Electromagnetic Inversion for Noninvasive Specific Absorption Rate Characterization

by

Mario Lucien Phaneuf

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

MASTER OF SCIENCE

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

Abstract

This thesis investigates the use of the electromagnetic (EM) inversion framework for noninvasive specific absorption rate (SAR) characterization. Under this framework, the SAR characterization problem is formulated as an inverse problem where EM fields external to the SAR phantom are measured and then used to infer the internal SAR distribution. The advantages of the noninvasive approach include a potentially accelerated measurement, the option to use solid and inhomogeneous phantoms, and the ability to use existing near-field measurement hardware. An overview of contemporary SAR measurements and direct acceleration methods is presented to motivate the noninvasive approach. Following this, a classification of possible approaches to the noninvasive SAR characterization problem is presented and analyzed for the first time. From these approaches, two methods are chosen to be evaluated in this thesis. The first method is known as the phantom surface current method, which replaces the SAR phantom with equivalent surface currents which radiate in free space and can be used to obtain the internal SAR distribution. The noninvasive SAR problem is thus transformed into a near-field measurement of two independent antennas. A near-field antenna measurement algorithm is adapted for the SAR problem by including some novel numerical techniques. The performance is evaluated for 2D transverse magnetic (TM) problems. In addition, a fully vectorial 3D algorithm is developed and preliminarily investigated. The second method is proposed for the first time and is referred to as the simultaneous inversion (SI) method. This method is built upon the phantom surface current method by also considering the dependency between the two equivalent antennas and using this relationship as a regularization term. It is argued that this formulation is more resistant to measurement noise when compared to existing methods. The theoretical framework motivating this proposition is also presented. The noise resistance of the SI method and the theoretical framework are verified for 2D TM problems using synthetic measurement data perturbed by synthetic white noise.

Acknowledgements

First of all, I would like to thank my academic advisor Dr. Puyan Mojabi for his help, support, and friendship over the past couple of years. I truly enjoyed my time as his graduate student and I look forward to continuing as his PhD student.

I would like to thank the rest of my M.Sc. committee, Dr. Colin Gilmore and Dr. Ahmed Ashraf, for taking the time to review and improve my work.

I would also like to extend my gratitude to my friends and colleagues at the university of Manitoba. I would like to personally thank Chaitanya Narendra and Trevor Brown for their help with adapting the 3D near-field antenna measurement algorithm for the noninvasive SAR problem and also Nozhan Bayat for his help with the ANSYS® HFSS™ simulations.

Lastly, I would like to thank the National Sciences and Engineering Research Council of Canada and the University of Manitoba GETS program for their financial support.

To my family

Table of Contents

List of Figures

List of Tables

Important Abbreviations

Abbreviation	Description
2D	Two-dimensional
3D	Three-dimensional
CG	Conjugate gradient
DUT	Device under test
EM	Electromagnetic
FF	Far-field
PEC	Perfect electric conductor
RWG	Rao-Wilton-Glisson
SAR	Specific absorption rate
SRM	Source reconstruction method
SVD	Singular value decomposition
TM	Transverse magnetic

Important Symbols

Symbol	Description
$\hat{x}, \hat{y}, \hat{z}$	Unit vectors in the x, y, and z directions
$\vec{r}$	Position vector to a general observation point
$\vec{r}'$	Position vector to a general source point
j	Imaginary unit ($j^2 = -1$)
$\vec{E}$	Electric field intensity
$\vec{H}$	Magnetic field intensity
$\vec{J}$	Electric current density
$\vec{M}$	Magnetic current density
$\vec{A}$	Magnetic vector potential
$\vec{F}$	Electric vector potential
k	Wavenumber
ε	Complex dielectric permittivity
μ	Magnetic permeability
λ	Wavelength
η	Wave impedance
∇	Gradient operator
$\nabla\cdot$	Divergence operator
$\nabla\times$	Curl operator
∇^2	Laplacian operator
$\vec{\nabla}^2$	Vector Laplacian operator
$\|.\|_2$	L_2 norm
$(.)^H$	Hermitian operator

Chapter 1

Introduction

This introductory chapter will present the necessary background material in order to understand this research thesis. In particular, the general description and motivation for the research is presented. Broadly speaking, this research is interested in the use of electromagnetic (EM) inversion methods for the purpose of noninvasive specific absorption rate (SAR) measurements. The pros and cons of various noninvasive SAR approaches to the problem will be presented as part of a classification of approaches. Some of these methods will be examined with two-dimensional (2D) and three-dimensional (3D) simulation results. Finally, a novel noninvasive SAR measurement method is presented and is proposed to be particularly resistant to measurement noise.

1.1 Safety of Microwave Radiation

Since the advent of communication technologies relying on microwave radiation, the question of safety has been posed. Microwaves occupy a portion of the electromagnetic spectrum and thus are in the same family of physical phenomena such as X-rays and UV radiation. The dangers (and acceptable limits) of these latter forms of radiation are well known [1] and so it follows that microwaves may have similar dangers.

Fortunately, the biological effects of microwaves and X-rays are different. X-rays, and other small wavelength radiation, have enough energy to break the bonds of molecules in the human body. This group of EM radiation is referred to as *ionizing* radiation, on account of its chemically ionizing properties. This sort of damage can cause horrific effects on the human body, through simple burns to aggressive cancers [1]. On the other hand, larger wavelength radiation is *non-ionizing*, since the individual photon energy is dependent on wavelength only.[1] The matter in the human body can absorb this radiation but it is not enough to break molecular bonds. The molecules instead vibrate and reradiate the energy in the form of heat [3]. Of course, if too much heat is produced, it can damage the human body by literally "cooking" it.

Regulating bodies all around the world have since researched these heating effects and have imposed human exposure limits to microwave radiation. All devices and technologies which emit microwave radiation must be tested to verify compliance. Although exposure limits may vary from country to country, the universal measure used for most technologies is known as the *specific absorption rate*, or SAR.

The use of the SAR as a safety measure is only used for a certain range of frequencies (within which most current communication technologies lie). In Canada, the SAR measure is valid for frequencies in the 100 kHz to 6 GHz range [4]. Between the frequencies of 3 kHz to 10 MHz, microwave heating is minimal, so it is sufficient to measure the induced electric field only [4]. For frequencies above 6 GHz, heating occurs primarily in the skin, and thus it is sufficient to measure the incident power density of the radiating device [4]. Since the majority of current communication technologies use the SAR measure, we limit

[1] The energy E of a photon is $E = hf$, where h is the Planck constant and f is the frequency [2].

the discussion in this thesis to this domain.

1.1.1 Specific Absorption Rate

The specific absorption rate is a measure in watts per kilogram (W/kg) of the electromagnetic energy (or power) absorbed by human tissue or other absorbing materials. It is a useful quantity because it only requires knowledge of the electrical properties of human tissue (which are generally known) and the electric field intensity (which can be measured). The SAR value over an infinitesimal volume (i.e., a point) can be calculated with

$$SAR(\vec{r}) = \frac{\sigma(\vec{r})|\vec{E}(\vec{r})|^2}{\rho(\vec{r})}, \tag{1.1}$$

where $\sigma(\vec{r})$ is the electrical conductivity in Siemens per meter (S/m), $\vec{E}(\vec{r})$ is the electric field in Volts per meter (V/m), and $\rho(\vec{r})$ is the density in kilograms per cubic meter (kg/m^3).

The more conventional way of recording the SAR is to take the average over a volume, calculated as:

$$SAR = \frac{1}{V}\int_V \frac{\sigma(\vec{r})|\vec{E}(\vec{r})|^2}{\rho(\vec{r})} dV. \tag{1.2}$$

Conventional SAR limits use the latter interpretation of the SAR and are always referred to a specific averaging volume. The important takeaway from these equations is that for a known absorbing material (which is the case for the human body[2]), a SAR measurement is actually just an electric field measurement.

In Canada, there are prescribed SAR limits for the head, the torso, and the limbs. In the head and the torso, the SAR limit is 1.6 (W/kg) averaged over 1 g [4]. In the limbs, the limit climbs to 4 (W/kg) averaged over 10 g [4].

[2] As will be shown, the properties of the human body are assumed to be known when performing a SAR measurement.

1.2 SAR Measurements

One of the final steps in the design process for a microwave radiator is to perform a SAR measurement to ensure that the exposure level is below the prescribed limits. The SAR distribution in the human head or body depends on many factors (see Section 1.2.4); thus, it is imperative that the measurement is performed in a testing environment which is nearly identical to the real case.

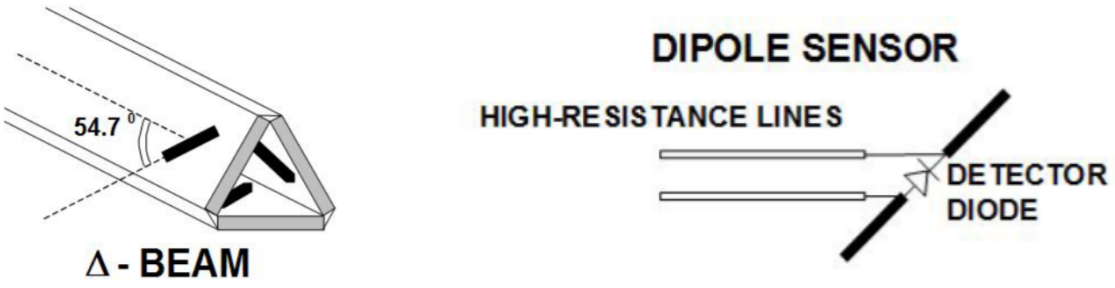
1.2.1 Contemporary Measurements and Standards

The standard contemporary measurement method consists of a mechanical scanning electric field probe which performs measurements in a homogeneous liquid (known as a *phantom*) as a stand-in for the human body or head. Two main standards exist which outline the recommended practice of SAR measurements: IEEE Std 1528 [5] and IEC 62209-1 [6]. The information in Section 1.2 is a brief summary of key aspects in IEEE Std 1528 [5] © 2011 IEEE. Note that this standard only covers measurements for the head.

