13, an analysis of the event-intervals synthesized by ECGsyn from the tachogram of Fig. A.13(a) is shown in Fig. A.22. Note the linear relationship between the smaller event intervals and the RR-intervals throughout the figure.

Specifically, the linear relationship in Fig. A.22(d) is not unexpected, since it was briefly described in the original paper [140] in the context of the QT-interval. However, this was considered acceptable following the reasoning of QT-correction. In fact, the analysis presented

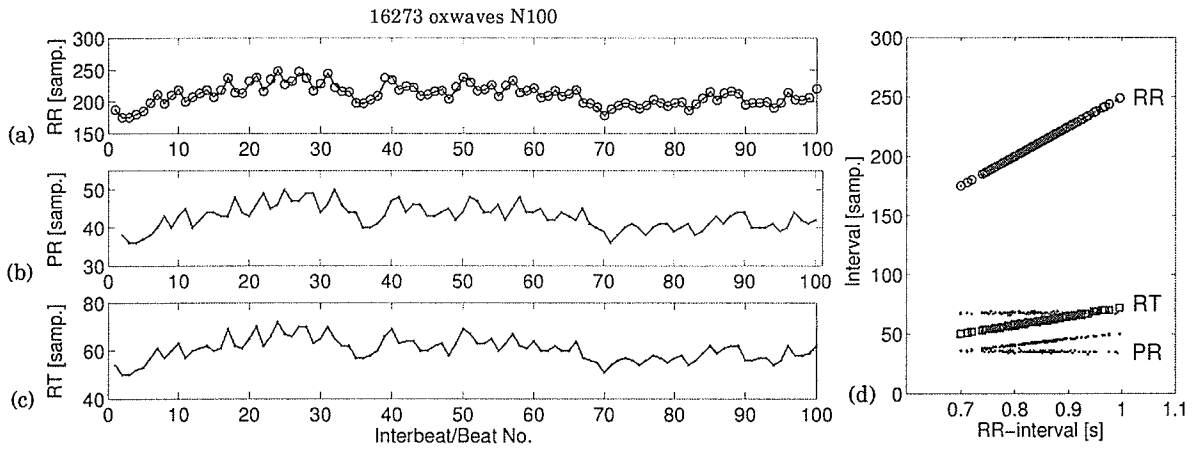


Fig. A.22 The event-interval behaviour of 100 beats synthesized by ECGsyn from the natural tachogram `sel16273.pu0`. (a) RR-interval sequence output (circles) and input (dots); (b) PR-interval sequence; (c) RT-interval sequence; (d) event-interval relationship to HRV: output ECG RR-intervals (circles), RT-intervals (squares), and PR-intervals (crosses), whereas the event-intervals of the input dataset are marked by dots. There is an artifactual linear relationship at the output.

here demonstrates that it should be considered as an artifact of the modelling process, since the original ECG may exhibit complex event-interval behaviours such as uncorrelated QT-variability, or hysteresis. The behaviour of these artifacts clearly indicates that the ECGsyn model will not correctly transcribe the time domain behaviour of the input ECG at the output.

Experiments with the same dataset above have demonstrated that FMPI on the tachogram, by itself, does not remedy this problem. PR- and RT- intervals appear linearly related to heart rate, even with a Hermite-based FMPI scheme. However, this is a simple matter of insufficient data. By including PR- and RT-intervals at the input, as in (A.21), the FMPI-based IHR signal can preserve these shorter event-intervals.

The extra constraints provided by the smaller intervals greatly improves the dynamics of the output event-intervals, as demonstrated by Fig. A.23. A visualization of the improved time series behaviour of the final synthesized surrogate ECG is shown in Fig. A.14(c), matching

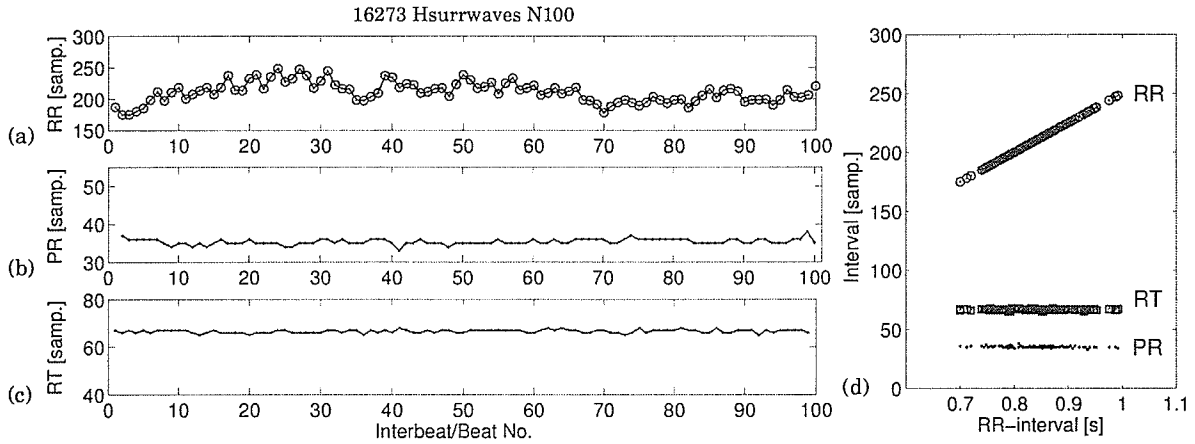


Fig. A.23 The event-interval behaviour of 100 beats synthesized by surrogate ECG from the natural tachogram `sel16273.pu0`. (a) RR-interval sequence output (circles) and input (dots); (b) PR-interval sequence; (c) RT-interval sequence; (d) event-interval relationship to HRV: output ECG RR-intervals (circles), RT-intervals (squares), and PR-intervals (crosses), whereas the event-intervals of the input dataset are marked by dots. By including smaller event-interval information in FMPI, there is a substantial improvement in interval fidelity.

better the natural behaviour displayed in panel (a).

Furthermore, it is interesting to see the effect of the smaller event-intervals on the IHR time series. As shown in Fig. A.24, the IHR becomes, itself cyclostationary. There is a parametric artifact in the peaks and valleys, though, because the angular values for the ECGfm model, the θ_i, are preassigned and not data-driven. Thus the IHR between the P- and R-waves will be excessively high if $|\theta_P|$ is too large. (This was not a concern with RR-intervals, because these must have length 2π.)

The conclusion to be drawn from these experiments is that the ECGfm algorithm has improved the dynamical fidelity of the synthesized ECG as compared to the ECGsyn algorithm. The time domain event-interval behaviour of the ECGfm/FMPI model matches the original data without introducing linear artifacts to the interval relationships.

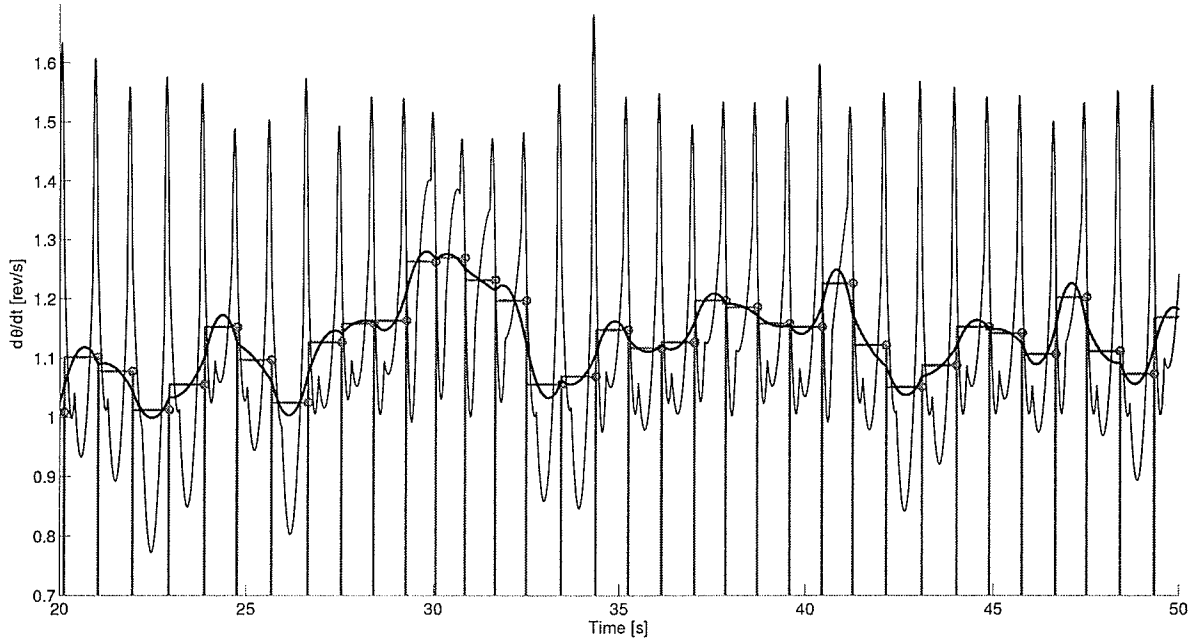


Fig. A.24 Details of IHR series: Example of enhanced Hermite interpolation; Hermite (thick continuous curve) and linear (thick semicontinuous curve) FMPI time series from natural tachogram (as in Fig. A.21(b)). For comparison, the IHR resulting from inclusion of PR- and RT-intervals with default angular parameters θ_i is shown (thin curve). RT- and TR-intervals are typically concave up, while PR-intervals are typically concave down. Here, the IHR becomes cyclostationary, but the magnitude of the extrema are excessively influenced by the choice of θ_i.

A.4 ECGfm: Important Code

Though all thesis code is available in electronic form at `www.ee.umanitoba.ca/~kinsner/projects/`, Listing 1 contains the main ECGfm algorithm implemented as a `Matlab` function.

