Kalman Filtering Beyond Gaussian Innovation Processes

by

Tiancheng Gao

A Thesis submitted to The Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Electrical and Computer Engineering

University of Manitoba

Winnipeg

December 2022

Nothing great in the world has ever been accomplished without passion.

George Wilhelm Friedrich Hegel

Abstract

Estimating time-varying signals becomes particularly challenging under non Gaussian innovation processes such as sparse and rapidly time-varying noise dynamics. In this thesis, by building upon the recent progress in the approximate message passing (AMP) algorithms, the vector AMP (VAMP) algorithm is unified with the Kalman filter (KF) into a common message passing framework that we coin VAMP-KF. The advantage of VAMP-KF is that it does not restrict the innovation dynamics to have a specific structure (e.g., same support over time when the innovation is sparse), thereby accounting for uncorrelated noise dynamics without the need of explicit innovation correlation modelling. For the sake of theoretical performance prediction, we conduct a state evolution (SE) analysis of the proposed algorithm and show its consistency with the asymptotic empirical mean-squared error (MSE). Numerical results on various rapidly time-varying innovation dynamics (e.g., with different sparsity rates) demonstrate unambiguously the effectiveness of the proposed VAMP-KF algorithm and its superiority over state of-the-art algorithms both in terms of reconstruction accuracy and computational complexity.

Contents

List of Figures

List of Abbreviations

AMP	Approximate message passing
AWGN	Additive white Gaussian noise
EP	Expectation propagation
i.i.d.	Independent and identically distributed
KF	Kalman filter
LMMSE	Linear minimum mean square error
MMSE	Minimum mean square error
MSE	Mean-square error
NRMSE	Normalized root mean-square error
SE	State evolution
SNR	Signal to noise ratio
SP	Sum product
SVD	Singular value decomposition
TNRMSE	Time-averaged normalized root mean-square error
VAMP	Vector approximate message passing
VAMP-KF	Vector approximate message passing Kalman filter

Acknowledgement

First and foremost, I would like to express my deepest gratitude to my advisor, Professor Faouzi Bellili. His constant support and advice are what made this thesis possible. I am also thankful to Professor Amine Mezghani who has shaped many of the core ideas of this work. His constant kindness and patience liberated me from frustration and built my confidence over time.

I would like to thank my colleague Mohamed Akrout, with whom I had great discussions and who was of a great help to complete this thesis.

My special thanks go to Professor Ekram Hossain who took time out of his busy schedule to evaluate this thesis.

Finally, I dedicate this work to my family for their endless support and encouragement.

Chapter 1

Introduction

1.1 Background and motivation

We consider the problem of estimating a collection of T state vectors $\{\boldsymbol{x}_t\}_{t=1}^{t=T}$ from their noisy observation vectors $\{\boldsymbol{y}_t\}_{t=1}^{t=T}$. As shown in Fig. 1.1, each measurement vector $\boldsymbol{y}_t$ at time step t is obtained as

$$\boldsymbol{y}_t = \boldsymbol{H}_t\, \boldsymbol{x}_t + \boldsymbol{w}_t. \tag{1.1}$$

Here, $\boldsymbol{H}_t \in \mathbb{C}^{M\times N}$ is the measurement matrix and $\boldsymbol{w}_t \in \mathbb{C}^M$ is an additive white Gaussian noise vector whose entries are assumed to be independent and identically distributed (i.i.d.) with mean zero and variance γ_{w}^{-1}, i.e., $w_{ij} \sim \mathcal{CN}(w_{ij}; 0, \gamma_{\mathsf{w}}^{-1})$. Moreover, each state vector $\boldsymbol{x}_t \in \mathbb{C}^N$ evolves from time step $t-1$ to time step t according to the state process equation:

$$\boldsymbol{x}_t = \alpha\, \boldsymbol{F}_t\, \boldsymbol{x}_{t-1} + \sqrt{1-\alpha^2}\, \boldsymbol{v}_t. \tag{1.2}$$

In (1.2), $\boldsymbol{F}_t \in \mathbb{C}^{N\times N}$ is the state transition matrix, and $\alpha \in [0,1]$ is a memory factor balancing the noise and signal contributions in the state evolution process. Moreover, $\boldsymbol{v}_t \in \mathbb{C}^M$ is an additive noise vector that represents the error in the state evolution

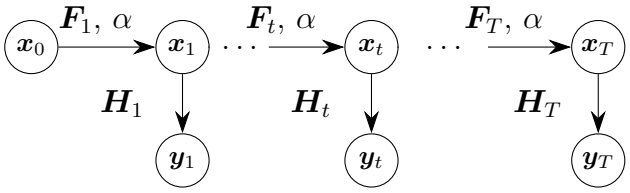


Figure 1.1: State-space model representation of the process and state equations given in (1.1) and (1.2).

dynamics which is commonly referred to as an *innovation*. Here, we assume that i) the signal and innovation vectors, $\boldsymbol{x}_t$ and $\boldsymbol{v}_t$, have the same average component-wise power for all $t \in \{0, \ldots, T\}$, and ii) the transition matrix $\boldsymbol{F}_t$ evolves the signal $\boldsymbol{x}_{t-1}$ without altering the signal power. That is to say:

$$\tfrac{1}{N} \left\|\boldsymbol{x}_t\right\|_2^2 = \tfrac{1}{N} \left\|\boldsymbol{v}_t\right\|_2^2, \qquad \forall\, t \in \{0, \ldots, T\}, \tag{1.3a}$$

$$\tfrac{1}{N} \left\|\boldsymbol{F}_t\, \boldsymbol{x}_{t-1}\right\|_2^2 = \tfrac{1}{N} \left\|\boldsymbol{x}_{t-1}\right\|_2^2, \quad \forall\, t \in \{1, \ldots, T\}. \tag{1.3b}$$

The assumptions in (1.3) are needed to avoid having the power of the signal $\boldsymbol{x}_t$ in (1.2) accumulate and dominate the innovation $\boldsymbol{v}_t$ with time.

When the initial state, $\boldsymbol{x}_0$, the measurement noise, $\boldsymbol{w}_t$, and the innovation, $\boldsymbol{v}_t$, are Gaussian distributed, the standard Kalman filter (KF) provides the optimal recursive linear minimum mean-square error (LMMSE) estimator to the state $\boldsymbol{x}_t$ for all $t \in \{1, \ldots, T\}$ [1]. However, for non-Gaussian (e.g., heavy-tailed or sparse) noise vectors, the degradation in the performance of the KF is rather severe. Since the introduction of the KF in [2], a plethora of filtering techniques have been developed to improve the performance of the KF for state estimation in the presence of non-Gaussiance noise.

In general, there are two distinct (yet possibly overlapping) scenarios accounting for non-Gaussian noise processes:

- *non-Gaussian measurement noise* $\boldsymbol{w}_t$: some natural phenomena (e.g., atmospheric noise, lightning spikes and ice cracking) as well as engineering applications (e.g.,

electronic devices, lasers, relay switching) can be more accurately characterized by heavy-tailed non-Gaussian measurement noise models [3].

- *non-Gaussian innovation* $\boldsymbol{v}_t$: realistic target tracking scenarios in radar and robotics rely on accurate target motion models. The latter must therefore accurately capture the uncertainty in the target maneuvers with non-Gaussian innovation distributions [4]. The innovation can also be sparse in many applications such as parallel magnetic resonance imaging [5] and direction-of-arrival estimation [6] in wireless communication[1].

1.2 Prior work

Previous studies have addressed the aforementioned non-Gaussianity of either the innovation $\boldsymbol{v}_t$ or the measurement noise $\boldsymbol{w}_t$, or both of them combined. Here, we only consider the innovation $\boldsymbol{v}_t$ to be a non-Gaussian vector and assume that the measurement noise $\boldsymbol{w}_t$ in (1.1) is an additive white Gaussian noise (AWGN). This assumption follows the widely used AWGN model in most engineering systems in signal processing, control, and communication.

To put our method for solving the state estimation problem with non-Gaussian innovations in proper perspective, we classify in Fig. 1.2 the existing algorithmic solutions in the open literature into three distinct categories:

- *Monte-Carlo methods* which resort to sequential Monte-Carlo sampling to approximate intractable probability distributions [19] at the cost of high computational complexity. Specifically, particle filters (PFs) can sample and evolve multiple instances of the time-varying signal, called "particles", through the process and observation equa-

[1]These applications assume a sparse state $\boldsymbol{x}_t$ in (1.2) which can be obtained only under the assumption of a sparse process noise $\boldsymbol{v}_t$ with a fixed support.

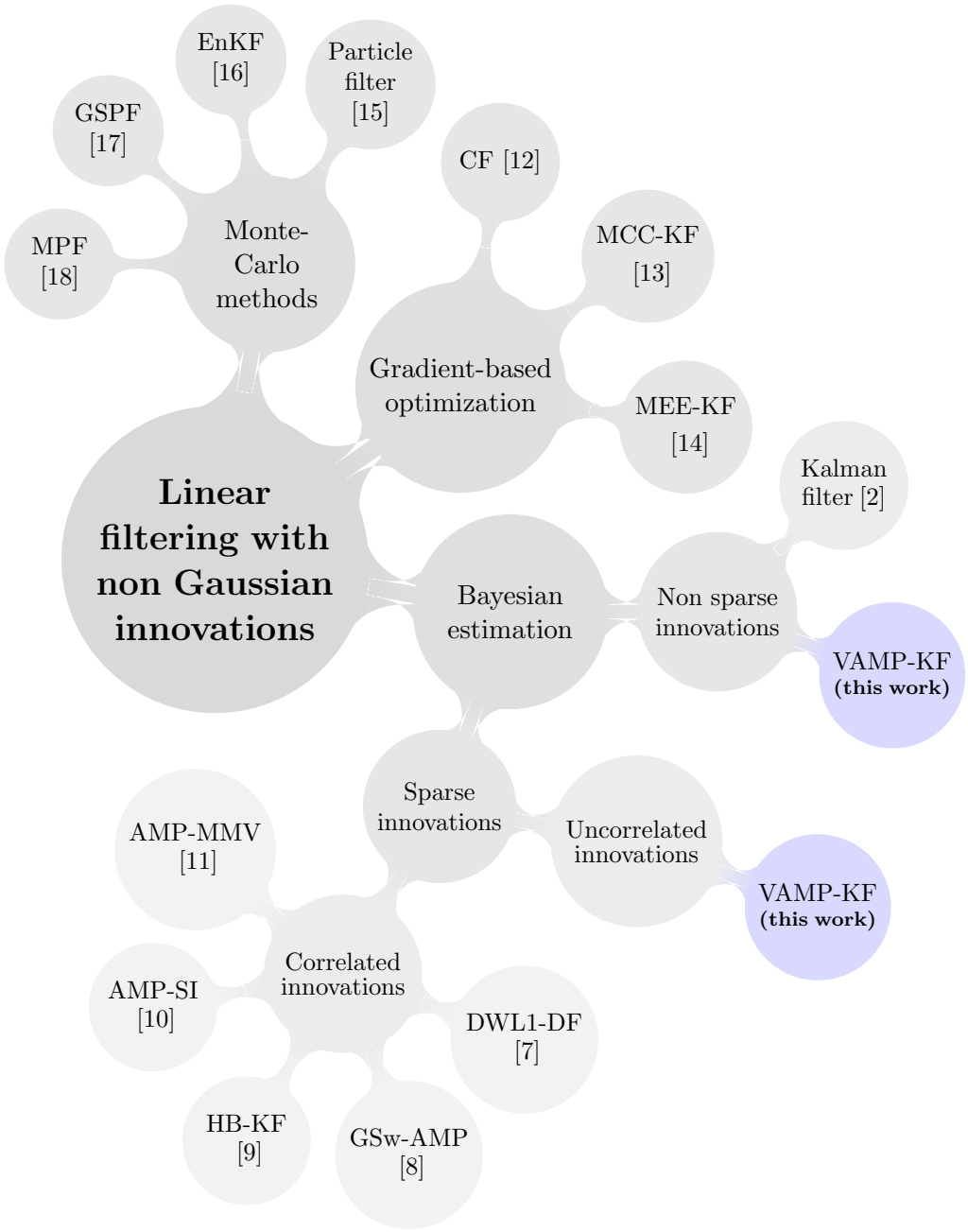


Figure 1.2: Prior work taxonomy of the algorithmic solutions for the linear filtering problem with non-Gaussian innovations. Our proposed algorithm VAMP-KF is highlighted in purple twice as it can handle both non-sparse innovations as well as sparse uncorrelated ones.

tions (1.1)–(1.2) [15]. Multiple PF variants were developed to approximate the posterior distribution of each particle as a mixture of Gaussian distributions [17]. The number of particles, however, grows exponentially with the dimension of the state-space model, thereby limiting their use even for medium dimensions (e.g., $N, M \sim 10$) [20].

- *Gradient-based optimization* which formulates the optimal filtering as a weighted-least-square problem [21] regardless of the innovation distribution. A cross-correntropy cost function was used in [12–14] to reduce the uncertainty induced by i) the measurement noise $\boldsymbol{w}_t$ between the $\boldsymbol{y}_t$ and $\boldsymbol{H}\,\boldsymbol{x}_t$, and ii) the innovation $\boldsymbol{v}_t$ between $\boldsymbol{x}_t$ and $\boldsymbol{F}\,\boldsymbol{x}_{t-1}$. However, this approach does not fully exploit the noise structure especially when the latter leads to non-convex cost functions. In this case, these algorithms converge to sub-optimal first-order stationary points.

- *Bayesian estimation*: this probabilistic approach to optimal filtering captures the uncertainty about the signal $\boldsymbol{x}_t$, the measurement noise $\boldsymbol{w}_t$, and the innovation $\boldsymbol{v}_t$ by assuming their prior distributions families to be known [22]. Research efforts within the compressed sensing (CS) framework [23] have introduced an elegant approach for recovering sparse signals from compressed measurements. To ensure that the signal $\boldsymbol{x}_t$ is sparse, existing studies assume that the innovation $\boldsymbol{v}_t$ is also sparse. To promote sparsity, Bayesian estimation is possible when the CS problem is relaxed using ℓ_1-regularization [7] or extended to optimize the element-wise variances of the innovations, $v_{i,t}$, using sparse bayesian learning (SBL) [9]. More recently, the approximate message passing (AMP) algorithm [24] has been adopted and extended to handle sparse innovations [8,10,11]. These algorithms exclusively focus on correlated innovations over time by restricting the innovation process to have a fixed innovation support and (possibly) a slow time-varying non-zero coefficients.

A major shortcoming of linear filtering methods with sparse innovations is its fragility with respect to the i.i.d. assumption of the innovation dynamics. In response, this thesis

addresses two limitations evident in the available Bayesian estimation literature:

- *binding the innovation sparsity with the signal sparsity*: existing methods consider sparse innovations $\boldsymbol{v}_t$ because the goal is to track sparse signals $\boldsymbol{x}_t$. In other words, the innovation sparsity is not studied for its own merit but rather as a constraint to ensure sparse signal evolution. This is done by enforcing the same support for $\boldsymbol{v}_t$ and $\boldsymbol{x}_t$ in (1.2).

- *inability to address uncorrelated innovations*: existing models for correlated innovations confine their evolution to a fixed support and focus on slowly time-varying non-zero coefficients. However, many applications such as mmWave communication are unfit to these assumptions where the sparse innovation contribution in the angular domain is not deterministic. In such cases, opportunities to model correlation patterns across time-varying innovations are rather limited and not available in practice.

To overcome all the aforementioned limitations, we devise a new algorithm, coined *VAMP-KF*, along with its state evolution analysis to solve the state-space model depicted in Fig. 1.1 under *rapidly* time-varying sparse innovations. Here, we use the adjective "rapidly" to emphasize the ability of VAMP-KF to estimate the signal while allowing both the support and the amplitude of the non-zero innovation elements to independently change from one time step to another.

1.3 Contributions

This thesis builds upon both Kalman filter (KF) [2] and vector approximate message passing (VAMP) [25] algorithms to estimate a time-varying signal measured through an AWGN process according to (1.1) while evolving under rapidly time-varying sparse innovations according to (1.2). VAMP-KF offers the best of the two worlds, combining

the sparse recovery performance of VAMP with the optimal signal tracking of KF under Gaussian message passing. In doing so, we turn Gaussian messages passing (a.k.a., expectation propagation) in VAMP from a foe to a friend of KF, thereby combining their advantages into a unified algorithmic framework.

VAMP-KF relies entirely on message passing starting from the combined factor graphs of KF and VAMP. Different from existing algorithms for sparse innovations, VAMP-KF enjoys the following keys features and benefits:

- It decouples the signal sparsity from the innovation sparsity so as to properly exploit the prior information about the innovation process.
- It handles uncorrelated innovation processes which are a broader class of practical applications beyond sparse innovation processes.
- It strikes a proper balance between reconstruction performance and computational complexity by sidestepping the matrix inverse operation in KF by virtue of a scalar expectation propagation approximation, i.e., covariance matrices are reduced to identity matrices scaled by the average of the diagonal elements of the full covariances matrices.
- It comes with theoretical performance guarantees based on the state evolution analysis, established in Chapter 4, that validates its superiority over state-of-the-art algorithms.

1.4 Organization and Notations

The rest of the thesis is organized as follows. We introduce the relevant preliminaries of the Kalman filter and the VAMP algorithm in Chapter 2. In Chapter 3, we combine

VAMP with Kalman filter by establishing the messages exchanged between their respective factor graphs to obtain the algorithmic steps of VAMP-KF. Chapter 4 derives the state evolution equations that predict the empirical asymptotic MSE of VAMP-KF. Exhaustive numerical results are presented in Chapter 5, from which we draw out some concluding remarks regarding possible extensions of the proposed algorithm.

Notations: We use Sans Serif fonts (e.g., x) for random variables and Serif fonts (e.g., x) for their realizations. We use boldface lowercase letters for random vectors and their realizations (e.g., x and $\boldsymbol{x}$) and boldface uppercase letters for random matrices and their realizations (e.g., X and $\boldsymbol{X}$). Vectors are in column-wise orientation by default. For $k = k_{\min}, \ldots, k_{\max}$, we denote by $\boldsymbol{x}_{k_{\min}:k_{\max}}$ the set of vectors $\{\boldsymbol{x}_k\}_{k=k_{\min}}^{k_{\min}}$.

Given any matrix $\boldsymbol{X}$, we use $\boldsymbol{x}_i$ and x_{ij} to denote its ith column and ijth entry, respectively. We also denote the ith component of a vector $\boldsymbol{x}$ as $[\boldsymbol{x}]_i$ or x_i. We denote the kth canonical basis vector in $\mathbb{R}^N$ as $\boldsymbol{e}_k = [0, \cdots, 0, 1, 0, \cdots, 0]^\mathsf{T}$, which has a single 1 in position k. The operator $\mathrm{diag}(\boldsymbol{X})$ stacks the diagonal elements of $\boldsymbol{X}$ in a vector while $\mathrm{Diag}(\boldsymbol{x})$ returns the diagonal matrix created from the vector $\boldsymbol{x}$. The operation $\boldsymbol{x}^2$ returns a vector whose components are the square those of $\boldsymbol{x}$. We use $\mathbf{I}_N$ and $\mathbf{1}_N$ to denote the $N \times N$ identity and all-ones matrix, respectively. We also use $p_\mathsf{x}(x;\boldsymbol{\theta})$, $p_\mathsf{x}(\boldsymbol{x};\boldsymbol{\theta})$, and $p_\mathsf{X}(\boldsymbol{X};\boldsymbol{\theta})$ to denote the probability density function (pdf) of random variables/vectors/matrices; as being parameterized by a parameter vector $\boldsymbol{\theta}$. Moreover, $\mathcal{CN}(\boldsymbol{x};\widehat{\boldsymbol{x}},\boldsymbol{R})$ stands for the complex multi-variate Gaussian pdf of any random vector x with mean $\widehat{\boldsymbol{x}}$ and covariance matrix $\boldsymbol{R}$. The Bernoulli distribution with probability p is denoted as $\mathcal{B}(p)$. We use $\sim$ and $\propto$ as short-hand notations for "distributed according to" and "proportional to", respectively. We also use $\mathbb{E}[\mathsf{x}|d(\boldsymbol{x})]$ and $\mathrm{Cov}[\mathsf{x}|d(\boldsymbol{x})]$ to denote the expectation and the covariance matrix of $\mathsf{x} \sim d(\boldsymbol{x})$, respectively, and $\delta(\boldsymbol{x})$ denotes the Dirac delta distribution. In addition, $\langle \boldsymbol{x} \rangle$ and $\langle \boldsymbol{X} \rangle$ return the (empirical) mean of vectors and matrices, i.e., $\langle \boldsymbol{x} \rangle \triangleq \frac{1}{N}\sum_{i=1}^{N} x_i$ for $\boldsymbol{x} \in \mathbb{R}^N$ and $\langle \boldsymbol{X} \rangle \triangleq \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} x_{ij}$ for $\boldsymbol{X} \in \mathbb{R}^{N\times M}$. The symbol $\odot$ denotes the Hadamard (i.e., elemen-

twise) product between any two vectors/matrices and we refer to the Frobenius norm of any matrix $\boldsymbol{X}$ by $\|\boldsymbol{X}\|_{\mathrm{F}}$. Finally, for any multi-variate function $\boldsymbol{f}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M)$, we denote by $\boldsymbol{f}'(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M)$ a vector constituted by the diagonal elements of the partial derivative of the function $\boldsymbol{f}$ with respect to its first parameter $\boldsymbol{x}_1$, i.e:

$$\boldsymbol{f}'(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M) = \operatorname{diag}\left(\frac{\partial \boldsymbol{f}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M)}{\partial \boldsymbol{x}_1}\right).$$

Chapter 2

Preliminaries

In this chapter, we introduce factor graphs and the sum-product algorithm in Section 2.1, and then present Kalman filter and VAMP algorithms in Section 2.2 and Sections 2.3, respectively. We then unify VAMP and Kalman filter in Chapter 3 to infer signals evolving under rapidly time-varying sparse noise dynamics.

2.1 Factor Graphs and the Sum-Product Algorithm

A factor graph is a bipartite graph that represents the factorization of a "global" function into a product of "local" functions [26]. The term "bipartite" refers to the fact that the graph involves two types of nodes only, namely variable nodes and factor nodes. For example, a global function $f(x_1, x_2, x_3)$ which factorizes as

$$f(x_1, x_2, x_3) = f_1(x_1) f_2(x_1, x_2) f_3(x_2, x_3) \tag{2.1}$$

can be represented by the factor graph given in Fig. 2.1. There, the variable nodes $\{x_1,\ x_2,\ x_3\}$ and the factor nodes $\{f_1,\ f_2,\ f_3\}$ are connected by the so-called "edges" based on the relationship expressed in (2.1). The factor graph in Fig. 2.1 has a "tree"

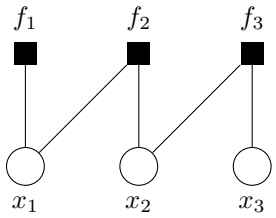


Figure 2.1: A factor graph example.

structure because it has no cycles. The nodes f_1 and x_3 are the so-called "leaf" nodes because each of them is connected to only one edge.