The first step in a standard compliant measurement is what is known as an *area scan*, where the scanning probe measures the electric field in a relatively coarse grid of points inside the phantom. This allows the user to quickly pinpoint the general area which contains the peak SAR. The next step is the *zoom scan*, where the scanning probe measures the electric field in a relatively fine grid of points around the peak SAR. Unlike the first scan, the grid of points is only a subset of the total phantom volume and corresponds to the required averaging volume to evaluate (1.2).

1.2.2 Scanning Probe

The scanning probe is responsible for recording the actual electric field. An ideal probe would be electrically small (to measure at a point), isotropic, non-coupling (to not disturb



(a) Orthogonal dipoles **(b)** Electric circuit

Fig. 1.1: Electric field probe construction [5, 7]. © [2013] IEEE

the internal field), and would have a linear response to the square of the magnitude of the electric field (so that the measurement is linearly proportional to the SAR).

In practice, the probe is in fact three separate dipole antennas which each measure one vector component of the field. The dipole antennas are loaded with diodes which are made to operate in the square-law region of the response. The dipole feed transmission line creates an effective low-pass filter which results in a DC voltage measurement which is proportional to the square of the electric field magnitude. See Figure 1.1 for the visualization of these details.

The complete probe must be constructed carefully to limit any disturbances (i.e., coupling) to the underlying field. Coupling is limited by suspending the probe in a solid dielectric which mimics the phantom material and forms the body of the probe. The equivalent probe circuit is also designed to have a large input resistance to further reduce coupling to the electromagnetic field.

1.2.3 Phantom Properties

In order to replicate the effect of the head or human body, the phantom should closely resemble the true case. For the case of the head, the shape is chosen such that it is rep-

resentative of the majority of human heads and results in a conservative SAR estimate.[3] A practical head phantom consists of a plastic shell representing half of the human head, which is filled with a homogeneous, liquid, and lossy dielectric. The dielectric must necessarily be a liquid in order to allow direct probe measurements. The dielectric properties are chosen such that the resulting SAR measurement is a conservative estimate for the true inhomogeneous case. That is, the SAR in the phantom should be a conservative estimate to the actual SAR that would occur in the actual human head. Since dielectric properties are dependent on the frequency, different phantom dielectrics must be used accordingly.

Table 1.1 below shows some phantom liquid parameter values for frequencies in the range of 750 Mhz to 1500 Mhz.

Table 1.1: Dielectric properties of phantom liquid in the 750 to 1500 MHz range. © [2013] IEEE

Frequency (MHz)	**Relative permittivity**	**Conductivity** σ (S/m)
750	41.9	0.89
835	41.5	0.90
900	41.5	0.97
1450	40.5	1.20
1500	40.4	1.23

1.2.4 SAR Factors

An important aspect of SAR measurements is the sensitivity to many factors. Indeed, the averaged SAR measurement and even the SAR distribution inside the phantom can change because of non-obvious factors [8]. These include the phantom size and shape [9], the position and orientation of the device under test (DUT) [10], the human hand [11], any objects

[3] In practice, a large adult head absorbs the most energy due to stronger coupling effects. A phantom of this size provides a conservative SAR estimate for men, women, and children.

worn by the user [12, 13], the operating frequency [5], and DUT modulation schemes [5].

The IEEE Std 1528 recommends that separate SAR measurements should be performed for all usable frequency bands, all modulation schemes, and a set of predetermined device orientations. Each of these measurements must obtain a SAR value which is below the limits set by the appropriate regulating body.

1.2.5 Limitations

The main limitations of contemporary SAR measurements are the inability to use inhomogeneous phantoms and the relatively slow measurements which need to be repeated many times. These are all consequences of using a single electric probe and having to measure the field at each point within the phantom.

Since the recommended SAR measurement requires the probe to measure the field *inside* the phantom, the phantom material must be aqueous. Although the phantom liquid parameters are chosen such that the resulting SAR is equivalent to the real case, the frequency dependence of the inhomogeneous head means that different phantom liquids must be used for different operating frequencies. This will further contribute to the overall measurement time for a given device under test.

As mentioned in Section 1.2.4, an individual SAR measurement must be done for different frequency bands, modulation schemes, and spatial configurations of the DUT. As new devices continue to become more sophisticated (more frequencies of operation, more modulation schemes), the total number of measurements continues to grow. In turn, the overall measurement time also increases.

Due to these limitations, the SAR measurement can become quite a large bottleneck in the design process for devices which utilize microwave radiation.

1.3 Acceleration Methods

Due to the importance of SAR measurements and the ever-increasing acquisition times, there has been a push towards developing acceleration methods. In general, these methods will increase the speed of measurements at the cost of some accuracy. These methods can be categorized into two broad groups: field reconstruction methods and rapid measurements. Both of these methods produce the same number of data points and differ only on whether each point is actually measured. Field reconstruction techniques accelerate SAR measurements by cutting the number of measurement points and instead using some sort of field reconstruction technique to infer the field at the remaining points. This technique is an approximation to the true field and will affect the accuracy of the final SAR value in proportion to the degree of the approximation. One form of field reconstruction is to use a fitting function that can be evaluated from the knowledge of the field at a few points. This fitting function is non-trivial and the merits of this method have been demonstrated by Kanda et al. in [14]. Another form of field reconstruction is to use the well-known forward propagation technique to reconstruct the SAR distribution from a planar measurement [15]. Rapid measurements involve using a probe array instead of a single probe. It is clear that this can reduce the total measurement time by increasing the rate at which the required measurement points are obtained. Of course, we can combine field reconstruction techniques with probe array measurements in order to further reduce the measurement time [16]. In fact, a commercial probe-array SAR measurement system using forward propagation has already been developed and provides very fast measurements [17].

1.4 Motivation for Research

While the acceleration methods presented are effective (some are even commercially available), they still rely on generally expensive probe systems and homogeneous liquid phantoms. As technology continues to progress, the number of devices with communication capabilities is ever increasing. The prospect of purchasing the probe system or outsourcing the SAR measurement can be an unattractive option.

The motivation for this research was to explore the possibility of using existing measurement devices, such as those used for antenna measurements, in order to measure the SAR indirectly by measuring the fields external to the phantom and DUT. Since the goal is to infer the cause (internal fields) from the effect (external fields), the problem can be formulated as an electromagnetic inverse problem. There are two general categories of EM inverse problems: inverse scattering problems, and inverse source problems. The former is used to determine unknown objects (e.g., medical imaging problems [18]), while the latter is used to determine the electromagnetic sources of an unknown active source (e.g., near-field antenna measurements [18]). For the SAR characterization problem, we use the inverse source framework to determine the effective DUT currents in the presence of the known phantom. Due to the electromagnetic uniqueness and equivalence theorems, there are multiple ways of formulating the inverse source problem. A classification of these methods will be presented, which will motivate the two noninvasive SAR algorithms presented in this thesis.

The noninvasive SAR approach can be a much more attractive option, especially in industries which already rely on existing antenna measurement systems. Furthermore, this approach would allow for the use of inhomogeneous solid phantoms which can be potentially more accurate (since each major part of the head can be represented separately) and easier to store and maintain (e.g., not susceptible to liquid evaporation). Of course, the trade-off

with the noninvasive approach is the relatively complex algorithms required to relate the externally measured fields to the fields induced inside the phantom. In other words, we reduce the hardware complexity by increasing the software complexity.

1.5 Contributions of Thesis Work

The contributions of this thesis work include a classification of noninvasive SAR approaches, a noninvasive SAR measurement algorithm with novel techniques, and a separate novel noninvasive SAR algorithm accompanied by theoretical work which argues that the method is particularly resistant to noise. Note that some of the algorithms only concern themselves with the fields on the surface of the phantom and do not actually obtain the internal SAR distribution. (Due to the electromagnetic uniqueness theorem, it is sufficient to know the fields on the surface of the phantom in order to obtain the internal fields as well as the SAR.)

Classification of Noninvasive SAR Methods

The first contribution of this thesis is a classification of potential noninvasive SAR measurement approaches. In particular, the advantages and disadvantages of these approaches are presented and analyzed. This classification motivates the first noninvasive algorithm which is an adaptation of an existing inverse algorithm. The limitations of this method in the context of the overarching classification will motivate the novel algorithm which is developed and presented in this thesis.

Noninvasive SAR Algorithm with Novel Adaptations

Both a 2D and 3D algorithm are implemented in MATLAB®[4] which view the noninvasive SAR problem as an antenna characterization problem by using the source reconstruction

[4]MATLAB® is a registered trademark of The MathWorks, Inc.

method (SRM). The 2D algorithm includes some novel numerical techniques which were developed specifically for the additional challenges of the noninvasive SAR problem. The 2D algorithm was tested using synthetic MATLAB data and simulation data obtained through the ANSYS® HFSS™ software. The algorithm requires magnitude and phase measurements and reconstructs the internal SAR distribution. The 3D algorithm was adapted from existing MATLAB code used for near-field antenna measurements. The algorithm was tested with simulation data from ANSYS HFSS. The algorithm requires magnitude and phase measurements and reconstructs the phantom surface fields only. In addition, none of the novel techniques were implemented in the 3D algorithm.

Novel Noninvasive SAR Algorithm

A novel noninvasive SAR algorithm is developed and presented which is argued to be more resistant to measurement noise than the alternatives. The theoretical framework to make this argument is also developed and presented. A 2D version of the algorithm was implemented in MATLAB which requires magnitude and phase measurements to reproduce the phantom surface fields. The algorithm is compared to the alternative methods for both noiseless and noisy synthetic measurement data.

Refereed Journal Papers

1. M. Phaneuf and P. Mojabi, "Electromagnetic Inversion for Noninvasive Specific Absorption Rate Characterization". Submitted November 2019.

Refereed Conference Papers

1. M. Phaneuf and P. Mojabi, "Evaluation of an Inversion Algorithm for Noninvasive Specific Absorption Rate Applications," *2019 IEEE Int. Symposium on Antennas and Propagation and USNC-URSI RSM*, Atlanta, GA, USA, 2019, pp. 533-534.