Listing 1 (ECGfm.m) The `Matlab` code of the ECGfm surrogate ECG algorithm.

```
function [s,timevector,ihr,Rpeaks] = ECGfm(Rsequence,sfecg,sfint,ti,ai↓
→      ,bi,scaling,method,ihrparam,flag)
% ECGfm: Two-step process to produce a synthetic ECG from beat annotations
%      s = ECGfm             - Use stochastic HRV to create Rsequence (MCTS03 default)
%  or s = ECGfm(Rvector) - Fit given vector of R-markers [in seconds]
%  or s = ECGfm(struct)  - Fit given struct of beat annotations [in seconds]
```

```
%           where struct has fields 'R' or 'RR', ('PR' and 'RT' optional)
%
% ECGfm gives the following outputs: [s,t,ihr,peaks] = ECGfm(...)
%   s: ECG times series
%   t: timepoints of ECG samples
%   ihr: instantaneous heart rate time series
%   peaks: ECGsyn wave category of sample;
%            [P,Q,R,S,T] = [1,2,3,4,5],  otherwise 0
%
% ECGfm takes the following inputs:
%   ... = ECGfm([Rin],[sfecg],[sfint],[ti],[ai],[bi],[scaling],[method],[ihrparam],[↓
→      ihrflag])
% These are (default parameters values shown in []):
% Rin: vector of input R-markers [default ECGSYN.m synthesis] in units of time, or
%       struct with fields 'R' or 'RR' ('PR' and 'RT' optional) in units of time
% sfecg: ECG sampling frequency [256 Hertz]
% sfint: Internal sampling frequency [2*sfecg]
%       Order of extrema: - see MCTS03 [P Q R S T]
% ti = angles of extrema - see MCTS03 [-70 -15 0 15 100] degrees
% ai = z-position of extrema - see MCTS03 [1.2 -5 30 -7.5 0.75]
% bi = Gaussian width of peaks - see MCTS03 [0.25 0.1 0.1 0.1 0.4]
% scaling: use default ECGSYN.m rescaling [0 (none)]
% method: interpolation method identifier {'linear','hermite'} [hermite]
% ihrparam: if method == linear:
%                   ihrparam = offset for linear interpolations in units of pi (i.e., ↓
→      between 0 and 2) [0 rrInterp]
%           : if method == hermite: >> PR intervals (scalar or sequence) in units of ↓
→      time [0]
%                   ihrparam =  PR intervals (scalar or vector) in units of time [0]
% ihrflag: prevents numerical integration (ihr only) if set to 'ihr' ['']
%
% NOTE: ECGfm is best used in conjunction with s2t.m and structanno.m functions
%      ... = ECGfm(s2t(structanno(..,)),[sfecg])
% see s2t, structanno

% Copyright (c) 2008 by Michael Potter, All Rights Reserved
% See M. Potter and W. Kinsner, "Improved event-interval reconstruction in synthetic ↓
→      electrocardiograms,"
% IEEE Trans Biomed Eng, (accepted May 21, 2008).
% Contact M. Potter (m.potter@ieee.org)

% Modified from the original ECGSYN.m algorithm...
%   Copyright (c) 2003 by Patrick McSharry & Gari Clifford, All Rights Reserved
%   See IEEE Transactions On Biomedical Engineering, 50(3), 289-294, March 2003.
%   Contact P. McSharry (patrick@mcsharry.net) or G. Clifford (gari@mit.edu)
%
%   ecgsyn.m and derivsecgsynbase.m are freely availble from Physionet -
%   http://www.physionet.org/ - please report any bugs to the authors above.
```

```
%  This program is free software; you can redistribute it and/or modify
%  it under the terms of the GNU General Public License as published by
%  the Free Software Foundation; either version 2 of the License, or
%  (at your option) any later version.
%
%  This program is distributed in the hope that it will be useful,
%  but WITHOUT ANY WARRANTY; without even the implied warranty of
%  MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.  See the
%  GNU General Public License for more details.
%
%  You should have received a copy of the GNU General Public License
%  along with this program; if not, write to the Free Software
%  Foundation, Inc., 59 Temple Place, Suite 330, Boston, MA  02111-1307  USA

%%%%%%%%%%%%%%%(1) Argument processing
% set parameter default values
RRsrc = ''; %default source is argument
if nargin < 1,
    RRsrc = 'MCTS default'; %if no Rsequence argument, use MCTS spectral model
    Rsequence = [];
elseif isempty(Rsequence) || isscalar(Rsequence),
    if ~isstruct(Rsequence),
        RRsrc = 'MCTS default'; %if Rsequence argument empty or integer, use ↓
→           MCTS spectral model
    end
end
if nargin < 2 || isempty(sfecg),
   sfecg = 256; %default sampling frequency
end
if nargin < 3 || isempty(sfint),
   sfint = sfecg*2; %default NDM-sampling frequency
end
if nargin < 4 || isempty(ti),
    %      P    Q   R   S   T
   ti = [-70 -15 0 15 100]; %default angles of x-Waves
end
if nargin < 5 || isempty(ai), % z position of attractor
      %  P    Q    R    S    T
   ai = [1.2 -5 30 -7.5 0.75]; %default amplitude of x-Wave point-attractors
end
if nargin <6 || isempty(bi), % Gaussian width of each attractor
    %     P     Q    R    S    T
   bi = [0.25 0.1 0.1 0.1 0.4]; %default falloff (width) of x-Wave point-↓
→       attractors
end
if nargin <7 || isempty(scaling),
   SCALE = 0;  %default ECG scaling is none
```

```
96  else,
       SCALE = scaling;
    end
    if nargin <8 || isempty(method),
        method='hermite';
101 end
    if nargin <9 || isempty(ihrparam),
        ihrparam=[];
    end
    if nargin <10
106     flag ='';
    end

    %determine RR-sequence source / generate if necessary
    if isempty(RRsrc),
111    RRsrc = 'passed';
    elseif strcmp(RRsrc,'MCTS default'),
       %McSharry et al. ECGSYN.m default parameters
       flo = 0.1; fhi = 0.25; flostd = 0.01; fhistd = 0.01; lfhfratio = ↓
→         0.5; sampfreqrr = 1;
       %hrmean = 60; hrstd = 1; N=256; %ECGSYN.m default
116    hrmean = 68; hrstd = 6; %our modification and below
       if Rsequence > 0, %i.e., positive scalar
           N = Rsequence;
       else
           N = 60; %our default
121    end
       trr = 1/sampfreqrr; tstep = 1/sfecg; rrmean = (60/hrmean); Nrr = ↓
→         2^(ceil(log2(N*rrmean/trr)));
       RR = RRprocessMCTS(flo, fhi, flostd, fhistd, lfhfratio, hrmean, ↓
→         hrstd, sampfreqrr, Nrr);
       Rsequence = cumsum(RR); %convert interbeat-intervals to beat-timepoints
    end
126

    %some verbose...
    fid = 1;
    fprintf(fid,'\nECGsurrogate.m....\n'); %
    fprintf(fid,'Using %s RR sequence\n',RRsrc); %
131

    %determine ihrparam purpose
    if strcmp(method,'hermite'),
    %    PRints = floor(abs(ihrparam*sfecg));  %express in number of samples >= 0
        PRints = ihrparam;  %express in time
136 elseif strcmp(method,'linear'),
        theta0 = ihrparam; %units of pi
    end

    %deal with vector argument
```

```
if ~isstruct(Rsequence),
    Rsequence.R = Rsequence;
end

%deal with missing R-sequence field
if ~isfield(Rsequence,'R') && isfield(Rsequence,'RR'),
    Rsequence.R = [0; makecol(cumsum(Rsequence.RR))];
end

%%%%%%%%%%%%%%(2) Preprocessing
Rsequence.R = makecol(Rsequence.R); %needs vector
N = length(Rsequence.R); %needs vector
RR = diff(Rsequence.R); %needs vector
rrmean = mean(RR);
if rrmean > 10,
    warning('Rsequence input should be in units of time...not samples'↓
→       );
    disp('System is in pause... use <C>-C to exit, or press any key to↓
→       continue');
    pause
end
ti = ti*pi/180;  % convert X-wave angles to radians
%ti(1) %debug

% adjust NDM-extrema parameters for mean heart rate
hrmean = 60/rrmean;
hrfact =  sqrt(hrmean/60);
hrfact2 = sqrt(hrfact);
bi = hrfact*bi;
ti = [hrfact2 hrfact 1 hrfact hrfact2].*ti;
%ti(1) %debug

% check that sfint is an integer multiple of sfecg
q = round(sfint/sfecg);
qd = sfint/sfecg;
if q ~= qd
   error(['Internal sampling frequency (sfint) must be an integer↓
→       multiple ' ...
'of the ECG sampling frequency (sfecg). Your current choices are:' ↓
→    ...
'sfecg = ' int2str(sfecg) ' and sfint = ' int2str(sfint) '.']);
end

%more verbose
fprintf(fid,'Approximate number of heart beats: %d\n',N);
fprintf(fid,'RR-interval mean: %d s\n',rrmean);
fprintf(fid,'Heart rate mean: %d bpm\n',hrmean);
```

```
%%%%%%%%%%%%%%%(3) IHR processing using RRmodule: stage 1
%timevector interval begins at first R-peak
dt = 1/sfint; Nt = floor((Rsequence.R(end)-Rsequence.R(1))/dt); %needs ↓
→    vector
timevector = [0:dt:Nt*dt]; %time vector

%some verbose...
fprintf(fid,'Using␣%s␣IHR␣interpolation\n',method);
disp('IHR␣being␣interpolated...');

%instantaneous rr and ihr
switch method,
    case 'linear',
        if isempty(theta0),
            theta0 = 0;
        end
        fprintf(fid,'Linear␣IHR␣offset:␣%d␣pi␣radians\n',theta0);
        if theta0==0,
            [ihr,thetadotinv,Rpeaks] = rrInterp(Rsequence.R,timevector↓
→               );
        elseif theta0,
            [ihr,thetadotinv,Rpeaks] = OffsetInterp(Rsequence.R,↓
→               timevector,theta0);
        end
    case 'hermite',
        %PRints = floor(PRints*sfecg);  %convert time into samples
        [ihr,thetadotinv,Rpeaks] = HermiteInterp(Rsequence,timevector,↓
→           PRints,ti);
end
rrn = 30./ihr; %derivecgsyn uses units of sec/pi
disp('.....IHR␣interpolation␣done.');

if strcmp('ihr',flag), %only IHR required, ignore integrating ECG
    %downsample from sfint to sfecg
    % --problem with downsampling is that Rpeaks is already determined at sfint
        %timevector = timevector(1:q:end);
        %ihr = ihr(1:q:end); -
    %complete rest of arguments
        s=[]; %ecg
    fprintf(fid,'ECGsurrogate.m␣done.\n');
    return %exit
end

%more verbose...
fprintf(fid,'Surrogate␣ECG␣sampled␣at␣%d␣Hz\n',sfecg);
fprintf(fid,'Internal␣sampling␣frequency:␣%g\n',sfint);
fprintf(fid,'␣␣␣␣␣␣P␣␣Q␣␣R␣␣S␣␣T\n');
```

```
    fprintf(fid,'ti = [%1.4f %1.4f %1.4f %1.4f %1.4f] radians\n',ti(1),ti↓
→       (2),ti(3),ti(4),ti(5));
    fprintf(fid,'ai = [%g %g %g %g %g]\n',ai(1),ai(2),ai(3),ai(4),ai(5));
    fprintf(fid,'bi = [%g %g %g %g %g]\n',bi(1),bi(2),bi(3),bi(4),bi(5));
    %%%%%%%%%%%%%%%(4) ECG Modulation using NDM-module: stage 2
    TRANSIENTCONTROL = 1; %flag for switching pre-convergence on/off
    AMPLITUDECORRECTION = 1;%nonzero flag turns on HR amplitude compensation
    if AMPLITUDECORRECTION,
        disp('*****Using amplitude correction in ECGsyn dynamics!!*****');
        acorr = 1;
    end
    if TRANSIENTCONTROL,
        % integrate system using fourth order Runge-Kutta on
        % repeated first beat for convergence
        fprintf(fid,'Integrating dynamical system for transient control↓
→           ...\n');
        x0 = [1,0,0.04];
        Nbeg = 1; Nend = Rpeaks(find(Rpeaks>1,1,'first'))-1; %these points ↓
→           mark first beat
        Nrpt = 20; %number of times to repeat beat
        %now integrate this constant beat multiple times
        [T,X0] = ode45(@derivsecgsynbase,[0 (Nrpt*Nend-1)/sfint],x0,[],↓
→           repmat(rrn(Nbeg:Nend),1,Nrpt),sfint,ti,ai,bi,acorr);
    %     [T,X0] = ode45(@derivsecgsynbase,[0 20-dt],x0,[],repmat(rrn(1),20*sfint,1),sfint,↓
→       ti,ai,bi,acorr);
        fprintf(fid,'.....Transient control done.\n');
        x0 = [1,0,max(X0(end-fix(length(X0)/4):end,3))];
    %     x0 = [1,0,max(X0(end-2*sfint:end,3))];
    else
        x0 = [1,0,0.04]; %defalut initial condition (ECGSYN.m)
    end

    % integrate system using fourth order Runge-Kutta
    fprintf(fid,'Integrating dynamical system...\n');
    [timevector,X0] = ode45(@derivsecgsynbase,timevector,x0,[],rrn,sfint,↓
→       ti,ai,bi,acorr);
    fprintf(fid,'.....Dynamical system integration done.\n');

    %%%%%%%%%%%%%%%(5) Postprocessing
    % downsample to required sfecg
    X = X0(1:q:end,:);
    timevector = timevector(1:q:end);
    ihr = ihr(1:q:end);

    % extract R-peaks times
    ipeaks = detectpeaks(X, ti, sfecg);
    Rpeaks = ipeaks; %find(ipeaks==3); %temp! do I want struct, vector,... ??
```

```
% OPTIONAL: Scaling of output signal
if SCALE,
% Scale signal to lie between -0.4 and 1.2 mV (ECGSYN.m default)
    disp('*****ECG␣output␣using␣MCTS␣scaling!!*****');
    z = X(:,3);
    zmin = min(z);
    zmax = max(z);
    zrange = zmax - zmin;
    z = (z - zmin)*(1.6)/zrange -0.4;
else,
    z = X(:,3);
end

s = z;
fprintf(fid,'ECGsurrogate.m␣done.\n');

return %--end of main routine
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%SUBFUNCTIONS
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [ihr,thetadotinv,Rpeaks] = rrInterp(Rtimes,time);
%my fix of MCTS code
Rtimes = Rtimes-Rtimes(1);
thetadotinv2pi = diff([Rtimes]);
%create time series -- "tdi" = "theta-dot-inv"
[tdi,Rpeaks] = boxcar(thetadotinv2pi,Rtimes(2:end),time);
Rpeaks = [1, Rpeaks];
thetadotinv = thetadotinv2pi/2;
ihr = 60./tdi;
return

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [ihr,thetadotinv,Rpeaks] = OffsetInterp(Rtimes,time,theta0);
%theta0 is in units of pi (number between 0 and 2)
if nargin <3,
    theta0 = 1.5;
end
N = length(Rtimes);
rr1 = makecol(zeros(1,N+1));
Rtimes = [0; Rtimes];
rr0 = diff(Rtimes);
%initialize offset matrix problem
if theta0 <= 1,
    a1 = 2-theta0;
    a2 = theta0;
    b = [0; makecol(rr0)];
```

```
else,
    a1 = theta0;
    a2 = 2-theta0;
    b = [0; flipud(makecol(rr0))];
end
A = diag(a2*ones(size(rr0)),-1) + diag(a1*ones(size(rr1)),0);
e1 = [1; zeros(N,1)];
%offset matrix solution
temp = a1*(A\e1);
%temp = (A\e1);
rr1 = A\b;
free = mean(rr0)/2; %max likelihood assumption
if theta0 <= 1,
    rr1 = rr1 + free*temp;
else,
    rr1 = flipud(rr1 + free*temp);
end
thetadotinv = rr1;
Tntheta0 = Rtimes + theta0*rr1;

%create time series -- "tdipi" = "theta-dot-inv*pi"
if min(thetadotinv) > 0, %only if its a true solution
    [tdipi,Disc,Rpeaks] = boxcar(thetadotinv,Tntheta0,time,Rtimes);
    ihr = 30./tdipi;
    Rpeaks = [1; Rpeaks];
else
    tdipi=[];
    Disc=[];
    ihr =[];
    Rpeaks =[];
end

return

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [ihr,thetadotinv,Rpeaks] = HermiteInterp(Rtimes,time,PRints,thetain);
if ~isempty(PRints),
    Rtimes = Rtimes.R-Rtimes.R(1);
    Ptheta = thetain(1)/pi; %scale radians into units of pi radians (<0)
    [temp,Rpeaks] = boxcar(diff([makecol(Rtimes)]),Rtimes,time);
    Rpeaks = [1; makecol(Rpeaks)]; %returns peaks if 'ihr only' is set
    if isscalar(PRints)
        if PRints==0,
            fprintf('...ignoring␣PR␣interval\n');
            %define R-wave points
            Interptimes = [makecol(Rtimes)]; %Rpeak time positions in units of time
```

```
            Interpthetas = 2*[0:length(Interptimes)-1]; %Rpeak theta ↓
→               positions in units of pi
        else,
            fprintf('...constant␣PR␣interval:␣%d␣\n',PRints);
            %define R-wave points
            Rtimes = [makecol(Rtimes)]; %Rpeak time positions in units of ↓
→               time
            Rthetas = 2*[0:length(Rtimes)-1]; %Rpeak theta positions in ↓
→               units of pi
            %define P-wave points
            Ptimes = Rtimes-PRints; %Ppeak time positions in units of time
            Pthetas = Rthetas+Ptheta; %Ppeak theta positions in units of pi
            % stack points
            Interptimes = [makerow(Ptimes); makerow(Rtimes)];
            Interpthetas = [makerow(Pthetas); makerow(Rthetas)];
            % sort points
            Interptimes = makecol(Interptimes(:));
            Interpthetas = makecol(Interpthetas(:));
            %remove early points (negative time/angle)
            Interptimes(1) = [];
            Interpthetas(1) = [];
        end
    else %vector PRints
        fprintf('...mean␣PR␣interval:␣%d␣\n',mean(PRints));
        %arrange vector lengths
        Rtimes = makecol(Rtimes);
        PRints = makecol(PRints);
        L = min(length(PRints),length(Rtimes));
        Ptimes = Rtimes(1:L)-PRints(1:L); %Ppeak time positions in units of ↓
→           time
        %define R-wave points
        Rtimes = [0; Rtimes(1:L)]; %Rpeak time positions in units of time
        Rthetas = 2*[0:length(Rtimes)-1]; %Rpeak theta positions in units of ↓
→           pi
        %define P-wave points
        Pthetas = Rthetas(2:end)+Ptheta; %Ppeak theta positions in units of ↓
→           pi
        % initialize
        Interptimes = zeros(2*L+1,1);
        Interpthetas = zeros(2*L+1,1);
        % arrange
        Interptimes(2:2:end) = Ptimes;
        Interptimes(1:2:end) = Rtimes;
        Interpthetas(2:2:end) = Pthetas;
        Interpthetas(1:2:end) = Rthetas;
    end
else %struct argument with embedded intervals
    waves = Rtimes;
```

```
    if isfield(waves,'R'),
        Rtimes = makecol(waves.R-min(waves.R)); %start at R-wave at 0
        Rthetas = makecol(2*[0:length(Rtimes)-1]);
        Interptimes = Rtimes;
411     Interpthetas = Rthetas;
        maxtime = max(Rtimes);
        [temp,Rpeaks] = boxcar(diff([Rtimes]),Rtimes,time);
        Rpeaks = [1; makecol(Rpeaks)];  %returns peaks if 'ihr only' is set
    else
416     error('struct argument requires an ''R'' field');
    end
    if isfield(waves,'PR'),
        disp('...constraining PR intervals');
        if isscalar(waves.PR),
421         L = length(Rtimes(2:end));
            Ptimes = Rtimes(2:L+1)-waves.PR;
            Pthetas = Rthetas(2:L+1)+thetain(1)/pi;
        else
            L = min(length(Rtimes(2:end)),length(waves.PR));
426         Ptimes = Rtimes(2:L+1)-makecol(waves.PR(1:L));
            Pthetas = Rthetas(2:L+1)+thetain(1)/pi;
        end
        Interptimes = [Interptimes; Ptimes];
        Interpthetas = [Interpthetas; Pthetas];
431     maxtime = min(maxtime,Rtimes(L+1));
    end
    if isfield(waves,'RT'),
        disp('...constraining RT intervals')
        if isscalar(waves.RT),
436         L = length(Rtimes(1:end));
            Ttimes = Rtimes(1:L)+waves.RT;
            Tthetas = Rthetas(1:L)+thetain(5)/pi;
        else
            L = min(length(Rtimes(1:end)),length(waves.RT));
441         Ttimes = Rtimes(1:L)+makecol(waves.RT(1:L));
            Tthetas = Rthetas(1:L)+thetain(5)/pi;
        end
        Interptimes = [Interptimes; Ttimes];
        Interpthetas = [Interpthetas; Tthetas];
446     maxtime = min(maxtime,Ttimes(end));
    end
    mintime = Interptimes(1);
    mintheta = Interpthetas(1);
    Interptimes = sort(Interptimes);
451 Interpthetas = sort(Interpthetas);
    %crop to appropriate interval beg
    Interptimes(Interptimes<mintime) = [];
    Interpthetas(Interpthetas<mintheta) = [];
```

```
        %crop to appropriate interval end
        indx = find(Interptimes>maxtime);
        Interptimes(indx) = [];
        Interpthetas(indx) = [];
    end
    if ismonotonic(Interptimes) && ismonotonic(Interpthetas),
        theta = interp1(Interptimes,Interpthetas,time,'pchip');
        ihr = 30*gradient(theta,time);
        thetadotinv = [];
    else
        error('non-monotonic␣hermite␣interpolation')
    end
    return;

    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    function [tseries,Dindx,Rpeaks] = boxcar(values,disct,time,Rtimes);
    %     [Y,C] = BOXCAR(A,D,[X]);
    % create boxcar (analog) signal Y from constant values A and
    % discontinuity times D against time vector X.
    %     Vector C are the indices to the discontinuity leading edges.

    %CREATED: 2006/03/29 -Michael Potter

    if nargin == 0,
        disp('usage:␣Y␣=␣boxcar(Amp,Disc,[X])');
        return;
    elseif nargin < 4,
        Rtimes = disct;
    end

    K = length(disct);
    if max(time) > max(disct),
        if length(values) < K+1,
            disp('Warning!␣Too␣few␣amplitude␣values!');
            disp('Ignoring␣final␣elements␣of␣time␣vector');
    %        time(find(time>max(disct))) = [];
            time(time>max(disct)) = []; %suggested as faster
         else
            disct(K+1) = max(time)+eps;
        end
    end
    N = length(time);
    Dindx = ones(1,K);
    Rpeaks = ones(size(Rtimes));
    tseries = zeros(size(time));
    j=1; k=1;
    for i=1:N,
        if time(i) > disct(j),  %if pass discontinuity
```

```
            Dindx(j)=i-1;        %   identify discontinuity
            j = j+1;             %   switch to next interval
        end
        if time(i) > Rtimes(k),  %if pass Rtime
            Rpeaks(k)=i-1;       %   identify Rpeak
            k = k+1;             %   switch to next beat number
        end
        if j > length(values),
%           pause;
        else,
            tseries(i) = values(j); %current value is value of interval
        end
    end
    if j <= length(Dindx),
        Dindx(j:end)=[];
    end
    return;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function rr = RRprocessMCTS(flo, fhi, flostd, fhistd, lfhfratio, hrmean, hrstd, sfrr, n)
w1 = 2*pi*flo;
w2 = 2*pi*fhi;
c1 = 2*pi*flostd;
c2 = 2*pi*fhistd;
sig2 = 1;
sig1 = lfhfratio;
rrmean = 60/hrmean;
rrstd = 60*hrstd/(hrmean*hrmean);

df = sfrr/n;
w = [0:n-1]'*2*pi*df;
dw1 = w-w1;
dw2 = w-w2;

Hw1 = sig1*exp(-0.5*(dw1/c1).^2)/sqrt(2*pi*c1^2);
Hw2 = sig2*exp(-0.5*(dw2/c2).^2)/sqrt(2*pi*c2^2);
Hw = Hw1 + Hw2;
Hw0 = [Hw(1:n/2); Hw(n/2:-1:1)];
Sw = (sfrr/2)*sqrt(Hw0);

ph0 = 2*pi*rand(n/2-1,1);
ph = [ 0; ph0; 0; -flipud(ph0) ];
SwC = Sw .* exp(j*ph);
x = (1/n)*real(ifft(SwC));

xstd = std(x);
ratio = rrstd/xstd;
```

```
rr = rrmean + x*ratio;
return


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function ind = detectpeaks(X, thetap, sfecg)
N = length(X);
irpeaks = zeros(N,1);

theta = atan2(X(:,2),X(:,1));
ind0 = zeros(N,1);
for i=1:N-1
   a = ( (theta(i) <= thetap) & (thetap <= theta(i+1)) );
   j = find(a==1);
   if ~isempty(j)
      d1 = thetap(j) - theta(i);
      d2 = theta(i+1) - thetap(j);
      if d1 < d2
         %ind0(i) = j; %original
         ind0(i) = min(j); %MP reduces dual values
      else
         %[j theta(i) thetap theta(i+1)] %debug only
         %ind0(i+1) = j; %original
         ind0(i+1) = min(j); %MP reduces dual values
      end
   end
end

d = ceil(sfecg/64);
d = max([2 d]);
ind = zeros(N,1);
z = X(:,3);
zmin = min(z);
zmax = max(z);
zext = [zmin zmax zmin zmax zmin];
sext = [1 -1 1 -1 1];
for i=1:5
   clear ind1 Z k vmax imax iext;
   ind1 = find(ind0==i);
   n = length(ind1);
   Z = ones(n,2*d+1)*zext(i)*sext(i);
   for j=-d:d
      k = find( (1 <= ind1+j) & (ind1+j <= N) );
      Z(k,d+j+1) = z(ind1(k)+j)*sext(i);
   end
   [vmax, ivmax] = max(Z,[],2);
   iext = ind1 + ivmax-d-1;
   ind(iext) = i;
```

```
end
return

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function dxdt = derivsecgsynbase(t,x,rr,sfint,ti,ai,bi,AMPLITUDECORRECTION)
% dxdt = derivsecgsyn(t,x,flag,rr,sampfreq,ti,ai,bi)
% ODE file for generating the synthetic ECG
% This file provides dxdt = F(t,x) taking input paramters:
% rr: rr process
% sfint: Internal sampling frequency [Hertz]
% Order of extrema: [P Q R S T]
% ti = angles of extrema [radians]
% ai = z-position of extrema
% bi = Gaussian width of peaks
% Copyright (c) 2003 by Patrick McSharry & Gari Clifford, All Rights Reserved
% See IEEE Transactions On Biomedical Engineering, 50(3), 289-294, March 2003.
% Contact P. McSharry (patrick AT mcsharry DOT net) or
% G.D. Clifford (gari AT mit DOT edu)

%    This program is free software; you can redistribute it and/or modify
%    it under the terms of the GNU General Public License as published by
%    the Free Software Foundation; either version 2 of the License, or
%    (at your option) any later version.
%
%    This program is distributed in the hope that it will be useful,
%    but WITHOUT ANY WARRANTY; without even the implied warranty of
%    MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.  See the
%    GNU General Public License for more details.
%
%    You should have received a copy of the GNU General Public License
%    along with this program; if not, write to the Free Software
%    Foundation, Inc., 59 Temple Place, Suite 330, Boston, MA  02111-1307  USA
%
% ecgsyn.m and its dependents are freely availble from Physionet -
% http://www.physionet.org/ - please report any bugs to the authors above.

xi = cos(ti);
yi = sin(ti);
ta = atan2(x(2),x(1));
r0 = 1;
a0 = 1.0 - sqrt(x(1)^2 + x(2)^2)/r0;
ip = 1+floor(t*sfint);
w0 = pi/rr(ip);

fresp = 0.25;
```

```
    zbase = 0; %removed default oscillation '0.005*sin(2*pi*fresp*t);' to match records ↓
→       better

646 dx1dt = a0*x(1) - w0*x(2);
    dx2dt = a0*x(2) + w0*x(1);

    dti = rem(ta - ti, 2*pi);
    dx3dt = - sum(ai.*dti.*exp(-0.5*(dti./bi).^2)) - 1.0*(x(3) - zbase);
651
    if nargin <8,
        AMPLITUDECORRECTION = 1;
    end
    if AMPLITUDECORRECTION, %nonzero flag turns on HR amplitude compensation
656     dxdt = [dx1dt; dx2dt; dx3dt/rr(ip)];
    else,
        dxdt = [dx1dt; dx2dt; dx3dt];
    end
```