When the global function is a joint probability density function of some random variables, the factor graph representation is a useful tool for inferring the marginal densities of subsets of those random variables. This can be achieved by the sum-product algorithm, which propagates messages along the edges of the graph. With the denotations in Fig. 2.1, the message that propagates along the edge between f_i and x_j is denoted either by $\mu_{f_i \to x_j}$ or $\mu_{x_j \to f_i}$ depending on the direction of propagating. The message updating rules of the sum-product algorithm are summarized based on the illustrative factor graphs in Fig. 2.2 as the following:

$$\mu_{f \to x}(x) \propto \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x, x_1, \ldots, x_n) \prod_{i=1}^{n} \mu_{x_i \to f}(x_i)\, dx_1 \ldots dx_n, \tag{2.2a}$$

$$\mu_{x \to f}(x) \propto \prod_{i=1}^{n} \mu_{f_i \to x}(x), \tag{2.2b}$$

$$\mu_{f_{\text{leaf}} \to x}(x) = f_{\text{leaf}}(x), \tag{2.2c}$$

$$\mu_{x_{\text{leaf}} \to f} = 1, \tag{2.2d}$$

in which (2.2a)-(2.2d) evaluate messages that depart from an arbitrary factor node, an arbitrary variable node, a leaf factor node, and a leaf variable node, and correspond to Figs. 2.2a-2.2d, respectively. Following the rules, the marginal of any variable node x

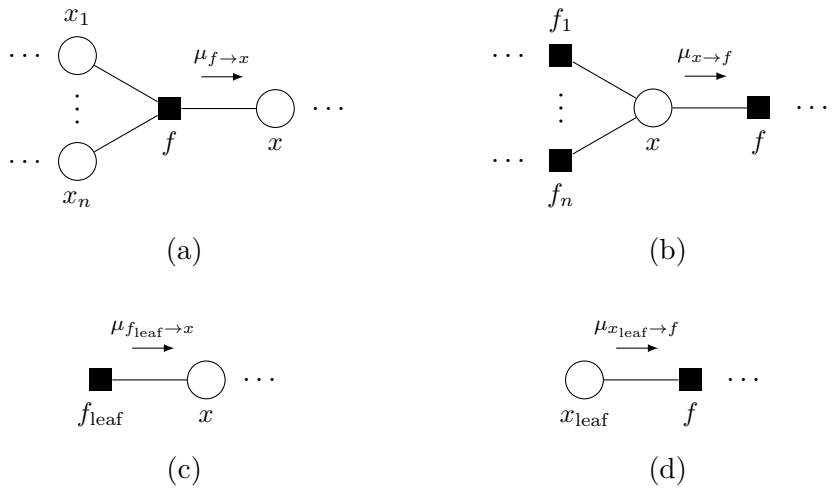


Figure 2.2: Messages that depart from (a) an arbitrary factor node, (b) an arbitrary variable node, (c) a leaf factor node, and (d) a leaf variable node, which correspond to (2.2a)-(2.2d), respectively.

can be computed as

$$p(x) \propto \prod_{i \in \mathcal{E}(x)} \mu_{f_i \to x}(x), \tag{2.3}$$

where $\mathcal{E}(x)$ denotes the set of indices of all the factor nodes that are neighbours of x–this means that the marginal x is equal to the product of all incoming messages from its neighbouring factor nodes.

2.2 Kalman filtering

We consider the discrete-time linear stochastic system that evolves with every time index $t \in \{1, \ldots, T\}$ according to the following state-space model:

$$\boldsymbol{x}_t = \boldsymbol{F}_t \, \boldsymbol{x}_{t-1} + \boldsymbol{B}_t \, \boldsymbol{u}_t + \boldsymbol{v}_t, \qquad \text{(process equation)} \tag{2.4a}$$

$$\boldsymbol{y}_t = \boldsymbol{H}_t \, \boldsymbol{x}_t + \boldsymbol{w}_t, \qquad \text{(measurement equation)} \tag{2.4b}$$

where $\boldsymbol{x}_t \in \mathbb{C}^N$ is the state vector to be estimated, $\boldsymbol{u}_t \in \mathbb{C}^N$ is an input control vector, $\boldsymbol{y}_t \in \mathbb{C}^M$ is the measurement vector, $\boldsymbol{F}_t \in \mathbb{C}^{N\times N}$ is the state transition matrix, $\boldsymbol{B}_t \in \mathbb{C}^{N\times N}$ is an input transition matrix, and $\boldsymbol{H}_t \in \mathbb{C}^{M\times N}$ is the measurement matrix. Moreover, $\boldsymbol{v}_t \in \mathbb{C}^N$ and $\boldsymbol{w}_t \in \mathbb{C}^M$ are the innovation process and measurement noise vectors with zero-mean and covariance matrices $\boldsymbol{Q}_t$ and $\boldsymbol{R}_t$, respectively, i.e., $\boldsymbol{v}_t \sim \mathcal{CN}(\boldsymbol{v}_t; \boldsymbol{0}, \boldsymbol{Q}_t)$ and $\boldsymbol{w}_t \sim \mathcal{CN}(\boldsymbol{w}_t; \boldsymbol{0}, \boldsymbol{R}_t)$.

Using the process and measurement equations in (2.4), the joint density of all states $\boldsymbol{x}_{0:T}$ and measurements $\boldsymbol{y}_{1:T}$ factorizes as a Markov chain model as depicted in Fig. 2.3. Given the set of observations $\boldsymbol{y}_{1:t}$, the Kalman filter generates an optimal estimate of the state $\boldsymbol{x}_t$, which we denote $\widehat{\boldsymbol{x}}_t$, that minimises the mean square error (MSE), i.e.,

$$\widehat{\boldsymbol{x}}_t^{\bigstar} = \underset{\widehat{\boldsymbol{x}}_t}{\operatorname{argmin}} \ \mathbb{E}\Big[\|\boldsymbol{x}_t - \widehat{\boldsymbol{x}}_t\|_2^2 \Big]. \tag{2.5}$$

When both $\boldsymbol{v}_t$ and $\boldsymbol{w}_t$ are Gaussian, the Kalman filter solves (2.5) recursively by means of an exact Bayesian recursive procedure [27]. The latter is a forward-only message passing algorithm (i.e., messages flow from top to bottom in the factor graph of Fig. 2.3): that is to say Kalman filter does not pass messages backward to the previous states [27]. The Kalman filter alternates between the following two update steps [1]:

- *measurement update (prediction step)*: it consists in performing a one-step-ahead-prediction based on the observations $\boldsymbol{y}_{1:t-1}$ by computing the posterior mean and covariance of the state vector at time index t as follows:

$$\begin{aligned}
\widehat{\boldsymbol{x}}_{t|t-1} &= \mathbb{E}\big[\boldsymbol{x}_t \big|\, p_{\mathbf{x}_t|\mathbf{y}_{1:t-1}}(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1})\big] \\
&= \boldsymbol{F}_t\, \widehat{\boldsymbol{x}}_{t-1|t-1} + \boldsymbol{B}_t\, \boldsymbol{u}_t, \qquad (2.6a) \\
\widehat{\boldsymbol{\Sigma}}_{t|t-1} &= \mathbb{E}\big[(\boldsymbol{x}_t - \widehat{\boldsymbol{x}}_{t|t-1})\,(\boldsymbol{x}_t - \widehat{\boldsymbol{x}}_{t|t-1})^{\mathsf{H}} \big|\, p_{\mathbf{x}_t|\mathbf{y}_{1:t-1}}(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1})\big] \\
&= \boldsymbol{F}_t\, \widehat{\boldsymbol{\Sigma}}_{t-1|t-1}\, \boldsymbol{F}_t^{\mathsf{H}} + \boldsymbol{Q}_t. \qquad (2.6b)
\end{aligned}$$

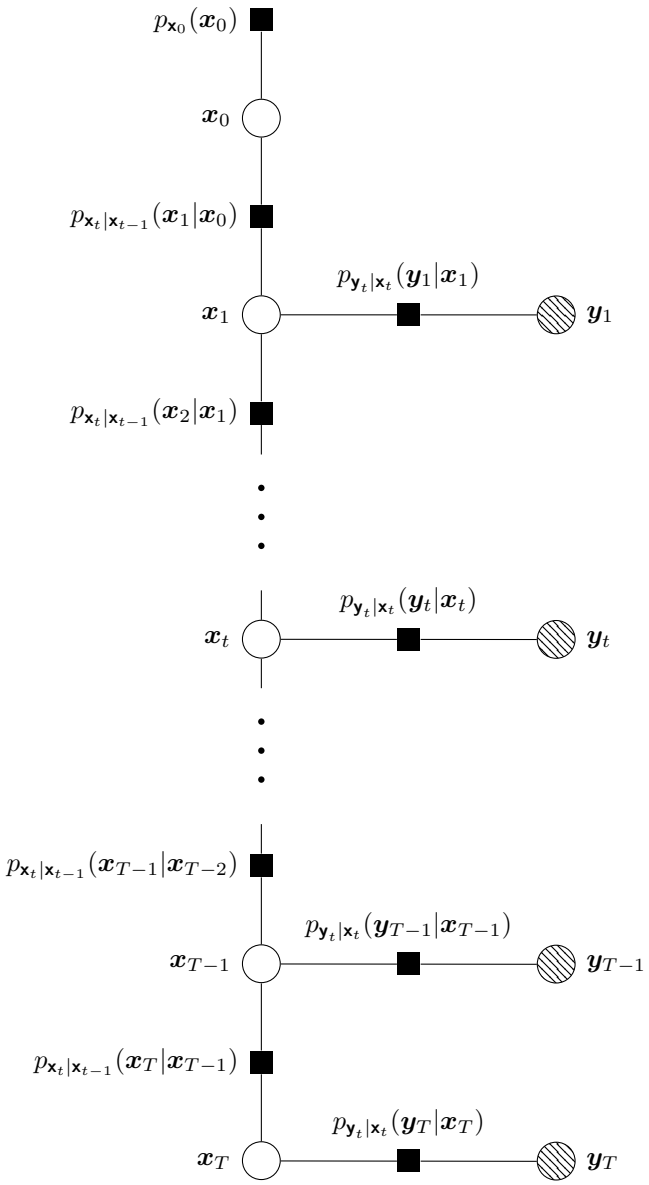


Figure 2.3: Factor graph of Kalman filter as a Markov chain model: starting with an initial state $\boldsymbol{x}_0$ with a prior density $p_{\mathbf{x}_0}(\boldsymbol{x}_0)$, the state evolves over time according to the transition density $p_{\mathbf{x}_t|\mathbf{x}_{t-1}}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) = \mathcal{CN}(\boldsymbol{x}_t; \boldsymbol{F}_t\,\boldsymbol{x}_{t-1} + \boldsymbol{B}_t\,\boldsymbol{u}_t, \boldsymbol{Q}_t)$. At every time step $t > 0$, the state $\boldsymbol{x}_t$ is observed through the measurement $\boldsymbol{y}_t$ under the joint density $p_{\mathbf{y}_t|\mathbf{x}_t}(\boldsymbol{y}_t|\boldsymbol{x}_t) = \mathcal{CN}(\boldsymbol{y}_t; \boldsymbol{H}_t\,\boldsymbol{x}_t, \boldsymbol{R}_t)$.

- *time update (estimation step)*: this updates the mean and covariance of the state using the predicted mean and covariance in (2.6) as well as the new measurement observation $\boldsymbol{y}_t$ as follows:

$$\widehat{\boldsymbol{x}}_{t|t} = \widehat{\boldsymbol{x}}_{t|t-1} + \boldsymbol{K}_t\big(\boldsymbol{y}_t - \boldsymbol{H}_t\,\widehat{\boldsymbol{x}}_{t|t-1}\big), \tag{2.7a}$$

$$\widehat{\boldsymbol{\Sigma}}_{t|t} = \widehat{\boldsymbol{\Sigma}}_{t|t-1} - \boldsymbol{K}_t\,\boldsymbol{H}_t\,\widehat{\boldsymbol{\Sigma}}_{t|t-1}, \tag{2.7b}$$

  in which the Kalman gain matrix $\boldsymbol{K}_t$ is given by

$$\boldsymbol{K}_t = \widehat{\boldsymbol{\Sigma}}_{t|t-1}\,\boldsymbol{H}_t^{\mathsf{H}}\Big(\boldsymbol{H}_t\,\widehat{\boldsymbol{\Sigma}}_{t|t-1}\,\boldsymbol{H}_t^{\mathsf{H}} + \boldsymbol{R}_t\Big)^{-1}. \tag{2.8}$$

Given a density $p_{\mathbf{x}_0}(\boldsymbol{x}_0) = \mathcal{CN}(\boldsymbol{x}_0; \widehat{\boldsymbol{x}}_{0|0}, \widehat{\boldsymbol{\Sigma}}_{0|0})$ of the initial state $\boldsymbol{x}_0$, Kalman filter alternates between (2.6) and (2.7) to find the optimal linear minimum MSE estimate given in (2.5).

2.3 Vector Approximate Message Passing

Consider recovering an unknown vector, $\boldsymbol{x} \in \mathbb{C}^N$, from its noisy observation:

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{w}, \tag{2.9}$$

where $\boldsymbol{H} \in \mathbb{C}^{M\times N}$ is the measurement matrix and $\boldsymbol{w} \in \mathbb{C}^M$ is an additive white Gaussian noise vector whose entries are assumed to be independent and identically distributed (i.i.d.) with mean zero and variance γ_{w}^{-1}, i.e., $w_{ij} \sim \mathcal{CN}(w_{ij}; 0, \gamma_{\mathsf{w}}^{-1})$.

2.3.1 The VAMP algorithm

Given a prior distribution $p_{\mathsf{x}}(\boldsymbol{x})$ on the signal $\boldsymbol{x}$ in (2.9), VAMP resorts to an approximation of the belief propagation algorithm, namely expectation propagation (EP) [28,29]. The latter is derived according to a factor graph with vector-valued variable nodes as shown in Fig. 2.4. There, after splitting $\boldsymbol{x}$ into two separate variables $\boldsymbol{x}^+$ and $\boldsymbol{x}^-$, the joint density of $\boldsymbol{x}$ and $\boldsymbol{y}$ is factorized as follows:

$$p_{\mathsf{x},\mathsf{y}}(\boldsymbol{x},\boldsymbol{y}) = p_{\mathsf{x}}(\boldsymbol{x}^+)\,\delta(\boldsymbol{x}^+ - \boldsymbol{x}^-)\,\mathcal{CN}(\boldsymbol{y};\boldsymbol{H}\boldsymbol{x}^-,\gamma_{\mathsf{w}}^{-1}\mathbf{I}_M). \tag{2.10}$$

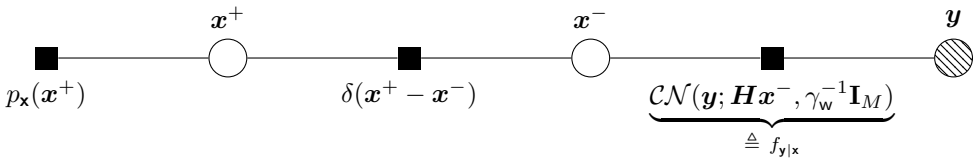


Figure 2.4: Factor graph of the VAMP algorithm.

The expectation propagation of messages between the variables nodes, $\boldsymbol{x}^+$ and $\boldsymbol{x}^-$, and the factor nodes, $p_{\mathsf{x}}(\boldsymbol{x}^+)$ and $f_{\mathsf{y}|\mathsf{x}}$, is based on the two following rules:

i) *EP approximation*: given the factor node $p_{\mathsf{x}}(\boldsymbol{x}^+)$ and the incoming extrinsic Gaussian belief $\mu_{\delta\to\boldsymbol{x}^+}(\boldsymbol{x}^+) = \mathcal{CN}\big(\boldsymbol{x}^+;\widehat{\boldsymbol{x}}_{\mathsf{e}}^-,\gamma_{\boldsymbol{x}_{\mathsf{e}}^-}^{-1}\mathbf{I}_N\big)$ from the factor node $\delta(\boldsymbol{x}^+ - \boldsymbol{x}^-)$, the posterior sum-product (SP) belief is expressed as

$$b_{\mathsf{sp}}(\boldsymbol{x}^+) \propto p_{\mathsf{x}}(\boldsymbol{x}^+)\,\mathcal{CN}\big(\boldsymbol{x}^+;\widehat{\boldsymbol{x}}_{\mathsf{e}}^-,\gamma_{\boldsymbol{x}_{\mathsf{e}}^-}^{-1}\mathbf{I}_N\big)$$

and is approximated by EP as the Gaussian belief $\mathcal{CN}\big(\boldsymbol{x}^+;\widehat{\boldsymbol{x}}_{\mathsf{p}}^+,\gamma_{\boldsymbol{x}_{\mathsf{p}}^+}^{-1}\mathbf{I}_N\big)$. The pos-

terior mean $\widehat{\boldsymbol{x}}_{\mathsf{p}}^{+}$ and precision $\gamma_{\boldsymbol{x}_{\mathsf{p}}^{+}}^{-1}$ are given by

$$\widehat{\boldsymbol{x}}_{\mathsf{p}}^{+} = \mathbb{E}\left[\boldsymbol{x}^{+} \,\middle|\, b_{\mathsf{sp}}(\boldsymbol{x}^{+})\right], \tag{2.11a}$$

$$\gamma_{\boldsymbol{x}_{\mathsf{p}}^{+}}^{-1} = \left\langle \mathrm{diag}\big(\mathrm{Cov}\left[\boldsymbol{x}^{+} \,\middle|\, b_{\mathsf{sp}}(\boldsymbol{x}^{+})\right]\big)\right\rangle, \tag{2.11b}$$

Due to the approximation of the covariance matrix with a scaled identity matrix, the EP belief approximation in (2.11) is a special variant of the EP algorithm [28], namely scalar EP [30].

ii) *Extrinsic belief computation*: Given $\mu_{\delta\to\boldsymbol{x}^{+}} = \mathcal{CN}\big(\boldsymbol{x}^{+}; \widehat{\boldsymbol{x}}_{\mathsf{e}}^{-}, \gamma_{\boldsymbol{x}_{\mathsf{e}}^{-}}^{-1}\mathbf{I}_N\big)$ as the incoming Gaussian belief from the factor node $\delta(\boldsymbol{x}^{+} - \boldsymbol{x}^{-})$ and $\mathcal{CN}\big(\boldsymbol{x}^{+}; \widehat{\boldsymbol{x}}_{\mathsf{p}}^{+}, \gamma_{\boldsymbol{x}_{\mathsf{p}}^{+}}^{-1}\mathbf{I}_N\big)$ as the posterior Gaussian belief computed in (2.11), the mean and precision of the outgoing extrinsic Gaussian belief $\mu_{\boldsymbol{x}^{+}\to\delta} = \mathcal{CN}\big(\boldsymbol{x}^{+}; \widehat{\boldsymbol{x}}_{\mathsf{e}}^{+}, \gamma_{\boldsymbol{x}_{\mathsf{e}}^{+}}^{-1}\mathbf{I}_N\big)$ are given by [25]:

$$\gamma_{\boldsymbol{x}_{\mathsf{e}}^{+}} = \gamma_{\boldsymbol{x}_{\mathsf{p}}^{+}} - \gamma_{\boldsymbol{x}_{\mathsf{e}}^{-}}, \tag{2.12a}$$

$$\widehat{\boldsymbol{x}}_{\mathsf{e}}^{+} = \gamma_{\boldsymbol{x}_{\mathsf{e}}^{+}}^{-1}\left(\gamma_{\boldsymbol{x}_{\mathsf{p}}^{+}}\,\widehat{\boldsymbol{x}}_{\mathsf{p}}^{+} - \gamma_{\boldsymbol{x}_{\mathsf{e}}^{-}}\,\widehat{\boldsymbol{x}}_{\mathsf{e}}^{-}\right), \tag{2.12b}$$

which correspond to lines 7–8 of Algorithm 1.

Note that the same computations in (2.11)–(2.12) are also carried out for the messages exchanged between the variable node $\boldsymbol{x}^{-}$ and the variable node $f_{\mathbf{y}|\mathbf{x}}$. In this case, one would simply substitute in (2.11)–(2.12) the variable node $\boldsymbol{x}^{+}$ with $\boldsymbol{x}^{-}$, and the factor node $p_{\mathbf{x}}(\boldsymbol{x}^{+})$ with $f_{\mathbf{y}|\mathbf{x}}$.

By employing the aforementioned rules, VAMP runs iteratively according to the algorithmic steps outlined in Algorithm 1 by alternating between the minimum-MSE (MMSE) estimation of $\boldsymbol{x}^{+}$ and the linear MMSE (LMMSE) estimation of $\boldsymbol{x}^{-}$.

Algorithm 1 VAMP (SVD form)

Require : measurement matrix $\boldsymbol{H} \in \mathbb{C}^{M\times N}$, measurement vector $\boldsymbol{y} \in \mathbb{C}^{M}$, MMSE denoising function $\boldsymbol{g}_{\mathsf{x}^+}(\cdot)$, noise precision γ_{w}, and the maximum number of iterations K.

1: Compute economic SVD $\boldsymbol{H} = \overline{\boldsymbol{U}}_{\boldsymbol{H}} \operatorname{Diag}(\overline{\boldsymbol{s}}_{\boldsymbol{H}}) \overline{\boldsymbol{V}}_{\boldsymbol{H}}^{\mathsf{H}}$ with $\overline{\boldsymbol{U}}_{\boldsymbol{H}}^{\mathsf{H}} \overline{\boldsymbol{U}}_{\boldsymbol{H}} = \mathbf{I}_R$, $\overline{\boldsymbol{V}}_{\boldsymbol{H}}^{\mathsf{H}} \overline{\boldsymbol{V}}_{\boldsymbol{H}} = \mathbf{I}_R$, $\overline{\boldsymbol{s}}_{\boldsymbol{H}} \in \mathbb{R}_+^R$, and $R = \operatorname{rank}(\boldsymbol{H})$.

2: Compute $\widetilde{\boldsymbol{y}} = \operatorname{Diag}(\overline{\boldsymbol{s}}_{\boldsymbol{H}})^{-1} \overline{\boldsymbol{U}}_{\boldsymbol{H}}^{\mathsf{H}} \boldsymbol{y}$.

3: Initialize $\widehat{\boldsymbol{x}}_{\mathsf{e},0}^{-}$ and $\gamma_{\boldsymbol{x}_{\mathsf{e}}^-,0} \geq 0$.