1.6 Thesis Outline

This thesis consists of 5 chapters, including this introduction where we presented the necessary background information about SAR measurements. The motivation for this research, as well as the contributions were also presented.

The second chapter will focus on the classification of noninvasive SAR measurements. The different methods of approaching the problem will be described and their motivation will be presented. Furthermore, the challenges and pitfalls of such methods will be described.

The third chapter will present the formulation of the first noninvasive SAR measurement method which is an adaptation of an existing inverse algorithm. Solutions for the challenges presented in the previous chapter, which include novel numerical techniques, will be incorporated into the formulation. The 2D and 3D simulation results will also be presented in this chapter.

The fourth chapter will begin with the theoretical framework which motivates the novel noninvasive SAR algorithm proposed in this thesis. In particular, this algorithm is argued to be more resistant to measurement noise. Simulation results for the 2D case will be presented to reinforce this argument. The final chapter will be a summary of the thesis and will include future work and considerations towards the realization of a practical noninvasive SAR measurement algorithm.

Chapter 2

Classification of Noninvasive SAR Measurements

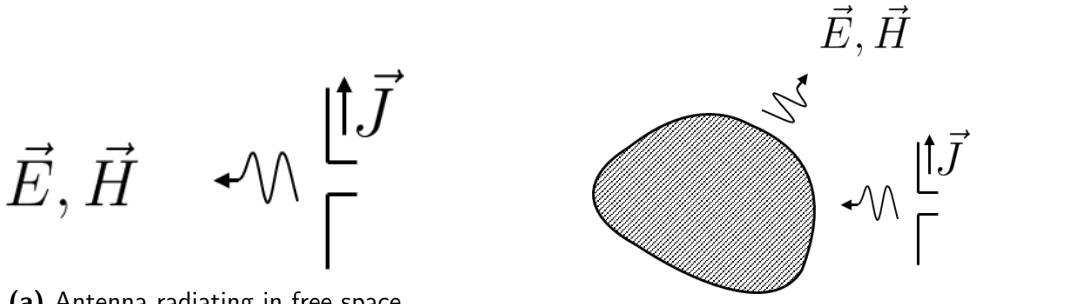
In this chapter, the noninvasive SAR characterization problem is analyzed. We begin by considering the general wave equation, from which the potential approaches are intuitively found. Following this discussion, a formal classification of the methods is presented, which includes the advantages and disadvantages of each method. Finally, a discussion on the general difficulties of inverse problems will be held, which will further motivate the novel algorithm proposed in Chapter 4.

2.1 Wave Equation Discussion

It is insightful to first consider the free-space electric field wave equation which can be expressed as

$$\vec{\nabla}^2 \vec{E}(\vec{r},t) - \frac{1}{c_0^2}\frac{\partial \vec{E}(\vec{r},t)}{\partial t^2} = 0, \tag{2.1}$$

where c_0 is the speed of light in a vacuum, $\vec{r}$ is the position vector, t is time, and $\vec{\nabla}^2$ is the vector Laplacian operator. If we instead go to the frequency domain with a time



(a) Antenna radiating in free space

(b) Antenna radiating in the presence of a dielectric scatterer

Fig. 2.1: Two cases for the wave equation analysis.

dependency of $e^{j\omega t}$ (where $j^2 = -1$ and ω is the radial frequency), the equation becomes the Helmholtz equation

$$\vec{\nabla}^2\vec{E}(\vec{r}) + k_0^2\vec{E}(\vec{r}) = 0, \tag{2.2}$$

where we have introduced the free-space wavenumber $k_0 = \omega/c_0$. This can be extended to include charge density ρ and current density $\vec{J}$ as field sources in an arbitrary and isotropic medium as

$$\vec{\nabla}^2\vec{E}(\vec{r}) + k^2(\vec{r})\vec{E}(\vec{r}) = \frac{1}{\varepsilon(\vec{r})}\nabla\rho(\vec{r}) + j\omega\mu(\vec{r})\vec{J}(\vec{r}), \tag{2.3}$$

where $k(\vec{r})$ is the wavenumber which is dependent on the background material, $\varepsilon(\vec{r})$ is the dielectric permittivity, and $\mu(\vec{r})$ is the magnetic permeability. From this point on, we will make the assumption $\rho = 0$, which will simplify the analysis of (2.3). Note that the general conclusions of this analysis remain true even in the case $\rho \neq 0$.

Consider the two cases illustrated in Figure 2.1: an antenna radiating in free space and an antenna radiating in the presence of a lossy non-magnetic dielectric scatterer (i.e., the scatterer is characterized by a complex permittivity). In the first case, the governing wave equation is

$$\vec{\nabla}^2\vec{E}(\vec{r}) + k_0^2\vec{E}(\vec{r}) = j\omega\mu_0\vec{J}(\vec{r}). \tag{2.4}$$

In the second case, the governing wave equation is

$$\vec{\nabla}^2\vec{E}(\vec{r}) + k^2(\vec{r})\vec{E}(\vec{r}) = j\omega\mu_0\vec{J}(\vec{r}). \tag{2.5}$$

Let us define the field in the first case as the *incident* field

$$\vec{E}(\vec{r})|_{\text{CASE 1}} = \vec{E}^{\text{inc}}(\vec{r}), \tag{2.6}$$

and the field in the second case as the *total* field

$$\vec{E}(\vec{r})|_{\text{CASE 2}} = \vec{E}^{\text{tot}}(\vec{r}). \tag{2.7}$$

We will now introduce the field quantity known as the *scattered* field, defined as

$$\vec{E}^{\text{scat}}(\vec{r}) = \vec{E}^{\text{tot}}(\vec{r}) - \vec{E}^{\text{inc}}(\vec{r}). \tag{2.8}$$

With these new definitions in hand, we start by subtracting (2.4) from (2.5). This difference becomes

$$\begin{aligned}
0 &= \vec{\nabla}^2\big(\vec{E}^{\text{tot}}(\vec{r}) - \vec{E}^{\text{inc}}(\vec{r})\big) + k^2(\vec{r})\vec{E}^{\text{tot}}(\vec{r}) - k_0^2\vec{E}^{\text{inc}}(\vec{r}),\\
0 &= \vec{\nabla}^2\vec{E}^{\text{scat}}(\vec{r}) + k^2(\vec{r})\vec{E}^{\text{tot}}(\vec{r}) - k_0^2\big(\vec{E}^{\text{tot}}(\vec{r}) - \vec{E}^{\text{scat}}(\vec{r})\big),\\
0 &= \vec{\nabla}^2\vec{E}^{\text{scat}}(\vec{r}) + k_0^2\vec{E}^{\text{scat}}(\vec{r}) + \big(k^2(\vec{r}) - k_0^2\big)\vec{E}^{\text{tot}}(\vec{r}).
\end{aligned} \tag{2.9}$$

What we have at this point is almost a wave equation for the scattered field. Next, we introduce the dielectric contrast $\chi(\vec{r})$,

$$\chi(\vec{r}) = \frac{k^2(\vec{r}) - k_0^2}{k_0^2} = \frac{\varepsilon(\vec{r}) - \varepsilon_0}{\varepsilon_0} = \varepsilon_r(\vec{r}) - 1, \tag{2.10}$$

where ε_r is the complex relative dielectric permittivity. Lastly, we introduce the contrast source current density $\vec{w}(\vec{r})$,

$$\vec{w}(\vec{r}) = \chi \vec{E}^{\text{tot}}(\vec{r}). \tag{2.11}$$

With these new definitions, we can finally write a wave equation for the scattered field:

$$\vec{\nabla}^2 \vec{E}^{\text{scat}}(\vec{r}) + k_0^2 \vec{E}^{\text{scat}}(\vec{r}) = -k_0^2 \vec{w}(\vec{r}). \tag{2.12}$$

The solution to these is the convolution of the sources $k_0^2 \vec{w}(\vec{r})$ and the free space green's function $\overline{\overline{G}}(\vec{r}, \vec{r}')$,

$$\vec{E}^{\text{scat}}(\vec{r}) = \overline{\overline{G}}(\vec{r}, \vec{r}') \circledast (k_0^2 \vec{w}(\vec{r})), \tag{2.13}$$

where $\overline{\overline{G}}$ is the solution to the following homogeneous wave equation:

$$\vec{\nabla}^2 \overline{\overline{G}}(\vec{r}, \vec{r}') + k_0^2 \overline{\overline{G}}(\vec{r}, \vec{r}') = -\overline{\overline{I}} \delta(\vec{r} - \vec{r}'), \tag{2.14}$$

where $\overline{\overline{I}}$ is the identity dyad and δ denotes the Dirac delta function. Equation 2.12 is a powerful result since it is an equation which directly relates the external field (through $\vec{E}^{\text{scat}}$)[1] to the internal electric field in the dielectric (though $\vec{w}$). (Recall that the goal of a noninvasive SAR algorithm is to obtain the internal SAR (or field) distribution in the phantom from external field measurements.) This is the ideal case for a potential noninvasive SAR method.

The contrast source formulation is an EM inverse method used primarily in medical imaging problems (e.g., [19]). In these problems, both the total field $\vec{E}^{\text{tot}}(\vec{r})$ and the contrast $\chi(\vec{r})$ are unknowns in a nonlinear equation. In the SAR problem, the contrast is known, and the modified problem becomes a linear equation for the unknown internal fields $\vec{E}^{\text{tot}}(\vec{r})$.

[1] The scattered field can be obtained by performing two measurements: one in the presence of the phantom ($\vec{E}^{\text{tot}}$), and another without the phantom ($\vec{E}^{\text{inc}}$). The external scattered field is the difference of these two measurements.

2.1.1 Limitations of the Contrast Source Formulation

The derivation of (2.9) necessitates a very critical assumption regarding the DUT and scatterer system; this is the presumption that the currents $\vec{J}$ do not change in the presence of the scatterer. This is an incorrect assumption in the SAR case as the DUT is affected (i.e., becomes loaded) by the presence of the human body [20, 21]. In (2.9), there would need to be a term corresponding to the difference of the DUT currents which arises from this loading effect.