A.5 Summary

This appendix has covered the details of the novel ECGfm surrogate ECG algorithm used to synthesize noise-free uncontaminated ECG signals. The technique of using a two-module cyclostationary decomposition for ECG synthesis has been described. The details of the integrator module, as well as a comparison to the original ECGsyn system have been presented. The mechanisms of ECGfm's superior performance at reducing amplitude artifacts has been demonstrated with synthetic and natural tachograms.

As well, considerable attention in this appendix has been made to present FMPI for the inverse modelling of instantaneous rate from discrete event-interval measurements. Comparisons of FMPI to the IPFM model and tachogram resampling have demonstrated the importance of this general technique at preserving the time domain characteristics of event-intervals. Other comparisons have been considered in [178, Potter (2008)]. In particular, experiments on synthetic and natural tachograms have determined that ECGfm outperforms ECGsyn at

preserving the time domain features of a template ECG.

Last, this appendix has demonstrated in Sec. A.2.4 a unique visualization of the ECG attractors that are featured heavily in this work. By using the hidden IHR variable $\omega(t)$ as an embedding coordinate of a reconstructed ECG attractor, an interpretation of the sensitivity of MFA to the distribution of the IHR has been discussed.

Since the ECGfm algorithm described here has made some significant contributions to ECG modelling, IHR analysis, and the MFA of ECG attractors, there is a considerable amount of recommended follow up. This includes

(a) further analysis of natural recordings to identify an optimal amplitude correction;

(b) application of the FMPI approach to 3D ECGsyn extensions such the cardiac dipole of [192, (2007)];

(c) options for the tuning of parameters, similar to [191, (2007)],

(d) confirmation that the MFA of an ECG attractor is independent of recorded lead, and essentially dependent on the IHR signal;

(e) consideration of analytic differentiation and the development of higher-order monotonic interpolation methods; and

(f) coding improvements.

The FMPI model for IHR could also benefit from further comparison to other models of instantaneous rate, such as the *heart timing signal* [139, (2003)], the *heart instantaneous frequency* [195, (2004)], and *IPFM demodulation* [205, (1997)]. These comparisons could focus on time domain, as well as spectral, performance.

Appendix B

Summary of Literature on Fetal ECG Separation

B.1 Summary of Table Notation

Experiment types are broken down into three subgroups:

(a) real fECG data, identified by *real noisy data* (RND);

(b) simulated but not noise-free data, identified by *simulated but noisy data* (SND) (*e.g.*, data is manipulated ECG recordings);

(c) simulated noise-free fECG data, identified by *simulated noise-free data* (SFD) (*e.g.*, data is generated from a mathematical ECG model); and

(d) simulated analytical data, identified by *simulated analytical data* (SAD) (*e.g.*, data is generated from a (non-ECG) mathematical model).

Multiple datasets will appear with slashes (*e.g.*, SAD/RND). One particular dataset from [51, (2000)] is used repeatedly and so will be identified by RND^{L}.

Algorithm types are broken down as:

(a) ACM — adaptive correlation matrix

(b) AF — adaptive filtering

(c) ANN — artificial neural network

(d) ANC — adaptive noise cancellation

(e) CDN — chaos denoising

(f) ICA — independent component analysis

(g) PCA — principal component analysis

(h) SVD — singular value decomposition

(i) WT — wavelet transform

Actual algorithm names may appear in brackets (*e.g.*, ICA(Xinfomax)), and joint algorithms can be identified by a hyphen (*e.g.*, WT-ICA). Multiple algorithms will appear with slashes (*e.g.*, PCA/ICA). If an alternative to a well-known algorithm is proposed and evaluated, it will be noted by a † (*e.g.*, JADE†).

Performance metrics are broken down as:

(a) API — Amari performance index

(b) ASI — Amari-squared index

(c) CCI — cross-correlation index

(d) ECG — some ECG-specific measure

(e) GAV — Given's angle variance

(f) HOS — higher-order statistics

(g) SNR — signal-to-noise ratio

(h) Viz — visual inspection

Actual performance metric names may appear in brackets (*e.g.*, ECG(QRS "trust factor")). Multiple metrics will appear with slashes (*e.g.*, SNR/ECG).

B.2 fECG Literature Summary

Table B.1: Summary of Literature on Fetal ECG Separation

Repr. Author	Year	Ref.	Exp. Type	Alg. Type	Perf. Metric	Conclusions	Significance
Widrow *et al.*	1975	[234]	RND	ANC	Viz of fetal R-waves	ANC reduces maternal contamination	Signal processing seminal paper for fECG separation
Widrow *et al.*	1982	[58]	RND	time-sequenced ANC	Viz of fetal R-waves	Time sequenced post-processing of two ANC channels produce clearer fetal R-wave contrast	Adaptive filtering post-processing gives better R-wave, but not better application
Vandewalle *et al.*	1987	[228]	RND	SVD	Viz of fetal R-waves	3 thoracic electrodes guarantee mECG cancellation; implementable batch or online with equivalent accuracy; changing abdominal positions changes noise in fetal estimates, not mECG suppression	Multiple electrode positions used; mECG is 3 channel, while fECG is not
Vandewalle *et al.*	1989	[30]	RND	online SVD	Viz of fetal R-waves	Practical implementation of fECG estimation is achievable	Describes first "real-time system" for fECG separation
Woolfson *et al.*	1990	[241]	SND	ANC/AF	ECG	Effect of added noise to beat averaging in simulated fECG is characterized for the methods	Uses scalp data for fECG simulation; measures R-wave detectability; methods assumes fECG and mECG bandwidth is 100 Hz and 25 Hz respectively; quantitative performance measure

Continued on Next Page...

Table B.1 – Continued

Author	Year	Ref.	Exp. Type	Alg. Type	Perf. Metric	Conclusions	Significance
Schreiber *et al.*	1996	[201]	SND	CDN/AF	Viz of fetal R-waves	Chaos denoising extracts fECG better than Wiener filtering	Embedding approach; characteristic SND methodology from adult ECG databases
Kanjilal *et al.*	1997	[106]	RND	SVD (1 channel)	Viz/SNR	Estimated signals are strongest periodic SVD component	Time-based approach
Schreiber *et al.*	1998	[185]	SND/-RND	CDN/AF	Viz of fetal R-waves	Generic approach applies to any lead; second pass clears noise from fECG	Delay is time-based approach; generality of procedure motivates generality of SQM
Barros *et al.*	2000	[18]	SND(ECG)/-RND(MEG)	AF/ICA(-FastICA†)	Viz	Adapting FastICA rule to converge to periodic components can isolate ECG from artifacts; Weiner initialization increases stability	Interesting adaptation to include a priori fECG information; particularly significant for evoked responses
De Lathauwer *et al.*	2000	[51]	SND/-RND^L	PCA/-ICA(Comon)	Viz/FHR detectability	Demonstrates ICA on fECG, superior to PCA; simulations demonstrate ICA is robust to conditions such as twins and ectopic beats	Seminal ICA-based fECG extraction; introduces commonly used dataset; ICA must allow multi-channel subspace-indeterminism
Khamene *et al.*	2000	[115]	SND/-RND^L/-RND	WT	SNR/CCI/Viz	WT can be applied successfully to multiple or single channel recordings	Multiscale approach; does two-part analysis with SNR measures
Nandi *et al.*	2000	[243]	RND^L	ANC/-JADE†	Viz	ANC too dependent on electrode position	First ICA comparison to ANC

Continued on Next Page...

Table B.1 – Continued

Author	Year	Ref.	Exp. Type	Alg. Type	Perf. Metric	Conclusions	Significance
Nandi *et al.*	2001	[244]	RND^{L}	ANC/-JADE†	Viz of fetal R-waves	ANC too dependent on electrode position	Discussion of ANC tap effects and number of electrodes
Mochimaru *et al.*	2002	[147]	RND	WT	Viz	High frequency sampling (5-20 kHz) allows FeCG separation and identification of some P and T waves	Demonstrates possibilities of high-sampling WT; first discussion of P and T detectability
Muller *et al.*	2002	[141]	SFD/-RND^{L}	ICA(JADE)/-BSS(TDSEP)	GAV-Bootstrap clustering	GAV-Bootstrap can measure consistency and indicate breakdown of specific ICA assumptions	Seminal paper on bootstrapping for ICA SQM; analytical theory
Jutten *et al.*	2003	[232]	SND	PCA/-sPCA/-ICA(JADE)	API†/CCI	ICA result improved if raw channels are preprocessed to reduce dimensionality	Considers effect of indeterminate number of sources
Jutten *et al.*	2003	[230]	RND	ICA(JADE)/-ACM	Viz	ACM isolates high-frequency noise better than JADE; WT denoising efficient in improving fECG quality; choice of wavelet needs to be analyzed	Uses Cardoso (ACM) non-iid approach [169]; 1 kHz sampling of fECG can visualize P and T waves
Principe	2003	[136]	SND/-RND	ICA(Mermaid/-Infomax/-FastICA)	SNR/ECG/Viz	FHR detectability as performance measure identifies Infomax worst, Mermaid best; Mermaid more data efficient than FastICA	Attempt at practical SQM; disconnect between performance measures in the SND and RND; doesn't use Xinfomax

Continued on Next Page...