4: **for** $k = 0, 1, \ldots, K$ **do**

▷ MMSE estimation of $\boldsymbol{x}^+$

5: $\quad \widehat{\boldsymbol{x}}_{\mathsf{p},k}^{+} = \boldsymbol{g}_{\mathsf{x}^+}\left(\widehat{\boldsymbol{x}}_{\mathsf{e},k}^{-}, \gamma_{\boldsymbol{x}_{\mathsf{e}}^-,k}\right)$

6: $\quad \gamma_{\boldsymbol{x}_{\mathsf{p}}^+,k} = \gamma_{\boldsymbol{x}_{\mathsf{e}}^-,k} \Big/ \left\langle \boldsymbol{g}'_{\mathsf{x}^+}\left(\widehat{\boldsymbol{x}}_{\mathsf{e},k}^{-}, \gamma_{\boldsymbol{x}_{\mathsf{e}}^-,k}\right)\right\rangle$

7: $\quad \gamma_{\boldsymbol{x}_{\mathsf{e}}^+,k} = \gamma_{\boldsymbol{x}_{\mathsf{p}}^+,k} - \gamma_{\boldsymbol{x}_{\mathsf{e}}^-,k}$

8: $\quad \widehat{\boldsymbol{x}}_{\mathsf{e},k}^{+} = \gamma_{\boldsymbol{x}_{\mathsf{e}}^+,k}^{-1}\left(\gamma_{\boldsymbol{x}_{\mathsf{p}}^+,k}\, \widehat{\boldsymbol{x}}_{\mathsf{p},k}^{+} - \gamma_{\boldsymbol{x}_{\mathsf{e}}^-,k}\, \widehat{\boldsymbol{x}}_{\mathsf{e},k}^{-}\right)$

▷ LMMSE estimation of $\boldsymbol{x}^-$

9: $\quad \boldsymbol{d}_k = \gamma_{\mathsf{w}} \operatorname{Diag}\left(\gamma_{\mathsf{w}}\, \overline{\boldsymbol{s}}_{\boldsymbol{H}}^2 + \gamma_{\boldsymbol{x}_{\mathsf{e}}^+,k}\, \mathbf{1}_R\right)^{-1} \overline{\boldsymbol{s}}_{\boldsymbol{H}}^2$

10: $\quad \gamma_{\boldsymbol{x}_{\mathsf{e}}^-,k+1} = \gamma_{\boldsymbol{x}_{\mathsf{e}}^+,k} \left\langle \boldsymbol{d}_k \right\rangle / \left(\frac{N}{R} - \left\langle \boldsymbol{d}_k \right\rangle\right)$

11: $\quad \widehat{\boldsymbol{x}}_{\mathsf{e},k+1}^{-} = \widehat{\boldsymbol{x}}_{\mathsf{e},k}^{+} + \frac{N}{R}\, \overline{\boldsymbol{V}}_{\boldsymbol{H}} \operatorname{Diag}\left(\frac{\boldsymbol{d}_k}{\langle \boldsymbol{d}_k \rangle}\right)\left(\widetilde{\boldsymbol{y}} - \overline{\boldsymbol{V}}_{\boldsymbol{H}}^{\mathsf{H}}\, \widehat{\boldsymbol{x}}_{\mathsf{e},k}^{+}\right)$

12: **end for**

13: **return** $\widehat{\boldsymbol{x}}_{\mathsf{p},K}^{+}$

MMSE estimation of $\boldsymbol{x}^+$

Given $\mu_{\delta\to\boldsymbol{x}^+} = \mathcal{CN}\big(\boldsymbol{x}^+; \widehat{\boldsymbol{x}}_{\mathsf{e}}^{-}, \gamma_{\boldsymbol{x}_{\mathsf{e}}^-}^{-1}\mathbf{I}_N\big)$ being the extrinsic Gaussian belief and under a separable prior $p_{\mathsf{x}}(\boldsymbol{x}^+) = \prod_{j=1}^{N} p_{\mathsf{x}_j}(x_j^+)$, the MMSE denoiser, $\boldsymbol{g}_{\mathsf{x}^+}(\cdot,\cdot)$ in line 5, along with its divergence, $\mathbf{g}'_{\mathsf{x}}(\cdot,\cdot)$ in line 6, are given by

$$\boldsymbol{g}_{\mathsf{x}^+}\left(\widehat{\boldsymbol{x}}_{\mathsf{e}}^{-}, \gamma_{\boldsymbol{x}_{\mathsf{e}}^-}\right) = \frac{\displaystyle\int \boldsymbol{x}^+\, p_{\mathsf{x}}\left(\boldsymbol{x}^+\right) \mathcal{CN}\left(\boldsymbol{x}^+; \widehat{\boldsymbol{x}}_{\mathsf{e}}^{-}, \gamma_{\boldsymbol{x}_{\mathsf{e}}^-}^{-1}\, \mathbf{I}_N\right) \mathrm{d}\boldsymbol{x}^+}{\displaystyle\int p_{\mathsf{x}}\left(\boldsymbol{x}^+\right) \mathcal{CN}\left(\boldsymbol{x}^+; \widehat{\boldsymbol{x}}_{\mathsf{e}}^{-}, \gamma_{\boldsymbol{x}_{\mathsf{e}}^-}^{-1}\, \mathbf{I}_N\right) \mathrm{d}\boldsymbol{x}^+}, \tag{2.13a}$$

$$\boldsymbol{g}'_{\mathsf{x}^+}\left(\widehat{\boldsymbol{x}}_{\mathsf{e}}^{-}, \gamma_{\boldsymbol{x}_{\mathsf{e}}^-}\right) = \operatorname{diag}\left(\frac{\partial \boldsymbol{g}_{\mathsf{x}^+}(\widehat{\boldsymbol{x}}_{\mathsf{e}}^{-}, \gamma_{\boldsymbol{x}_{\mathsf{e}}^-})}{\partial \widehat{\boldsymbol{x}}_{\mathsf{e}}^{-}}\right). \tag{2.13b}$$

LMMSE estimation of $\boldsymbol{x}^-$

Given $\mu_{\delta\to\boldsymbol{x}^-}(\boldsymbol{x}^-) = \mathcal{CN}\big(\boldsymbol{x}^-; \widehat{\boldsymbol{x}}^+_{\mathsf{e}}, \gamma^{-1}_{\boldsymbol{x}^+_{\mathsf{e}}}\mathbf{I}_N\big)$ as the extrinsic Gaussian belief, and $\boldsymbol{y}$ as the measurement vector accordingly in the AWGN model (2.9), the LMMSE estimator, $\boldsymbol{g}_{\mathsf{x}^-}(\cdot,\cdot)$, is obtained in closed form, along with its posterior variance as follows:

$$\begin{aligned}
\widehat{\boldsymbol{x}}^-_{\mathsf{p},k} &= \big(\gamma_{\mathsf{w}}\,\boldsymbol{H}^{\mathsf{H}}\boldsymbol{H} + \gamma_{\boldsymbol{x}^+_{\mathsf{e}},k}\,\mathbf{I}_N\big)^{-1}\big(\gamma_{\mathsf{w}}\,\boldsymbol{H}^{\mathsf{H}}\,\boldsymbol{y} + \gamma_{\boldsymbol{x}^+_{\mathsf{e}},k}\,\widehat{\boldsymbol{x}}^+_{\mathsf{e},k}\big) \\
&\triangleq \boldsymbol{g}_{\mathsf{x}^-}\left(\widehat{\boldsymbol{x}}^+_{\mathsf{e},k}, \gamma_{\boldsymbol{x}^+_{\mathsf{e}},k}\right) \qquad (2.14a)\\
\gamma^{-1}_{\boldsymbol{x}^-_{\mathsf{p}},k} &= \frac{1}{N}\mathrm{Tr}\left[\big(\gamma_{\mathsf{w}}\,\boldsymbol{H}^{\mathsf{H}}\boldsymbol{H} + \gamma_{\boldsymbol{x}^+_{\mathsf{e}},k}\,\mathbf{I}_N\big)^{-1}\right]. \\
&\triangleq \left\langle \boldsymbol{g}'_{\mathsf{x}^-}\left(\widehat{\boldsymbol{x}}^+_{\mathsf{e},k}, \gamma_{\boldsymbol{x}^+_{\mathsf{e}},k}\right)\right\rangle \qquad (2.14b)
\end{aligned}$$

To mitigate the high computational cost stemming from the matrix inversion in (2.14), the singular value decomposition (SVD) of the matrix $\boldsymbol{H} = \overline{\boldsymbol{U}}_{\boldsymbol{H}}\,\mathrm{Diag}(\overline{\boldsymbol{s}}_{\boldsymbol{H}})\,\overline{\boldsymbol{V}}^{\mathsf{H}}_{\boldsymbol{H}}$ can be used in (2.14) to reduce the complexity of each VAMP iteration from $\mathcal{O}(N^3)$ to $\mathcal{O}(NR)$, with R being the rank of $\boldsymbol{H}$. After further algebraic manipulations, the LMMSE estimation of $\boldsymbol{x}^-$ simplifies as stated in lines 9–11. We refer the reader to [25] for further details about the derivation of this step.

2.3.2 State evolution update equations of VAMP

We denote the true vector observed through the linear observation model in (2.9) as $x^{(0)} \in \mathbb{C}^N$, i.e.,

$$\boldsymbol{y} = \boldsymbol{H}\,\boldsymbol{x}^{(0)} + \boldsymbol{w}, \qquad \boldsymbol{w} \sim \mathcal{CN}(\boldsymbol{w}; \mathbf{0}, \gamma_{\mathsf{w}}^{-1}\,\mathbf{I}_M). \tag{2.15}$$

State evolution assumptions

In (2.15), the sensing matrix $\boldsymbol{H}$ is assumed to have a full SVD of the form

$$\boldsymbol{H} = \widetilde{\boldsymbol{U}}_{\boldsymbol{H}}\,\widetilde{\boldsymbol{S}}_{\boldsymbol{H}}\,\widetilde{\boldsymbol{V}}^{\mathsf{H}}_{\boldsymbol{H}}, \qquad \mathrm{diag}(\widetilde{\boldsymbol{S}}_{\boldsymbol{H}}) = \widetilde{\boldsymbol{s}}_{\boldsymbol{H}}, \tag{2.16}$$

where $\widetilde{\boldsymbol{U}}_{\boldsymbol{H}}$ and $\widetilde{\boldsymbol{V}}_{\boldsymbol{H}}$ are orthogonal matrices, and $\widetilde{\boldsymbol{s}}_{\boldsymbol{H}}$ is a vector containing the singular values of $\boldsymbol{H}$. To treat both skinny and fat matrices in a unified manner in (2.16), it is important to pad $\widetilde{\boldsymbol{U}}_{\boldsymbol{H}}$, $\widetilde{\boldsymbol{V}}_{\boldsymbol{H}}$, and $\widetilde{\boldsymbol{s}}_{\boldsymbol{H}}$ so as they all have the same dimension. For instance, let $\boldsymbol{H}$ be a fat $M \times N$ matrix (i.e. $M < N$). Its padded SVD of the form $\boldsymbol{U}_{\boldsymbol{H}} \operatorname{Diag}(\boldsymbol{s}_{\boldsymbol{H}})\, \boldsymbol{V}_{\boldsymbol{H}}^{\mathsf{H}}$ can be expressed as function of its full SVD in (2.16) as follows:

$$\boldsymbol{U}_{\boldsymbol{H}} = \begin{bmatrix} \widetilde{\boldsymbol{U}}_{\boldsymbol{H}} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N-M} \end{bmatrix}, \quad \boldsymbol{s}_{\boldsymbol{H}} = \begin{bmatrix} \widetilde{\boldsymbol{s}}_{\boldsymbol{H}} \\ \mathbf{0} \end{bmatrix}, \quad \boldsymbol{V}_{\boldsymbol{H}} = \widetilde{\boldsymbol{V}}_{\boldsymbol{H}}^{\mathsf{H}}. \tag{2.17}$$

The padded square matrix $\boldsymbol{H} = \boldsymbol{U}_{\boldsymbol{H}} \operatorname{Diag}(\boldsymbol{s}_{\boldsymbol{H}})\, \boldsymbol{V}_{\boldsymbol{H}}^{\mathsf{H}}$ is assumed to be right-rotationally invariant: that is to say the distribution of $\boldsymbol{H}$ is identical to that of $\boldsymbol{HQ}$ for any orthogonal matrix $\boldsymbol{Q}$. Moreover, it is assumed that the components of $\boldsymbol{s}_{\boldsymbol{H}}$, the extrinsic mean vector $\widehat{\boldsymbol{x}}_{\mathsf{e},k}^{-}$, and the true signal $\boldsymbol{x}^{(0)}$ converge empirically with second-order moments according to

$$\lim_{N\to\infty} \left\{ \left(s_{\boldsymbol{H},n}, x_n^{(0)}, \widehat{x}_{\mathsf{e},k,n}^{-} \right) \right\}_{n=1}^{N} \overset{PL(2)}{=} \left(\mathsf{s}_{\boldsymbol{H}}, \mathsf{x}^{(0)}, \widehat{\mathsf{x}}_{\mathsf{e},k}^{-} \right), \tag{2.18}$$

for some random variables $\mathsf{s}_{\boldsymbol{H}}, \mathsf{x}^{(0)}$, and $\widehat{\mathsf{x}}_{\mathsf{e},k}^{-}$, with density functions $p_{\mathsf{s}_{\boldsymbol{H}}}(s_{\boldsymbol{H}})$, $p_{\mathsf{x}^{(0)}}(x^{(0)})$, and $p_{\widehat{\mathsf{x}}_{\mathsf{e},k}^{-}}(\widehat{x}_{\mathsf{e},k}^{-})$, respectively. In (2.18), the notation "$LP(2)$" refers the empirical convergence with second-order moments (see Appendix B in [25]). Finally, the derivation of the state evolution equations at every iteration k is based on the following concentration of measure for the precision variables in the asymptotic regime:

$$\lim_{N\to\infty} \left(\gamma_{\boldsymbol{v}_{\mathsf{p}}^{+},k}, \gamma_{\boldsymbol{x}_{\mathsf{e}}^{+},k} \right) = \left(\overline{\gamma}_{\boldsymbol{x}_{\mathsf{p}}^{+},k}, \overline{\gamma}_{\boldsymbol{x}_{\mathsf{e}}^{+},k} \right), \tag{2.19a}$$

$$\lim_{N\to\infty} \left(\gamma_{\boldsymbol{x}_{\mathsf{p}}^{-},k}, \gamma_{\boldsymbol{x}_{\mathsf{e}}^{-},k} \right) = \left(\overline{\gamma}_{\boldsymbol{x}_{\mathsf{p}}^{-},k}, \overline{\gamma}_{\boldsymbol{x}_{\mathsf{e}}^{-},k} \right). \tag{2.19b}$$

State evolution equations

Based on the aforementioned state evolution assumptions and under the Bayes-optimal setting[1], the predicted MSE of the VAMP algorithm at iteration k is defined recursively through the so-called state evolution (SE) update equations:

$$\overline{\gamma}_{\boldsymbol{x}_{\mathsf{p}}^{+},k} = \frac{1}{\mathcal{E}_{\mathsf{x}^{+}}\left(\overline{\gamma}_{\boldsymbol{x}_{\mathsf{e}}^{-},k}\right)}, \quad \overline{\gamma}_{\boldsymbol{x}_{\mathsf{e}}^{+},k} = \overline{\gamma}_{\boldsymbol{x}_{\mathsf{p}}^{+},k} - \overline{\gamma}_{\boldsymbol{x}_{\mathsf{e}}^{-},k}, \tag{2.20a}$$

$$\overline{\gamma}_{\boldsymbol{x}_{\mathsf{p}}^{-},k} = \frac{1}{\mathcal{E}_{\mathsf{x}^{-}}\left(\overline{\gamma}_{\boldsymbol{x}_{\mathsf{e}}^{+},k}\right)}, \quad \overline{\gamma}_{\boldsymbol{x}_{\mathsf{e}}^{-},k+1} = \overline{\gamma}_{\boldsymbol{x}_{\mathsf{p}}^{-},k} - \overline{\gamma}_{\boldsymbol{x}_{\mathsf{e}}^{+},k}. \tag{2.20b}$$

In (2.20), the function $\mathcal{E}_{\mathsf{x}^{+}}\left(\gamma_{\boldsymbol{x}_{\mathsf{e}}^{-}}\right)$ stands for the component-wise MSE of the estimate $g_{\mathsf{x}^{+}}\left(\widehat{\mathsf{x}}_{\mathsf{e}}^{-}, \gamma_{\boldsymbol{x}_{\mathsf{e}}^{-}}\right)$ from the "true" $\mathsf{x}^{(0)}$, where the random variable $\widehat{\mathsf{x}}_{\mathsf{e}}^{-}$ is equal to $\mathsf{x}^{(0)}$ corrupted by a zero-mean Gaussian noise of precision $\gamma_{\boldsymbol{x}_{\mathsf{e}}^{-}}$, i.e., $\widehat{\mathsf{x}}_{\mathsf{e}}^{-} = \mathsf{x}^{(0)} + \mathsf{p}^{-}$ with $\mathsf{p}^{-} \sim \mathcal{CN}(p^{-}; 0, \gamma_{\boldsymbol{x}_{\mathsf{e}}^{-}}^{-1})$. Here, $g_{\mathsf{x}^{+}}\left(\widehat{\mathsf{x}}_{\mathsf{e}}^{-}, \gamma_{\boldsymbol{x}_{\mathsf{e}}^{-}}\right)$ is the component-wise denoiser of the separable MMSE vector denoiser $\boldsymbol{g}_{\mathsf{x}^{+}}\left(\widehat{\mathbf{x}}_{\mathsf{e}}^{-}, \gamma_{\boldsymbol{x}_{\mathsf{e}}^{-}}\right)$. On this account, it follows that [25]

$$\mathcal{E}_{\mathsf{x}^{+}}\left(\gamma_{\boldsymbol{x}_{\mathsf{e}}^{-}}\right) = \mathbb{E}\left[\left| g_{\mathsf{x}^{+}}\left(\widehat{\mathsf{x}}_{\mathsf{e}}^{-}, \gamma_{\boldsymbol{x}_{\mathsf{e}}^{-}}\right) - \mathsf{x}^{(0)} \right|^{2} \middle| p_{\mathsf{x}^{(0)},\mathsf{p}^{-}}(x^{(0)}, p^{-}) \right]. \tag{2.21}$$

Given the fact that $\mathsf{x}^{(0)}$ and p^{-} are independent, (2.21) becomes

$$\begin{aligned}\mathcal{E}_{\mathsf{x}^{+}}\left(\gamma_{\boldsymbol{x}_{\mathsf{e}}^{-}}\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} &\left| g_{\mathsf{x}^{+}}\left(x^{(0)} + p^{-}, \gamma_{\boldsymbol{x}_{\mathsf{e}}^{-}}\right) - x^{(0)} \right|^{2} \\ &\times\, p_{\mathsf{x}^{(0)}}(x^{(0)})\, \mathcal{CN}\left(p^{-}; 0, \gamma_{\boldsymbol{x}_{\mathsf{e}}^{-}}^{-1}\right) \mathrm{d}x^{(0)}\, \mathrm{d}p^{-}.\end{aligned} \tag{2.22}$$

[1]This setting is commonly assumed in state evolution analysis and it holds true if the algorithm at hand is optimum. That is to say the noise precision γ_{w} is known, and the elements of the MMSE estimation error $\widehat{\boldsymbol{x}}_{\mathsf{p}}^{+} - \boldsymbol{x}^{(0)}$ (resp., $\widehat{\boldsymbol{x}}_{\mathsf{p}}^{-} - \boldsymbol{x}^{(0)}$) have averaged variance equal to the one predicted by the algorithm, i.e., $\overline{\gamma}_{\boldsymbol{x}_{\mathsf{p}}^{+}}^{-1}$ (resp., $\overline{\gamma}_{\boldsymbol{x}_{\mathsf{p}}^{-}}^{-1}$).

For the LMMSE estimator, the function $\mathcal{E}_{\mathsf{x}^-}\left(\gamma_{\boldsymbol{x}_\mathsf{e}^+}\right)$ in (2.20) represents the component-wise average MSE defined as [25]:

$$\mathcal{E}_{\mathsf{x}^-}\left(\gamma_{\boldsymbol{x}_\mathsf{e}^+}\right) = \lim_{N\to\infty} \tfrac{1}{N}\, \mathbb{E}\Big[\big\| \boldsymbol{g}_{\mathsf{x}^-}(\widehat{\mathbf{x}}_\mathsf{e}^+, \gamma_{\boldsymbol{x}_\mathsf{e}^+}) - \boldsymbol{x}^{(0)} \big\|^2 \,\Big|\, p_{\mathbf{p}^+,\mathbf{w}}(\boldsymbol{p}^+, \boldsymbol{w}) \Big] \tag{2.23}$$

where

$$\widehat{\mathbf{x}}_\mathsf{e}^+ = \boldsymbol{x}^{(0)} + \mathbf{p}^+, \quad \mathbf{p}^+ \sim \mathcal{CN}(\boldsymbol{p}^+; \mathbf{0}, \gamma_{\boldsymbol{x}_\mathsf{e}^+}^{-1}\, \mathbf{I}_N), \tag{2.24a}$$

$$\mathbf{y} = \boldsymbol{H}\, \boldsymbol{x}^{(0)} + \mathbf{w}, \quad \mathbf{w} \sim \mathcal{CN}(\boldsymbol{w}; \mathbf{0}, \gamma_\mathsf{w}^{-1}\, \mathbf{I}_N). \tag{2.24b}$$

Note that the random vectors in (2.24b) are of size N instead of M due to the padded SVD described in (2.17). Finally, it was shown in [25] that (2.23) can be expressed as

$$\mathcal{E}_{\mathsf{x}^-}\left(\gamma_{\boldsymbol{x}_\mathsf{e}^+}\right) = \mathbb{E}\left[\frac{1}{\gamma_\mathsf{w}\, |\mathsf{s}_{\boldsymbol{H}}|^2 + \gamma_{\boldsymbol{x}_\mathsf{e}^+}} \,\middle|\, p_{\mathsf{s}_{\boldsymbol{H}}}(s_{\boldsymbol{H}}) \right]. \tag{2.25}$$

Chapter 3

The VAMP-KF Algorithm

Before delving into the derivation details, we first recall the state-space model given in (1.1)–(1.2):

$$\boldsymbol{x}_t = \alpha\, \boldsymbol{F}_t\, \boldsymbol{x}_{t-1} + \sqrt{1-\alpha^2}\, \boldsymbol{v}_t, \qquad \text{(process equation)} \tag{3.1a}$$

$$\boldsymbol{y}_t = \boldsymbol{H}_t\, \boldsymbol{x}_t + \boldsymbol{w}_t. \qquad \text{(measurement equation)} \tag{3.1b}$$

The VAMP-KF algorithm estimates the signal vector $\boldsymbol{x}_t$ in (3.1) at every time step t. Its algorithmic steps are outlined in Algorithm 2. Its factor graph is depicted in Fig. 3.1. For better illustration, the block diagram of Algorithm 2 is depicted in Fig. 3.2 whereby we show its different constituent blocks associated with the VAMP and Kalman filter modules, namely the MMSE and LMMSE denoisers as they interact through the extrinsic information (cf. Section 2.3.1).