Let us consider for a moment that this assumption is correct, and that we can solve the SAR characterization problem as a modified imaging problem. There are two main problems that result from this assumption. The first is the difficulty of solving the modified imaging problem. This sort of approach would result in a 3D distribution of unknown field points inside the phantom. In a regular imaging problem, we use many different transmitting antennas in order to increase the relative number of measurements. Since we are only concerned with a single transmitter in the case of SAR characterization, it can be very difficult numerically to solve this problem. The second problem arising from this assumption is the conflict with the original purpose of direct SAR measurements: we do not know our DUT currents in the presence of the human body. If the DUT remained unchanged, the SAR measurement could simply be a near-field antenna measurement. Since we know our phantom parameters exactly, we could simply solve the forward problem with the equivalent currents of the DUT using the commercial software of our choice and the need for an actual SAR measurement would be completely nullified.

With the knowledge that the DUT currents cannot be ignored, we must devise another approach for noninvasive SAR measurements. In general, there are two directions towards solving (2.3): either determine the exact green's function or formulate an equivalent prob-

lem with the free-space green's function.

A final consideration for this problem is the nature of our unknown DUT currents. We do not know the magnitude or the exact distribution of the currents, which is a major problem if we want to directly use (2.3).[2] One method of handling this is making an assumption on the current distribution by using the *surface equivalence theorem*. We will first take this discussion to the surface equivalence theorem before returning to the noninvasive SAR measurement approaches.

2.2 Surface Equivalence Theorem

The following discussion can be found in most textbooks on electromagnetics [22].

Let us start by considering a region (Region 1) bounded by a surface Γ containing arbitrary electric and magnetic currents $\vec{J}_1$ and $\vec{M}_1$, and arbitrary dielectric and magnetic material with permittivity ε and permeability μ as shown in Figure 2.2. Assume that Region 2 is free space and that it contains no sources. The surface equivalence theorem states that we can construct an equivalent problem for Region 2 by introducing *equivalent currents* on the surface Γ which account for the source and material effects in Region 1. The equivalence theorem is itself based upon the electromagnetic *uniqueness theorem*, which states that the electromagnetic fields in a region are uniquely known if we know the tangential fields on a closed boundary which bounds the region. As long as the currents are chosen to satisfy the electromagnetic boundary conditions, the fields in Region 2 remain the same. A consequence of this is that we can set arbitrary fields inside Region 1, and we can even replace the background material.

[2] Equivalently, we can use the vector potential wave equations.

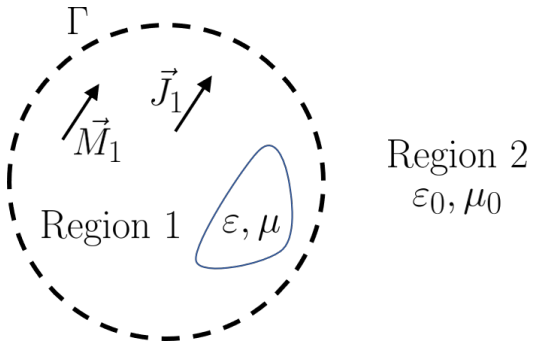


Fig. 2.2: Initial problem with arbitrary Region 1 and free-space Region 2.

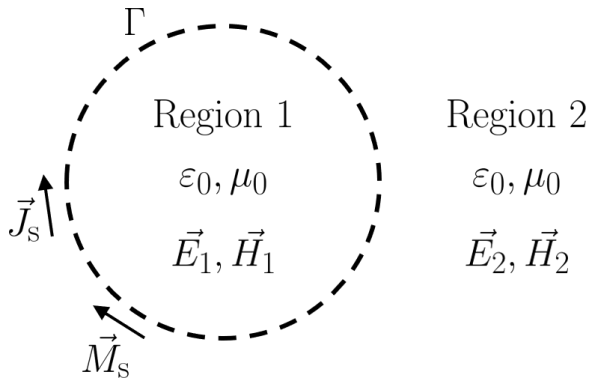


Fig. 2.3: Equivalent problem for Region 2 where the arbitrary sources and materials have been replaced with electric and magnetic surface currents $\vec{J}_S$ and $\vec{M}_S$.

Figure 2.3 shows the application of the equivalence theorem to our initial problem. Here we have introduced the fields $\vec{E}_1, \vec{H}_1$ in Region 1 and $\vec{E}_2, \vec{H}_2$ in Region 2. The fields in Region 1 are arbitrary, while the fields in Region 2 are the same fields that were in Figure 2.2. The electromagnetic boundary conditions state that the equivalent electric and magnetic surface currents $\vec{J}_s$ and $\vec{M}_s$ are related to the fields by

$$\vec{J}_s = \hat{n} \times (\vec{H}_2 - \vec{H}_1), \tag{2.15}$$

and

$$\vec{M}_s = -\hat{n} \times (\vec{E}_2 - \vec{E}_1), \tag{2.16}$$

where $\hat{n}$ is the unit outer normal on Γ, and "$\times$" denotes the vector cross product. A common choice for the fields in Region 1 is zero or null fields. This particular choice of fields is known as *Love's Equivalence Condition*. The equivalent currents are now related to the fields in Region 2 directly by

$$\vec{J}_s = \hat{n} \times \vec{H}_2, \tag{2.17}$$

and

$$\vec{M}_s = -\hat{n} \times \vec{E}_2. \tag{2.18}$$

This is a very powerful result, as it means we could use the free-space green's function[3] in the wave equation. Additionally, this allows us to handle the unknown physical distribution of DUT currents by introducing equivalent surface currents.

2.3 Noninvasive SAR Measurement Approaches

Now that we have covered the fundamental concepts, we can begin our discussion on the different SAR measurement approaches. As mentioned in Section 2.1.1, there are two broad approaches towards solving (2.3): finding and using the exact green's function (retaining the inhomogeneous background) or formulating an equivalent free-space problem.

2.3.1 Inhomogeneous Background Approach

Conceptually, the easiest way to approach the noninvasive SAR problem is just to find the green's function for the problem. Since the phantom parameters are known, it should be possible to determine the green's function and solve the wave equation.

As previously stated, it is virtually necessary to at least replace the DUT currents with

[3]We would like to emphasize that the scatterer in Figure 2.2 has been removed in Figure 2.3. The scattering effects are fully represented by the electric and magnetic surface currents $\vec{J}_S$ and $\vec{M}_S$, which allows for the use of the free-space green's function.

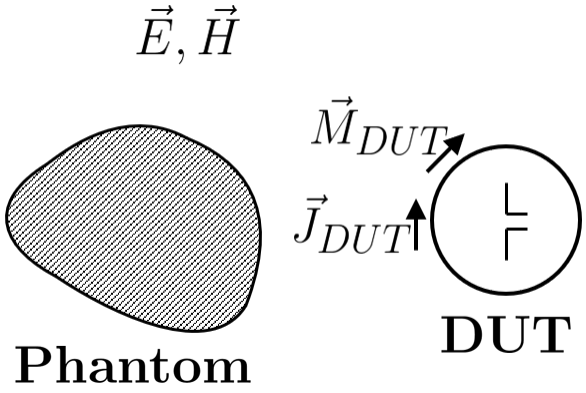


Fig. 2.4: Inhomogeneous background approach.

equivalent currents. The equivalent problem is illustrated in Figure 2.4. The external electric field is related to the DUT currents by the green's function and is governed by an equation of the form

$$\vec{E} = \mathcal{L}_k\big(\vec{J}_{\mathrm{DUT}}, \vec{M}_{\mathrm{DUT}}\big), \tag{2.19}$$

where the subscript k, as opposed to the free-space k_0, is emphasizing the modified green's function, and $\mathcal{L}_k$ is the operator which operates on the currents $\vec{J}_{\mathrm{DUT}}$ and $\vec{M}_{\mathrm{DUT}}$ to produce the external electric field $\vec{E}$. This equation can be discretized to form a matrix equation, which can then be solved for the equivalent DUT currents. After we obtain the DUT currents,[4] we can simply solve the forward problem for the fields inside the phantom. The SAR value can then be calculated directly from these fields. The advantages of this approach include the conceptual simplicity and relatively high accuracy from using the information contained in the phantom. The main disadvantage of this method is the non-triviality[5] of finding the green's function. It is for this reason that we shift our attention towards approaches which rely on the free-space green's function.

[4] Unlike with the contrast source approach, these DUT currents account for loading effects.

[5] In general, we would need to solve the forward problem [23], which can be computationally expensive in the 3D case.

2.3.2 Homogeneous (Free-Space) Background Approach

The second class of approaches rely on some form of equivalence theorem to reduce the effect of the phantom to equivalent electromagnetic sources. One example of such an approach has already been presented, which is the contrast source approach. The effect of the phantom (i.e., the scattered fields) is accounted for by the new contrast sources which radiate in free space. The advantages of this method include the direct relationship between the internal field distribution and the externally measured fields, and of course the numerical simplicity of using the free-space green's function. The disadvantages of this method include the inability to account for loaded DUT currents, and the necessity of solving for unknowns in a 3D volumetric distribution.

A second example of a free-space approach is the surface field approach. This method relies on the uniqueness theorem which states that we can obtain the fields in a source-free region (e.g., the phantom) if we know the fields on a surface bounding the region. This means that if we can obtain the fields on the surface of the phantom, we can obtain the internal field and SAR distribution. Using Love's condition, we can set equivalent currents on a surface bounding the phantom in a similar manner that was done for the DUT. It is then possible to use an *inverse source* or *source reconstruction* method (which comprise a subset of electromagnetic inversion techniques along with inverse scattering methods) to find the currents which produce the externally measured fields.