Table B.1 – Continued

Author	Year	Ref.	Exp. Type	Alg. Type	Perf. Metric	Conclusions	Significance
Sahambi	2003	[9]	SND	ICA(JADE/-Comon/-FastICA/-Infomax)	SNR/Viz	Equivalent performance from algorithms	Considers added Gaussian noise; performance analysis very weak
Muller *et al.*	2003	[77]	SFD/-RND^L	ICA(JADE)/-BSS-TDSEP	GAV-Bootstrap clustering	ICA(JADE) (non-Gaussianity) more reliable than TDSEP (non-whiteness) and SEPA-GAUS (non-stationarity)	Generalizes (non-white) bootstrap to non-Gaussian bootstrap via additive Gaussian noise; fixed added noise SNR
Assaleh *et al.*	2004	[11]	RND^L	ANN	Viz	Nonlinear mapping of thoracic lead into abdominal cancelling signal	Simple, two-channel approach; not tested under generic noise conditions; theoretical go-between of CDN and PCA
Clifford *et al.*	2004	[42]	SFD (1-channel)	CDN/ICA	SNR/CCI	Denoising: ICA of embedding gives better SNR; CDN gives better linear correlation; CDN is better at preserving shape; ICA is better at preserving R-wave	(Not fECG); comparison of interesting methods
Clifford *et al.*	2004	[82]	RND	ICA(JADE)	Viz	Kurtosis and variance can be effective to identify artifactual ICA components in 3-channel ECG; results could be improved by quasiperiodic considerations	(Not fECG); simple and effective experiments; opportunities with 3-leads only

Continued on Next Page...

Table B.1 – Continued

Author	Year	Ref.	Exp. Type	Alg. Type	Perf. Metric	Conclusions	Significance
Grassberger *et al.*	2004	[215]	SAD/-RND	ICA(:)	API/HOS(MI)	MI is estimated directly so convergence is easier to interpret; noise robustness is characterized; MILCA and RADICAL perform best; ICA on embedding can resolve components even further	Slow but reliable; using delay embedding is superior "in principle": can distinguish different Gaussian signals by their autocorrelation
Sahambi *et al.*	2004	[166]	SND	ICA(JADE/-Infomax/-Comon)	SNR/ASI	Additive noise is tolerated as long as mixing is dominant	Use of coloured noise; example of 2×2 mixture experiment
Mochimaru *et al.*	2004	[148]	RND	PCA-WT/ICA-WT	Viz	P- and T-wave extraction of ICA-WT better than PCA-WT (wavelet-denoising applied)	Substantial (25 patient) dataset; high-frequency data
Foresta *et al.*	2005	[60]	RND^L	ICA-WT(WICA)	ECG	ICA is applied to visually selected subset of wavelet decompositions; evaluates ST waveform by (deflection T)/(height QRS)	Essentially WT denoising first, ICA second; demonstrates how subbands can be used as additional ICA channels; plenty of user input, opportunities for automation
Jafari *et al.*	2005	[99]	SAD/-SND/-RND^L	WT-ICA/-ICA	ASI	ICA in WT domain makes all features super-Gaussian; increased performance and speed in noisy conditions	Some good experiments worth expanding
Assaleh *et al.*	2005	[13]	RND^L	ANN	Viz	Nonlinear mapping of thoracic lead into abdominal cancelling signal	Identical to [11]

Continued on Next Page...

Table B.1 – Continued

Author	Year	Ref.	Exp. Type	Alg. Type	Perf. Metric	Conclusions	Significance
Jutten *et al.*	2006	[193]	RND^L	ICA(JADE)	SNR/ECG	ICA components relate to vectorcardiogram components; best remove baseline wander and symmetrize against isoelectric point; maternal ICA subspace dimensionality can exceed 3 (LdL-data is 4)	First discussion of clinical indices; uses beat averaging to demonstrate existence residual fECG features in all ICA components (including maternal subspace), which is a possible ICA SQM
Jutten *et al.*	2007	[192]	SFD	ICA(JADE)	Viz	3D model extends Kalman filtering opportunities with ECGsyn parameterization; methods can provide interpretations of ICA components	ECGsyn generalized to dipole model; seminal modelling of fECG system; only visual metrics
Sato *et al.*	2007	[196]	SND/-RND	ICA/-BSSR	Viz/ECG	Reference signal guides update; average P-waves can be extracted	RND includes scalp electrode during delivery; twin simulation; compares average fECG beat between estimates and scalp signal; polar and bipolar references required
Assaleh	2007	[12]	SFD/-RND^L	ANN/AF	SNR/Viz	Fuzzy ANN improves over other ANN [13] when fECG component is weaker	Two components only; no attempt at practical SQM; requires study of robustness (1 dataset only)

B.3 Summary

This appendix has presented a summary of the fECG separation literature in tabular form. This hopefully contributes a useful resource for the interested readers.

Appendix C

Details of the Direct MS Algorithm from Correlation Partitions

C.1 Overview

This appendix provides details and code for (i) the calculation of a correlation partition (CP), and (ii) the transformation of that CP into the Hölder and Mandelbrot entropies directly. In fact, the transformation will be described for all three types of entropies, (*i.e.*, H_q, γ_q, and Υ_q). Derivation of the entropies themselves is presented in Sec. 4.7.3.

The next section will deal with the overall design of the algorithm and the modular components that are used. Then, Sec. C.2.2 presents the box-assist technique suggested by Kantz *et al.* [108, NTSA, (1997)] to speed up the otherwise $O(N^2)$ search for comparisons with neighbours. The code for the box assist and the partition calculation itself is provided in Sec. C.3. Last, Sec. C.4 presents the algorithm for converting the partition function into the MFA scaling entropies.

C.2 Correlation Partition Algorithm

The calculation of the correlation partition is achieved by the function call

```
[Sout,scales] = CIHptn(data,scales,TW,lag,emdim,metr,p,verbose)
```

and proceeds in five steps. The first step is a preprocessing of the algorithm that (i) checks for correct input and applies optional default values, and (ii) confirms that the choice of scales

Table C.1 Default Values of CIHptn.m Parameters

Parameter	W_T	Δn	N_e	Metric	p	Verbosity
Default Value	1	1	max(2, `size(data,2)`)	'che'	1	0

provided at the input enables the box-assist technique shown in Fig. 6.8. This is implemented in the `Matlab` function. The second step is to embed the data into an array for analysis and create pointers for reading from the array. This is implemented at the beginning of the `CIHptn_mex.c` executable called by the `Matlab` function. The third step is to establish the box-assist indexing by passing twice through the data. This is implemented in the C executable. The fourth step is the neighbour comparison that assigns integer values to the CP itself. This is the bulk of the C executable, and it is designed to use the box-assist indexing and process all scales in parallel. The final step is the postprocessing of the CP into the proper form for the entropy calculation. In particular this involves the normalization of the CP from integer values into the proper probabilities (*i.e.*, < 1), and restructuring the array. This is implemented in the final part of the `Matlab` function.

C.2.1 Function Call and Preprocessing

The minimal call for the calculation of a CP is

```
[Sout,scales] = CIHptn(data,scales)
```

and the default values for the other parameters identified in Table C.1 are applied during the preprocessing routine. This can be convenient because the data input can already be in attractor form as an $N_o \times N_e$ floating point array, or it can be a floating point vector to be lag-embedded during the algorithm execution.

The length of the data and the `scales` vector determine the size of the partition array. The

maximum value of the scales vector

$$\epsilon_g = \max_i \texttt{scales}(i) \tag{C.1}$$

is used to determine the grid-sizing for the box-assist indexing that is applied to the first two coordinates of the attractor.[1] To ensure that speed gains will occur, the range of the attractor (*i.e.*, max-min) is checked to be greater than 9 $\epsilon \times \epsilon$ boxes. If not, then the scales vector is reduced by a factor of 2

$$\texttt{scales}(i) \leftarrow \frac{1}{2}\texttt{scales}(i) \tag{C.2}$$

until the condition is satisfied.

After some error checks, the main C executable is called. Note, however, that the data is passed in transposed form to make the observations appear as columns because C is column-centric. Similarly, the executable output is transposed upon completion.

Furthermore, note that three possible executables can be called, depending on the level of verbosity desired. The standard C executable has limited standard output (verbose =0), while the next (verbose=1) writes more to screen, while the highest level (verbose=3) writes all successful comparisons to a file for debugging purposes. All calculations in this work use standard (fast) verbosity.

C.2.2 The Box-Assist Technique for Neighbour Comparison

As mentioned, there are two actions in the C executable that precede the main algorithmic loop. The first is the usual initialization of variables, pointers, and arrays (and, in particular, the arrangement of the embedded observations into an array). The second is the preparation

[1] One TISEAN implementation suggests using the first and last coordinate for box-indexing (*i.e.*, a "wide" box-assist) instead of the first two (*i.e.*, a "narrow" one). That is particularly efficient if only the results for the maximum embedding dimension are desired. If more than one embedding is to be evaluated in parallel, however, some comparisons will be lost in the smaller embeddings. Consequently, that is not advocated here.

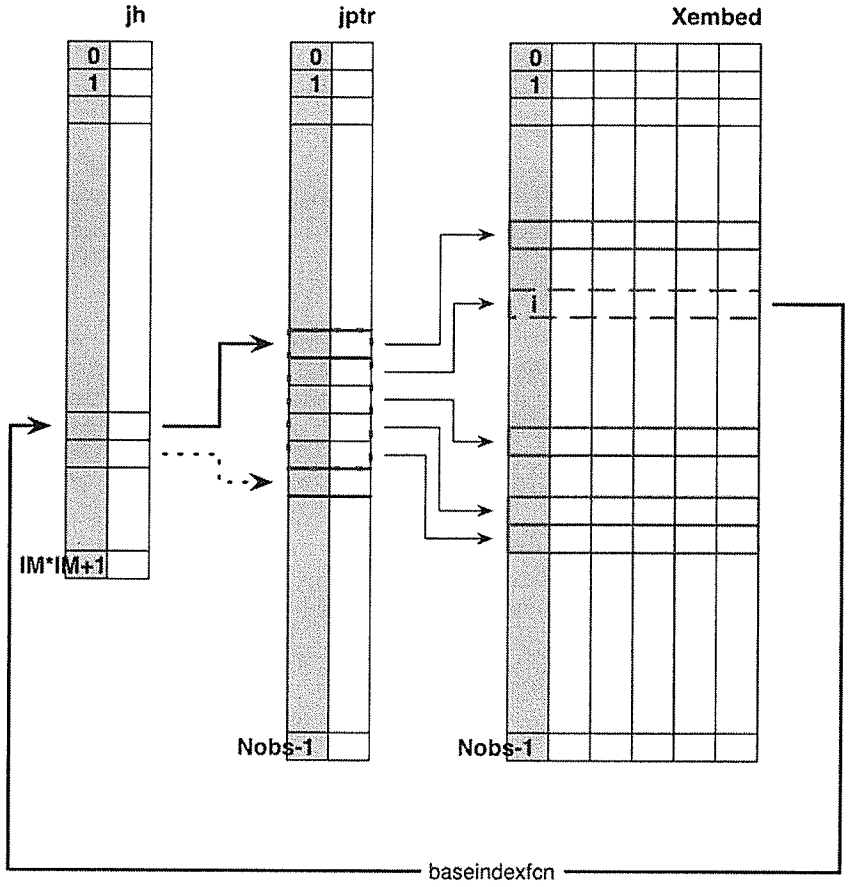


Fig. C.1 The indexing method of the CP box-assist indexes the 2D grid as a vector `jh`. The elements of this vector point to the first entry of the neighbour list in `jptr` for that box. The consecutive elements in `jptr` point to the location of the neighbours in the observation array. This mechanism allows the number of comparisons to be reduced in typical circumstances.

of the box-assist method for neighbour comparisons. The implementation described here is adapted from the code published in [108, NTSA, (1997)].

As described in Sec. 6.4.5, the box-assist technique assigns each embedded point into one of the boxes covering the attractor in 2D, so that only the points in the neighbouring boxes are considered for comparison. All other points will necessarily compare to zero. The box-assist method consists of (i) a mapping function from a 2D grid into an integer, (ii) a vector of integers the length of the complete 2D grid, and (iii) a vector of integers the length of the

attractor, as well as the attractor array itself. These elements, along with their relationships, are visualized in Fig. C.1 as `baseindexfcn`, `jh`, `jptr`, and `Xembed`.

The function of the `jptr` vector is to index all the points in the attractor by the boxes in which they reside. The function of the `jh` vector is to map the index of the 2D grid's boxes onto the `jptr` vector. This is done practically by identifying the `jptr` location of the first member in the chosen box. That is, if `jj0 = jh[j]` and `jj1 = jh[j+1]`, then all the members of the attractor array that are in the `j`th box are indexed by the elements in `jptr` from position `jj0` to `jj1-1`.

It follows that to find all the neighbours of a point `i`,

(a) the grid identity `(i0,i1)` of the point `i` is identified by division of its first two coordinates by ϵ_g;

(b) the `baseindexfcn` function maps the grid identity `(i0,i1)` into an index `j` of `jh`;

(c) then, for `jj` initialized to `jh[j]` and incrementing to `jh[j+1]-1`, the Theiler window is checked;

(d) if the Theiler window condition is satisfied, the distance between point `ii = jptr[jj]` and point `i` is calculated;

(e) this distance is checked against all scales: if the `(i,ii)`-interpoint distance is less than the `scales[m]`, then the CP entry `CP[i,m]` is incremented.

For each point `i`, the above procedure is repeated from b onward for each of the neighbouring boxes: `(i0-1,i1)`, `(i0+1,i1)`, `(i0,i1-1)`, `(i0,i1+1)`, `(i0-1,i1-1)`, `(i0-1,i1+1)` `(i0+1,i1-1)`, and `(i0+1,i1+1)`. Thereafter, the loop repeats for the next centre point, `i` ← `i+1`.

Remark C.1. Note that integer arrays are used here to control the nested indexing of the box-assist procedure. This follows the original implementation [108, NTSA (1997)] and reduces the complexity required in multilevel pointers. There is a bug, however, in the original `mbase2.c` code listed in [108] (though the `Fortran mbase2.f` is accurate). Two passes through the data are used to properly establish the `jh` and `jptr` arrays. On the first pass through the data the `jh` entries are used to count the number of points that fall into each box. Once finished, by updating `jh` with its cumulative sum, its entries now point to the *last* entry for that box in `jptr`. The `jptr` vector is then filled on the second pass, so that

```
jptr[jh[j]-1] = i                                    (**)
```

and the value of `jh[j]` is decremented. Care must be taken, however, to decrement properly so that it fills the vector from 0 to $N_o - 1$ and points to the correct index of `jptr`.[2] Once this second pass is complete, the two box-assist arrays will function as prescribed to reduce point by point comparisons.

C.2.3 Postprocessing and Output

The CP is theoretically an $N_o \times N_\epsilon \times \max N_e$ array which is stored as an $(N_\epsilon \times \max N_e) \times N_o$ array in the `CIHptn_mex.c` executable. It is postprocessed by transposition and column repositioning into an $(N_o \times N_\epsilon) \times \max N_e$ form. Its integer values are also normalized by the factor

$$N_o - 2 * W_T + 1; \tag{C.3}$$

[2]The assignment (**) cannot be replaced by `jptr[jh[j]--] = i` because the decrement is applied only after the assignment. It is only correct if it is applied beforehand as in

```
jh[j]--
jptr[jh[j]] = i
```

which is used here.

to become actual probabilities. The output returned by `CIHptn.m` is dependent on the number of embedding dimensions specified. If N_e is a scalar, then the output is CP as an $N_o \times N_\epsilon$ array. If N_e is a vector, then the output is a `Matlab` struct of same length as N_e with each component having the CP as a field filled by an $N_o \times N_\epsilon$ array.

C.3 CIHptn: Important Code

Though all thesis code is available in electronic form at `www.ee.umanitoba.ca/~kinsner/projects/`, Listing 2 contains the calling function for the `CIHptn.m` algorithm implemented in `Matlab`. The main calculation of the correlation partition is in Listing 3 on p. C-10 implemented as a `Matlab` mex-function coded in C. Supporting elements for the correlation partition engine, including the `mbase2` creation of the box-assist arrays, is coded in C as Listing 4 on p. C-21.