Algorithm 2 VAMP-KF

Require: Number of time steps T, process memory factor α, process matrix $\boldsymbol{F}_t \in \mathbb{C}^{N\times N}$, measurement matrix $\boldsymbol{H}_t \in \mathbb{C}^{M\times N}$, measurement vector $\boldsymbol{y}_t \in \mathbb{C}^M$, MMSE denoising function $\boldsymbol{g}_{\mathsf{v}^+,t}(\cdot)$, noise precision γ_{w}, mean $\widehat{\boldsymbol{x}}_{0|0}$ and precision $\gamma_{\boldsymbol{x}_{0|0}}$ for the initial state $\boldsymbol{x}_0$, and the number of iterations K.

for $t = 1, 2, \ldots, T$ **do**

  Compute economic SVD $\boldsymbol{H}_t = \overline{\boldsymbol{U}}_{\boldsymbol{H}} \operatorname{Diag}(\overline{\boldsymbol{s}}_{\boldsymbol{H}}) \overline{\boldsymbol{V}}_{\boldsymbol{H}}^{\mathsf{H}}$ with $\overline{\boldsymbol{U}}_{\boldsymbol{H}}^{\mathsf{H}} \overline{\boldsymbol{U}}_{\boldsymbol{H}} = \mathbf{I}_R$, $\overline{\boldsymbol{V}}_{\boldsymbol{H}}^{\mathsf{H}} \overline{\boldsymbol{V}}_{\boldsymbol{H}} = \mathbf{I}_R$, $\overline{\boldsymbol{s}}_{\boldsymbol{H}} \in \mathbb{R}_+^R$, and $R = \operatorname{rank}(\boldsymbol{H})$.

  $\overline{\boldsymbol{s}} = \sqrt{1-\alpha^2}\, \overline{\boldsymbol{s}}_{\boldsymbol{H}}$; $\widetilde{\boldsymbol{z}} = \operatorname{Diag}(\overline{\boldsymbol{s}})^{-1} \overline{\boldsymbol{U}}_{\boldsymbol{H}}^{\mathsf{H}} \left(\boldsymbol{y}_t - \alpha\, \boldsymbol{H}_t \boldsymbol{F}_t\, \widehat{\boldsymbol{x}}_{t-1|t-1}\right)$

  $\widetilde{\boldsymbol{y}} = \operatorname{Diag}(\overline{\boldsymbol{s}}_{\boldsymbol{H}})^{-1} \overline{\boldsymbol{U}}_{\boldsymbol{H}}^{\mathsf{H}}\, \boldsymbol{y}_t$; $\gamma_{\mathsf{z}} = \left\langle \operatorname{diag}\left(\alpha^2\, \gamma_{\boldsymbol{x}_{t-1|t-1}}^{-1}\, \boldsymbol{H}_t \boldsymbol{F}_t \boldsymbol{F}_t^{\mathsf{H}} \boldsymbol{H}_t^{\mathsf{H}} + \gamma_{\mathsf{w}}^{-1}\, \mathbf{I}_M\right)\right\rangle^{-1}$

  ▷ **VAMP module**

  Initialize $\widehat{\boldsymbol{v}}_{\mathsf{e},0}^-$ and $\gamma_{\boldsymbol{v}_{\mathsf{e}}^-,0} \geq 0$.

  for $k = 0, 1, \ldots, K$ **do**

    MMSE step

    $\widehat{\boldsymbol{v}}_{\mathsf{p},k}^+ = \boldsymbol{g}_{\mathsf{v}^+,t}\left(\widehat{\boldsymbol{v}}_{\mathsf{e},k}^-, \gamma_{\boldsymbol{v}_{\mathsf{e}}^-,k}\right)$

    $\gamma_{\boldsymbol{v}_{\mathsf{p}}^+,k} = \gamma_{\boldsymbol{v}_{\mathsf{e}}^-,k} \Big/ \left\langle \boldsymbol{g}'_{\mathsf{v}^+,t}\left(\widehat{\boldsymbol{v}}_{\mathsf{e},k}^-, \gamma_{\boldsymbol{v}_{\mathsf{e}}^-,k}\right)\right\rangle$

    $\gamma_{\boldsymbol{v}_{\mathsf{e}}^+,k} = \gamma_{\boldsymbol{v}_{\mathsf{p}}^+,k} - \gamma_{\boldsymbol{v}_{\mathsf{e}}^-,k}$

    $\widehat{\boldsymbol{v}}_{\mathsf{e},k}^+ = \gamma_{\boldsymbol{v}_{\mathsf{e}}^+,k}^{-1}\left(\gamma_{\boldsymbol{v}_{\mathsf{p}}^+,k}\, \widehat{\boldsymbol{v}}_{\mathsf{p},k}^+ - \gamma_{\boldsymbol{v}_{\mathsf{e}}^-,k}\, \widehat{\boldsymbol{v}}_{\mathsf{e},k}^-\right)$

    LMMSE step

    $\boldsymbol{d}_k = \gamma_{\mathsf{z}} \operatorname{Diag}\left(\gamma_{\mathsf{z}}\, \overline{\boldsymbol{s}}^2 + \gamma_{\boldsymbol{v}_{\mathsf{e}}^+,k}\, \mathbf{1}_R\right)^{-1} \overline{\boldsymbol{s}}^2$

    $\gamma_{\boldsymbol{v}_{\mathsf{e}}^-,k+1} = \gamma_{\boldsymbol{v}_{\mathsf{e}}^+,k} \left\langle \boldsymbol{d}_k\right\rangle \Big/ \left(\frac{N}{R} - \left\langle \boldsymbol{d}_k\right\rangle\right)$

    $\widehat{\boldsymbol{v}}_{\mathsf{e},k+1}^- = \widehat{\boldsymbol{v}}_{\mathsf{e},k}^+ + \frac{N}{R}\, \overline{\boldsymbol{V}}_{\boldsymbol{H}} \operatorname{Diag}\left(\frac{\boldsymbol{d}_k}{\left\langle \boldsymbol{d}_k\right\rangle}\right)\left(\widetilde{\boldsymbol{z}} - \overline{\boldsymbol{V}}_{\boldsymbol{H}}^{\mathsf{H}}\, \widehat{\boldsymbol{v}}_{\mathsf{e},k}^+\right)$

  end for

  ▷ **Kalman filter module**

  Prediction step

  $\widehat{\boldsymbol{x}}_{t|t-1} = \alpha\, \boldsymbol{F}_t\, \widehat{\boldsymbol{x}}_{t-1|t-1} + \sqrt{1-\alpha^2}\, \widehat{\boldsymbol{v}}_{\mathsf{e},K}^+$

  $\gamma_{\boldsymbol{x}_{t|t-1}} = \left(\alpha^2\, \gamma_{\boldsymbol{x}_{t-1|t-1}}^{-1} \left\langle \operatorname{diag}\left(\boldsymbol{F}_t \boldsymbol{F}_t^{\mathsf{H}}\right)\right\rangle + (1-\alpha^2)\, \gamma_{\boldsymbol{v}_{\mathsf{e}}^+,K}^{-1}\right)^{-1}$

  Estimation step

  $\boldsymbol{d}_{t|t} = \gamma_{\mathsf{w}} \operatorname{Diag}\left(\gamma_{\mathsf{w}}\, \overline{\boldsymbol{s}}_{\boldsymbol{H}}^2 + \gamma_{\boldsymbol{x}_{t|t-1}}\, \mathbf{1}_R\right)^{-1} \overline{\boldsymbol{s}}_{\boldsymbol{H}}^2$

  $\widehat{\boldsymbol{x}}_{t|t} = \widehat{\boldsymbol{x}}_{t|t-1} + \frac{N}{R}\, \overline{\boldsymbol{V}}_{\boldsymbol{H}} \operatorname{Diag}\left(\frac{\boldsymbol{d}_{t|t}}{\left\langle \boldsymbol{d}_{t|t}\right\rangle}\right)\left(\widetilde{\boldsymbol{y}} - \overline{\boldsymbol{V}}_{\boldsymbol{H}}^{\mathsf{H}}\, \widehat{\boldsymbol{x}}_{t|t-1}\right)$

  $\gamma_{\boldsymbol{x}_{t|t}} = \gamma_{\boldsymbol{x}_{t|t-1}} \left\langle \boldsymbol{d}_{t|t}\right\rangle \Big/ \left(\frac{N}{R} - \left\langle \boldsymbol{d}_{t|t}\right\rangle\right)$

end for

return $\left\{\widehat{\boldsymbol{x}}_{1|1}, \ldots, \widehat{\boldsymbol{x}}_{T|T}\right\}$

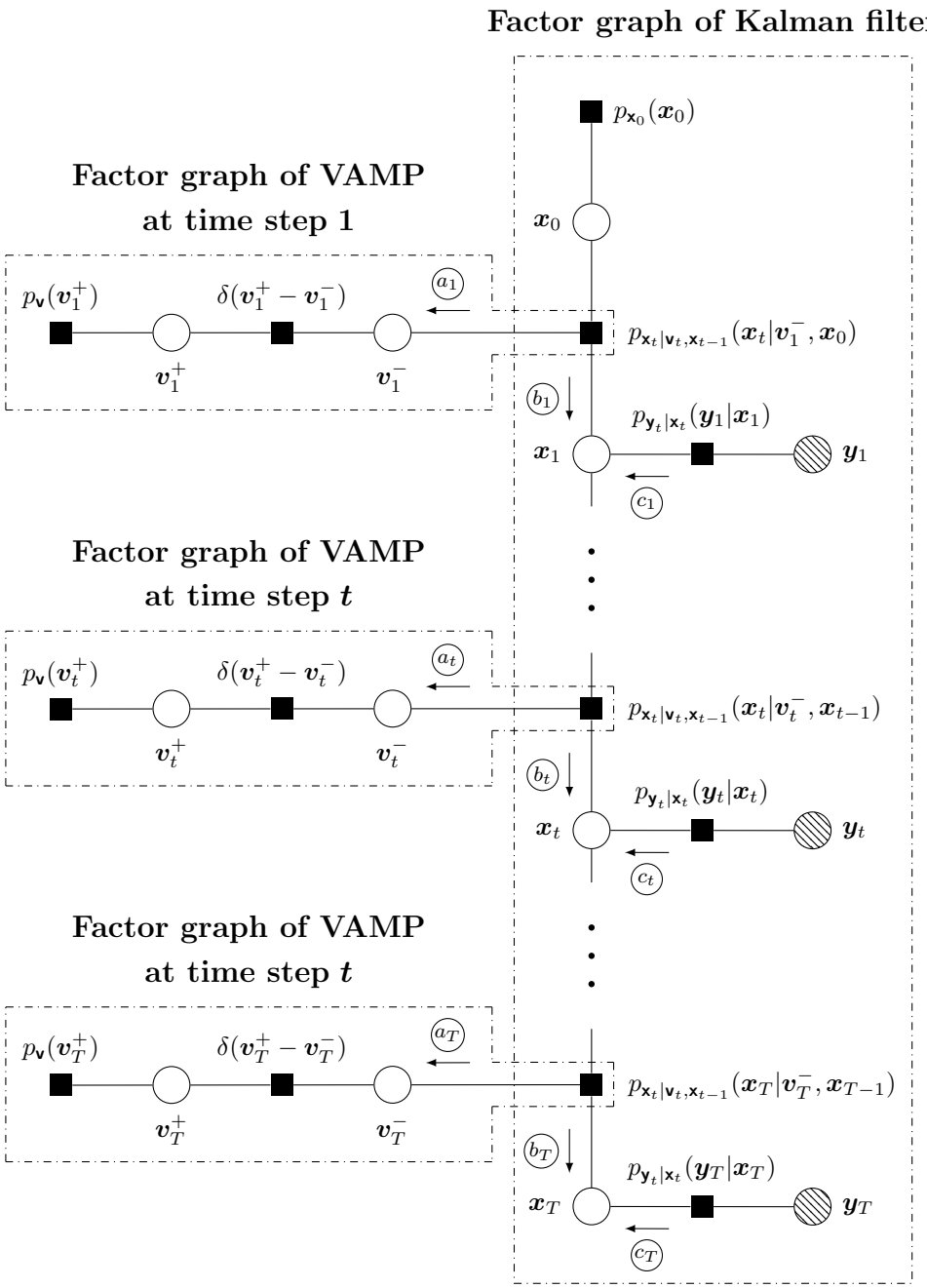


Figure 3.1: The factor graph of VAMP-KF representating the decomposition of the joint density given in (3.4).

3.1 Factor graph: where VAMP meets Kalman filter

The factor graph of the Kalman filter in Fig 2.3 presumes Gaussian innovations $\boldsymbol{v}_t$ in the process equation given in (2.4), i.e., $\boldsymbol{v}_t \sim \mathcal{CN}(\boldsymbol{v}_t; \boldsymbol{0}, \boldsymbol{Q}_t)$. This assumption gives rise to a Gaussian factor node representing the following transition density from $\boldsymbol{x}_t$ to $\boldsymbol{x}_{t-1}$ (cf. Section 2.2):

$$p_{\mathbf{x}_t|\mathbf{x}_{t-1}}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) = \mathcal{CN}(\boldsymbol{x}_t; \boldsymbol{F}_t\,\boldsymbol{x}_{t-1} + \boldsymbol{B}_t\,\boldsymbol{u}_t, \boldsymbol{Q}_t). \tag{3.2}$$

One of our key ideas in this thesis is to substitute the conventional Gaussian innovation prior $\boldsymbol{v}_t \sim \mathcal{CN}(\boldsymbol{v}_t; \boldsymbol{0}, \boldsymbol{Q}_t)$ with a Gaussian prior arising from running scalar EP using the VAMP algorithm. The latter inherently captures the rapidly time-varying sparse evolution of innovations. We start by factoring the joint pdf of all the observed and

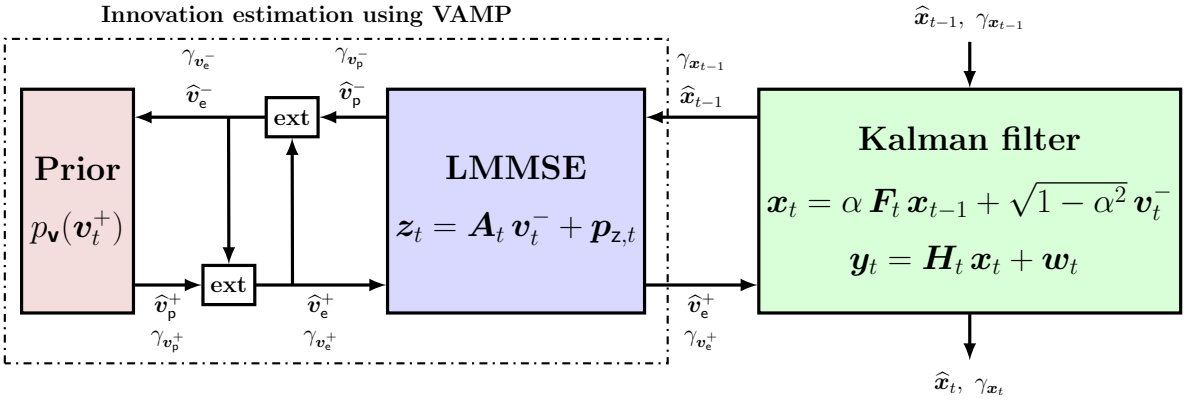


Figure 3.2: Block diagram of VAMP-KF at each time step $t > 0$ with its three modules: the innovation denoising module incorporating the prior information, $p_{\mathbf{v}}(\cdot)$, the LMMSE module, and the Kalman filter module. The denoising and LMMSE modules exchange extrinsic messages through the ext blocks. The color of each module matches the color of its algorithmic steps in Algorithm 2.

unobserved variables in (3.1) at each time step t as follows:

$$p_{\mathbf{y}_t,\mathbf{x}_t,\mathbf{v}_t,\mathbf{x}_{t-1}}(\boldsymbol{y}_t,\boldsymbol{x}_t,\boldsymbol{v}_t,\boldsymbol{x}_{t-1}) = p_{\mathbf{y}_t|\mathbf{x}_t}(\boldsymbol{y}_t|\boldsymbol{x}_t)\, p_{\mathbf{x}_t|\mathbf{v}_t,\mathbf{x}_{t-1}}(\boldsymbol{x}_t|\boldsymbol{v}_t,\boldsymbol{x}_{t-1})\, p_{\mathbf{x}_{t-1}}(\boldsymbol{x}_{t-1})\, p_{\mathbf{v}_t}(\boldsymbol{v}_t). \tag{3.3}$$

Then, we split the innovation variable $\boldsymbol{v}$ into two auxiliary variables, i.e., $\boldsymbol{v}^+$ and $\boldsymbol{v}^-$, thereby leading to the equivalent factorization:

$$\begin{aligned} &p_{\mathbf{y}_t,\mathbf{x}_t,\mathbf{v}_t^-,\mathbf{v}_t^+,\mathbf{x}_{t-1}}(\boldsymbol{y}_t,\boldsymbol{x}_t,\boldsymbol{v}_t^-,\boldsymbol{v}_t^+,\boldsymbol{x}_{t-1}) \\ &\qquad = p_{\mathbf{y}_t|\mathbf{x}_t}(\boldsymbol{y}_t|\boldsymbol{x}_t)\, p_{\mathbf{x}_t|\mathbf{v}_t^-,\mathbf{x}_{t-1}}(\boldsymbol{x}_t|\boldsymbol{v}_t^-,\boldsymbol{x}_{t-1})\, p_{\mathbf{x}_{t-1}}(\boldsymbol{x}_{t-1})\, \delta(\boldsymbol{v}_t^- - \boldsymbol{v}_t^+)\, p_{\mathbf{v}_t^+}(\boldsymbol{v}_t^+). \end{aligned} \tag{3.4}$$

In (3.4), the factor nodes pertaining to the conditional densities $p_{\mathbf{x}_t|\mathbf{v}_t^-,\mathbf{x}_{t-1}}(\boldsymbol{x}_t|\boldsymbol{v}_t^-,\boldsymbol{x}_{t-1})$ and $p_{\mathbf{y}_t|\mathbf{x}_t}(\boldsymbol{y}_t|\boldsymbol{x}_t)$ are specified according to the state-space model in (3.1) as

$$\begin{aligned} p_{\mathbf{x}_t|\mathbf{v}_t^-,\mathbf{x}_{t-1}}(\boldsymbol{x}_t|\boldsymbol{v}_t^-,\boldsymbol{x}_{t-1}) &= \delta\big(\boldsymbol{x}_t - \alpha\, \boldsymbol{F}_t\, \boldsymbol{x}_{t-1} - \sqrt{1-\alpha^2}\, \boldsymbol{v}_t^-\big), \\ p_{\mathbf{y}_t|\mathbf{x}_t}(\boldsymbol{y}_t|\boldsymbol{x}_t) &= \mathcal{CN}(\boldsymbol{y}_t; \boldsymbol{H}_t\, \boldsymbol{x}_t, \gamma_{\mathsf{w}}^{-1}\, \mathbf{I}_M). \end{aligned}$$

To lighten the notation, we will refer to the factor node pertaining to the density $p_{\mathbf{x}_t|\mathbf{v}_t^-,\mathbf{x}_{t-1}}(\boldsymbol{x}_t|\boldsymbol{v}_t^-,\boldsymbol{x}_{t-1})$ as $f_{t|t-1}$. To combine VAMP with the Kalman filter into a unified message passing framework, the factor graph of VAMP-KF in Fig. 3.1 illustrates the three key messages to be determined at every time step $t \in \{1,\ldots,T\}$:

- The message $\mu_{f_{t|t-1}\to\boldsymbol{v}_t^-}$ from the factor node $f_{t|t-1}$ to the variable node $\boldsymbol{v}_t^-$ is represented by $\textcircled{a_t}$. This message is involved in the LMMSE step of the VAMP module stated in lines 11–13 of Algorithm 2 to estimate $\boldsymbol{v}_t^-$ after incorporating the message $\textcircled{c_t}$.

- The message $\textcircled{b_t}$ from the factor node $f_{t|t-1}$ to the variable node $\boldsymbol{x}_t$ represents the prediction step of the Kalman filter given in (2.6). This step must account for the message coming from the variable node $\boldsymbol{v}_t^-$.

- The message $\textcircled{c_t}$ from the factor node $p_{\mathbf{y}_t|\mathbf{x}_t}(\boldsymbol{y}_t|\boldsymbol{x}_t)$ to the variable node $\boldsymbol{x}_t$ pertains to the time update step of the Kalman filter in (2.7). The step is carried out with also the involvement of $\textcircled{b_t}$.

To fully characterize the EP derivation of the proposed VAMP-KF algorithm, one must establish the analytical expressions of the aforementioned messages, which we do in the next section.

3.2 EP derivation of VAMP-KF

In this section, we detail the derivation of the messages exchanged between the VAMP module and the Kalman filter. Towards this goal, we depict in Fig. 3.3 the factor graph of VAMP-KF at time step t. Specifically, we focus on the derivation of the messages ③, ⑥, and ⑦, which correspond to the aforementioned messages $\textcircled{a_t}$, $\textcircled{b_t}$, and $\textcircled{c_t}$ in Fig. 3.1.