Coupled with the DUT surface currents, the equivalent problem would have two surfaces with equivalent currents, as shown in Figure 2.5. The external electric field would be related to the equivalent currents by an equation of the form

$$\vec{E} = \mathcal{L}_{k_0}\big(\vec{J}_{\mathrm{DUT}}, \vec{M}_{\mathrm{DUT}}, \vec{J}_{\mathrm{pha}}, \vec{M}_{\mathrm{pha}}\big), \tag{2.20}$$

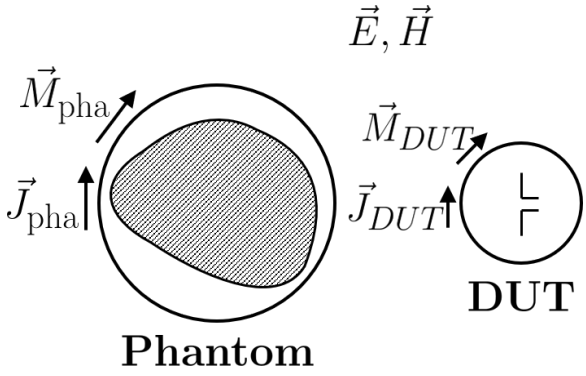


Fig. 2.5: Free-space background approach where the effects of the DUT and phantom are replaced by surface currents which radiate in free space to reproduce the external fields of the original system.

where the subscript k_0 is emphasizing that we are using the free-space green's function for the radiating sources. After obtaining the surface fields, the internal fields (and thus the SAR) can be obtained with any method of choice (e.g., the finite difference frequency domain (FDFD) formulation).

2.3.3 Hybrid Approach

Of course, there is also no reason why the two approaches can't be combined to make a hybrid approach. Since the DUT and phantom equivalent currents are related (specifically, the phantom surface fields are related to the DUT currents by the inhomogeneous green's function), an additional operator can be obtained which relates the currents

$$\big(\vec{J}_{\text{pha}}, \vec{M}_{\text{pha}}\big) = \mathcal{L}_L\big(\vec{J}_{\text{DUT}}, \vec{M}_{\text{DUT}}\big), \tag{2.21}$$

where the subscript L denotes the "link" between the currents. With such a relation, it is possible to reduce the surface field approach in (2.20) by substituting this relation:

$$\vec{E} = \mathcal{L}_{k_0}\big(\vec{J}_{\text{DUT}}, \vec{M}_{\text{DUT}}, \mathcal{L}_L(\vec{J}_{\text{DUT}}, \vec{M}_{\text{DUT}})\big). \tag{2.22}$$

Note that in this case, due to the uniqueness of EM fields, the equivalent operator in (2.22) (operating on the DUT currents) is the same as the one in (2.19) under infinite precision. However, the current link can be obtained more easily through alternative methods (such as the so-called PMCHW scattering formulation [24, 25]), as opposed to directly using the inhomogeneous green's function. This approach has previously been attempted using simulation and experimental data in [24, 25].

Alternatively, the current relation can be kept as a regularization[6] term at the optimization step of the algorithm. Thus, the surface fields are inverted simultaneously, while using the current relation as a filter to select a solution which meets both conditions. This regularized simultaneous inversion (SI) method will be presented in Chapter 4. To the best of the author's knowledge, this is the first use of the SI method for noninvasive SAR measurements. We also propose that such a method is particularly resistant to measurement noise.

2.4 Ill-Posedness of Inverse Problems

The main disadvantage of any noninvasive SAR measurement method is the issue of *ill-posedness*. Since we are trying to determine the *source* (e.g., DUT currents) from its external *effect* (i.e., externally measured field), we are dealing with what is known as an *inverse problem*. These types of problems are well-known to exhibit ill-posedness.

Ill-posed problems are mathematical problems which satisfy one or many of the following three conditions [26]:

1. No solution exists (*non-existence*).[7]

[6] Regularization refers to the process of filtering possible solutions to the original equation. See section 2.4 and 3.2.2 for more details.

[7] For example, given an allowable physical size, an antenna may not exist to create a desired radiation pattern. This is an example of using electromagnetic inversion for design purposes [27]. In characterization problems such as the SAR characterization problem, the solution always exists.

2. A given solution is not unique (*non-uniqueness*).

3. The solution is unstable, i.e., small changes in the known data can drastically change the obtained solution.

In the case of noninvasive SAR measurements, we are dealing with the latter two conditions. The issue of non-uniqueness arises when the total amount of information is insufficient for the problem (e.g., the number of measurements is less than the number of unknowns). The discrete operator representing the green's function is said to have a *null space*; that is, there exists a non-trivial number of current distributions which produce null fields at the measurement locations. While this can be a problem, it can be mitigated by simply taking more measurements or by using a coarser discretization scheme. Additionally, we can employ the strategy known as regularization, which usually involves imposing conditions on the unknown currents which reduces the total number of allowed solutions.

The issue of instability is perhaps the biggest challenge faced by the noninvasive SAR characterization problem. This behavior occurs due to the inherent smoothing properties [8] of the green's function operator; solutions with large spatial frequencies are "smoothed" such that the measured effect is smooth itself. This allows for many solutions which can produce the measured fields almost perfectly; thus, there is an effective or numerical null-space to the discrete green's function operator.

Numerically, it is easier to understand this phenomenon by looking at the singular value decomposition (SVD) of the ill-posed operator. For a discrete matrix problem $Ax = b$ with $A \in \mathbb{C}^{m \times n}$ and $m > n$, it can be shown that the least-squares solution is of the form [28]

$$x = \sum_{i=1}^{n} \frac{u_i^H b}{\sigma_i} v_i, \tag{2.23}$$

[8] This is the general property of Fredholm integrals of the first kind [28].

where u_i and v_i are the left and right singular vectors of A, σ_i are the singular values, and n is the number of columns in A. If this system represents an ill-posed problem, then the ratio between the largest and smallest singular values is large (this is also reflected by the condition number of the matrix A [28]). What this means is that for small changes in b, the effect on x can be amplified by many orders of magnitude. Since small singular values are associated with large spatial frequencies (this is the smoothing effect of the green's function), small changes in b can result in unreasonable solutions. Due to measurement noise, a completely different solution can be obtained for the exact same algorithm. It is for this fact that we say that there is an effective null-space to the problem. Furthermore, unlike the case with true non-uniqueness, the effective null-space cannot be removed with a sufficient number of measurement data. Regularization techniques must be employed in order to filter out "bad" solutions. Regularization can be thought as a process which modifies the terms in (2.23) such that the effect of noise is dampened. Note that over-regularization is generally not good since we are ignoring information in A.

The issue of instability can also be understood physically as the presence of non-radiating or evanescent waves. Due to noise and finite numerical precision, the measured field data will never record the evanescent fields. Since any linear combination of the evanescent waves will result in (almost) the same measured data, our solution will be very susceptible to instability or effective non-uniqueness. Much of the theory and techniques in the following chapters are focused on reducing these effects of ill-posedness.

Chapter 3

Surface Field Approach

The surface field approach described in the previous chapter is an attractive option since it is a free-space method and it also uses surfacic current elements. The disadvantage of this method is the increased ill-posedness of the equivalence principle formulation from not considering the dependence of the phantom currents to the DUT currents. In this chapter, we will cover all of the details regarding the free-space surface field approach for noninvasive SAR measurements. We will start from a more detailed problem statement and then move on to a description of all numerical techniques that were used for this portion of the research thesis. Next, we will present simulation results using synthetic MATLAB and simulated ANSYS HFSS data for 2D and 3D cases. Finally, the limitations of the method will be discussed, which will lead into the next chapter which will present the novel method proposed by this thesis.

3.1 Problem Statement

We start by introducing a more detailed surface field measurement approach which is shown in Figure 3.1. The phantom and DUT reconstruction surfaces[1] are denoted by Γ and Σ

[1] A reconstruction surface is a surface upon which the equivalent currents are found (i.e., reconstructed).

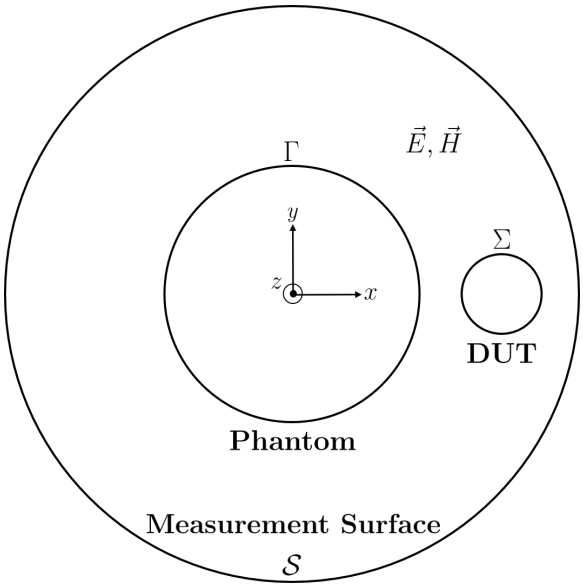


Fig. 3.1: Detailed surface field approach.

respectively. The surface upon which we perform the external measurements is denoted by $\mathcal{S}$ (which is external to Γ and Σ). All of the surfaces appear as circles/spheres, but there is no such limitation for the source reconstruction method. The surfaces Σ, Γ, and $\mathcal{S}$ can be arbitrary, although the first two are often closed surfaces.