Listing 2 (CIHptn.m) The `Matlab` code of the CIHptn algorithm for correlation partitions.

```
function [out,scales] = CIHptn(data,scales,TW,lag,emdim,metr,p,verbose↓
→    );

% CIHPTN.m is a fast mex-based calculation of the general
% q-based correlation-integral partition.
% CIHptn.m requires the mex-installation provided by "makeCIptn.m".
%
% usage: [S,E] = CIHptn(X,E,[T],[L],[D],[metr],[P],[V])
% output: S is a struct of length(emdim), with fields S.emdim, S.scales, and
%         S.Ptn, if emdim is a vector. Otherwise, S is a matrix equivalent
%         to the partition S.Ptn. (see GPA_Ptn.m);
%         E is a vector of scales (maybe changed from input E;
% input: X is a matrix of M-dimensional row observations;
%         E is a vector of scales;
%         T is the scalar value for the Theiler window (default 1);
%         L is the scalar value for the embedding lag (default 1);
%         D is a vector of embedding dimensions (default M);
%         metr is a string identifying the metric to be used, see pdist.m (default 'che↓
→       ');
%         P is the scalar value of the Minkowski ('min') metric order;
%         V is the scalar flag for verbose output: default 0 (none), 1 (screen), 2 (↓
→      screen & files);
```

```
% CIHptn.m - v.1.1
% CREATED:  2007-07-20 by Michael Potter
% MODIFIED: 2007-08-14 (M. Potter) !removed the zero-elimination for compatibility with↓
→       new PtnFcn
%             2007-09-21 (M. Potter) !added makecol initialization
%             2008-01-08 (M. Potter) !added error checks in wrapper

%argcheck
if nargin < 2,
    disp('usage: [out,scales] = CIHptn(data,scales,[T],[L],[Edim],[↓
→       metr],[p],[verbose]');
    return
end
if nargin < 3 || isempty(TW),
    TW = 1;
end
if nargin < 4 || isempty(lag),
    lag = 1;
end
if nargin < 5 || isempty(emdim),
    emdim = max(2,size(data,2));
end
if nargin < 6 || isempty(metr),
    metr = 'che';
end
if nargin < 7 || isempty(p),
    p = 1;
end
if nargin < 8 || isempty(verbose),
    verbose = 0;
end

%initialize
Ns = length(scales);
if isvector(data),
    data = makecol(data);
end

%check bestscale
mindata = min(data);
maxdata = max(data);
rangedata = maxdata - mindata;
if isscalar(rangedata),
    rangedata = [rangedata rangedata];
end
mdata = size(rangedata,2);
```

```
    numcells = ceil(rangedata(1)/max(scales))*ceil(rangedata(mdata)/max(↓
→       scales));
 66 while numcells < 10,
        warning('CIHptn:␣max␣scale␣is␣too␣large␣for␣box␣assist␣method');
        disp('...reducing␣scales');
        scales = scales/2;
        numcells = ceil(rangedata(1)/max(scales))*ceil(rangedata(mdata)/↓
→           max(scales));
 71 end
    disp(['attractor␣covers␣',num2str(numcells),'␣cells␣at␣max␣scale']);

    %the action
    ptn = CIHmexwrapper(data,scales,TW,lag,emdim,metr,p,verbose);
 76 Ncorr = size(ptn,1)-2*TW+1;
    ptn = ptn/Ncorr;

    %reorganize data by emdim then scale
    out = zeros(size(ptn));
 81 Memdim = max(emdim);
    for k=1:Memdim,
        out(:,(1:Ns)+(k-1)*Ns) = ptn(:,k+((1:Ns)-1)*Memdim);
    end

 86 %reorganize into struct if necessary
    temp = out;
    out = struct('emdim',cell(1,length(emdim)),'scales',cell(1,length(↓
→       emdim)),'Ptn',cell(1,length(emdim)));
    for k=1:length(emdim),
        out(k).emdim = emdim(k);
 91     out(k).scales = scales;
        out(k).Ptn = temp(:,(1:Ns)+(emdim(k)-1)*Ns);
        out(k).Ptn = makecol(num2cell(out(k).Ptn,1));
        % Keep zeros values - removed in PtnFcn
    % %      for j=1:length(out(k).Ptn),
 96 % %           out(k).Ptn{j}(out(k).Ptn{j}==0)=[];    %remove zero values
    % %      end
    end
    if length(emdim)==1,
        out = out(1).Ptn;
101 end
    return

    function ptn = CIHmexwrapper(data,scales,TW,lag,emdim,metr,p,verbose);
    if ~isinteger(TW),
106     error('CIHptn:CIHmexwrapper:InvalidTW','variable␣"TW"␣must␣be␣an␣↓
→           integer');
    end
    if ~isinteger(lag),
```

```
        error('CIHptn:CIHmexwrapper:InvalidLag','variable␣"lag"␣must␣be␣an↓
→           ␣integer');
    end
    if ~isinteger(emdim),
        error('CIHptn:CIHmexwrapper:InvalidEmdim','variable␣"emdim"␣must␣↓
→           be␣an␣integer');
    end
    if ~ischar(metr),
        error('CIHptn:CIHmexwrapper:InvalidMetr','variable␣"metr"␣must␣be␣↓
→           a␣string');
    end

    if verbose == 0,
        ptn = CIHptn_mex(data.',scales,TW,lag,emdim,metr,p).';
    elseif verbose == 1,
        ptn = CIHptn_mexv(data.',scales,TW,lag,emdim,metr,p).';
    elseif verbose == 2,
        ptn = CIHptn_mexfo(data.',scales,TW,lag,emdim,metr,p).';
    else
        error('CIHptn:CIHmexwrapper:InvalidVerbose','improper␣assignment␣↓
→           to␣variable␣"verbose"');
    end
    return
```

Listing 3 (CI_Hptn_mex.c) The correlation partition engine CI_Hptn_mex.c (Matlab mex-function coded in C).

```
/*
 * CI_Hptn_mex.c
 *
 * Calculates Heaviside correlation integral partition between observations.
 * Helper function to CIHptn.m
 *
 * This is a MEX-file for MATLAB.
 * Created: 2007-05-25 Michael Potter
 * Developed from TISEAN algorithms after the MATLAB file pdistmex.cpp (Revision: ↓
→    1.1.6.2)
 */

/* $Revision: 2.1 $  $Date: 2007/07/23
 * NOTE: This is a "narrow" implementation of the TISEAN mbase routine
 * That is, the box-assist method is applied on the first and second
 * embedding dimensions. This may be slower than a "wide" implementation,
 * but ensures that the partitions will be accurate for all embedding
 * dimensions 2 and greater (i.e., not the maximum only).
 */
/* $Revision: 2.0 $  $Date: 2007/07/19 */
```

```
/* $Revision: 1.0 $  $Date: 2007/05/25 */

  /*****************************PREAMBLE & OPTIONS*****************************/
#include "mex.h"
#include <math.h>
#include <string.h>

/* Compile options - set with the -D option during mex compile
 *
 *    - DEBUGLAYER: [1:6] !restricts out code for debugging.
 *         !5 uses all code.
 *         !6 removes extra fileout files
 *         #define DEBUGLAYER 6
 *    - verbose: [0,1] !print extra information to terminal
 *         #define verbose 1
 *    - FILEOUT: [0,1] !write arrays to file for later analysis
 *         #define FILEOUT 1
 */

#include "TiseanPorts.c"


  /*****************************HELPER FUNCTIONS*****************************/
/* Euclidean interpoint distance */
void eucdist(const double *x, const double *y, const int kmax, double ↓
→       *d)
{
    /*
     d = sqrt(sum((XI-YJ).^2,2));                % Euclidean
     */
    int k;
    double   theSum,compdiff;
    const double   *XI,*YI;
    double *dI;

    XI = x;
    YI = y;
    dI = d;
    theSum = 0;
    for (k=0;k<kmax;k++,XI++,YI++) { /* sum over multiple dimensions */
      /*debug:        //       mexPrintf("*XI =%f\n",(*XI));
      //       mexPrintf("*YI =%f\n",(*YI)); */
      compdiff = (*XI)-(*YI);
      theSum += compdiff*compdiff;
      *(dI++) = (double)sqrt(theSum);
    }
}
```

```
/* Chebychev distance */
void chedist(const double *x, const double *y, const int kmax, double ↓
→       *d)
{
    /*
     d = max(abs(XI-XJ),[],2);
     */
    int k;
    double   theMax,compdiff;
    const double   *XI,*YI;
    double *dI;

    XI = x;
    YI = y;
    dI = d;
    theMax = 0;
    for (k=0;k<kmax;k++,XI++,YI++) { /* max over multiple dimensions */
      /* debug:        //       mexPrintf("*XI =%f\n",(*XI));
      //       mexPrintf("*YI =%f\n",(*YI)); */
      compdiff = (double)fabs((*XI)-(*YI));
      if (compdiff>theMax)
————→theMax = compdiff;
      *(dI++) = theMax;
    }
}

/* Minkowski distance */
void mindist(const double *x, const double *y, const int kmax, const ↓
→       double arg, double *d)
{
    /*
     d = sum(abs((XI-XJ)).^arg,2).^(1/arg);  % Minkowski
     */
    int k;
    double   theSum,compdiff;
    double   argRecip = 1/arg;
    const double   *XI,*YI;
    double *dI;

    XI = x;
    YI = y;
    dI = d;
    theSum = 0;
    for (k=0;k<kmax;k++,XI++,YI++) { /* sum over multiple dimensions */
      /* debug:        //       mexPrintf("*XI =%f\n",(*XI));
      //       mexPrintf("*YI =%f\n",(*YI)); */
      compdiff = (double)fabs((*XI)-(*YI));
```

```
113         theSum += (double)pow(compdiff,arg);
            *(dI++) = (double)pow(theSum,argRecip);
        }
    }

118


    /* maximum element function - max(x(:)) */
    double maxelement(const double *x, const int m)
    {
123     /*
         d = max(abs(XI-XJ),[],2);
         */
        int i;
        const double *XI;
128     double Y,theMax;

        XI = x;
        theMax = *XI; /* start with i = 0 */
        XI++;
133     for (i=1; i<m; i++,XI++) {
          Y = (*XI);
          if (Y>theMax) {
    ————→theMax = Y;
          }
138     }
        return theMax;
    }


143

    /* observation constructor */
    void ConstrObservationMatrix(const double *xin, const int d, const int↓
→       lag, const int n, const int m, double *yout)
    {
        /*
148      map observations into m-dimensional phase space with lag lag
         assuming d-dimensional data, gives n observations.
         Arranged as m -by- n array.
         */
        int i,j,k;
153     int vbose = 0;
        const double *x,*XI,*XI0;
        double *y;

        /* pointer x runs around getting values */
158     x = xin;
        y = yout;
```

```
        /* pointer XI0 moves from col to col in x-array */
        XI0 = x;
        if (vbose) {
163       mexPrintf("construct␣data␣\n");
          mexPrintf("value1␣=%f\n",(*x)); x++;
          mexPrintf("value2␣=%f\n",(*x)); x++;
          mexPrintf("value3␣=%f\n",(*x)); x++;
          mexPrintf("value4␣=%f\n",(*x)); x++;
168     }

        for (j=0; j<n; j++) { /* j counts through all points = cols of output/input ↓
→           */
            x = XI0;
    ————→if (vbose)
173 ————→  mexPrintf("col␣value␣=%f\n",(*x));
    ————→XI = XI0;
    ————→k=0; /* k counts through rows of observations = phase space coordinates */
    ————→while (k<m) {
    ————→  for (i=0; i<d; i++) { /* i counts through rows of input = data ↓
→       coordinates */
178 ————→    if (k<m) {
    ————→      (*y) = (*x);
    ————→      if (vbose)
    ————→————→mexPrintf("␣␣␣␣transcribed␣value␣=%f\n",(*x));
    ————→      k++; y++; x++;
183 ————→    }
    ————→  }
    ————→  XI += lag * d; /* advance to "lag" input cols away */
    ————→  x = XI;
    ————→}
188 ————→XI0 += d; /* next input col j; y has advanced by m */
        }
    }


193 /* lagfunction */
    int lagfunction(const int mD, const int mO)
    {
        int numlag;

198     numlag = mO / mD;
        if (mO % mD == 0)
          numlag--;
        return numlag;
    }
203
```

```
    /*************************MAIN*****************************/
  /* the gateway function */
208 void mexFunction(int nlhs, mxArray *plhs[], int nrhs, const mxArray *↓
→     prhs[])
  {
      int     status,numCoords,numPoints,numScales,numObs,numLags,↓
→        numCols;
      int     lag,Twindow,embeddimmax,*currmatrixdims;
      int     i,j,k,m;
213   char    metric[4];
      FILE    *distfid;
      double      *x,scalarArg;
      double      *eDim;
      double      *epsvec,*epsvec0,*xi,*ptn;
218   mxArray *Xembed,*dist;
    /* expect calling function as "CI_Hptn_mex(data,epsvec,Twindow,lag,embeddims,metric,↓
→       params)"  */

    /*************************ARGUMENTS*************************/

223 /*  check for proper number of arguments */
    if (nrhs<5) {
        mexErrMsgIdAndTxt("CI_Hptn_mex:TooFewInputs",
                          "Two␣input␣arguments␣required.");
    } else if(nlhs>1) {
228     mexErrMsgIdAndTxt("CI_Hptn_mex:TooManyOutputs",
                          "Too␣many␣output␣arguments.");
    }

  /* Check the type of the input array --  only works with double */
233   if (!mxIsDouble(prhs[0])) {
        mexErrMsgIdAndTxt("CI_Hptn_mex:BadInputType",
                          "CI_Hptn␣only␣supports␣real␣DOUBLE␣data.");
    }

238   if (verbose) {
      mexPrintf("CI_Hptnmex␣[CIHptn␣helper␣function]...\n");
      mexPrintf("␣CIHptnmex:␣...processing␣arguments\n");
    }

243 /*  Deal with Input Data */
    /*  create a pointer to the input matrix x */
    x = mxGetPr(prhs[0]);

    /*  get the dimensions of the matrix input x */
248 numCoords = mxGetM(prhs[0]);
    numPoints = mxGetN(prhs[0]);
    mexPrintf("␣Input␣numCoords␣=%d\n",numCoords);
```

```
        mexPrintf("␣Input␣numPoints␣=%d\n",numPoints);
        if (FILEOUT) {
253       WriteOut(x,numCoords,numPoints,"DataIn.out");  /* write out a text ↓
→             file copy for debugging  */
          x = mxGetPr(prhs[0]);
        }

      /*  Deal with scales eps-vector */
258   /*  create a pointer to the scales vector (eps-vector) */
        epsvec = mxGetPr(prhs[1]); /* this is a pointer */
        epsvec0 = epsvec;                  /* keep a record of first element pointer↓
→           */
        numScales = mxGetNumberOfElements(prhs[1]);
        mexPrintf("␣Analyzing␣%d␣scales␣concurrently\n",numScales);
263     if (verbose) {
          mexPrintf("␣epsvec(1)␣=%f\n",(*epsvec));
        }

      /*  Deal with Theiler window */
268   /* Check to make sure the lag is a real scalar. */
        if (!mxIsDouble(prhs[2]) || mxIsComplex(prhs[2]) || mxGetN(prhs↓
→           [2])*mxGetM(prhs[2]) != 1) {
          mexErrMsgTxt("Twindow␣must␣be␣a␣scalar.");
        }
      /*  create Twindow variable */
273     Twindow = (int)mxGetScalar(prhs[2]); /* this is a pointer */
        mexPrintf("␣Twindow␣=%d\n",Twindow);

      /*  Deal with lag */
      /* Check to make sure the lag is a real scalar. */
278     if (!mxIsDouble(prhs[3]) || mxIsComplex(prhs[3]) || mxGetN(prhs↓
→           [3])*mxGetM(prhs[3]) != 1) {
          mexErrMsgTxt("Lag␣␣must␣be␣a␣scalar.");
        }
      /*  create lag variable */
        lag = (int)mxGetScalar(prhs[3]);
283     mexPrintf("␣lag␣=%d\n",lag);

      /*  Deal with embedding dimension vector */
      /*  Only care about the maximum embedding dimension for this algorithm */
      /*  create a pointer to the integer embeddims array */
288     eDim = mxGetPr(prhs[4]);
        embeddimmax = (int)maxelement(eDim,mxGetNumberOfElements(prhs[4]))↓
→           ;
        mexPrintf("␣embeddimmax␣=%d\n",embeddimmax);

      /*  Deal with the metric */
293     status = mxGetString(prhs[5],metric,4);
```

```
        if (verbose) {
          mexPrintf("␣Calling␣%s␣distance␣metric",metric);  /* newline is ↓
→             observed below */
        }

298   /* Deal with extra arg  */
        if (nrhs>6 && !mxIsEmpty(prhs[6])) {
            if (mxGetNumberOfElements(prhs[6]) == 1) {  /*scalar case */
                scalarArg = (double)mxGetScalar(prhs[6]);
            } else {
303             mexErrMsgIdAndTxt("CI_Hptn_mex:MixedInputTypes",
                                  "Additional␣input␣arguments␣must␣be␣the␣↓
→                                    same␣class␣as␣X.");
            }
        } else {
          scalarArg = 1.0; /* default Minkowski metric is 'taxicab' */
308     }

        if (verbose) {
          if(strcmp(metric,"min") == 0)
    ————→mexPrintf("␣with␣p-exponent␣%f\n",scalarArg);
313       else
    ————→mexPrintf("\n");
        }

      /*************************PREPARATIONS*************************/
318     if (DEBUGLAYER > 0) {
          if (verbose) {
    ————→mexPrintf("␣CIHptnmex:␣...preparing␣observation␣arrays\n");
          }

323       /* PREPARE OBSERVATION (EMBEDDED) MATRIX */
          /*  arrange working matrix as (edims) -by- (numObs) // numObs < numPoints */
          numLags = lagfunction(numCoords,embeddimmax);
          if (verbose) {
    ————→mexPrintf("␣numLags=␣%d\n",numLags);
328       }
          numObs = numPoints - numLags * lag;
          if ( numObs * (embeddimmax) >= INT_MAX) {
            mexErrMsgIdAndTxt("CI_Hptn_mex:EmbedMatrixTooLarge",
                              "Embedded␣Observations␣matrix␣has␣more␣↓
→                                elements␣than␣the␣maximum␣allowed␣size␣↓
→                                in␣MATLAB.");
333       }
          if ( numObs < 1) {
            mexErrMsgIdAndTxt("CI_Hptn_mex:EmbedMatrixTooSmall",
                              "Embedded␣Observations␣matrix␣has␣no␣↓
→                                elements␣with␣lag␣settings.");
```

```
        }
338     Xembed = mxCreateNumericMatrix(embeddimmax, numObs, mxGetClassID↓
→           (prhs[0]), mxREAL);
        xi = mxGetPr(Xembed);
        if (verbose) {
    ————→currmatrixdims = (int*)mxGetDimensions(prhs[0]);
    ————→mexPrintf("␣input␣dims=␣%d␣%d\n",(*currmatrixdims++),(*↓
→       currmatrixdims));
343     }
        currmatrixdims = (int*)mxGetDimensions(Xembed);
        mexPrintf("␣Embedding␣is␣%d-dimensional␣for␣%d␣points\n",(*↓
→           currmatrixdims++),(*currmatrixdims));
        ConstrObservationMatrix(x, numCoords, lag, numObs, embeddimmax, ↓
→           xi); /* actively fills the matrix */
        if (FILEOUT) {
348 ————→xi = mxGetPr(Xembed);
    ————→WriteOut(xi,embeddimmax,numObs,"Xembed.out");  /* write out a text↓
→       file copy for debugging  */
    ————→distfid = fopen("dist.out","w");
        }

353
        /*  PREPARE OUTPUT (PARTITION) MATRIX */
        /* arrange output matrix as (edims-per-eps1 |...| edims-per-epsN) -by- (↓
→           numPoints) */
        numCols = numScales * embeddimmax;
        if (numCols * numObs >= INT_MAX) {
358       mexErrMsgIdAndTxt("CI_Hptn_mex:OutputTooLarge",
                            "Partition␣matrix␣has␣more␣elements␣than␣the↓
→                               ␣maximum␣allowed␣size␣in␣MATLAB.");
        }
        plhs[0] = mxCreateNumericMatrix(numCols, numObs, mxGetClassID(↓
→           prhs[0]), mxREAL);
        ptn = mxGetPr(plhs[0]); /* pointer to output matrix */
363     if (verbose) {
    ————→currmatrixdims = (int*)mxGetDimensions(plhs[0]);
    ————→mexPrintf("␣output␣dims=␣%d␣%d\n",(*currmatrixdims++),(*↓
→       currmatrixdims));
        }

368     if (DEBUGLAYER > 1) {
    ————→/* PREPARE TISEAN BOX-ASSIST METHODS */
    ————→int const IM = 100;
    ————→int const II = IM * IM * IM * IM;
    ————→int Njh = (IM) * (IM)+1;
373 ————→double grideps = maxelement(epsvec,numScales);
    ————→int *jh, *jptr;
    ————→/*    int jh[Njh]; int jptr[numObs]; */
```

```
    ————→jh = (int*)mxCalloc(Njh, sizeof(int));
    ————→jptr = (int*)mxCalloc(numObs, sizeof(int));
378
    ————→if (verbose) {
    ————→  mexPrintf("␣CIHptnmex:␣...preparing␣box-assist\n");
    ————→}

383 ————→xi = mxGetPr(Xembed);
    ————→mbase2(numObs, embeddimmax, xi, jh, jptr, grideps, &II, &IM); ↓
→       /* call database function */

    ————→if (DEBUGLAYER > 2) {

388 ————→  /* PREPARE FOR MAIN LOOP */
    ————→  int jj,kk,jk,boxjj,boxkk,s;
    ————→  int y;
    ————→  double *xj,*xj0,*d,*d0,*ptn0;
    ————→  dist = mxCreateNumericMatrix(1, numCols, mxGetClassID(prhs↓
→       [0]), mxREAL); /* distance vector */
393 ————→  d = mxGetPr(dist);     /* running pointer through distance vector */
    ————→  xi = mxGetPr(Xembed); /* pointer to position of point xi */
    ————→  xj = mxGetPr(Xembed); /* pointer to position of point xj */
    ————→  xj0 = xj;    /* first element in Xembed (first observation xj) */
    ————→  d0 = d;      /* first element in dist vector */
398 ————→  ptn0 = ptn; /* first element in output partition matrix */

    ————→  /*************************MAIN LOOP*****************************/
    ————→  if (verbose) {
    ————→    mexPrintf("␣CIHptnmex:␣...beginning␣main␣loop\n");
403 ————→  }


    ————→  for (i=0; i<numObs; i++) {
    ————→    if (verbose) {
408 ————→      mexPrintf("␣␣Initial␣point␣xi␣=␣%d\n",i);
    ————→    }
    ————→    /* for all points j - use box reduction of jh and jptr */
    ————→    jj = (int)(*(xi)/grideps);                              /*  ...↓
→       identifies the points row box */
    ————→    kk = (int)(*(xi+1)/grideps);            /*  ... identifies the↓
→       points col box */
413 ————→                                            /* Note: the above is ↓
→       a "narrow" box-assist */
    ————→    /* find comparison worthy points */
    ————→    for (j=jj-1; j<jj+2; j++) {                        /*  So for ↓
→       neighbouring row boxes */
    ————→      /* debug: if (verbose) mexPrintf("j =%d\n",j); */
```

```
    ————→        for (k=kk-1; k<kk+2; k++) {                    /*  ... and ↓
→       for neighbouring col boxes */
418 ————→————→/* debug:  if (verbose)  mexPrintf("k =%d\n",k); */
    ————→————→jk = baseindexfcn(j, k, &II, &IM);               /*  get this↓
→        neighbouring box */
    ————→————→for (y=jh[jk]; y<jh[jk+1]; y++) {                /*  and ↓
→       consider all points in it */

    ————→————→  if (fabs(jptr[y]-i) >= Twindow) {              /* ... ↓
→       consider Twindow correction */
423 ————→————→     /* have found a comparison worhty point */
    ————→————→     xj = xj0 + jptr[y]*embeddimmax;             /* advance ↓
→       xj pointer to that column in Xembed */
    ————→————→     /* debug: if (verbose) mexPrintf("Comparison point xj = %d\n",↓
→       jptr[y]); */

    ————→————→     if (DEBUGLAYER > 3) {
428 ————→————→       d = d0; /* restart at beginning of dist vector */
    ————→————→       /* Calculate distances for all embedding dimensions (length ↓
→       embeddimmax) */
    ————→————→       if (strcmp(metric,"euc") == 0)
    ————→————→————→eucdist(xi,xj,(int)embeddimmax,d);
    ————→————→       else if(strcmp(metric,"min") == 0)
433 ————→————→————→mindist(xi,xj,(int)embeddimmax,scalarArg,d);
    ————→————→       else if(strcmp(metric,"che") == 0)
    ————→————→————→chedist(xi,xj,(int)embeddimmax,d);
    ————→————→       if (FILEOUT) {  /* writeout distance vector to file */
    ————→————→————→/* d=d0; */
438 ————→————→————→for (m=0; m<embeddimmax; m++, d++) {
    ————→————→————→  fprintf(distfid,"%14.13e␣",*d);
    ————→————→————→}
    ————→————→————→fprintf(distfid,"\n");
    ————→————→       }
443
    ————→————→       if (DEBUGLAYER > 4) {

    ————→————→————→/* Check distances against scales and sum correlation ↓
→       integral  */
    ————→————→————→epsvec = epsvec0; /* restart at first scale */
448 ————→————→————→ptn = ptn0;       /* restart at first scale of point ↓
→       i */
    ————→————→————→for (s=0; s<numScales; s++, epsvec++) { /* sum ↓
→       over scales (slow)  */
    ————→————→————→  /* debug: mexPrintf("eps is currently =%f\n",(*epsvec))↓
→       ; */
    ————→————→————→  d = d0; /* restart at beginning of dist vector */
    ————→————→————→  for (m=0; m<embeddimmax; m++, d++, ptn++) { ↓
→       /* sum over embedding dimensions  */
```

```
                              /* debug; mexPrintf("edim is currently =%d\n",m+1); */
                              if (*d < *epsvec) /* Heaviside correlation integral */
                                (*ptn) += 1;
                            }
                          }
                        } /* DEBUGLAYER 5 */
                        /* end of correlation integral update for point j */
                      } /* DEBUGLAYER 4 */
                    } /* endif of point j Twindow test */
                  } /* end of point j */
              }     /* end of neighbour box col */
            }       /* end of neighbour box row */
            xi+=embeddimmax;        /* next point i */
            ptn0+=numCols;    /* next point i */
          } /* end of main loop */
          if (verbose) {
            mexPrintf(" CIHptnmex: ... main loop completed\n");
            mexPrintf("CIHptnmex: ... done.\n");
          }

      /* destroy memory allocations before exit */
          /*        mxDestroyArray(dist); */
        } /* DEBUGLAYER 3 */
          /*        mxFree(jh); */
          /*        mxFree(jptr); */
      } /* DEBUGLAYER 2 */
      if (FILEOUT)
        fclose(distfid);
      /*      mxDestroyArray(Xembed); */
    } else { /* DEBUGLAYER 1 */
      plhs[0] = mxCreateNumericMatrix(1, 1, mxGetClassID(prhs[0]), mxREAL);
    } /* DEBUGLAYER 0 */
}
```