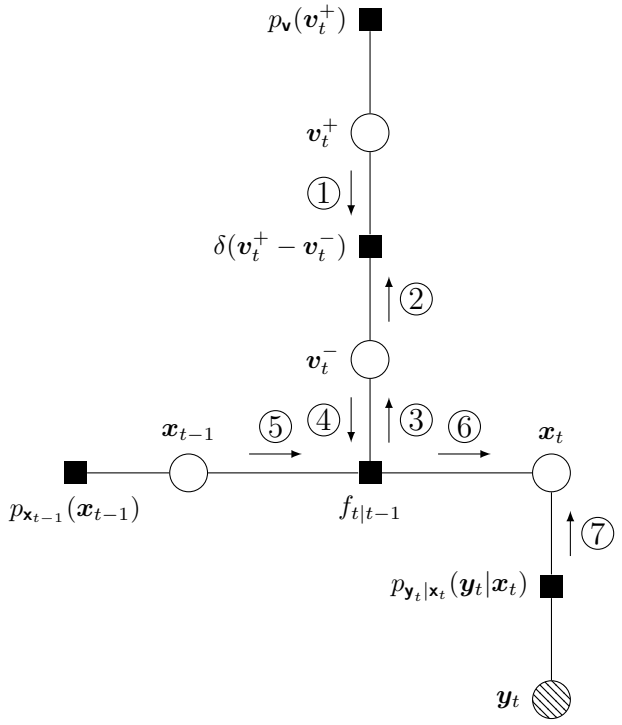
3.2.1 Derivation of the LMMSE estimate of the innovation

We now detail the derivation of the message $\mu_{f_{t|t-1}\to\boldsymbol{v}_t^-}$ which is denoted as message ③ in Fig. 3.3. The latter is obtained as the marginal of the density resulting from the product of messages ⑤ and ⑦, and the factor node $f_{t|t-1}$, i.e.,

$$\mu_{f_{t|t-1}\to\boldsymbol{v}_t^-} \propto \int_{-\infty}^{\infty} \underbrace{p_{\mathbf{x}_{t-1}}(\boldsymbol{x}_{t-1})}_{⑤} \int_{-\infty}^{\infty} \underbrace{p_{\mathbf{y}_t|\mathbf{x}_t}(\boldsymbol{y}_t|\boldsymbol{x}_t)}_{⑦} \underbrace{p_{\mathbf{x}_t|\mathbf{v}_t,\mathbf{x}_{t-1}}(\boldsymbol{x}_t|\boldsymbol{v}_t^-,\boldsymbol{x}_{t-1})}_{f_{t|t-1}} \,\mathrm{d}\boldsymbol{x}_t \,\mathrm{d}\boldsymbol{x}_{t-1}.$$

Equivalently, we have

$$\mu_{f_{t|t-1}\to\boldsymbol{v}_t^-} \propto \int_{-\infty}^{\infty} \mathcal{CN}\left(\boldsymbol{x}_{t-1}; \widehat{\boldsymbol{x}}_{t-1|t-1}, \gamma^{-1}_{\boldsymbol{x}_{t-1|t-1}} \mathbf{I}_N\right) \int_{-\infty}^{\infty} \mathcal{CN}\left(\boldsymbol{y}_t; \boldsymbol{H}_t\,\boldsymbol{x}_t, \gamma_{\mathsf{w}}^{-1}\,\mathbf{I}_M\right) \times \delta\left(\boldsymbol{x}_t - \alpha\,\boldsymbol{F}_t\,\boldsymbol{x}_{t-1} - \sqrt{1-\alpha^2}\,\boldsymbol{v}_t^-\right) \mathrm{d}\boldsymbol{x}_t \,\mathrm{d}\boldsymbol{x}_{t-1}. \quad (3.6)$$



$\mu_{\boldsymbol{v}_t^+ \to \delta}$	①	$\mathcal{CN}(\boldsymbol{v}_t^+; \widehat{\boldsymbol{v}}_{\mathsf{e},k}^+, \gamma_{\boldsymbol{v}_\mathsf{e}^+,k}^{-1} \mathbf{I}_N)$
$\mu_{\boldsymbol{v}_t^- \to \delta}$	②	$\mathcal{CN}(\boldsymbol{v}_t^-; \widehat{\boldsymbol{v}}_{\mathsf{e},k}^-, \gamma_{\boldsymbol{v}_\mathsf{e}^-,k}^{-1} \mathbf{I}_N)$
$\mu_{f_{t\|t-1} \to \boldsymbol{v}_t^-}$	③	$\mathcal{CN}(\boldsymbol{z}_t; \boldsymbol{A}_t \boldsymbol{v}_t^-, \gamma_{\mathsf{z},t}^{-1} \mathbf{I}_M)$
$\mu_{\boldsymbol{v}_t^- \to f_{t\|t-1}}$	④	$\mathcal{CN}(\boldsymbol{v}_t^-; \widehat{\boldsymbol{v}}_{\mathsf{e},K}^+, \gamma_{\boldsymbol{v}_\mathsf{e}^+,K}^{-1} \mathbf{I}_N)$
$\mu_{\boldsymbol{x}_{t-1} \to f_{t\|t-1}}$	⑤	$\mathcal{CN}(\boldsymbol{x}_{t-1}; \widehat{\boldsymbol{x}}_{t-1\|t-1}, \gamma_{\boldsymbol{x}_{t-1\|t-1}}^{-1} \mathbf{I}_N)$
$\mu_{f_{t\|t-1} \to \boldsymbol{x}_t}$	⑥	$\mathcal{CN}(\boldsymbol{x}_t; \widehat{\boldsymbol{x}}_{t\|t-1}, \gamma_{\boldsymbol{x}_{t\|t-1}}^{-1} \mathbf{I}_N)$
$\mu_{p_{\mathbf{y}_t\|\mathbf{x}_t} \to \boldsymbol{x}_t}$	⑦	$\mathcal{CN}(\boldsymbol{y}_t; \boldsymbol{H}_t \boldsymbol{x}_t, \gamma_{\mathsf{w}}^{-1} \mathbf{I}_M)$

Figure 3.3: Factor graph of VAMP-KF at time step t.

Using standard manipulations, the inner integral is expressed in a closed form and (3.6) becomes

$$\mu_{f_{t|t-1} \to \boldsymbol{v}_t^-} \propto \int_{-\infty}^{\infty} \mathcal{CN}\left(\boldsymbol{x}_{t-1}; \widehat{\boldsymbol{x}}_{t-1|t-1}, \gamma_{\boldsymbol{x}_{t-1|t-1}}^{-1} \mathbf{I}_N\right) \times \mathcal{CN}\left(\alpha \boldsymbol{H}_t \boldsymbol{F}_t \boldsymbol{x}_{t-1}; \boldsymbol{y}_t - \boldsymbol{A}_t \boldsymbol{v}_t^-, \gamma_{\mathsf{w}}^{-1} \mathbf{I}_M\right) \mathrm{d}\boldsymbol{x}_{t-1}, \tag{3.7}$$

with $\boldsymbol{A}_t \triangleq \sqrt{1-\alpha^2}\,\boldsymbol{H}_t$. Here, the first belief of the integrand function is in the signal dynamic space $\mathbb{C}^N$ while the second belief is in the measurement space $\mathbb{C}^M$. To find the integral in (3.7), we transform the signal $\boldsymbol{x}_{t-1}$ to the projected signal $\alpha\,\boldsymbol{H}_t\,\boldsymbol{F}_t\,\boldsymbol{x}_{t-1}$, thereby yielding

$$\begin{aligned}\mu_{f_{t|t-1}\to\boldsymbol{v}_t^-} \propto \int_{-\infty}^{\infty} &\mathcal{CN}\big(\alpha\,\boldsymbol{H}_t\,\boldsymbol{F}_t\,\boldsymbol{x}_{t-1};\alpha\,\boldsymbol{H}_t\,\boldsymbol{F}_t\,\widehat{\boldsymbol{x}}_{t-1|t-1},\alpha^2\,\gamma^{-1}_{\boldsymbol{x}_{t-1|t-1}}\,\boldsymbol{H}_t\,\boldsymbol{F}_t\,\boldsymbol{F}_t^{\mathsf{H}}\,\boldsymbol{H}_t^{\mathsf{H}}\big)\\ &\times\mathcal{CN}(\alpha\,\boldsymbol{H}_t\,\boldsymbol{F}_t\,\boldsymbol{x}_{t-1};\boldsymbol{y}_t-\boldsymbol{A}_t\,\boldsymbol{v}_t^-,\gamma_{\mathsf{w}}^{-1}\,\mathbf{I}_M)\,\mathrm{d}\boldsymbol{x}_{t-1}.\end{aligned}\tag{3.8}$$

Using to the scaled covariance approximation from scalar EP, we combine the two beliefs of the integrand function in (3.8) based on the multiplication rule of Gaussian densities to obtain the message $\mu_{f_{t|t-1}\to\boldsymbol{v}^-}$ (represented by ③ in Fig. 3.3) as

$$\mu_{f_{t|t-1}\to\boldsymbol{v}_t^-} \propto \mathcal{CN}\big(\boldsymbol{z}_t;\boldsymbol{A}_t\,\boldsymbol{v}_t^-,\gamma_{\mathsf{z},t}^{-1}\,\mathbf{I}_M\big),\tag{3.9}$$

with

$$\boldsymbol{z}_t = \boldsymbol{y}_t - \alpha\,\boldsymbol{H}_t\boldsymbol{F}_t\,\widehat{\boldsymbol{x}}_{t-1|t-1},\tag{3.10a}$$

$$\gamma_{\mathsf{z},t} = \Big\langle \mathrm{diag}\Big(\alpha^2\,\gamma^{-1}_{\boldsymbol{x}_{t-1|t-1}}\,\boldsymbol{H}_t\boldsymbol{F}_t\boldsymbol{F}_t^{\mathsf{H}}\boldsymbol{H}_t^{\mathsf{H}}+\gamma_{\mathsf{w}}^{-1}\,\mathbf{I}_M\Big)\Big\rangle^{-1}.\tag{3.10b}$$

The posterior message of $\boldsymbol{v}_t^-$ is given by the product of the message $\mu_{\delta\to\boldsymbol{v}^-} = \mathcal{CN}\big(\boldsymbol{v}^-;\widehat{\boldsymbol{v}}_{\mathsf{e}}^+,\gamma^{-1}_{\boldsymbol{v}_{\mathsf{e}}^+}\mathbf{I}_N\big)$ (represented by ① in Fig. 3.3) with message ③ in (3.9). Using standard manipulations (as shown in Appendix A of [25]), the posterior mean and

precision of the LMMSE estimator of $\boldsymbol{v}_t^-$ are obtained as

$$\begin{aligned}\widehat{\boldsymbol{v}}_{\mathsf{p},k}^- &= \left(\gamma_{\mathsf{z},t}\,\boldsymbol{A}_t^{\mathsf{H}}\boldsymbol{A}_t + \gamma_{\boldsymbol{v}_{\mathsf{e}}^+,k}\,\mathbf{I}_N\right)^{-1}\left(\gamma_{\mathsf{z},t}\,\boldsymbol{A}_t^{\mathsf{H}}\,\boldsymbol{z}_t + \gamma_{\boldsymbol{v}_{\mathsf{e}}^+,k}\,\widehat{\boldsymbol{v}}_{\mathsf{e},k}^+\right), \\ &\triangleq \boldsymbol{g}_{\mathbf{v}^-,t}\left(\widehat{\boldsymbol{v}}_{\mathsf{e},k}^+, \gamma_{\boldsymbol{v}_{\mathsf{e}}^+,k}, \boldsymbol{z}_t, \gamma_{\mathsf{z},t}\right) \qquad (3.11a)\\ \gamma_{\boldsymbol{v}_{\mathsf{p}}^-,k} &= \gamma_{\boldsymbol{v}_{\mathsf{e}}^+,k}/\langle \boldsymbol{g}'_{\mathbf{v}^-,t}(\widehat{\boldsymbol{v}}_{\mathsf{e},k}^+, \gamma_{\boldsymbol{v}_{\mathsf{e}}^+,k}, \boldsymbol{z}_t, \gamma_{\mathsf{z},t})\rangle \\ &= N/\mathrm{Tr}\left[\left(\gamma_{\mathsf{z},t}\,\boldsymbol{A}_t^{\mathsf{H}}\boldsymbol{A}_t + \gamma_{\boldsymbol{v}_{\mathsf{e}}^+,k}\,\mathbf{I}_N\right)^{-1}\right]. \qquad (3.11b)\end{aligned}$$

Note here the high similarity between the obtained LMMSE estimate of $\boldsymbol{v}_t^-$ in (3.11) with the one of VAMP in (2.14). Finally, the extrinsic message represented by ② in Fig. 3.3 can be obtained based on the extrinsic belief computation rule (cf. Section 2.3) as follows:

$$\gamma_{\boldsymbol{v}_{\mathsf{e}}^-,k+1} = \gamma_{\boldsymbol{v}_{\mathsf{p}}^-,k} - \gamma_{\boldsymbol{v}_{\mathsf{e}}^+,k}, \qquad (3.12a)$$

$$\widehat{\boldsymbol{v}}_{\mathsf{e},k+1}^- = \gamma_{\boldsymbol{v}_{\mathsf{e}}^-,k+1}^{-1}\left(\gamma_{\boldsymbol{v}_{\mathsf{p}}^-,k}\,\widehat{\boldsymbol{v}}_{\mathsf{p},k}^- - \gamma_{\boldsymbol{v}_{\mathsf{e}}^+,k}\,\widehat{\boldsymbol{v}}_{\mathsf{e},k}^+\right). \qquad (3.12b)$$

Similarly to VAMP, to mitigate the high computational cost of the matrix inversion in (3.11) and (3.12), we compute the SVD of $\boldsymbol{A}_t$ to obtain the steps outlined in lines 11–13 of Algorithm 2, which are similar to lines 9–11 of Algorithm 1.

3.2.2 Derivation of the VAMP-KF prediction step

Once the innovation vector $\boldsymbol{v}_t$ has been estimated through VAMP, the prediction step of the Kalman filter combines the incoming message from the variable node $\boldsymbol{v}_t^-$ (represented by ④ in Fig. 3.3) with both the prior message $p_{\mathbf{x}_{t-1}}(\boldsymbol{x}_{t-1})$ (represented by message ⑤)

and the factor node $f_{t|t-1}$ using the sum product rule:

$$\begin{aligned}\underbrace{\mu_{f_{t|t-1}\to\boldsymbol{x}_t}}_{\textcircled{6}} &\propto \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \underbrace{\mu_{\boldsymbol{v}_t^-\to f_{t|t-1}}}_{\textcircled{4}}\, \underbrace{p_{\mathbf{x}_{t-1}}(\boldsymbol{x}_{t-1})}_{\textcircled{5}}\, \underbrace{p_{\mathbf{x}_t|\mathbf{v}_t^-,\mathbf{x}_{t-1}}(\boldsymbol{x}_t|\boldsymbol{v}_t^-,\boldsymbol{x}_{t-1})}_{f_{t|t-1}}\, \mathrm{d}\boldsymbol{v}_t^-\, \mathrm{d}\boldsymbol{x}_{t-1} \\ &= \int_{-\infty}^{\infty} \mathcal{CN}\left(\boldsymbol{x}_{t-1}; \widehat{\boldsymbol{x}}_{t-1|t-1}, \gamma^{-1}_{\boldsymbol{x}_{t-1|t-1}}\,\mathbf{I}_N\right) \int_{-\infty}^{\infty} \mathcal{CN}\left(\boldsymbol{v}_t^-; \widehat{\boldsymbol{v}}^+_{\mathsf{e},K}, \gamma^{-1}_{\boldsymbol{v}^+_{\mathsf{e}},K}\,\mathbf{I}_N\right) \\ &\qquad\qquad \times \delta\left(\boldsymbol{x}_t - \alpha\,\boldsymbol{F}_t\,\boldsymbol{x}_{t-1} - \sqrt{1-\alpha^2}\,\boldsymbol{v}_t^-\right)\,\mathrm{d}\boldsymbol{v}_t^-\,\mathrm{d}\boldsymbol{x}_{t-1}.\end{aligned} \tag{3.13}$$

Explicitly expressing the inner integral in (3.13) yields

$$\begin{aligned}\mu_{f_{t|t-1}\to\boldsymbol{x}_t} \propto \int_{-\infty}^{\infty} &\mathcal{CN}\left(\boldsymbol{x}_{t-1}; \widehat{\boldsymbol{x}}_{t-1|t-1}, \gamma^{-1}_{\boldsymbol{x}_{t-1|t-1}}\,\mathbf{I}_N\right) \\ &\times \mathcal{CN}\Big(\alpha\,\boldsymbol{F}_t\,\boldsymbol{x}_{t-1}; \boldsymbol{x}_t - \sqrt{1-\alpha^2}\,\widehat{\boldsymbol{v}}^+_{\mathsf{e},K}, (1-\alpha^2)\,\gamma^{-1}_{\boldsymbol{v}^+_{\mathsf{e}},K}\,\mathbf{I}_N\Big)\,\mathrm{d}\boldsymbol{x}_{t-1}.\end{aligned} \tag{3.14}$$

Here, we follow the same argument we used to proceed from (3.7) to (3.8). In (3.14), we transform the signal $\boldsymbol{x}_{t-1}$ in the first belief to the projected signal $\alpha\,\boldsymbol{F}_t\,\boldsymbol{x}_{t-1}$, thereby yielding

$$\begin{aligned}\mu_{f_{t|t-1}\to\boldsymbol{x}_t} \propto\!\!\!\propto \int_{-\infty}^{\infty} &\mathcal{CN}\Big(\alpha\,\boldsymbol{F}_t\,\boldsymbol{x}_{t-1}; \alpha\,\boldsymbol{F}_t\,\widehat{\boldsymbol{x}}_{t-1|t-1}, \alpha^2\,\gamma^{-1}_{\boldsymbol{x}_{t-1|t-1}}\,\boldsymbol{F}_t\,\boldsymbol{F}_t^{\mathsf{H}}\Big) \\ &\times \mathcal{CN}\Big(\alpha\,\boldsymbol{F}_t\,\boldsymbol{x}_{t-1}; \boldsymbol{x}_t - \sqrt{1-\alpha^2}\,\widehat{\boldsymbol{v}}^+_{\mathsf{e},K}, (1-\alpha^2)\,\gamma^{-1}_{\boldsymbol{v}^+_{\mathsf{e}},K}\,\mathbf{I}_N\Big)\,\mathrm{d}\boldsymbol{x}_{t-1}.\end{aligned} \tag{3.15}$$

By combining the two beliefs of the integrand function in (3.15) and then finding the integral, we obtain the expression of the message $\mu_{f_{t|t-1}\to\boldsymbol{x}_t}$ (represented by $\textcircled{6}$ in Fig. 3.3) as

$$\mu_{f_{t|t-1}\to\boldsymbol{x}_t} \propto \mathcal{CN}\left(\boldsymbol{x}_t; \widehat{\boldsymbol{x}}_{t|t-1}, \gamma^{-1}_{\boldsymbol{x}_{t|t-1}}\,\mathbf{I}_N\right), \tag{3.16}$$

with

$$\begin{aligned}\widehat{\boldsymbol{x}}_{t|t-1} &= \alpha\, \boldsymbol{F}_t\, \widehat{\boldsymbol{x}}_{t-1|t-1} + \sqrt{1-\alpha^2}\, \widehat{\boldsymbol{v}}^{+}_{\mathsf{e},K}, \\ &\triangleq \boldsymbol{g}_{\mathbf{x}_{t|t-1}}\left(\widehat{\boldsymbol{x}}_{t-1|t-1}, \widehat{\boldsymbol{v}}^{+}_{\mathsf{e},K}\right) \qquad (3.17a)\\ \gamma_{\boldsymbol{x}_{t|t-1}} &= \left(\alpha^2\, \gamma^{-1}_{\boldsymbol{x}_{t-1|t-1}}\, \left\langle \mathrm{diag}\left(\boldsymbol{F}_t \boldsymbol{F}_t^{\mathsf{H}}\right)\right\rangle + (1-\alpha^2)\, \gamma^{-1}_{\boldsymbol{v}^{+}_{\mathsf{e}},K}\right)^{-1}, \qquad (3.17b)\end{aligned}$$

which correspond to lines 15-16 of Algorithm 2. Note that (3.17) could also be derived using the standard prediction step of the Kalman filter given in (2.6) by setting $\widehat{\boldsymbol{\Sigma}}_{t-1|t-1} = \gamma^{-1}_{\boldsymbol{x}_{t-1|t-1}}\, \mathbf{I}_N$, $\boldsymbol{Q}_t = (1-\alpha^2)\, \gamma^{-1}_{\boldsymbol{v}^{+}_{\mathsf{e}},K}\, \mathbf{I}_N$, $\boldsymbol{B}_t = \mathbf{I}_N$, and $\boldsymbol{u}_t = \sqrt{1-\alpha^2}\, \widehat{\boldsymbol{v}}^{+}_{\mathsf{e},K}$, with the state transition matrix being $\alpha\, \boldsymbol{F}_t$.

3.2.3 Derivation of the VAMP-KF time update step

In this step, the Kalman filter calibrates the prediction estimate in (3.16) with the measurement vector $\boldsymbol{y}_t$. It does so by multiplying the message ⑥ in (3.16) with the density $p_{\mathbf{y}_t|\mathbf{x}_t}(\boldsymbol{y}_t|\boldsymbol{x}_t)$ pertaining to the factor node $f_{\mathbf{y}|\mathbf{x}}$ (represented by ⑦ in Fig. 3.3), i.e.,

$$p_{\mathbf{x}_t}(\boldsymbol{x}_t) \propto \underbrace{\mathcal{CN}(\boldsymbol{x}_t; \widehat{\boldsymbol{x}}_{t|t-1}, \gamma^{-1}_{\boldsymbol{x}_{t|t-1}}\, \mathbf{I}_N)}_{⑥}\, \underbrace{\mathcal{CN}(\boldsymbol{y}_t; \boldsymbol{H}_t\, \boldsymbol{x}_t, \gamma^{-1}_{\mathsf{w}}\, \mathbf{I}_N)}_{⑦}. \qquad (3.18)$$

Based on the multiplication rule of Gaussian densities, (3.18) simplifies to

$$p_{\mathbf{x}_t}(\boldsymbol{x}_t) = \mathcal{CN}\left(\boldsymbol{x}_t; \widehat{\boldsymbol{x}}_{t|t}, \gamma^{-1}_{\boldsymbol{x}_{t|t}}\, \mathbf{I}_N\right), \qquad (3.19)$$

where $\widehat{\boldsymbol{x}}_{t|t}$ and $\gamma_{\boldsymbol{x}_{t|t}}$ are given by

$$\begin{aligned}\widehat{\boldsymbol{x}}_{t|t} &= \widehat{\boldsymbol{x}}_{t|t-1} + \boldsymbol{K}_t\, (\boldsymbol{y}_t - \boldsymbol{H}_t\, \widehat{\boldsymbol{x}}_{t|t-1}), \qquad (3.20a)\\ \gamma_{t|t} &= \left\langle \mathrm{diag}\left(\gamma^{-1}_{\boldsymbol{x}_{t|t-1}}\, \mathbf{I}_N - \gamma^{-1}_{\boldsymbol{x}_{t|t-1}}\, \boldsymbol{K}_t\, \boldsymbol{H}_t\right)\right\rangle^{-1} \qquad (3.20b)\end{aligned}$$

with the Kalman gain matrix being

$$\boldsymbol{K}_t = \gamma_{\boldsymbol{x}_{t|t-1}}^{-1} \boldsymbol{H}_t^{\mathsf{H}} \left(\gamma_{\boldsymbol{x}_{t|t-1}}^{-1} \boldsymbol{H}_t \boldsymbol{H}_t^{\mathsf{H}} + \gamma_{\mathsf{w}}^{-1} \mathbf{I}_M \right)^{-1}.$$

Note how the time updated mean and precision in (3.20) coincide with the Kalman filter update step given in (2.7) by setting $\widehat{\boldsymbol{\Sigma}}_{t|t-1} = \gamma_{\boldsymbol{x}_{t|t-1}}^{-1} \mathbf{I}_N$ and $\boldsymbol{R}_t = \gamma_{\mathsf{w}}^{-1} \mathbf{I}_M$. This comes with no surprise since the Kalman filter in VAMP-KF is used under the scalar EP approximation. The time updated estimate in (3.20) can be rewritten as (see Appendix I)

$$\begin{aligned}
\widehat{\boldsymbol{x}}_{t|t} &= \left(\gamma_{\mathsf{w}} \boldsymbol{H}_t^{\mathsf{H}} \boldsymbol{H}_t + \gamma_{\boldsymbol{x}_{t|t-1}} \mathbf{I}_N \right)^{-1} \left(\gamma_{\mathsf{w}} \boldsymbol{H}_t^{\mathsf{H}} \boldsymbol{y}_t + \gamma_{\boldsymbol{x}_{t|t-1}} \widehat{\boldsymbol{x}}_{t|t-1} \right) \\
&\triangleq \boldsymbol{g}_{\mathbf{x}_{t|t}}(\widehat{\boldsymbol{x}}_{t|t-1}, \gamma_{\boldsymbol{x}_{t|t-1}}), \qquad (3.21a) \\
\gamma_{\boldsymbol{x}_{t|t}} &= \gamma_{\boldsymbol{x}_{t|t-1}} / \langle \boldsymbol{g}'_{\mathbf{x}_{t|t}}(\widehat{\boldsymbol{x}}_{t|t-1}, \gamma_{\boldsymbol{x}_{t|t-1}}) \rangle \\
&= N / \mathrm{Tr} \left[\left(\gamma_{\mathsf{w}} \boldsymbol{H}_t^{\mathsf{H}} \boldsymbol{H}_t + \gamma_{\boldsymbol{x}_{t|t-1}} \mathbf{I}_N \right)^{-1} \right]. \qquad (3.21b)
\end{aligned}$$

It worth mentioning that, under the scalar EP approximation, the time updated estimation in (3.21) has the same form as the LMMSE estimation steps in (3.11) of VAMP-KF as well as (2.14) of VAMP.