Next, we introduce the equivalent currents denoted by $\vec{J}_\Gamma$, $\vec{J}_\Sigma$, $\vec{M}_\Gamma$, and $\vec{M}_\Sigma$ as shown in Figure 3.2. These equivalent currents are related to the true fields by the following relations:

$$\vec{J}_\Gamma(\vec{r}) = \hat{n} \times \vec{H}(\vec{r}) \quad \forall \vec{r} \in \Gamma, \tag{3.1}$$

$$\vec{J}_\Sigma(\vec{r}) = \hat{n} \times \vec{H}(\vec{r}) \quad \forall \vec{r} \in \Sigma, \tag{3.2}$$

$$\vec{M}_\Gamma(\vec{r}) = -\hat{n} \times \vec{E}(\vec{r}) \quad \forall \vec{r} \in \Gamma, \tag{3.3}$$

$$\vec{M}_\Sigma(\vec{r}) = -\hat{n} \times \vec{E}(\vec{r}) \quad \forall \vec{r} \in \Sigma. \tag{3.4}$$

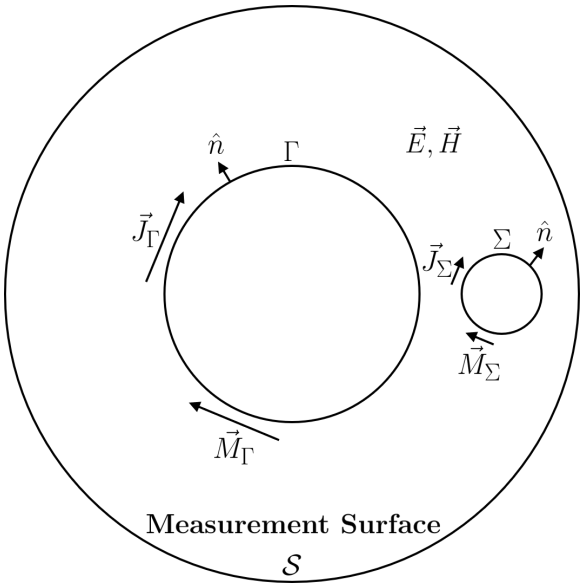


Fig. 3.2: Detailed surface field approach with equivalent currents.

By the equivalence principle, these currents radiate to produce the external fields. The total external fields can be decomposed into a component due to the electric current $\vec{J}$, and another component due to the magnetic current $\vec{M}$ such that

$$\vec{E}(\vec{r}) = \vec{E}_J(\vec{r}) + \vec{E}_M(\vec{r}), \tag{3.5}$$

$$\vec{H}(\vec{r}) = \vec{H}_J(\vec{r}) + \vec{H}_M(\vec{r}). \tag{3.6}$$

To obtain the electric and magnetic fields, we first find the vector potentials associated with the currents. If we use the Lorentz gauge for the potentials, they can be found with the following (free-space) integral equations [29]:

$$\vec{A}(\vec{r}) = \frac{\mu_0}{4\pi} \iiint_V \vec{J}(\vec{r}) \frac{e^{-jk_0|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|} dv', \tag{3.7}$$

$$\vec{F}(\vec{r}) = \frac{\varepsilon_0}{4\pi} \iiint_V \vec{M}(\vec{r}) \frac{e^{-jk_0|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|} dv'. \tag{3.8}$$

These reduce to surface integrals in the case of surface currents:

$$\vec{A}(\vec{r}) = \frac{\mu_0}{4\pi} \iint_S \vec{J}_S(\vec{r}) \frac{e^{-jk_0|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|} dS', \tag{3.9}$$

$$\vec{F}(\vec{r}) = \frac{\varepsilon_0}{4\pi} \iint_S \vec{M}_S(\vec{r}) \frac{e^{-jk_0|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|} dS'. \tag{3.10}$$

The fields can then be found by evaluating [29]:

$$\vec{E}_J(\vec{r}) = -j\omega\mu_0\vec{A}(\vec{r}) - j\frac{1}{\omega\mu_0\varepsilon_0}\nabla(\nabla\cdot\vec{A}(\vec{r})), \tag{3.11}$$

$$\vec{E}_M(\vec{r}) = -\frac{1}{\varepsilon_0}\nabla\times\vec{F}, \tag{3.12}$$

$$\vec{H}_J(\vec{r}) = \frac{1}{\mu_0}\nabla\times\vec{A}, \tag{3.13}$$

$$\vec{H}_M(\vec{r}) = -j\omega\vec{F}(\vec{r}) - j\frac{1}{\omega\mu_0\varepsilon_0}\nabla(\nabla\cdot\vec{F}(\vec{r})), \tag{3.14}$$

where "∇", "$\nabla\times$", and "$\nabla\cdot$" denote the gradient, curl, and divergence operators. Of course, we can also insert the vector potential expressions directly in order to obtain the final field equations:

$$\vec{E}_J(\vec{r}) = -\frac{j\eta_0}{4\pi k_0} \iint_{S'} \left\{ (k_0^2 + \nabla^2) \left(\vec{J}_S(\vec{r}') \frac{e^{-jk_0|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|} \right) \right\} dS', \tag{3.15}$$

$$\vec{E}_M(\vec{r}) = -\frac{1}{4\pi}\nabla\times \iint_{S'} \left\{ \vec{M}_S(\vec{r}') \frac{e^{-jk_0|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|} \right\} dS', \tag{3.16}$$

$$\vec{H}_J(\vec{r}) = \frac{1}{4\pi}\nabla\times \iint_{S'} \left\{ \vec{J}_S(\vec{r}') \frac{e^{-jk_0|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|} \right\} dS', \tag{3.17}$$

$$\vec{H}_M(\vec{r}) = -\frac{j}{4\pi k_0\eta_0} \iint_{S'} \left\{ (k_0^2 + \nabla^2) \left(\vec{M}_S(\vec{r}') \frac{e^{-jk_0|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|} \right) \right\} dS'. \tag{3.18}$$

For this research thesis, we were also interested in two-dimensional (2D) TM problems (fields invariant to the z-axis with an E_z component only as well as H_x and H_y components only). In the 2D realm, the green's function becomes a zeroth-order Hankel function of the

second kind (denoted by H_0^2) and the currents are of the form

$$\vec{J} = J_z\hat{z}, \tag{3.19}$$

$$\vec{M} = M_x\hat{x} + M_y\hat{y}. \tag{3.20}$$

Thus, the potentials become

$$\vec{A}(\vec{\rho}) = \hat{z}\frac{\mu_0}{4j}\int_l J_z(\vec{\rho}')H_0^2(k_0|\vec{\rho}-\vec{\rho}'|)dl', \tag{3.21}$$

$$\vec{F}(\vec{\rho}) = \frac{\varepsilon_0}{4j}\int_l (M_x(\vec{\rho}')\hat{x} + M_y(\vec{\rho}')\hat{y})H_0^2(k_0|\vec{\rho}-\vec{\rho}'|)dl', \tag{3.22}$$

where $\vec{\rho}$ is the radial position vector in the x-y plane. If we use this green's function and take advantage of the invariant z dimension, the field equations can be further simplified into [30]

$$\vec{E}_J(\vec{\rho}) = \hat{\mathrm{z}}\frac{-\omega\mu_0}{4}\int_{l'} J_z(\vec{\rho}')H_0^2(k_0\rho)dl', \tag{3.23}$$

$$\vec{E}_M(\vec{\rho}) = \hat{\mathrm{z}}\frac{1}{4j}\int_{l'}\left\{ - M_x(\vec{\rho}')k_0H_1^2(k_0\rho)\frac{(y-y')}{\rho} + M_y(\vec{\rho}')k_0H_1^2(k_0\rho)\frac{(x-x')}{\rho}\right\}dl', \tag{3.24}$$

$$\vec{H}_J(\vec{\rho}) = \frac{1}{4j}\int_{l'}\left\{ - \hat{\mathrm{x}}J_z(\vec{\rho}')k_0H_1^2(k_0\rho)\frac{(y-y')}{\rho} + \hat{\mathrm{y}}J_z(\vec{\rho}')k_0H_1^2(k_0\rho)\frac{(x-x')}{\rho}\right\}dl', \tag{3.25}$$

$$\vec{H}_M(\vec{\rho}) = \frac{-\omega\varepsilon_0}{4}\int_{l'}\left\{\hat{\mathrm{x}}M_x(\vec{\rho}')H_0^2(k_0\rho) + \hat{\mathrm{y}}M_y(\vec{\rho}')H_0^2(k\rho)\right\}dl', \tag{3.26}$$

where the scalar ρ represents $|\rho - \vec{\rho}'|$. We would like to use the field equations that we have obtained in order to create a system of equations that can be solved for the unknown currents. We begin by performing external field measurements and setting these to the left-hand side of the field equations. Next, the unknown currents are approximated with

appropriate basis functions as

$$\vec{J} \approx \sum_{1}^{n} c_n \vec{f}_n(\vec{r}), \vec{M} \approx \sum_{1}^{n} d_n \vec{g}_n(\vec{r}), \tag{3.27}$$

where c_n, d_n are scalar constants and $\vec{f}_n$, $\vec{g}_n$ are vector basis functions. This allows us to replace the continuous integral in the field equations with a discrete sum composed of integral terms and unknown basis function coefficients. We then construct the system of linear equations (if the measured fields are electric fields only) as

$$\begin{bmatrix} E^{\text{meas}} \end{bmatrix} = \begin{bmatrix} A_{c_n} & A_{d_n} \end{bmatrix} \begin{bmatrix} c_n \\ d_n \end{bmatrix} \tag{3.28}$$

or the more general representation

$$b = Ax. \tag{3.29}$$

3.2 Numerical treatment

Now that we have a matrix equation $Ax = b$, we can address the method for obtaining the unknown currents in vector x. If the matrix A is square and non-singular, x can be found exactly with any matrix equation solving algorithm. In our case, we do not typically have a square matrix and thus we will instead seek to minimize the cost function over x,

$$\min_{x} \left\{ ||Ax - b||_2^2 \right\}, \tag{3.30}$$

where $\|.\|_2$ denotes the L_2 norm. We will be using the conjugate gradient (CG) iterative algorithm as a means to solve (3.30). The CG method is an iterative algorithm for solving

a system of linear equations

$$Ax = b, \tag{3.31}$$

with the constraint that A is a symmetric, positive-definite (non-zero positive eigenvalues) matrix [31]. In the case of our noninvasive SAR measurement, the matrices that we deal with are typically overdetermined (due to the use of regularization terms). The most appropriate solution that we can obtain is the one that is the closest to satisfying each equation, i.e., the solution which provides the least-squares solution [31]. We can show the explicit least-squares solution by first expanding the norm in (3.30),

$$\begin{aligned} ||Ax - b||_2^2 &= (Ax - b)^H (Ax - b), \\ &= x^H A^H Ax - x^H A^H b - b^H Ax + b^H b, \\ &= f(x), \end{aligned} \tag{3.32}$$

where the superscript "H" denotes the Hermitian operator (complex conjugate transpose). The minimum of $f(x)$ is the value of x that satisfies $\partial f(x)/\partial x = 0$. Using the identity [32]

$$\frac{\partial (v^H u)}{\partial v} = \frac{\partial (u^H v)}{\partial v} = u^H, \tag{3.33}$$

we can evaluate the derivative as

$$\begin{aligned} \frac{\partial f(x)}{\partial x} &= x^H A^H A + x^H A^H A - b^H A - b^H A, \\ 0 &= x^H A^H A - b^H A, \\ x^H A^H A &= b^H A, \\ &\text{or} \\ A^H Ax &= A^H b. \end{aligned} \tag{3.34}$$

The least-squares solution can be shown to be the solution to the modified system of equations in (3.34) [31]. Coincidentally, the matrix $A^H A$ is symmetric positive definite [31] so we can use the CG algorithm to obtain the least-squares solution.