Listing 4 (Tiseanports.c) Supporting elements for the correlation partition engine (including the box-assist components) ported from the Tisean Fortran code [108] into C.

```
/* Ports of Tisean utilities for fas neighbour searching
 * Created 2007/05/30 by Michael Potter
 * GPL2 License
 * /


/* WriteOut fcn - for debug usage */
```

```
    void WriteOut(const double *y, const int m, const int n, const char ↓
→       str[])
 9  {
        int i;
        const double *x;
        FILE *fid;
        x = y;
14      fid = fopen(str,"w");
        for (i=0; i<n*m; i++, x++) {
          fprintf(fid,"%14.13e ",*x);
          if (i%m == m-1)
    ————→fprintf(fid,"\n");
19      }
        fclose(fid);
    }

    /* WriteOut fcn - for debug usage */
24  void WriteOutLong(const int *y, const int m, const int n, const char ↓
→       str[])
    {
        int i;
        const int *x;
        FILE *fid;
29      x = y;
        fid = fopen(str,"w");
        for (i=0; i<n*m; i++, x++) {
          fprintf(fid,"%d ",*x);
          if (i%m == m-1)
34  ————→fprintf(fid,"\n");
        }
        fclose(fid);
    }

39

    /* initialize matrices  */
    void SetAllValues(int *x, const int y, const int kmax)
    {
        int k;
44

        /*     XI = x;  */
        for (k=0;k<kmax;k++,x++){ /* all elements in array */
          *(x) = y;
        }
49  }

    /* Indexing function for base routines */
    int baseindexfcn(int a, int b, int const *II, int const *IM)
    {
```

```
    return ((*IM) * ((a + (*II)) %(*IM)) + (b + (*II)) %(*IM));
}


/* Tisean port: "mbase2" function for fast neighbour searching - narrow version */
void mbase2(const int nmax, const int mmax, const double *y, int jh[],↓
→       int jptr[], const double eps, const int *II, const int *IM)
{
/* NOTE: This is a "narrow" implementation of the TISEAN mbase routine
 * That is, the box-assist method is applied on the first and second
 * embedding dimensions. This may be slower than a "wide" implementation,
 * but ensures that the partitions will be accurate for all embedding
 * dimensions 2 and greater (i.e., not the maximum only).
 */

  int i,n,Njh,BIF;
  int  vbose = verbose;
  const double *y1,*y2;
  FILE *bif_fid;


/*NOTE  no value for id prevents "base" from being called when mmax =1. Call function ↓
→     after embedding. */
/*        --FORTRAN
        do 10 i=0,im*im
        10        jh(i)=0   */
  /*      --Cpp         */
  Njh = (int)(*IM) * (*IM)+1;
  /* initialize to zero done at mxCalloc - below unnecessary*/
  /*  SetAllValues(jh,0,Njh);
      SetAllValues(jptr,0,nmax); */
  if (FILEOUT) {
      if (DEBUGLAYER < 6) {
          WriteOutLong(jh,1,Njh,"jh0.out");        /* write out a text file ↓
→             copy for debugging */
          WriteOutLong(jptr,1,nmax,"jptr0.out");   /* write out a text file ↓
→             copy for debugging */
      }
  }
  /*       --FORTRAN
      do 20 n=1,nmax                                          ! make histogram
        i=im*mod(int(y(n,1)/eps)+ii,im)+mod(int(y(n,mmax)/eps)+ii,im)
      20      jh(i)=jh(i)+1   */
  /*        --Cpp      */
  y1 = y;             /* first point */

  if (vbose) {
      mexPrintf("␣grideps␣=␣%f\n",eps);
```

```
    }
    if (FILEOUT) {
        bif_fid = fopen("bif.out","w");
    }

    for (n=0; n<nmax; n++, y1 += mmax) {
      y2 = y1+1;                          /* y1 is first coord, y2 is next in Xembed ↓
→         matrix */
                                          /* Note: the above is a "narrow" box-assist -↓
→                                             see top */
      BIF = baseindexfcn((int)(*y1/eps), (int)(*y2/eps), II, IM);
      if (FILEOUT) {
    ————→fprintf(bif_fid,"%d\n",BIF);
    ————→/*   debug:    mexPrintf("y1/eps, y2/eps = %f %f\n",*y1/eps,*y2/eps);
    ————→       mexPrintf("BIF = %d\n",BIF); */
      }
      jh[BIF]+=1;
    }
    if (FILEOUT) {
        if (DEBUGLAYER < 6) {
            WriteOutLong(jh,1,Njh,"jh1.out");  /* write out a text file copy ↓
→               for debugging */
        }
    }
    /*      --FORTRAN
        do 30 i=1,im*im                                                  ! accumulate ↓
→           it
        30      jh(i)=jh(i)+jh(i-1)   */
    /*      --Cpp           */
    for (i=1; i<Njh; i++)                 /* start at 1, not zero because of ↓
→       difference */
      jh[i] += jh[i-1];
    if (FILEOUT) {
        if (DEBUGLAYER < 6) {
            WriteOutLong(jh,1,Njh,"jh2.out");  /* write out a text file copy ↓
→               for debugging */
        }
    }
    /*      --FORTRAN
         do 40 n=(mmax-1)*id+1,nmax                          ! fill list of pointers
           i=im*mod(int(y(n,1)/eps)+ii,im)+mod(int(y(n,mmax)/eps)+ii,im)
           jptr(jh(i))=n
        40      jh(i)=jh(i)-1   */
    /*      --Cpp           */
    y1 = y;
    for (n=0; n<nmax; n++) {
      y2 = y1 + 1;                  /* Note: "narrow" box-assist */
      i = baseindexfcn((int)(*y1/eps), (int)(*y2/eps), II, IM);
```

```
        jh[i]--;                    /* decrease it first to keep 0:N-1 indexing */
        jptr[jh[i]] = n;
        /* jh[i]--;                 //...not here */
        /*    jptr[*(jh[i])--] = y1; // doesn't work because it increments poast ↓
→           assignment */
144     y1 += mmax;
      }
      if (FILEOUT) {
        WriteOutLong(jh,1,Njh,"jh3.out");  /* write out a text file copy for ↓
→           debugging */
        WriteOutLong(jptr,1,nmax,"jptr.out");  /* write out a text file copy for ↓
→           debugging */
149     fclose(bif_fid);
      }
    }


154 /* Tisean port: "mbase2" function for fast neighbour searching - wide version */
    void mbase2wide(const int nmax, const int mmax, const double *y, int ↓
→       jh[], int jptr[], const double eps, const int *II, const int *IM)
    {
    /* NOTE: This is a "wide" implementation of the TISEAN mbase routine
     * That is, the box-assist method is applied on the first and last
159  * embedding dimensions. This may be faster than a "narrow" implementation,
     * but  the partitions may be inaccurate for all embedding dimensions less
     * than the maximum.
     */

164   int i,n,Njh,BIF;
      int  vbose = verbose;
      const double *y1,*y2;
      FILE *bif_fid;

169
    /*NOTE  no value for id prevents "base" from being called when mmax =1. Call function ↓
→       after embedding. */
    /*       --FORTRAN
           do 10 i=0,im*im
           10       jh(i)=0   */
174   /*      --Cpp         */
      Njh = (int)(*IM) * (*IM)+1;
      /* initialize to zero done at mxCalloc - below unnecessary*/
      /*  SetAllValues(jh,0,Njh);
           SetAllValues(jptr,0,nmax); */
179   if (FILEOUT) {
        if (DEBUGLAYER < 6) {
          WriteOutLong(jh,1,Njh,"jh0.out");        /* write out a text file ↓
→             copy for debugging */
```

```
            WriteOutLong(jptr,1,nmax,"jptr0.out");  /* write out a text file ↓
→               copy for debugging */
        }
184 }
    /*      --FORTRAN
        do 20 n=1,nmax                                          ! make histogram
          i=im*mod(int(y(n,1)/eps)+ii,im)+mod(int(y(n,mmax)/eps)+ii,im)
        20      jh(i)=jh(i)+1   */
189 /*       --Cpp      */
    y1 = y;             /* first point */

    if (vbose) {
        mexPrintf("␣grideps␣=␣%f\n",eps);
194 }
    if (FILEOUT) {
        bif_fid = fopen("bif.out","w");
    }

199 for (n=0; n<nmax; n++, y1 += mmax) {
      y2 = y1+mmax-1;                       /* y1 is first coord, y2 is last in ↓
→          Xembed matrix */
                                            /* The above is a "wide" box-assist - ↓
→                                               see Note at top */
      BIF = baseindexfcn((int)(*y1/eps), (int)(*y2/eps), II, IM);
      if (FILEOUT) {
204 ————→fprintf(bif_fid,"%d\n",BIF);
    ————→/*   debug:     mexPrintf("y1/eps, y2/eps = %f %f\n",*y1/eps,*y2/eps);
    ————→        mexPrintf("BIF = %d\n",BIF); */
        }
      jh[BIF]+=1;
209 }
    if (FILEOUT) {
        if (DEBUGLAYER < 6) {
            WriteOutLong(jh,1,Njh,"jh1.out");  /* write out a text file copy ↓
→               for debugging */
        }
214 }
    /*      --FORTRAN
        do 30 i=1,im*im                                         ! accumulate ↓
→           it
        30      jh(i)=jh(i)+jh(i-1)   */
    /*       --Cpp          */
219 for (i=1; i<Njh; i++)               /* start at 1, not zero because of ↓
→       difference */
      jh[i] += jh[i-1];
    if (FILEOUT) {
        if (DEBUGLAYER < 6) {
```

```
                WriteOutLong(jh,1,Njh,"jh2.out");  /* write out a text file copy ↓
→                   for debugging */
224         }
        }
        /*         --FORTRAN
              do 40 n=(mmax-1)*id+1,nmax                          ! fill list of pointers
                i=im*mod(int(y(n,1)/eps)+ii,im)+mod(int(y(n,mmax)/eps)+ii,im)
229             jptr(jh(i))=n
              40      jh(i)=jh(i)-1   */
        /*         --Cpp          */
        y1 = y;
        for (n=0; n<nmax; n++) {
234       y2 = y1 + mmax-1;                          /* Note: "wide" box-assist */
          i = baseindexfcn((int)(*y1/eps), (int)(*y2/eps), II, IM);
          jh[i]--;                      /* decrease it first to keep 0:N-1 indexing */
          jptr[jh[i]] = n;
          /* jh[i]--;                  //...not here */
239       /*     jptr[*(jh[i])--] = y1; // doesn't work because it increments poast ↓
→             assignment */
          y1 += mmax;
        }
        if (FILEOUT) {
          WriteOutLong(jh,1,Njh,"jh3.out");  /* write out a text file copy for ↓
→             debugging */
244       WriteOutLong(jptr,1,nmax,"jptr.out");  /* write out a text file copy for ↓
→             debugging */
          fclose(bif_fid);
        }
    }
```