Similarly to the LMMSE step of VAMP in (2.14) and the prediction step in Section 3.2.1, we sidestep the high computational cost of the matrix inversion in (3.21) by relying on the SVD of $\boldsymbol{H}_t$, thereby leading to lines 17-19 in Algorithm 2.

3.3 Complexity

The computational cost per time step t of VAMP-KF is dominated by the SVD in line 2 in Algorithm 2 and is of order $\mathcal{O}(NMR)$ with $R = \min(N, M)$. This can be further decreased to $O(NM \log(R))$ by modern approaches [31]. The computation of

the extrinsic statistics from the posterior statistics only involves arithmetic operations which take $\mathcal{O}(1)$ per entry under the assumption of "bounded values" according to the uniform cost criterion [32]. Note that VAMP-KF involves scalar inverses only as shown in lines 4, 11, 16, and 17. It is also notable how the SVD offers more than just a fast algorithmic implementation. It connects VAMP to Kalman filter by using the same SVD in their LMMSE steps. Further, it removes the need for inversing the Kalman gain matrix in (2.8) due to the scalar EP approximation.

Chapter 4

State Evolution

To understand the behavior of the proposed VAMP-Kalman algorithm in the asymptotic regime, we derive its state evolution (SE) equations which recursively track its asymptotic empirical MSE. We first start by stating the necessary assumptions in connections to those of VAMP described in Section 2.3.2. We then provide close-form expressions of the MSE associated to the MMSE and LMMSE steps of VAMP-KF.

4.1 State evolution assumptions

At every time step $t \in [1, \dots, T]$, we denote the true vectors, which satisfy the state space model given in (1.1)–(1.2), as $\boldsymbol{v}_t^{(0)}$, $\boldsymbol{x}_t^{(0)}$, and $\boldsymbol{x}_{t-1}^{(0)}$, i.e.,

$$\boldsymbol{x}_t^{(0)} = \alpha \, \boldsymbol{F}_t \, \boldsymbol{x}_{t-1}^{(0)} + \sqrt{1-\alpha^2} \, \boldsymbol{v}_t^{(0)}, \tag{4.1a}$$

$$\boldsymbol{y}_t = \boldsymbol{H}_t \, \boldsymbol{x}_t^{(0)} + \boldsymbol{w}_t. \tag{4.1b}$$

We suppose that the true signal $\boldsymbol{x}_{t-1}^{(0)}$ in (4.1a) has already been estimated at time step $t-1$ with mean $\widehat{\boldsymbol{x}}_{t-1|t-1}$ and precision $\gamma_{\boldsymbol{x}_{t-1|t-1}}$. It is common practice in the approximate message passing literature to assume that the estimate $\widehat{\boldsymbol{x}}_{t-1|t-1}$ behaves

like an AWGN-corrupted version of the true signal $\boldsymbol{x}_{t-1}^{(0)}$ with precision $\gamma_{\boldsymbol{x}_{t-1|t-1}}^{-1}$. That is to say:

$$\widehat{\boldsymbol{x}}_{t-1|t-1} = \boldsymbol{x}_{t-1}^{(0)} + \boldsymbol{e}_{\boldsymbol{x}_{t-1|t-1}}. \tag{4.2}$$

with $\boldsymbol{e}_{\boldsymbol{x}_{t-1|t-1}} \sim \mathcal{CN}(\boldsymbol{e}_{\boldsymbol{x}_{t-1|t-1}}; \mathbf{0}, \gamma_{\boldsymbol{x}_{t-1|t-1}}^{-1} \mathbf{I}_N)$. By injecting (4.2) into (4.1) and after some algebraic manipulations we obtain

$$\underbrace{\boldsymbol{y}_t - \alpha\, \boldsymbol{H}_t\, \boldsymbol{F}_t\, \widehat{\boldsymbol{x}}_{t-1|t-1}}_{\boldsymbol{z}_t} = \underbrace{\sqrt{1-\alpha^2}\, \boldsymbol{H}_t}_{\boldsymbol{A}_t}\, \boldsymbol{v}_t^{(0)} + \underbrace{\alpha\, \boldsymbol{H}_t\, \boldsymbol{F}_t\, \boldsymbol{e}_{\boldsymbol{x}_{t-1|t-1}} + \boldsymbol{w}_t}_{\boldsymbol{p}_{\mathsf{z},t}}. \tag{4.3}$$

In (4.3), $\boldsymbol{p}_{\mathsf{z},t}$ is a realization of a Gaussian random variable with the density $\mathbf{p}_{\mathsf{z},t} \sim \mathcal{CN}\big(\boldsymbol{p}_{\mathsf{z},t}; \mathbf{0}, \alpha^2\, \gamma_{\boldsymbol{x}_{t-1|t-1}}^{-1}\, \boldsymbol{H}_t \boldsymbol{F}_t \boldsymbol{F}_t^{\mathsf{H}} \boldsymbol{H}_t^{\mathsf{H}} + \gamma_{\mathsf{w}}^{-1}\big)$. Using the scalar EP approximation, its covariance is reduced to an identity scaled by the the average of the diagonal elements $\gamma_{\mathsf{z},t}$, i.e., $\mathbf{p}_{\mathsf{z},t} \sim \mathcal{CN}(\boldsymbol{p}_{\mathsf{z},t}; \mathbf{0}, \gamma_{\mathsf{z},t}^{-1}\, \mathbf{I}_N)$ with

$$\gamma_{\mathsf{z},t} = \Big(\alpha^2\, \gamma_{\boldsymbol{x}_{t-1|t-1}}^{-1}\, \langle \mathrm{diag}\left(\boldsymbol{H}_t \boldsymbol{F}_t \boldsymbol{F}_t^{\mathsf{H}} \boldsymbol{H}_t^{\mathsf{H}}\right)\rangle + \gamma_{\mathsf{w}}^{-1}\Big)^{-1}. \tag{4.4}$$

Similarly to the assumptions presented in Section 2.3.2, $\boldsymbol{F}_t$, $\boldsymbol{H}_t$ and $\boldsymbol{H}_t\, \boldsymbol{F}_t$ are modeled as large right-rotationally invariant random matrices with full SVD of the forms:

$$\begin{aligned}
\boldsymbol{F}_t &= \widetilde{\boldsymbol{U}}_{\boldsymbol{F},t}\, \widetilde{\boldsymbol{S}}_{\boldsymbol{F},t}\, \widetilde{\boldsymbol{V}}_{\boldsymbol{F},t}^{\mathsf{H}}, && \mathrm{diag}\left(\widetilde{\boldsymbol{S}}_{\boldsymbol{F},t}\right) = \widetilde{\boldsymbol{s}}_{\boldsymbol{F},t}, \\
\boldsymbol{H}_t &= \widetilde{\boldsymbol{U}}_{\boldsymbol{H},t}\, \widetilde{\boldsymbol{S}}_{\boldsymbol{H},t}\, \widetilde{\boldsymbol{V}}_{\boldsymbol{H},t}^{\mathsf{H}}, && \mathrm{diag}\left(\widetilde{\boldsymbol{S}}_{\boldsymbol{H},t}\right) = \widetilde{\boldsymbol{s}}_{\boldsymbol{H},t}, \\
\boldsymbol{H}_t\, \boldsymbol{F}_t &= \widetilde{\boldsymbol{U}}_{\boldsymbol{HF},t}\, \widetilde{\boldsymbol{S}}_{\boldsymbol{HF},t}\, \widetilde{\boldsymbol{V}}_{\boldsymbol{HF},t}^{\mathsf{H}}, && \mathrm{diag}\left(\widetilde{\boldsymbol{S}}_{\boldsymbol{HF},t}\right) = \widetilde{\boldsymbol{s}}_{\boldsymbol{HF},t},
\end{aligned}$$

and are padded as in (2.17), thereby leading to the following padded SVDs:

$$\boldsymbol{U}_{\boldsymbol{F},t}\,\mathrm{Diag}\left(\boldsymbol{s}_{\boldsymbol{F},t}\right)\,\boldsymbol{V}_{\boldsymbol{F},t}^{\mathsf{H}}, \tag{4.6a}$$

$$\boldsymbol{U}_{\boldsymbol{H},t}\,\mathrm{Diag}\left(\boldsymbol{s}_{\boldsymbol{H},t}\right)\,\boldsymbol{V}_{\boldsymbol{H},t}^{\mathsf{H}}, \tag{4.6b}$$

$$\boldsymbol{U}_{\boldsymbol{HF},t}\,\mathrm{Diag}\left(\boldsymbol{s}_{\boldsymbol{HF},t}\right)\,\boldsymbol{V}_{\boldsymbol{HF},t}^{\mathsf{H}}. \tag{4.6c}$$

As a result, all the matrices involved in (4.6) are square matrices of the same size denoted as $N \times N$ in the rest of this section.

Besides, it is assumed that the components of the vectors of singular values $\boldsymbol{s}_{\boldsymbol{F},t}$, $\boldsymbol{s}_{\boldsymbol{H},t}$, $\boldsymbol{s}_{\boldsymbol{HF},t}$, the true signal $\boldsymbol{v}_t^{(0)}$, and the extrinsic mean $\widehat{\boldsymbol{v}}_{\mathsf{e},k}^{-}$ converge empirically with second-order moments as

$$\lim_{N\to\infty} \left\{s_{\boldsymbol{F},t,n}\right\}_{n=1}^{N} \overset{PL(2)}{=} \mathsf{s}_{\boldsymbol{F},t}, \tag{4.7a}$$

$$\lim_{N\to\infty} \left\{s_{\boldsymbol{H},t,n}\right\}_{n=1}^{N} \overset{PL(2)}{=} \mathsf{s}_{\boldsymbol{H},t}, \tag{4.7b}$$

$$\lim_{N\to\infty} \left\{s_{\boldsymbol{HF},t,n}\right\}_{n=1}^{N} \overset{PL(2)}{=} \mathsf{s}_{\boldsymbol{HF},t}, \tag{4.7c}$$

$$\lim_{N\to\infty} \left\{\left(v_{t,n}^{(0)}, \widehat{v}_{\mathsf{e},k,n}^{-}\right)\right\}_{n=1}^{N} \overset{PL(2)}{=} \left(\mathsf{v}_t^{(0)}, \widehat{\mathsf{v}}_{\mathsf{e},k}^{-}\right), \tag{4.7d}$$

for some random variables $\mathsf{s}_{\boldsymbol{F},t}$, $\mathsf{s}_{\boldsymbol{H},t}$, $\mathsf{s}_{\boldsymbol{HF},t}$, $\mathsf{v}_t^{(0)}$, and $\widehat{\mathsf{v}}_{\mathsf{e},k}^{-}$ with corresponding densities $p_{\mathsf{s}_{\boldsymbol{F},t}}(s_{\boldsymbol{F}})$, $p_{\mathsf{s}_{\boldsymbol{H},t}}(s_{\boldsymbol{H}})$, $p_{\mathsf{s}_{\boldsymbol{HF},t}}(s_{\boldsymbol{HF}})$, $p_{\mathsf{v}_t^{(0)}}(v^{(0)})$, and $p_{\widehat{\mathsf{v}}_{\mathsf{e},k}^{-}}(\widehat{v}_{\mathsf{e}}^{-})$.

Finally, the derivation of the state evolution equations at every iteration k and time step t is based on the following concentration of measure for the precision variables in

the asymptotic regime:

$$\lim_{N\to\infty}\left(\gamma_{\boldsymbol{v}_{\mathsf{p}}^{+},k},\gamma_{\boldsymbol{v}_{\mathsf{e}}^{+},k}\right)=\left(\overline{\gamma}_{\boldsymbol{v}_{\mathsf{p}}^{+},k},\overline{\gamma}_{\boldsymbol{v}_{\mathsf{e}}^{+},k}\right), \tag{4.8a}$$

$$\lim_{N\to\infty}\left(\gamma_{\boldsymbol{v}_{\mathsf{p}}^{-},k},\gamma_{\boldsymbol{v}_{\mathsf{e}}^{-},k}\right)=\left(\overline{\gamma}_{\boldsymbol{v}_{\mathsf{p}}^{-},k},\overline{\gamma}_{\boldsymbol{v}_{\mathsf{e}}^{-},k}\right), \tag{4.8b}$$

$$\lim_{N\to\infty}\left(\gamma_{\boldsymbol{x}_{t|t-1}},\gamma_{\boldsymbol{x}_{t|t}}\right)=\left(\overline{\gamma}_{\boldsymbol{x}_{t|t-1}},\overline{\gamma}_{\boldsymbol{x}_{t|t}}\right). \tag{4.8c}$$

4.2 State evolution equations

Based on the aforementioned assumptions in Section 4.1 and under the Bayes-optimal setting [1], the predicted MSE of VAMP-KF in Algorithm 2 is defined recursively in the asymptotic regime (i.e., as $N\to\infty$) through the state evolution (SE) equations presented in Algorithm 3.

4.2.1 Asymptotic MSE for the MMSE estimation step of $\boldsymbol{v}_t^+$

Due to the i.i.d. assumption on the components of the innovation vector $\boldsymbol{v}_t^+$, its MMSE estimate in lines 7-10 of Algorithm 2 is obtained by following the same derivation steps used in the VAMP algorithm and reviewed in (2.21) and (2.22). In other words, the MSE on $\boldsymbol{v}_t^+$ is given by the function

$$\begin{aligned}\mathcal{E}_{\mathsf{v}^{+},t}\left(\gamma_{\boldsymbol{v}_{\mathsf{e}}^{-}}\right)&=\mathbb{E}\left[\left|g_{\mathsf{v}^{+},t}\left(\widehat{\mathsf{v}}_{\mathsf{e}}^{-},\gamma_{\boldsymbol{v}_{\mathsf{e}}^{-}}\right)-\mathsf{v}_t^{(0)}\right|^2\,\middle|\,p_{\mathsf{v}_t^{(0)},\mathsf{p}^{-}}\left(v^{(0)},p^{-}\right)\right]\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left|g_{\mathsf{v}^{+},t}\left(v^{(0)}+p^{-},\gamma_{\boldsymbol{v}_{\mathsf{e}}^{-}}\right)-v^{(0)}\right|^2 p_{\mathsf{v}_t^{(0)}}\left(v^{(0)}\right)\mathcal{CN}\left(p^{-};0,\gamma_{\boldsymbol{v}_{\mathsf{e}}^{-}}^{-1}\right)\\&\qquad\times\,\mathrm{d}v^{(0)}\,\mathrm{d}p^{-},\end{aligned} \tag{4.9}$$

[1] This means that the noise precision γ_{w} is known, and the estimation errors of $\widehat{\boldsymbol{v}}_{\mathsf{e}}^{+}$, $\widehat{\boldsymbol{v}}_{\mathsf{e}}^{-}$, and $\widehat{\boldsymbol{x}}_{t|t}$ are equal to the predicted values $\overline{\gamma}_{\boldsymbol{v}_{\mathsf{p}}^{+}}^{-1}$, $\overline{\gamma}_{\boldsymbol{v}_{\mathsf{p}}^{-}}^{-1}$, and $\overline{\gamma}_{\boldsymbol{x}_{t|t}}^{-1}$, respectively.

Algorithm 3 VAMP-Kalman State Evolution

Require : Noise precision γ_{w}, precision $\overline{\gamma}_{0|0}$, error functions $\mathcal{E}_{\mathsf{v}^+,t}\big(\gamma_{\boldsymbol{v}_{\mathsf{e}}^-}\big)$, $\mathcal{E}_{\mathsf{v}^-,t}\big(\gamma_{\boldsymbol{v}_{\mathsf{e}}^+},\gamma_{\mathsf{z}}\big)$, $\mathcal{E}_{\mathsf{x}_{t|t-1}}\big(\gamma_{\boldsymbol{x}_{t-1|t-1}},\gamma_{\boldsymbol{v}_{\mathsf{e}}+}\big)$, and $\mathcal{E}_{\mathsf{x}_{t|t}}\big(\gamma_{\boldsymbol{x}_{t|t-1}}\big)$, and the number of iterations K.

1: **for** $t = 1, \ldots, T$ **do**

2: Initialize $\overline{\gamma}_{\boldsymbol{v}_{\mathsf{e}}^-,0}$

3: Compute $\overline{\gamma}_{\mathsf{z},t} = \frac{1}{\alpha^2\,\overline{\gamma}_{\boldsymbol{x}_{t-1|t-1}}^{-1}\,\mathbb{E}\left[|\mathsf{s}_{\boldsymbol{HF},t}|^2\middle|p_{\mathsf{s}_{\boldsymbol{HF},t}}(s_{\boldsymbol{HF}})\right]+\gamma_{\mathsf{w}}^{-1}}$

4: **for** $k = 1, \ldots, K$ **do**

▷ compute the analytical posterior and extrinsic precision of v^+

5: $\overline{\gamma}_{\boldsymbol{v}_{\mathsf{p}}^+,k} = \frac{1}{\mathcal{E}_{\mathsf{v}^+,t}\left(\overline{\gamma}_{\boldsymbol{v}_{\mathsf{e}}^-,k}\right)}$

6: $\overline{\gamma}_{\boldsymbol{v}_{\mathsf{e}}^+,k} \;=\; \overline{\gamma}_{\boldsymbol{v}_{\mathsf{p}}^+,k} \;-\; \overline{\gamma}_{\boldsymbol{v}_{\mathsf{e}}^-,k}$

▷ compute the analytical posterior and extrinsic precision of v^-

7: $\overline{\gamma}_{\boldsymbol{v}_{\mathsf{p}}^-,k} = \frac{1}{\mathcal{E}_{\mathsf{v}^-,t}\left(\overline{\gamma}_{\boldsymbol{v}_{\mathsf{e}}^+,k},\overline{\gamma}_{\mathsf{z},t}\right)}$

8: $\overline{\gamma}_{\boldsymbol{v}_{\mathsf{e}}^-,k+1} \;=\; \overline{\gamma}_{\boldsymbol{v}_{\mathsf{p}}^-,k} \;-\; \overline{\gamma}_{\boldsymbol{v}_{\mathsf{e}}^+,k}$

9: **end for**

▷ compute the analytical precision of $\mathsf{x}_{t|t-1}$

10: $\overline{\gamma}_{\boldsymbol{x}_{t|t-1}} = \frac{1}{\mathcal{E}_{\mathsf{x}_{t|t-1}}\left(\overline{\gamma}_{\boldsymbol{x}_{t-1|t-1}},\overline{\gamma}_{\boldsymbol{v}_{\mathsf{e}}+,K}\right)}$

▷ compute the analytical precision of $\mathsf{x}_{t|t}$

11: $\overline{\gamma}_{\boldsymbol{x}_{t|t}} = \frac{1}{\mathcal{E}_{\mathsf{x}_{t|t}}\left(\overline{\gamma}_{\boldsymbol{x}_{t|t-1}}\right)}$

12: **end for**

13: **return** $\left\{\overline{\gamma}_{\boldsymbol{x}_{1|1}}, \ldots, \overline{\gamma}_{\boldsymbol{x}_{T|T}}\right\}$

which is involved in line 5 of Algorithm 3. In (4.9), $g_{\mathsf{v}^+,t}\big(\widehat{\mathsf{v}}_{\mathsf{e}}^-,\gamma_{\boldsymbol{v}_{\mathsf{e}}^-}\big)$ represents the component-wise denoiser of the separable MMSE vector denoiser $\boldsymbol{g}_{\mathsf{v}^+,t}\big(\widehat{\boldsymbol{\mathsf{v}}}_{\mathsf{e}}^-,\gamma_{\boldsymbol{v}_{\mathsf{e}}^-}\big)$ as in line 7 of Algorithm 2, given the random variable $\widehat{\mathsf{v}}_{\mathsf{e}}^- = \mathsf{v}_t^{(0)} + \mathsf{p}^-$ with $\mathsf{p}^- \sim \mathcal{CN}\big(p^-;0,\gamma_{\boldsymbol{v}_{\mathsf{e}}^-}^{-1}\big)$ being independent of $\mathsf{v}_t^{(0)}$.