3.2.1 Conjugate Gradient Method

The following discussion is based on [31].

In general, iterative methods find the minimum to (3.30) by iteratively "searching" along a direction in the space of x. The general update equation for x at each iterative step can be written as

$$x_{i+1} = x_i + \alpha_i d_i, \tag{3.35}$$

where d_i is the current search direction and α_i is the distance along d_i, or the step size. The classification and performance of iterative methods generally depends on the choice of directions d_i. The CG algorithm is part of the family of methods known as Krylov subspace methods. A Krylov subspace is a vector space whose span is defined by the matrix-vector products of a square matrix A and vector b:

$$\text{span}\{b, Ab, A^2 b, ...A^r b\}. \tag{3.36}$$

The special property of these vectors is that they are all orthogonal as long as the subspace dimension is less than or equal to the rank of A. (Although the matrix and vector can be arbitrary, in the context of the CG algorithm they are the actual A and b in (3.31).)

The CG algorithm uses the Krylov subspace vectors as the search directions. This makes the algorithm quite efficient since with an appropriate step size, we only need to search along a given direction once, and we are guaranteed to converge within a finite number of

iterations (provided that the original matrix is symmetric positive definite).

The aspect which separates the CG algorithm from other Krylov subspace methods is the choice of the initial search direction (b in (3.36)). This is chosen to be the initial residual for an initial guess x_0:

$$d_0 = r_0 = Ax_0 - b \tag{3.37}$$

It can be shown that the step size required to minimize the cost function in the search direction is given by

$$\alpha_i = \frac{r_i^H r_i}{d_i^H A d_i}. \tag{3.38}$$

It can also be shown that the next search direction can be obtained by

$$d_{i+1} = r_{i+1} + \beta_{i+1} d_i, \tag{3.39}$$

where

$$\beta_{i+1} = \frac{r_{i+1}^H r_{i+1}}{r_i^H r_i}, \tag{3.40}$$

$$r_{i+1} = r_i - \alpha_i A d_i. \tag{3.41}$$

At each iterative step, the error between consecutive solutions $|x_{i+1} - x_i|$ can be evaluated in order to check for convergence. If the difference is sufficiently small, we can stop the CG algorithm. An alternative check is to use $r_i^H r_i$, which is responsible for the step size and implicitly represents the difference between successive solutions. Note also that when this value becomes zero, (3.40) becomes singular so using this check is useful for preventing this.

If we use the CG algorithm to solve a least-squares problem with a non-square A, we seek to solve $A^H Ax = A^H b$. We can use the CG update equation if we substitute A with

$A^H A$, and b with $A^H b$. However, since we are dealing with an ill-posed A in the noninvasive SAR characterization problem, the product $A^H A$ should never be formed (this results in an unavoidable loss of information [28]). Instead, the effective matrix $A^H A$ should be implemented as successive matrix-vector products in the CG update equation.

3.2.2 Regularization

Since the original matrix equation in the noninvasive SAR problem is ill-posed, we require regularization in order to stabilize the obtained solution. There are many different regularization strategies which can be used. There are general regularizers, which primarily seek to effectively modify[2] the coefficients in (2.23), and more physical regularizers which limit the "kind" of possible solutions, effectively changing the singular vectors in (2.23).[3] The general regularizer is very effective at dampening noise, but is unbiased to the point that over-regularizing will not produce a good solution, since there is a loss of information. On the other hand, the physical regularizer does not necessarily prevent instability, but instead limits possible solutions to those allowed by physical theorems (which can lower instability by reducing the effective null-space of the problem). For this thesis, it was found that both are required in order to obtain a good solution for SAR applications.

The general regularizer that was chosen for this thesis is Krylov subspace regularization [33]. This is performed by truncating the CG algorithm early (i.e., before convergence). It can be shown that early truncation suppresses spurious portions of the unregularized solution [33] and is similar to truncating the spectral decomposition of x in (2.23). In this sense, early truncation generally produces a smooth solution, which is reflective of the physical reality for many inverse problems. For this thesis, the iteration at which the CG algorithm is truncated is chosen in an *ad-hoc* manner.

[2] In particular, we seek to reduce the coefficients corresponding to small singular values.
[3] This can be understood by considering the generalized SVD [28].

There is not one single physical regularizer which was used for this thesis. In this chapter, several physical regularizers have been used in order to refine the solution x. These will be described in the following section. In Chapter 4, the link between the surface currents is used as the physical regularizer.

3.3 Numerical Techniques

This section will cover the numerical techniques used for the surface current method, which include the physical regularizers. Some of these are novel techniques that were found to be useful for the surface current method.

3.3.1 Current Balancing

The first important numerical technique that has been used is known as *current balancing*. In general, the electric and magnetic fields (and by extension the equivalent magnetic and electric currents under Love's condition) differ in magnitude significantly. In the very far-field (FF) region, the wavefront of any radiation will be locally planar, and the ratio between the fields is the free-space impedance [29]:

$$\eta_0 = \frac{E_{FF}}{H_{FF}} \approx 120\pi \approx 377 \text{ Ohms}. \tag{3.42}$$

It follows that the fields will have a similar ratio in the near-field region. When solving for the unknown currents, it has been shown for antenna characterization that by balancing the currents, more accurate results are obtained [34]. What we mean by balancing is to modify the system of equations so that we solve for a scaled version of one of the sets of currents.

If we choose to scale the electric current, we scale the discrete J by η and then we scale the

coefficients by $1/\eta$ to maintain equality. Visually this looks like

$$\begin{bmatrix} A_J & A_M \end{bmatrix} \begin{bmatrix} J \\ M \end{bmatrix} = E^{\text{meas}} \rightarrow \begin{bmatrix} \frac{A_J}{\eta} & A_M \end{bmatrix} \begin{bmatrix} \eta J \\ M \end{bmatrix} = E^{\text{meas}}. \tag{3.43}$$

We can also choose to scale the magnetic currents instead, which looks like

$$\begin{bmatrix} A_J & A_M \end{bmatrix} \begin{bmatrix} J \\ M \end{bmatrix} = E^{\text{meas}} \rightarrow \begin{bmatrix} A_J & \eta A_M \end{bmatrix} \begin{bmatrix} J \\ \frac{M}{\eta} \end{bmatrix} = E^{\text{meas}}. \tag{3.44}$$

After solving for the scaled currents, the true values are obtained by simply undoing the scaling. It should be noted that for general scenarios, scaling one of the currents is equivalent to scaling the other.

3.3.2 Enforcement of Love's Condition

Throughout Chapters 2 and 3, an important concept was that of Love's condition — a specific choice of the arbitrary internal fields which allows for a direct relation between the equivalent currents and the true surface fields. However, in the absence of additional information, our iterative solver will not know which set of currents is the "true" solution. In fact, due to the ill-posedness of the problem, the solution that we obtain could be any

$$\vec{J} = \hat{\text{n}} \times (\vec{H} - \vec{H}_{\text{arbitrary}}), \tag{3.45}$$

$$\vec{M} = -\hat{\text{n}} \times (\vec{E} - \vec{E}_{\text{arbitrary}}), \tag{3.46}$$

which are just the initial boundary conditions for the general equivalence principle. It is clear that we need a mechanism to enforce Love's condition.

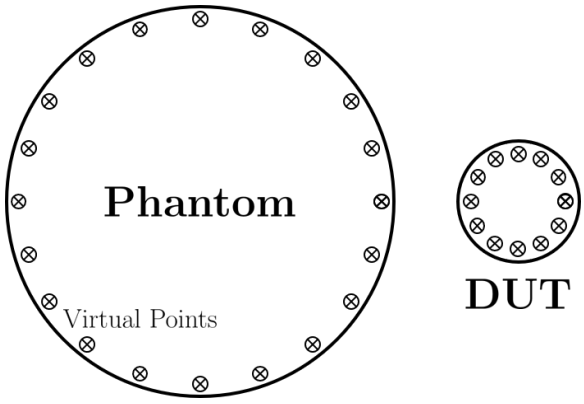


Fig. 3.3: Virtual measurement points for the enforcement of Love's condition.

Virtual Points

One method of doing this is to explicitly enforce Love's condition in the formulation. This is done by introducing virtual measurement points with null measured fields just inside the reconstruction surfaces (see Figure 3.3). With a sufficient number of points, the internal fields will be approximately zero by the uniqueness theorem. While these virtual measured fields do not reflect the true physical reality (i.e., in reality the internal fields are not zero), the true surface currents (which are equivalent to the tangential surface fields) do indeed satisfy this condition. These points and fields are treated just like the real measured data and thus we obtain an augmented matrix equation:

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} J \\ M \end{bmatrix} = \begin{bmatrix} E^{\text{meas}} \\ 0 \end{bmatrix}, \tag{3.47}$$

where "0" represents the virtual null field measurement points from the enforcement of Love's condition. Alternatively, we can consider these new measurement points and equations as a new term in our cost function,

$$\min_x \left\{ ||Ax - b||_2^2 \right\} \rightarrow \min_x \left\{ ||Ax - b||_2^2 + ||Bx||_2^2 \right\}, \tag{3.48}$$

where the matrix B is the discrete operator for the field at the virtual points. We can see from this form that the enforcement of Love's condition is actually a regularization term which is reducing the ill-posedness of the problem.