C.4 Converting the Partition into Entropies

As described in Sec. 4.7.3, the novelty of the proposed direct method is that the MS is calculated directly from the Hölder and Mandelbrot entropies defined in (4.140) and (4.141). These in turn are calculated from the correlation partition. These entropies have similar but different forms from their histogram partition counterparts (4.125) and (4.127). In fact, an analysis of the formulae demonstrates that the *both* forms of partitions can be harmonized as

follows. Let $Z_\mu(\epsilon)$ come from a histogram partitioning with n active cells. Define

$$\boldsymbol{m}_\epsilon = [p_1, \ldots, p_n] \tag{C.4}$$

$$\boldsymbol{w}_\epsilon = [p_1, \ldots, p_n] \tag{C.5}$$

as vectors indexed by i of length N_{grid}. Equivalently, for a correlation partition $\tilde{Z}_\mu(\epsilon)$, define

$$\boldsymbol{m}_\epsilon = [\tilde{p}_1, \ldots, \tilde{p}_N] \tag{C.6}$$

$$\boldsymbol{w}_\epsilon = [1/N, \ldots, 1/N] \tag{C.7}$$

as vectors indexed by i but of length N. By construction, then

$$\sum_i \boldsymbol{w}_\epsilon(i) = 1 \tag{C.8}$$

for both cases. It follows that the MS scaling entropies $\gamma_q(q, \epsilon)$ and $\Upsilon_q(q, \epsilon)$ (and the RS scaling entropy H_q) can be written exclusively and uniformly in terms of these vectors. Specifically, these are enhanced to three other variables using the vector of q values supplied by the user as

$$\boldsymbol{m}_\epsilon^{[q]} = \boldsymbol{m}_\epsilon^{(q-1)} \qquad \forall q \tag{C.9}$$

$$Z_\epsilon^{[q]} = \sum_i \boldsymbol{w}_\epsilon \,.*\, \boldsymbol{m}_\epsilon^{[q]} \qquad \forall q \tag{C.10}$$

$$\boldsymbol{r}_\epsilon^{[q]} = \boldsymbol{m}_\epsilon^{[q]} \,./\, Z_\epsilon^{[q]} \qquad \forall q \tag{C.11}$$

$$\tag{C.12}$$

where the "dot-operations" indicate element-wise operations after the `Matlab` syntax. This enhancement to correlation partitioning is implemented in the `Matlab` function `PtnFcn.m`.

As a consequence of the new vector variables, the definitions of the MFA scaling entropies

can be harmonized for both histogram and correlation partitions as

$$H_q(q,\epsilon) = -\log Z_\epsilon^{[q]} ./ (1-q) \qquad \forall q \tag{C.13}$$

$$\gamma_q(q,\epsilon) = \sum_i \boldsymbol{w}_\epsilon .* \boldsymbol{r}_\epsilon^{[q]} .* \log \boldsymbol{m}_\epsilon \qquad \forall q \tag{C.14}$$

$$\Upsilon_q(q,\epsilon) = \sum_i \boldsymbol{w}_\epsilon .* \boldsymbol{r}_\epsilon^{[q]} .* \log(\boldsymbol{m}_\epsilon .* \boldsymbol{r}_\epsilon^{[q]}) \qquad \forall q \tag{C.15}$$

These are calculated in the `RSpectrum.m` and `calcMSentropies.m` functions.

C.5 Summary

This appendix has covered the details of the implementation of the novel direct calculation of the Mandelbrot spectrum from a correlation partition. This technique was used in this work for the analysis of MFA feature convergence. Explicit descriptions of (i) the box-assist method from [108], (ii) the definition of the correlation partition, and (iii) the conversion of the correlation partition into the three scaling entropies has been provided in harmonized form. The `Matlab` and supporting C code has been included in Sec. C.3, but is also available electronically from `www.ee.umanitoba.ca/~kinsner/projects/`. Note especially that a bug was identified in the original TISEAN `mbase2.c` algorithm [108], but is corrected in this implementation.

The direct calculation of the MS from CP has been verified [177, Potter and Kinsner (2007)], and is a tool ready for application in MFA. Future study is recommended, however, of (i) the significance of the lack of symmetry in the Mandelbrot entropy, and (ii) the sensitivity of the MFA spectra to embedding dimensions.

Appendix D

WAVELET-BASED MFA

D.1 Overview

This appendix covers the background on the wavelet-based approach to time series MFA. The *wavelet-transform modulus-maxima* (WTMM) approach for the definition of a mutifractal partition function described here was eliminated from the experimental procedure to simplify the scope of the work. It is included for completeness since the background on MFA scaling in Ch. 4 was designed in such a way as to unify the "fractal function" approach of the WTMM with the traditional attractor-based "fractal density" approach used in Ch. 6 and App. C in a measure theoretic paradigm.

D.2 WTMM: MFA of Fractal Signals

Recall Sec. 4.6.1 considered an MFA characterization of ECG via attractor reconstruction. Here, an alternative method for MFA treats the ECG signal $\boldsymbol{x}(t)$ as a fractal function. As with all MFA, what is required is a coarse-graining of a measure and a partition function. The *wavelet formalism* directly identifies *wavelet coefficients* as coarse-grained measures [154, (1991)].

The *continuous wavelet transform* (WT) is a well-known signal processing technique [129] that maps functions in $x(t) \in L^2(\mathbb{R})$ into a surface $\mathcal{W}_\psi[x](a,t)$ over $\mathbb{R}^+ \times \mathbb{R}$. Defined as the

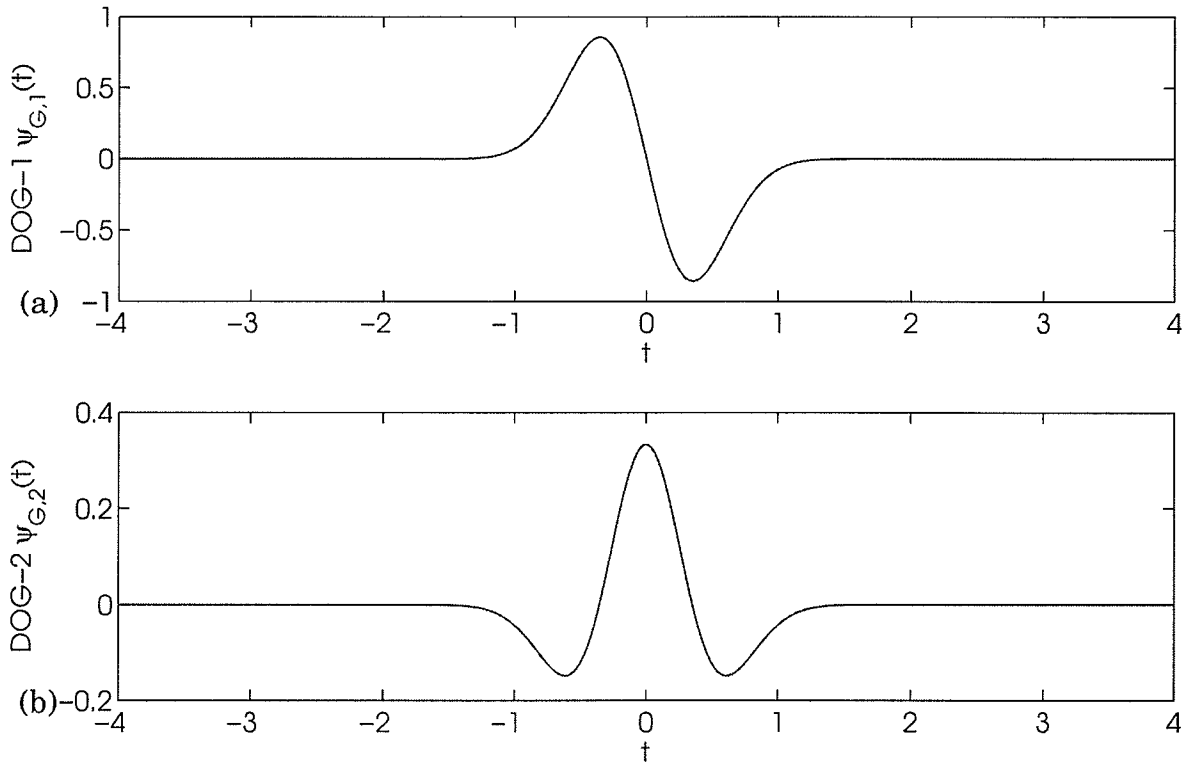


Fig. D.1 DOG smoothing wavelets: (a) DOG-1, $\psi_{g,1}$; (b) DOG-2, $\psi_{g,2}$.

convolution with a scaled and translated wavelet function ψ, the WT is

$$\mathcal{W}_\psi[x](a,t) = \frac{1}{a}\int_{\mathbb{R}} x(t')\psi\left(\frac{t'-t}{a}\right)dt' \ . \tag{D.1}$$

This has many utilities depending on the form of ψ. The ABM-technique relies on *smoothing wavelets* such as *derivatives of the Gaussian* (DOG) function,

$$\psi_{g,n} \propto \frac{d^n}{dt^n}e^{-t^2} \tag{D.2}$$

some of which are shown in Fig. D.1

For a fixed a, the curve $\mathcal{W}_\psi[x](a,t)$ is a smoothed version of the original signal $x(t)$, Fig. D.2. As such, it contains a lot of redundancy. However, by changing the value of the scale a, the "window-size", or scale, of the smoothing is changed. This "wavelet zoom"

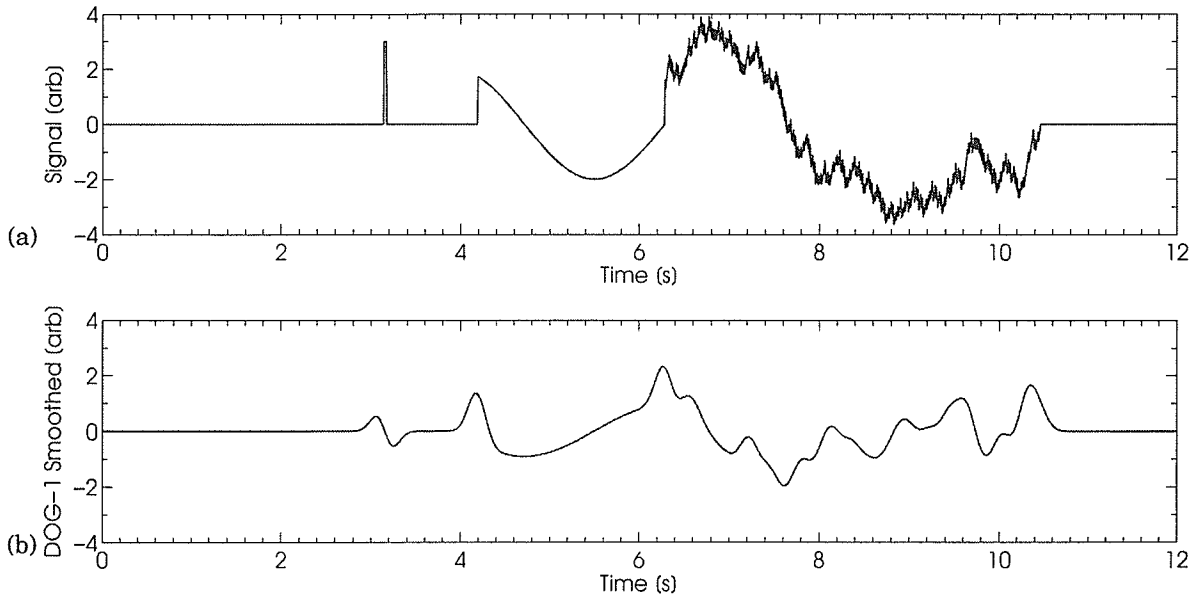


Fig. D.2 Wavelet smoothing: (a) signal $x(t)$; (b) smoothed by DOG-1 WT

feature [129, Mallat (1998)] for functions is very similar in principle to the coarse-graining of measures by ϵ-coverings. In fact, Mallat and Zhong [131, 1992] have demonstrated that the redundancy of the WT surface could be reduced by looking not on the full support $(a, t) \in \mathbb{R}^+ \times \mathbb{R}$, but the support of the WTMM, which contains all the information about the singularities in the function $x(t)$, and can often reconstruct the entire signal [129, Mallat (1998)]. In the discussions on *measure singularities* in Ch. 4, Hölder exponents, and their relationship to non-differentiability, were discussed. Here, a function-based analog is used to quantify the non-differentiability of the function $x(t)$ by the *Lipschitz-Hölder exponent*, h.