4.2.2 Asymptotic MSE for the LMMSE estimation step of $\boldsymbol{v}_t^-$

The LMMSE estimate of $\boldsymbol{v}_t^-$ in VAMP-KF is given in lines 11-13 of Algorithm 2. There, the LMMSE denoiser $\boldsymbol{g}_{\mathsf{v}^-,t}\big(\widehat{\boldsymbol{v}}_{\mathsf{e}}^+, \gamma_{\boldsymbol{v}_{\mathsf{e}}^+}, \boldsymbol{z}, \gamma_{\mathsf{z}}\big)$ given in (3.11a) has a similar expression as the LMMSE denoiser $\boldsymbol{g}_{\mathsf{x}^-}\big(\widehat{\boldsymbol{x}}_{\mathsf{e},k}^+, \gamma_{\boldsymbol{x}_{\mathsf{e}}^+,k}\big)$ of VAMP given in (2.14a). As a result, similarly to (2.25), in the large system limit the component-wise MSE of the LMMSE step is given by

$$\mathcal{E}_{\mathsf{v}^-,t}\big(\gamma_{\boldsymbol{v}_{\mathsf{e}}^+}, \gamma_{\mathsf{z}}\big) = \mathbb{E}\left[\frac{1}{\gamma_{\mathsf{z}}\,(1-\alpha^2)\,|\mathsf{s}_{\boldsymbol{F},t}|^2 + \gamma_{\boldsymbol{v}_{\mathsf{e}}^+}}\middle| p_{\mathsf{s}_{\boldsymbol{F},t}}(s_{\boldsymbol{F}})\right], \tag{4.10}$$

which is used in line 7 of Algorithm 3. In (4.10), the precision γ_{z} defined in (4.4) converges in the asymptotic regime to the variable $\overline{\gamma}_{\mathsf{z},t}$ given by

$$\begin{aligned}
\overline{\gamma}_{\mathsf{z},t} &= \lim_{N\to\infty} \gamma_{\mathsf{z},\mathsf{t}} \\
&= \lim_{N\to\infty} \frac{1}{\alpha^2\,\gamma_{\boldsymbol{x}_{t-1|t-1}}^{-1}\,\frac{1}{N}\mathrm{Tr}\left[\boldsymbol{H}_t\,\boldsymbol{F}_t\,\boldsymbol{F}_t^{\mathsf{H}}\,\boldsymbol{H}_t^{\mathsf{H}}\right] + \gamma_{\mathsf{w}}^{-1}} \\
&\overset{\text{(a)}}{=} \lim_{N\to\infty} \frac{1}{\alpha^2\,\gamma_{\boldsymbol{x}_{t-1|t-1}}^{-1}\,\frac{1}{N}\mathrm{Tr}\left[\boldsymbol{S}_{\boldsymbol{HF},t}\,\boldsymbol{S}_{\boldsymbol{HF},t}^{\mathsf{H}}\right] + \gamma_{\mathsf{w}}^{-1}} \\
&\overset{\text{(b)}}{=} \frac{1}{\alpha^2\,\overline{\gamma}_{\boldsymbol{x}_{t-1|t-1}}^{-1}\,\mathbb{E}\big[|\mathsf{s}_{\boldsymbol{HF},t}|^2 \,\big|\, p_{\mathsf{s}_{\boldsymbol{HF},t}}(s_{\boldsymbol{HF}})\big] + \gamma_{\mathsf{w}}^{-1}}.
\end{aligned} \tag{4.11}$$

In (4.11), (a) follows from the SVD in (4.6), while (b) results from the empirical convergence given in (4.7).

4.2.3 Asymptotic MSE for the VAMP-KF prediction step

Here, we consider the MSE pertaining to the prediction step of VAMP-Kalman stated in lines 15-16 of Algorithm 2. We first use the state-space model in (1.1) and (1.2) as

well as the prediction step of VAMP-KF in (3.17) to write $\widehat{\boldsymbol{x}}_{t|t-1}$ and $\boldsymbol{x}_t^{(0)}$ as follows:

$$\begin{aligned}\widehat{\boldsymbol{x}}_{t|t-1} &= \alpha\, \boldsymbol{F}_t\, \widehat{\boldsymbol{x}}_{t-1|t-1} + \sqrt{1-\alpha^2}\; \widehat{\boldsymbol{v}}_{\mathsf{e}}^{+},\\ &= \boldsymbol{g}_{\mathsf{x}_{t|t-1}}(\widehat{\boldsymbol{x}}_{t-1|t-1}, \widehat{\boldsymbol{v}}_{\mathsf{e}}^{+}) \qquad (4.12a)\\ \boldsymbol{x}_t^{(0)} &= \alpha\, \boldsymbol{F}_t\, \boldsymbol{x}_{t-1}^{(0)} + \sqrt{1-\alpha^2}\, \boldsymbol{v}_t^{(0)}. \qquad (4.12b)\end{aligned}$$

In the large system limit, we consider the MSE between the predicted signal $\widehat{\mathsf{x}}_{t|t-1}$ provided by the denoiser $\boldsymbol{g}_{\mathsf{x}_{t|t-1}}(\widehat{\mathsf{x}}_{t-1|t-1}, \widehat{\mathsf{v}}_{\mathsf{e}}^{+})$ given in (3.17a), and the true vector $\boldsymbol{x}_t^{(0)}$. That is to say

$$\mathcal{E}_{\mathsf{x}_{t|t-1}}\big(\gamma_{\boldsymbol{x}_{t-1|t-1}}, \gamma_{\boldsymbol{v}_{\mathsf{e}^+}}\big) = \lim_{N\to\infty} \frac{1}{N}\, \mathbb{E}\Big[\big\| \boldsymbol{g}_{\mathsf{x}_{t|t-1}}(\widehat{\mathsf{x}}_{t-1|t-1}, \widehat{\mathsf{v}}_{\mathsf{e}}^{+}) - \boldsymbol{x}_t^{(0)} \big\|^2 \,\Big|\, p_{\mathsf{p}^+,\mathsf{q}}(\boldsymbol{p}^+, \boldsymbol{q}) \Big], \quad (4.13)$$

where the random vectors $\widehat{\mathsf{v}}_{\mathsf{e}}^{+}$ and $\widehat{\mathsf{x}}_{t-1|t-1}$ are given by

$$\begin{aligned}\widehat{\mathsf{v}}_{\mathsf{e}}^{+} &= \boldsymbol{v}_t^{(0)} + \mathsf{p}^+, \qquad\qquad \mathsf{p}^+ \sim \mathcal{CN}(\boldsymbol{p}^+; \boldsymbol{0}, \gamma_{\boldsymbol{v}_{\mathsf{e}}^+}^{-1}\, \mathbf{I}_N),\\ \widehat{\mathsf{x}}_{t-1|t-1} &= \boldsymbol{x}_{t-1}^{(0)} + \mathsf{q}, \quad \mathsf{q} \sim \mathcal{CN}(\boldsymbol{q}; \boldsymbol{0}, \gamma_{\boldsymbol{x}_{t-1|t-1}}^{-1}\, \mathbf{I}_N),\end{aligned}$$

with p^+ and q being independent. By injecting (4.12) into (4.13), it follows:

$$\begin{aligned}&\mathcal{E}_{\mathsf{x}_{t|t-1}}\big(\gamma_{\boldsymbol{x}_{t-1|t-1}}, \gamma_{\boldsymbol{v}_{\mathsf{e}^+}}\big)\\ &\quad= \lim_{N\to\infty} \frac{1}{N}\, \mathbb{E}\Big[\big\| \alpha\, \boldsymbol{F}_t\, \mathsf{q} + \sqrt{1-\alpha^2}\; \mathsf{p}^+ \big\|^2 \,\Big|\, p_{\mathsf{p}^+,\mathsf{q}}(\boldsymbol{p}^+, \boldsymbol{q}) \Big]\\ &\quad\overset{(c)}{=} \lim_{N\to\infty} \frac{1}{N}\, \mathbb{E}\Big[\big\| \alpha\, \boldsymbol{F}_t\, \mathsf{q} \big\|^2 \,\Big|\, p_{\mathsf{q}}(\boldsymbol{q}) \Big] + \lim_{N\to\infty} \frac{1}{N}\, \mathbb{E}\Big[\big\| \sqrt{1-\alpha^2}\; \mathsf{p}^+ \big\|^2 \,\Big|\, p_{\mathsf{p}^+}(\boldsymbol{p}^+) \Big]\\ &\quad\overset{(d)}{=} \lim_{N\to\infty} \frac{\alpha^2}{N}\, \gamma_{\boldsymbol{x}_{t-1|t-1}}^{-1}\, \mathrm{Tr}\big[\boldsymbol{F}_t\, \boldsymbol{F}_t^{\mathsf{H}} \big] + (1-\alpha^2)\, \gamma_{\boldsymbol{v}_{\mathsf{e}}^+}^{-1}\\ &\quad\overset{(e)}{=} \lim_{N\to\infty} \frac{\alpha^2}{N}\, \gamma_{\boldsymbol{x}_{t-1|t-1}}^{-1}\, \mathrm{Tr}\big[\boldsymbol{S}_{\boldsymbol{F},t}\, \boldsymbol{S}_{\boldsymbol{F},t}^{\mathsf{H}} \big] + (1-\alpha^2)\, \gamma_{\boldsymbol{v}_{\mathsf{e}}^+}^{-1}\\ &\quad\overset{(f)}{=} \alpha^2\, \gamma_{\boldsymbol{x}_{t-1|t-1}}^{-1}\, \mathbb{E}\big[|\mathsf{s}_{\boldsymbol{F},t}|^2 \,\big|\, p_{\mathsf{s}_{\boldsymbol{F},t}}(s_{\boldsymbol{F}}) \big] + (1-\alpha^2)\, \gamma_{\boldsymbol{v}_{\mathsf{e}}^+}^{-1}. \qquad (4.15)\end{aligned}$$

In (4.15), (c) follows from the fact that both p^+ and q are zero-mean and independent; (e) results from the use of the SVD in (4.6); (f) stems from the empirical convergence stated in (4.7). The MSE function established in (4.15) is in line 10 of Algorithm 3.

4.2.4 Asymptotic MSE for the VAMP-KF time update step

The time update step in lines 17-19 of Algorithm 2 corresponds to the LMMSE estimation of $\boldsymbol{x}_{t|t}$ from $\boldsymbol{x}_{t|t-1}$. For this reason, its asymptotic MSE induced by the denoiser $\boldsymbol{g}_{\mathsf{x}_{t|t}}(\cdot,\cdot)$ in (3.21) has the same form as the LMMSE denoiser $\boldsymbol{g}_{\mathsf{x}^-}\big(\widehat{\boldsymbol{x}}^+_{\mathsf{e}}, \gamma_{\boldsymbol{x}^+_{\mathsf{e}}}\big)$ of VAMP given in (2.14). Similarly to the asymptotic MSE in (2.25), the MSE of the time update step computed in line 11 of Algorithm 3 is given by

$$\mathcal{E}_{\mathsf{x}_{t|t}}\big(\gamma_{\boldsymbol{x}_{t|t-1}}\big) = \mathbb{E}\left[\frac{1}{\gamma_{\mathsf{w}}\,|\mathsf{s}_{\boldsymbol{H},t}|^2 + \gamma_{\boldsymbol{x}_{t|t-1}}}\,\middle|\, p_{\mathsf{s}_{\boldsymbol{H},t}}(s_{\boldsymbol{H}})\right]. \tag{4.16}$$

4.3 Pseudo-code

Under all the aforementioned assumptions and derived MSEs, we now describe in Algorithm 2 our main result, which is the recursive update equations for the SE of VAMP-KF. Note here that at convergence one must have equality between the posterior variances $\overline{\gamma}_{\boldsymbol{v}^+_{\mathsf{p}},\infty}$ and $\overline{\gamma}_{\boldsymbol{v}^-_{\mathsf{p}},\infty}$ of the innovation variables in lines 5 and 7, respectively, i.e., $\overline{\gamma}^{-1}_{\boldsymbol{v}^+_{\mathsf{p}},\infty} = \overline{\gamma}^{-1}_{\boldsymbol{v}^-_{\mathsf{p}},\infty}$.

At each time step t, we run the state evolution of VAMP stated in lines 5–8 for K iterations to obtain the MSE of the innovation estimation, i.e. $\overline{\gamma}^{-1}_{\boldsymbol{v}^+_{\mathsf{p}},K} = \overline{\gamma}^{-1}_{\boldsymbol{v}^-_{\mathsf{p}},K}$. We then use the latter to compute the MSE of the prediction and estimation, $\overline{\gamma}_{\boldsymbol{x}_{t|t-1}}$ and $\overline{\gamma}_{\boldsymbol{x}_{t|t}}$, in lines 10 and 11, respectively. Finally the MSE of the time update step $\overline{\gamma}^{-1}_{\boldsymbol{x}_{t|t}}$ is used in the next time step $t+1$ to compute the precision $\boldsymbol{z}_t$, $\overline{\gamma}_{\mathsf{z},t}$ of the variable in line 3 before running VAMP SE.

Chapter 5

Numerical Results

In this section, we assess the performance behavior of the proposed VAMP-KF algorithm and benchmark it against the standard Kalman filter [2], as well as EnKF [16] and MCC-KF [13] that were developed specifically to cope with non-Gaussian innovations. In all simulations, we fix the state transition matrix $\boldsymbol{F}_t$ and the measurement matrix $\boldsymbol{H}_t$ during the entire simulation, i.e., for all $t \in \{1, \ldots, T\}$. We also set the number of time steps to $T = 20$, the number of VAMP iterations to $K = 500$, and perform $N_{\mathrm{MC}} = 100$ Monte-Carlo trials for different values of the SNR:

$$\mathrm{SNR} \;=\; 10 \, \log_{10}\!\left(\frac{\|\boldsymbol{H}_t \, \boldsymbol{x}_t\|_2^2}{\|\boldsymbol{w}\|_2^2} \right). \tag{5.1}$$

The performance metric considered throughout our tests is the time-averaged normalized root MSE (TNRMSE) between the true signals $\{\boldsymbol{x}_t\}_{t=1}^{t=T}$ and their estimates $\{\widehat{\boldsymbol{x}}_t\}_{t=1}^{t=T}$. The TNRMSE metric is defined as follows:

$$\mathrm{TNRMSE}\Big(\{\boldsymbol{x}_t\}_{t=1}^{t=T}, \{\widehat{\boldsymbol{x}}_t\}_{t=1}^{t=T} \Big) \triangleq \frac{1}{T} \sum_{t=1}^{T} \sqrt{\frac{\|\boldsymbol{x}_t - \widehat{\boldsymbol{x}}_t\|_2^2}{\|\boldsymbol{x}_t\|_2^2}}.$$

We examine the TNRMSE by varying one of the variables of the state-space model given in (1.1) and (1.2) at a time:

- the SNR in (5.1) by appropriately varying the precision γ_{w} of the measurement noise $\boldsymbol{w}_t$.

- the sparsity rate ρ (i.e., the percentage of the non-zero components) of the innovation vector $\boldsymbol{v}_t$. Specifically, we consider innovation vectors as realizations generated from a Bernoulli-Gaussian density:

$$p_{\mathbf{v}_t}(\boldsymbol{v}_t) = \rho\,\delta(\boldsymbol{v}_t) + (1-\rho)\,\mathcal{CN}\left(\boldsymbol{v}_t; \mathbf{0}, \mathbf{I}_N\right). \tag{5.2}$$

- the memory factor α of the process equation given in (1.2) which controls the amount by which the time evolution of the signal $\boldsymbol{x}_{t-1}$ contributes to the signal $\boldsymbol{x}_t$.

- the system dimension determined as the ratio M/N between the measurement dimension M and the signal dimension N. The measurement equation given in (1.1) can be underdetermined or overdetermined, i.e., $M < N$ or $M > N$, respectively. Note that the measurement equation remains solvable even when it is underdetermined due to the sparsity of $\boldsymbol{x}_t$. The latter can be controlled by properly adjusting both the memory factor α and the sparsity rate ρ of the innovation vector $\boldsymbol{v}_t$.

5.1 Benchmark Algorithms

In this section, we describe a number of existing signal reconstruction approaches that we will benchmark against VAMP-KF.

- *Standard Kalman filter [2]*: the standard Kalman filter (standard KF) is the optimal LMMSE estimator under time-varying Gaussian innovations. While our experiments

center around non-Gaussian innovations, we kept the standard KF as a mismatched benchmark for convenience reasons which will become apparent later in this section.

- *EnKF [16]*: the ensemble KF (EnKF) is a Monte Carlo variant of the standard KF that exploits the higher-order statistics of $\boldsymbol{x}_t$. This is to be opposed to the standard KF which uses second-order signal statistics only under non-Gaussian innovations. By exploiting the correlation over time between the observation noise $\boldsymbol{w}_t$ and the signal estimate, EnKF approximates the signal using a finite set of randomly selected samples. However, EnKF suffers from high computational cost that is inherent to most of the statistical sampling methods, thereby making it unpractical for real-time and/or mission-critical applications.

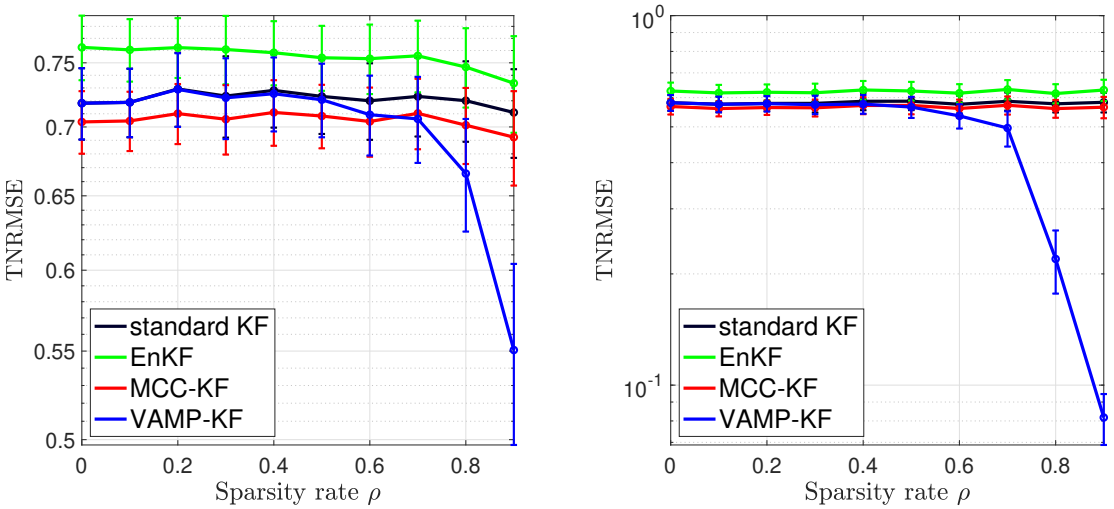
- *MCC-KF [13]*: the maximum correntropy criterion KF (MCC-KF) is a gradient-based optimization method that improves the robustness of the standard KF against non-Gaussian innovations in linear dynamical systems. Unlike the standard KF that minimizes the MSE, MCC-KF maximizes the correntropy as the optimality criterion. By doing so, it also exploits the higher-order statistics of $\boldsymbol{x}_t$ as opposed to the standard KF which uses second-order signal statistics only under non-Gaussian innovations.

Since the proposed VAMP-KF algorithm is geared towards uncorrelated innovation processes, we do not benchmark with the Bayesian estimation methods that cope with correlated sparse innovations presented in Fig. 1.2. Because these methods focus on correlated innovations over time by restricting the innovation process to have a fixed innovation support and (possibly) a slow time-varying non-zero coefficients, they are inappropriate benchmarks unless VAMP-KF algorithm incorporates the correlation of innovations as a side information in the derivation of its messages. Such scenario will be pursued in a future work.

5.2 Performance versus sparsity rate ρ

We examine the performance of VAMP-KF as a function of the sparsity rate ρ in the Bernoulli-Gaussian prior of the innovation vector given in (5.2). We vary the sparsity rate ρ from 0 to 0.9 and set the memory factor α to 0.7.

Figs. 5.1a-5.1b show the TNRSME as a function of the sparsity rate ρ for the SNR values 5 dB and 20 dB, respectively, with $M = 100$ and $N = 200$. There, it is seen that at lower sparsity rate values (i.e., innovations are almost Gaussian), the performance of VAMP-KF coincides with that of the standard KF. Specifically, when the time-varying innovation vectors are exactly Gaussian (i.e., $\rho = 0$), VAMP-KF behaves identically as the standard Kalman filter. However, as we increase the sparsity rate ρ, the innovation vectors deviate from the Gaussian case and a significant performance improvement of VAMP-KF is observed as compared to other algorithms.

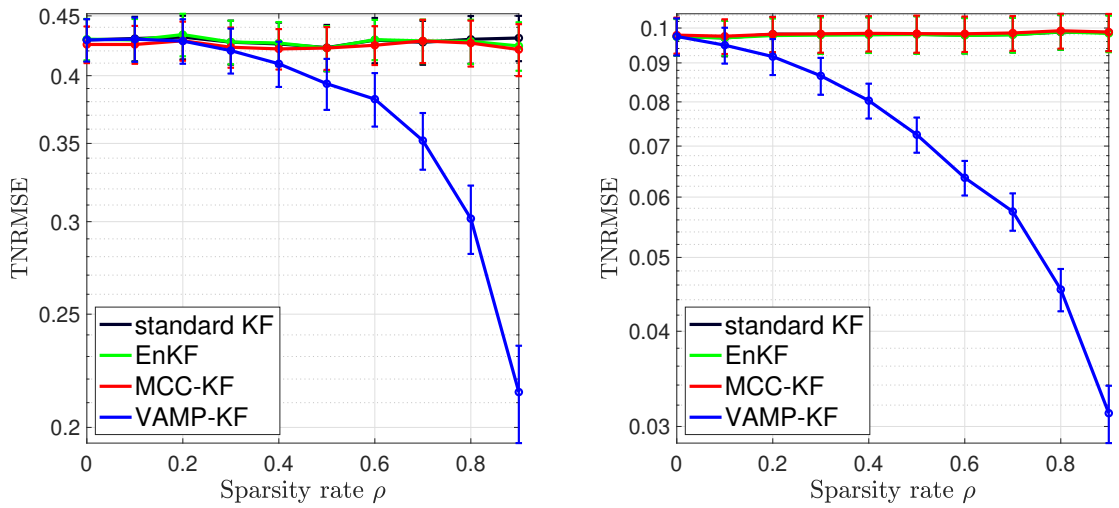


(a) SNR = 5 [dB]

(b) SNR = 20 [dB]

Figure 5.1: TNRMSE of VAMP-KF, standard KF, EnKF, and MCC-KF vs. the sparsity rate ρ of the Bernoulli-Gaussian prior of the innovation with $N = 200$, $M = 100$, and $\alpha = 0.7$ at (a) SNR = 5 [dB] and (b) SNR = 20 [dB].

Similar results can also be observed in Figs. 5.2a-5.2b, where the same experiment is repeated for an overdetermined measurement system, i.e., $M > N$.



(a) SNR = 5 [dB] (b) SNR = 20 [dB]

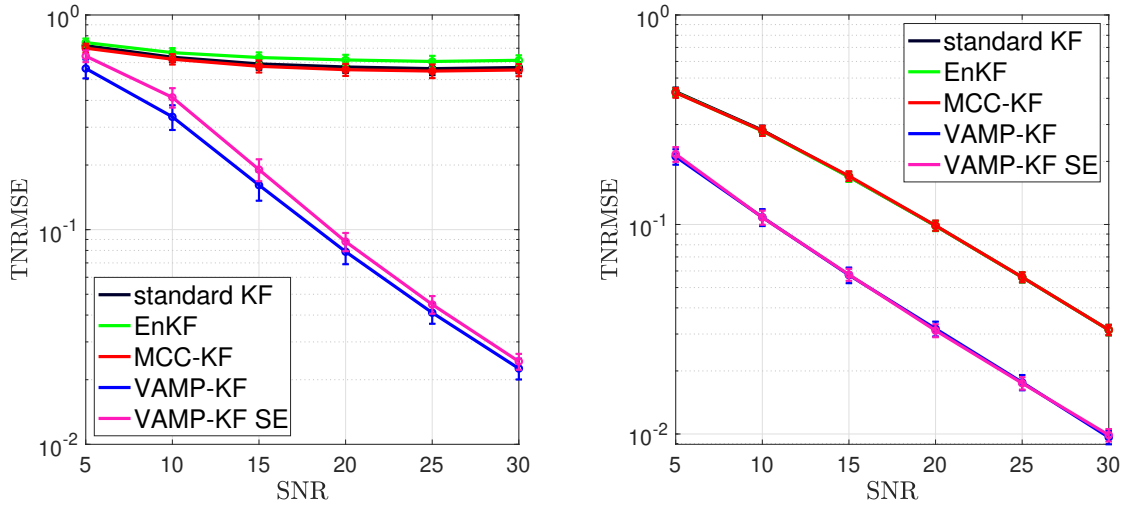
Figure 5.2: TNRMSE of VAMP-KF, standard KF, EnKF, and MCC-KF vs. the sparsity rate ρ of the Bernoulli-Gaussian prior of the innovation with $N = 100$, $M = 200$, and $\alpha = 0.7$ at (a) SNR = 5 [dB] and (b) SNR = 20 [dB].

5.3 Performance Versus SNR

We now investigate how VAMP-KF performs at SNR levels between 5 dB and 30 dB. The memory factor α and the sparsity rate ρ are fixed to 0.7 and 0.9, respectively.

The results are presented in Figs. 5.3a and 5.3b for the underdetermined and overdetermined cases, respectively. There, it is observed that VAMP-KF outperforms all benchmark algorithms at every SNR value. Moreover, it is seen that the empirical TNRMSE of VAMP-KF is accurately predicted by the analytical state evolution recursion thereby corroborating the theoretical analysis we conducted in Chapter 4. This endows the proposed algorithm with offline design guidelines when applied to different

engineering problems in practice.



(a) $N = 200$, $M = 100$ (b) $N = 100$, $M = 200$

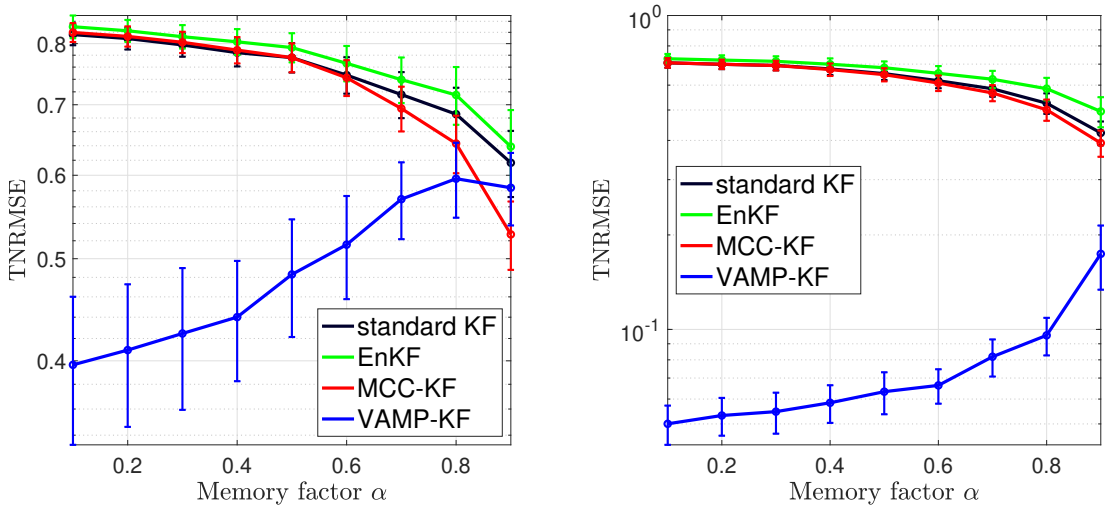
Figure 5.3: TNRMSE of VAMP-KF, standard KF, EnKF, and MCC-KF vs. the SNR with $\alpha = 0.7$, and sparsity rate $\rho = 0.9$ of the Bernoulli-Gaussian prior of the innovation when (a) $N = 200$, $M = 100$ and (b) $N = 100$, $M = 200$.

The small gap between the empirical TNRMSE of VAMP-KF and the one predicted by its state evolution analysis, at lower SNR values, is due to two reasons. The first one is the approximation of the state evolution equation using Monte-Carlo averaging because the integral in (4.9) under the Bernoulli-Gaussian prior in (5.2) cannot be evaluated by closed-form expressions. The second reason is due to the finite size effects as confirmed in [25] in the low-SNR regime.

5.4 Performance Versus memory factor α

To study how VAMP-KF performs as a function of the memory factor α, we vary the latter from 0.1 to 0.9 and fix the sparsity rate ρ at 0.9. We also consider low-SNR (i.e. 5 dB) and high-SNR (i.e., 20 dB) regimes as well as underdetermined and overdetermined

measurement systems. The results are presented in Figs. 5.4a-5.5b.



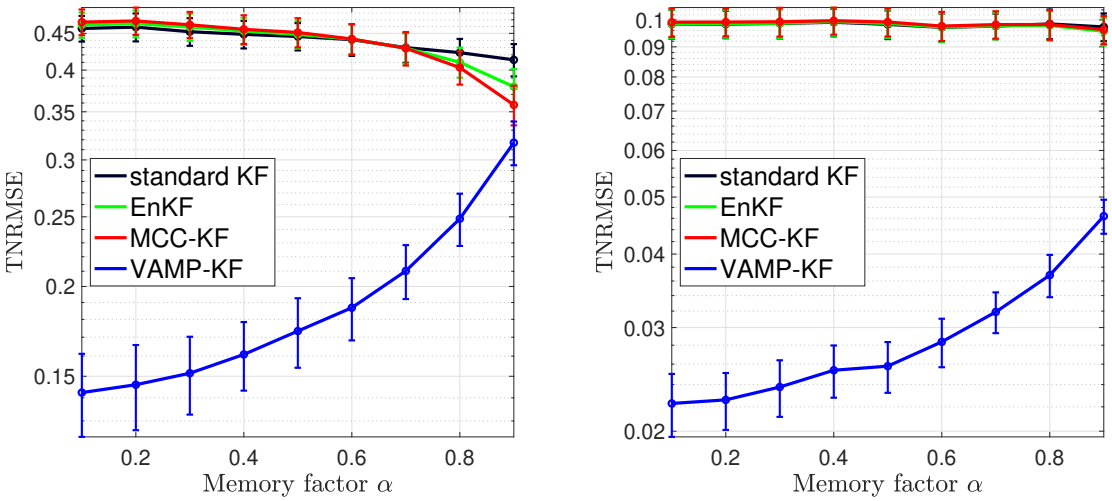
(a) SNR = 5 [dB]

(b) SNR = 20 [dB]

Figure 5.4: TNRMSE of VAMP-KF, standard KF, EnKF, and MCC-KF vs. the memory factor α with $N = 200$, $M = 100$, and sparsity rate $\rho = 0.9$ of the Bernoulli-Gaussian prior of the innovation at (a) SNR = 5 [dB] and (b) SNR = 20 [dB].

For the underdetermined case, Figs. 5.4a-5.4b show the performance for low and high SNR values, respectively. Given SNR = 5dB in Fig. 5.4a, for high values of α, the temporal evolution of the system signal $\boldsymbol{x}_t$ is heavily determined by the previous signal $\boldsymbol{x}_{t-1}$ and has little dependency on the innovation $\boldsymbol{v_t}$. In this case, all algorithms exhibit comparable performance. As α becomes smaller, the evolution of the signal $\boldsymbol{x}_t$ mainly depends on the innovation $\boldsymbol{v}_t$ and hence the sparsity of the innovation can be better exploited. In this scenario, VAMP-KF outperforms by far En-KF and MCC-KF which are deemed to be able to deal with non-Gaussian innovation processes in the state-space model. This observation is justified by the fact that the Bayesian estimation of the innovation by VAMP-KF is tailored to handle sparse estimation problems, which are known to be highly non-convex problems. Given SNR = 20dB in Fig. 5.4b, similar results can be observed except that VAMP-KF outperforms all the other algorithms for

the whole range of α. For the overdetermined case in Figs. 5.5a-5.5b, similar results are obtained, thereby confirming the superiority of VAMP-KF over existing state-of-the-art algorithms.



(a) SNR = 5 [dB]

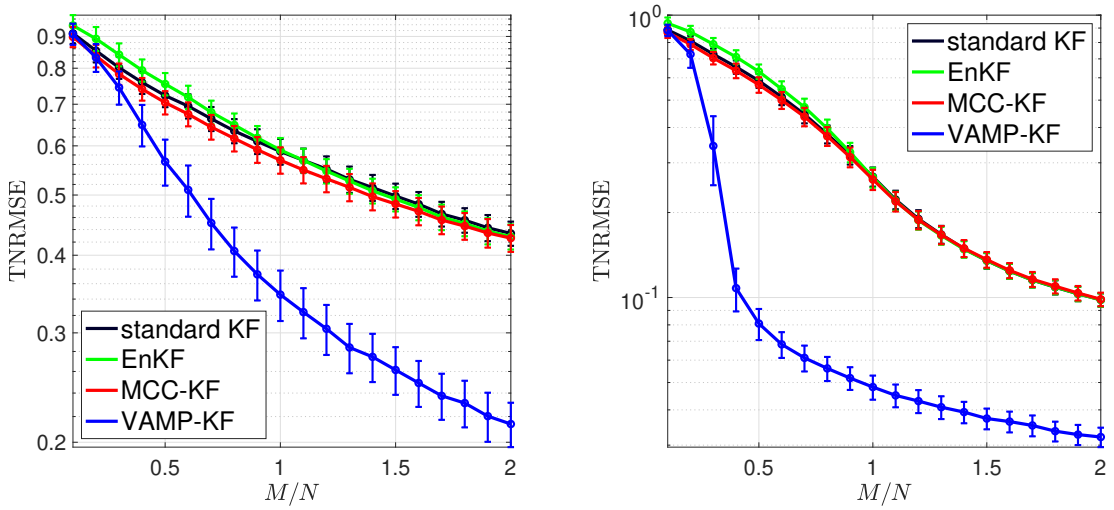
(b) SNR = 20 [dB]

Figure 5.5: TNRMSE of VAMP-KF, standard KF, EnKF, and MCC-KF vs. the memory factor α with $N = 100$, $M = 200$, and sparsity rate $\rho = 0.9$ of the Bernoulli-Gaussian prior of the innovation at (a) SNR = 5 [dB] and (b) SNR = 20 [dB].

5.5 Performance Versus sampling rate M/N

We finally investigate how the ratio of the number of measurement over the signal dimension, namely the sampling rate M/N, affects the performance of VAMP-KF. For this reason, we sweep the sampling rate M/N from 0.1 to 2 with a step size of 0.1. We set the memory factor α and the sparsity rate ρ to 0.7 and 0.9, respectively. Figs. 5.6a and 5.6b illustrate the performance for low and high SNR values, respectively. Starting from $M/N = 0.1$, when little measurement is provided to recover the signal, all algorithms exhibit the same performance. As M/N becomes larger and hence more measurements

are provided, VAMP-KF shows a significant advantage in the TNRMSE performance. While VAMP-KF outperforms EnKF and MCC-KF under both low-SNR and high-SNR regimes, the gap between them is not inherently related to the SNR value, but rather to the inability of the EnKF and MCC-KF to solve highly non-convex estimation problems with gradient-based optimization or sampling strategies, respectively. This is to be opposed to VAMP-KF which exploits the prior knowledge about the innovation process by employing VAMP for estimation under sparsity constraints.



(a) SNR = 5 [dB]

(b) SNR = 20 [dB]

Figure 5.6: TNRMSE of VAMP-KF, standard KF, EnKF, and MCC-KF vs. the sampling rate M/N with $\alpha = 0.7$, and sparsity rate $\rho = 0.9$ of the Bernoulli-Gaussian prior of the innovation at (a) SNR = 5 [dB] and (b) SNR = 20 [dB].

Chapter 6

Conclusion and Future Directions

In this work, we introduced a new algorithm, dubbed VAMP-KF, to estimate a time varying signal under rapidly time-varying unstructured innovation dynamics based on the approximate message passing paradigm. We do so by combining the VAMP and the standard Kalman filter algorithms in a unified Bayesian framework. After recalling the factor graph of each one of them, we determined the messages to be established to merge them into a fully message passing approach that incorporates the separable prior of the innovation within the estimation procedure. Our numerical results demonstrated that VAMP-KF exhibits the best reconstruction performance under non Gaussian innovation processes as compared to existing state-of-the-art algorithms. Additionally, we derived the state evolution update equations of VAMP-KF which theoretically predict its mean-square error.

This work can be extended towards many research directions to either improve the performance or cope with a wider range of problems, including but not limited to the following:

- As illustrated in Chapter 2, the standard Kalman filter that we incorporate is a forward-only algorithm, in which messages only propagate to the "future" with

respect to time t. However, if the estimation result is not required instantly at each time step t, efforts can be dedicated to propagate messages back to the "past" so as to further diminish the estimation error.

- The state-space model given in (1.1)–(1.2) is linear. Extending the proposed algorithm to handle non-linearity is an important step. In this case, we would consider the benchmark algorithms that handle non-linear cases such as the extended Kalman filter (EKF) [33].

- As discussed in Chapter 5, when the innovations are correlated with a known correlation model, research efforts can be devoted to exploit this correlation as a side information to tailor the proposed algorithm to those correlated scenarios.

Appendix A

Derivation of the time updated estimate

In this appendix, we show how to derive (3.21) starting from (3.20). We first state the well-known matrix inversion lemma (a.k.a., Woodbury matrix identity):

Lemma 1. *Given three arbitrary complex matrices,* $\boldsymbol{P}$, $\boldsymbol{B}$, and $\boldsymbol{R}$, *of suitable sizes, we have*

$$\boldsymbol{P}\boldsymbol{B}^{\mathsf{H}}\left(\boldsymbol{B}\boldsymbol{P}\boldsymbol{B}^{\mathsf{H}}+\boldsymbol{R}\right)^{-1}=\left(\boldsymbol{P}^{-1}+\boldsymbol{B}^{\mathsf{H}}\boldsymbol{R}^{-1}\boldsymbol{B}\right)^{-1}\boldsymbol{B}^{\mathsf{H}}\,\boldsymbol{R}^{-1}. \tag{A.1}$$

Using Lemma 1 with $\boldsymbol{P}=\gamma_{\boldsymbol{x}_{t|t-1}}^{-1}\,\mathbf{I}_N$, $\boldsymbol{B}=\boldsymbol{H}_t$, and $\boldsymbol{R}=\gamma_{\mathsf{w}}^{-1}\,\mathbf{I}_N$, the Kalman gain matrix $\boldsymbol{K}_t$ in (3.20) becomes

$$\boldsymbol{K}_t=\gamma_{\mathsf{w}}\,\left(\gamma_{\boldsymbol{x}_{t|t-1}}\,\mathbf{I}_N+\gamma_{\mathsf{w}}\,\boldsymbol{H}_t^{\mathsf{H}}\,\boldsymbol{H}_t\right)^{-1}\,\boldsymbol{H}_t^{\mathsf{H}}. \tag{A.2}$$

By injecting (A.2) back into (3.20a), we obtain

$$\begin{aligned}\widehat{\boldsymbol{x}}_{t|t}&=\widehat{\boldsymbol{x}}_{t|t-1}+\gamma_{\mathsf{w}}\,\left(\gamma_{\boldsymbol{x}_{t|t-1}}\,\mathbf{I}_N+\gamma_{\mathsf{w}}\,\boldsymbol{H}_t^{\mathsf{H}}\,\boldsymbol{H}_t\right)^{-1}\,\boldsymbol{H}_t^{\mathsf{H}}\,\left(\boldsymbol{y}_t-\boldsymbol{H}_t\,\widehat{\boldsymbol{x}}_{t|t-1}\right)\\&=\left(\gamma_{\boldsymbol{x}_{t|t-1}}\,\mathbf{I}_N+\gamma_{\mathsf{w}}\,\boldsymbol{H}_t^{\mathsf{H}}\,\boldsymbol{H}_t\right)^{-1}\left(\gamma_{\mathsf{w}}\,\boldsymbol{H}_t^{\mathsf{H}}\,\boldsymbol{y}_t+\gamma_{\boldsymbol{x}_{t|t-1}}\,\widehat{\boldsymbol{x}}_{t|t-1}\right),\end{aligned} \tag{A.3}$$

which corresponds to (3.21a). Finally, by taking the derivative of (A.3) w.r.t. $\widehat{\boldsymbol{x}}_{t|t-1}$ and applying the scalar EP approximation, we obtain the expression of the precision of the time update step given in (3.21b).

Bibliography

[1] S. S. Haykin and S. S. Haykin, *Kalman filtering and neural networks*. Wiley Online Library, 2001, vol. 284.

[2] R. E. Kalman, "A new approach to linear filtering and prediction problems," 1960.

[3] I. Pitas and A. N. Venetsanopoulos, *Nonlinear digital filters: principles and applications*. Springer Science & Business Media, 2013, vol. 84.

[4] S. Thrun, "Probabilistic robotics," *Communications of the ACM*, vol. 45, no. 3, pp. 52–57, 2002.

[5] D. Liang, L. Ying, and F. Liang, "Parallel mri acceleration using m-focuss," in *2009 3rd International Conference on Bioinformatics and Biomedical Engineering*. IEEE, 2009, pp. 1–4.

[6] G. Tzagkarakis, D. Milioris, and P. Tsakalides, "Multiple-measurement bayesian compressed sensing using gsm priors for doa estimation," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2010, pp. 2610–2613.

[7] A. S. Charles, A. Balavoine, and C. J. Rozell, "Dynamic filtering of time-varying sparse signals via ℓ_1 minimization," *IEEE Transactions on Signal Processing*, vol. 64, no. 21, pp. 5644–5656, 2016.

[8] C. K. Thomas and D. Slock, "Generalized swept approximate message passing based kalman filtering for dynamic sparse bayesian learning," in *2020 28th European Signal Processing Conference (EUSIPCO)*. IEEE, 2021, pp. 2065–2069.

[9] E. Karseras, K. Leung, and W. Dai, "Tracking dynamic sparse signals using hierarchical bayesian kalman filters," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 6546–6550.

[10] A. Ma, Y. Zhou, C. Rush, D. Baron, and D. Needell, "An approximate message passing framework for side information," *IEEE Transactions on Signal Processing*, vol. 67, no. 7, pp. 1875–1888, 2019.

[11] J. Ziniel and P. Schniter, "Efficient high-dimensional inference in the multiple measurement vector problem," *IEEE Transactions on Signal Processing*, vol. 61, no. 2, pp. 340–354, 2012.

[12] G. T. Cinar and J. C. Principe, "Hidden state estimation using the correntropy filter with fixed point update and adaptive kernel size," in *The 2012 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2012, pp. 1–6.

[13] R. Izanloo, S. A. Fakoorian, H. S. Yazdi, and D. Simon, "Kalman filtering based on the maximum correntropy criterion in the presence of non-gaussian noise," in *2016 Annual Conference on Information Science and Systems (CISS)*. IEEE, 2016, pp. 500–505.

[14] B. Chen, L. Dang, Y. Gu, N. Zheng, and J. C. Príncipe, "Minimum error entropy kalman filter," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 9, pp. 5819–5829, 2019.

[15] P. M. Djuric, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Miguez, "Particle filtering," *IEEE signal processing magazine*, vol. 20, no. 5, pp. 19–38, 2003.

[16] J. Čurn, D. Marinescu, G. Lacey, and V. Cahill, "Estimation with non-white gaussian observation noise using a generalised ensemble kalman filter," in *2012 IEEE International Symposium on Robotic and Sensors Environments Proceedings*. IEEE, 2012, pp. 85–90.

[17] J. H. Kotecha and P. M. Djuric, "Gaussian sum particle filtering," *IEEE Transactions on signal processing*, vol. 51, no. 10, pp. 2602–2612, 2003.

[18] P. M. Djurić and M. F. Bugallo, "Particle filtering for high-dimensional systems," in *2013 5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. IEEE, 2013, pp. 352–355.

[19] A. Doucet, N. d. Freitas, and N. Gordon, "An introduction to sequential monte carlo methods," in *Sequential Monte Carlo methods in practice*. Springer, 2001, pp. 3–14.

[20] T. Bengtsson, P. Bickel, B. Li *et al.*, "Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems," *Probability and statistics: Essays in honor of David A. Freedman*, vol. 2, pp. 316–334, 2008.

[21] A. E. Bryson and Y.-C. Ho, *Applied optimal control: optimization, estimation, and control*. Routledge, 2018.

[22] Y. Ho and R. Lee, "A bayesian approach to problems in stochastic estimation and control," *IEEE transactions on automatic control*, vol. 9, no. 4, pp. 333–339, 1964.

[23] Y. C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications*. Cambridge university press, 2012.

[24] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, 2009.

[25] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," *IEEE Transactions on Information Theory*, vol. 65, no. 10, pp. 6664–6684, 2019.

[26] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, 2001.

[27] K. P. Murphy, *Dynamic bayesian networks: representation, inference and learning*. University of California, Berkeley, 2002.

[28] T. P. Minka, "A family of algorithms for approximate bayesian inference," Ph.D. dissertation, Massachusetts Institute of Technology, 2001.

[29] M. Seeger, "Expectation propagation for exponential families," Tech. Rep., 2005.

[30] B. Cakmak and M. Opper, "Expectation propagation for approximate inference: Free probability framework," in *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2018, pp. 1276–1280.

[31] N. Halko, P.-G. Martinsson, and J. A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM review*, vol. 53, no. 2, pp. 217–288, 2011.

[32] A. V. Aho and J. E. Hopcroft, *The design and analysis of computer algorithms*. Pearson Education India, 1974.

[33] A. H. Jazwinski, *Stochastic processes and filtering theory*. Courier Corporation, 2007.