Definition D.1 (Lipschitz-Hölder Exponent, h)**.** Let $x(t)$ be a function on $\mathbb{R}$. Then x has a *Lipschitz-Hölder exponent* $h(t)$ at time t if for some polynomial $P_n(t)$ and some constant C

$$|x(t + \epsilon) - P_n(t + \epsilon)| \leq C\,|\epsilon|^{h(t)} \tag{D.3}$$

where $n = \lfloor h(t) \rfloor$. Written shorthand, $x(t + \epsilon) \sim |\epsilon|^{h(t)}$.

As observed before in the measure-theoretic case, differentiability and non-differentiability are quantified by this definition of scaling exponent. Note that most of the usual classes of smooth functions are characterized by a single Lipschitz-Hölder exponent on their entire support. If this single value applies everywhere, the spectrum of Lipschitz-Hölder exponents will be called *degenerate*. For example, polynomials of order n will have degenerate exponent spectrum equal to $h(t) = n$ for all t. Similarly, smooth functions $x(t) \in C^n(\mathbb{R})$ have $h(t) \geq n$. Analytic functions have a degenerate infinite spectrum, $h(t) = \infty$. That said, non-differentiability can now be quantified, with such examples as a discontinuity at t, $h(t) \leq 0$, or a Dirac Delta singularity, $h(t) = -1$. Integration of a function x has the rather straightforward property of increasing the exponent by one, $h_{\int x}(t) = h_x(t) + 1$.

In the presentation of Ch. 4, the MS of a measure was expressed as the curve of Hölder exponents and Hausdorff dimensions, $(\alpha, D_0(\alpha))$, where the Hausdorff dimension measures the equisingular set A_α, $D_0(\alpha) \equiv D_0(A_\alpha)$. The MS for a function is no different except the equisingular sets are now defined

$$A_h = \{ t \mid h(t) = h \} \tag{D.4}$$

resulting in a MS, $(h, D_0(h))$. Since the sets $A_h \subset \mathbb{R}$, it follows that $D_0(h) \leq 1$. The MS of a function therefore is expressed in a space like Fig. D.3.

Now the discussion will return to the WTMM in order to describe how the $(h, D_0(h))$ spectrum can be estimated. As mentioned, the modulus profile of the WT surface is smooth due to the nature of the smoothing wavelet, but local maxima exist in the modulus profile related to the singularities in the original signal x, Fig. D.4. By considering only these local maxima points of the surface, what is observed is the bifurcating "tree-like" effect as the wavelet scale a becomes finer, Fig. D.4(c). This support is the WTMM-tree, denoted $\mathcal{T}$. Note that at each scale, the WTMM-tree has finitely many branches. That is, for every scale a,

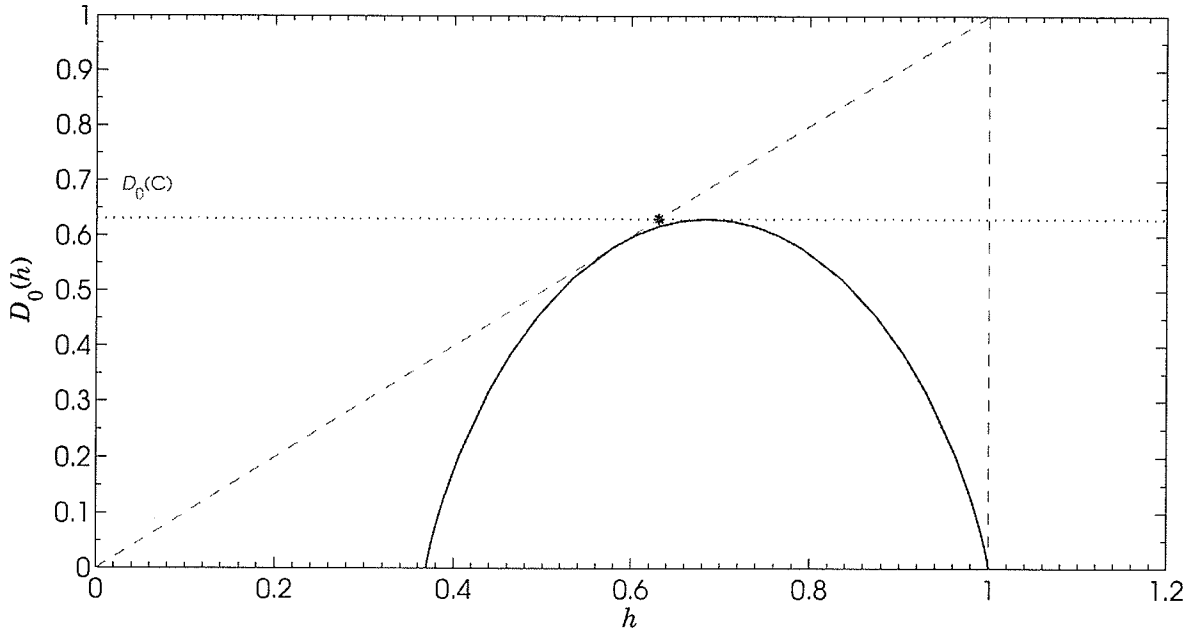


Fig. D.3 The WTMM MS of a function: $D_0(h)$ vs. h.

the WTMM of the WT profile has mapped the signal $x(t)$ into a finite sequence (*i.e.*, $\in l^0$). Functionally, this is exactly the same action as the measure coarse-graining, Fig. D.4(d). Thus for the characterization of the signal by scaling analysis, all that is required is a Boltzmann partition function defined from the finite sequences of the WTMM-tree, [154, ABM (1991)]

$$Z_B(x, q \mid a) = \sum_{t \in \mathcal{T}(a)} \left| \mathcal{W}_\psi[x](a,t) \right|^q \tag{D.5}$$

Then, exactly as in the previous derivation for the Hölder and Mandelbrot entropies in Ch. 4, the $(h, D_0(h))$ relationship can be estimated parametrically by q as [154, ABM (1991)]

$$h(q) \approx \frac{1}{\log a} \sum_{t \in \mathcal{T}(a)} \frac{\left| \mathcal{W}_\psi[x](a,t) \right|^q}{Z_B(x, q \mid a)} \log \left| \mathcal{W}_\psi[x](a,t) \right| \tag{D.6}$$

$$D_0(q) \approx \frac{1}{\log \epsilon} \sum_{t \in \mathcal{T}(a)} \frac{\left| \mathcal{W}_\psi[x](a,t) \right|^q}{Z_B(x, q \mid a)} \log \frac{\left| \mathcal{W}_\psi[x](a,t) \right|^q}{Z_B(x, q \mid a)} \tag{D.7}$$

The above is very similar to the expressions in Ch. 4 for the Boltzmann partition function from a histogram-based coarse-graining, except now the coarse-graining is parameterized by

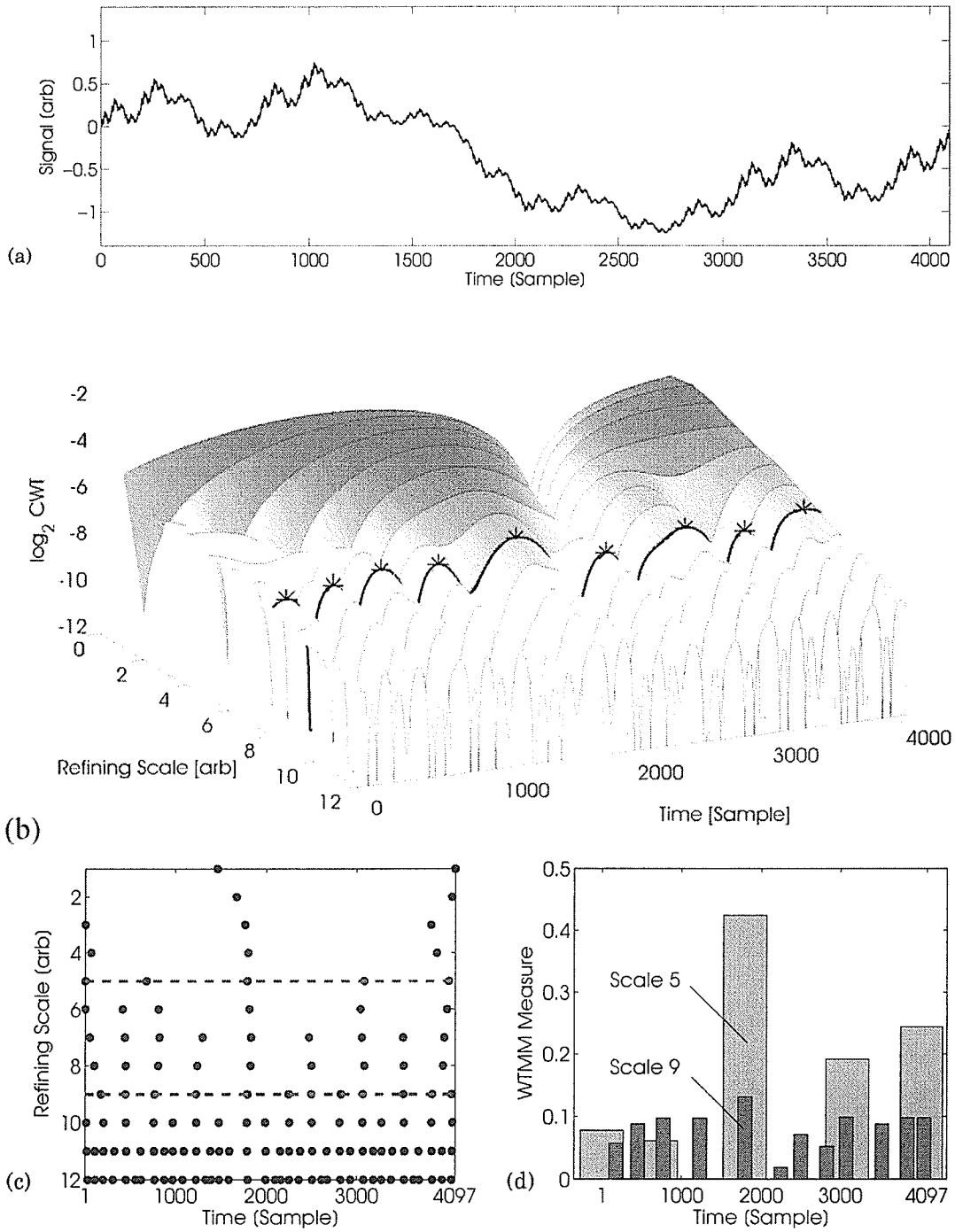


Fig. D.4 The WTMM partition process: (a) a non-differentiable signal $x(t)$; (b) WT surface, with some local maxima points identified; (c) WTMM tree (the position of all WTMM); (d) discrete WTMM measures

a instead of ϵ. Note however that here the *time-ordering* is an intrinsic element of this scaling analysis. Time-shuffling the signal changes the singular nature of the signal completely, since it is reduced essentially to white noise! Not only that, the MFA scaling by the WTMM formalism is characteristic of a particular time interval. It is of interest for future work to determine whether the cyclostationary nature of the ECG implies that a WTMM-MFA is essentially invariant when applied over many cycles (*i.e.*, is robust to the selection of time-windowing).

D.3 WTMM Calculation

Here a direct technique for multifractal analysis that considers the multifractal singularities (non-differentiable points) of the time series is considered, hereafter referred to as the *ABM-technique*.

Consider a real signal s. Then the continuous wavelet transform of s, $\mathcal{W}_\psi[s]$, is a surface on $\mathbb{R}^2$

$$\mathcal{W}_\psi[s](a,t) = \frac{1}{a}\int_{\mathbb{R}} s(t')\psi\left(\frac{t'-t}{a}\right)dt' \tag{D.8}$$

characteristic of the special L^2 function ψ called the *smoothing wavelet*. For further coverage of wavelet theory, the reader is referred to [129].

Mallat and Hwang [130, 1992] demonstrated that the redundancy of the WT could be reduced by looking not on the full support $\mathbb{R}^2$, but the support of the *wavelet transform modulus maxima* (WTMM)-tree. The modulus profile of the WT surface is smooth due to the nature of the smoothing wavelet, but local maxima exist in the modulus profile related to the singularities in the original signal s. By considering only these local maxima points of the surface, what is observed is the bifurcating "tree-like" effect as the wavelet scale a becomes finer. This WTMM-tree, $\mathcal{T}$, carries all the information of singularity scaling of the signal s, by defining

a partition function $Z_q(a)$ from $\mathcal{T}$ by [154][16]

$$Z_q(a) = \sum_{t \in \mathcal{T}} \left(\left| \mathcal{W}_\psi[s](a,t) \right| \right)^q \tag{D.9}$$

Research [217, Struzik, (2000)] has shown that local Hölder exponents estimated from the slope of $\log_2 \left| \mathcal{W}_\psi[s](a,t) \right|$ vs. $\log_2 a$ is unstable. Thus a global Hölder exponent is measured from Z_q by introducing the weights $W_\psi[s](a, t \in \mathcal{T})$

$$W_\psi[s](q,a,t) = \frac{\left(\left| \mathcal{W}_\psi[s](a,t) \right| \right)^q}{Z_q(a)} \quad ; \quad t \in \mathcal{T} \tag{D.10}$$

from which the global Hölder exponent

$$h(q) = \lim_{a \to 0} \frac{\sum_{t \in \mathcal{T}} W_\psi[s](q,a,t) \log \left| \mathcal{W}_\psi[s](a,t) \right|}{\log(a)} \tag{D.11}$$

is calculated.[1] Here, the reader should observe that a probability measure (Fig. D.4(d)) has been created on an unusual bifurcating support (Fig. D.4(c)) using the WTMM-tree.

The Hausdorff dimension of the subset with Hölder exponent h, A_h, is then calculated by the formula

$$D_0(q) = \lim_{a \to 0} \frac{\sum_{t \in \mathcal{T}} W_\psi[s](q,a,t) \log W_\psi[s](q,a,t)}{\log(a)} \tag{D.12}$$

related to (D.11) by the MS thermodynamic formalism.

D.4 Summary

This appendix briefly introduces the WTMM approach as a convenience to the interested reader. The WTMM approach to MFA is based on scaling partitions of modulus maxima coefficients from a smoothing wavelet transform. This approach was discarded from the present work to simplify the scope of the experiments. Research is required, however, to analyze the

[1] $a \to 0$ are the finer scales.

relationship of the WTMM-approach (which is applied along the time axis) to the traditional attractor approach (which requires an embedding and is applied in phase space). To the knowledge of the author, no connection between the contrasting approaches has been studied in the literature.