One last thing that needs to be addressed is the exact placement of the virtual points. In order to perfectly enforce Love's condition, we would need an infinite number of points[4] placed at an infinitesimal distance[5] inside of the reconstruction surface (in addition to infinitely discretized currents[6]). Due to finite numerical precision, we settle with placing a countable number of points at a sufficiently small distance from the reconstruction surface. A generally sufficient distance is $\lambda/10$ [35], but it can be shorter. The number of virtual points should generally be larger than the number of current basis functions [35].

The increased accuracy from using this method has already been demonstrated in the domain of antenna measurements [35].

Shadowing Effects

One of the disadvantages of enforcing Love's condition, however important it may be, is the introduction of *shadowing effects*. To understand what we mean by this, first recall that in typical SAR measurements, the DUT and the phantom are quite close to each other. Now let us consider the two nearest virtual points as shown in Figure 3.4. Due to the singular nature of the green's function, the nearest discretized currents will have the greatest effect on the field at the virtual points. There are essentially two possible solutions for the nearby currents: the true currents, or zero currents. Both cases will produce the null fields on the

[4] In order to have a null internal field by the uniqueness theorem, the field must be continuously zero on a closed surface (i.e., an infinite number of virtual points).

[5] In order to properly satisfy the boundary conditions, the null surface must be placed at an infinitesimal distance from the reconstruction surface.

[6] In order to have enough degrees of freedom to satisfy an infinite number of null field measurements, the currents need to be infinitely discretized.

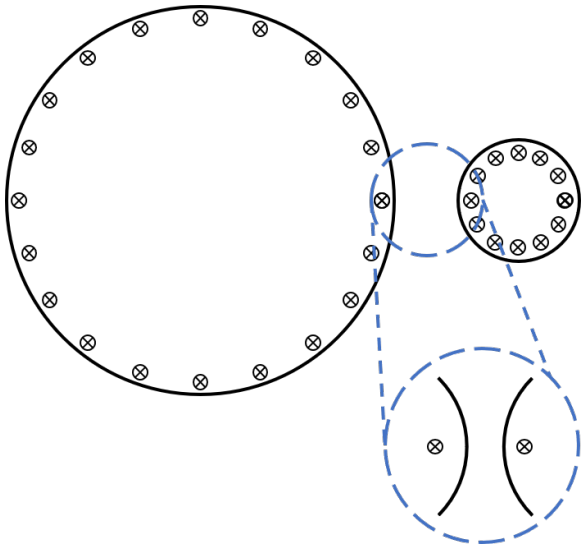


Fig. 3.4: Area on reconstruction surfaces which exhibits shadowing effects.

nearby virtual points.

When using a trivial initial guess of zero, the iterative solver will favor null currents on the portion of the surfaces which are close to each other. This results in depressed currents and fields at this location, hence the use of the term *shadowing*. Fortunately, there are a couple of techniques that can be used to reduce these effects. These are the novel techniques introduced for the surface current method.

Enforcement of Love's Condition for the Magnetic Field

The first technique to reduce shadowing effects is to add some redundancy to the enforcement of Love's condition. Under infinite precision, only considering the electric field is sufficient, but for this research thesis we found that enforcing null magnetic fields inside can reduce shadowing effects.[7] We simply use the operators in (3.13)–(3.14) and add additional

[7] When only enforcing Love's condition for the electric field, it was often found that the internal magnetic field was not a null field. The logical conclusion is to make this enforcement.

virtual measurement points. The augmented equations now look like

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \\ A_{31} & A_{32} \end{bmatrix} \begin{bmatrix} J \\ M \end{bmatrix} = \begin{bmatrix} E^{\text{meas}} \\ 0_E \\ 0_H \end{bmatrix}, \tag{3.49}$$

where the subscript E or H denotes the virtual measurement point for the electric field or magnetic field, respectively. Similarly to the case with the null electric field, the newly added equations are just like another additional term to the cost function,

$$\min_x \left\{ ||Ax - b||_2^2 + ||Bx||_2^2 + ||Cx||_2^2 \right\}, \tag{3.50}$$

where the matrix C is the discrete operator for the magnetic field at the virtual points.

Relaxation of Love's Condition on the DUT

The final technique to reduce shadowing is something that is only possible because of the surface field approach. Since we only really need the fields on the phantom surface, we can relax Love's condition on the DUT surface. While this does not necessarily reduce the shadowing effects as explained in the Figure 3.4, it does reduce shadowing by instead increasing the degrees of freedom to satisfy the null fields inside the phantom. The end result is a potential reduction in shadowing effects on the phantom.

3.3.3 Regularization Weights

At this point, we have added two new regularization terms: one which enforces Love's condition for the electric field, and another which does the same enforcement for the magnetic field. If we consider each term equally, each condition will be met equally. However, we only really want to meet the first term of the cost function, which is the real data. Thus, it is useful to weight these terms in order to favor the measured data. We introduce regularization

weights λ_B and λ_C such that the cost function becomes

$$\min_x \left\{ ||Ax - b||_2^2 + \lambda_B^2 ||Bx||_2^2 + \lambda_C^2 ||Cx||_2^2 \right\}. \tag{3.51}$$

Although there are automated methods of weighting regularization terms (see Section 4.2.1), we limit ourselves to choosing the weights in an *ad-hoc* way for this chapter.

3.4 Discretization

In this section, the discretization rules and strategies that were used for this thesis are discussed.

Current Basis Functions

In the 2D case, the basis functions chosen for the currents were pulse basis functions. In the 3D case, the currents are represented by the RWG basis functions [36].

Maximum Discretization Size

For general electromagnetic problems, the discretization size is typically $\lambda/10$, where λ is the smallest wavelength in the problem (in the case of multiple dielectrics). Since the surface current method is a free-space algorithm, the maximum discretization size is related to the free-space wavelength. However, the effect of the discretization size on the results was not thoroughly investigated for this thesis and we generally choose a size much smaller than $\lambda/10$ to avoid any issues that arise from insufficient discretization.

Minimum Number of Measurements

Since the noninvasive SAR measurement algorithm is equivalent to a two-antenna near-field measurement, the minimum number of measurements is likely similar to the antenna

measurement case [29, 37]. The effect of the number and spacing of measurement points was not investigated for this thesis.

3.5 Obtaining the Internal Fields

The last step in the noninvasive SAR measurement when using a surface field approach is to obtain the internal fields. Due to the uniqueness theorem, we can use any boundary value method that we want. For this thesis, we use the equivalence principle to formulate the equivalent problem under Love's condition in the region internal to Γ.

3.5.1 Equivalence Principle Formulation

Figure 3.5 shows the complementary equivalent problem for the fields internal to the phantom using a reconstruction Γ_2 which surrounds the phantom.

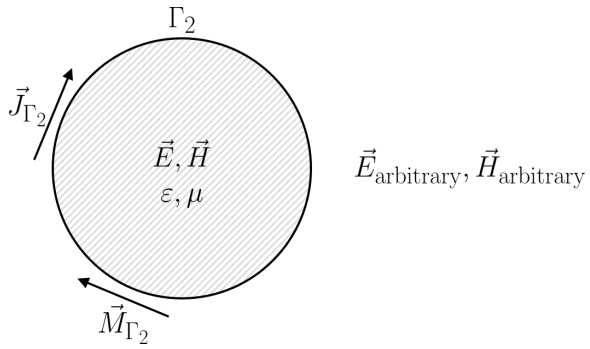


Fig. 3.5: Equivalent problem for the internal phantom fields.

The boundary conditions for this problem are

$$\vec{J}_{\Gamma_2} = \hat{n} \times (\vec{H}_{\text{arbitrary}} - \vec{H}), \tag{3.52}$$

$$\vec{M}_{\Gamma_2} = -\hat{n} \times (\vec{E}_{\text{arbitrary}} - \vec{E}), \tag{3.53}$$

where $\hat{n}$ is once again the unit outer normal on Γ_2. We set the arbitrary fields to the null

fields, invoking Love's condition to obtain

$$\vec{J}_{\Gamma_2} = \hat{n} \times (-\vec{H}), \tag{3.54}$$

$$\vec{M}_{\Gamma_2} = -\hat{n} \times (-\vec{E}). \tag{3.55}$$

If we choose $\Gamma_2 = \Gamma$, then the equivalent currents for this problem are just the negative of the currents that we solve for in our algorithm:

$$\vec{J}_{\Gamma_2} = \hat{n} \times (-\vec{H}) = -\vec{J}_{\Gamma}, \tag{3.56}$$

$$\vec{M}_{\Gamma_2} = -\hat{n} \times (-\vec{E}) = -\vec{M}_{\Gamma}. \tag{3.57}$$

By solving for the surface fields initially, we can just solve a forward problem in order to obtain the internal fields and thus the SAR distribution.

Under general circumstances, we require the green's function in the presence of the phantom; since this is what we were trying to avoid, solving the forward problem like this is not ideal. Fortunately, there is one way to avoid this. Since the fields outside of Γ are null fields by Love's condition, we can replace the free-space medium with the phantom material. As long as the reconstruction surface overlaps the physical phantom surface, the green's function is just the free-space version with the substitution $k_0 \rightarrow k$. We can then construct an operator similar to A, where the observation domain is inside the phantom, and evaluate a matrix-vector product to obtain the internal fields.

3.6 Results

To evaluate the free-space surface field approach, simulations were performed for 2D TM and fully vectorial 3D problems. The algorithm described in this chapter was fully implemented

for the 2D case and partially implemented for the 3D case. The results will reveal the strengths of the algorithm, as well as uncover the limitations which lead to the conclusion that this may not the preferred method for a noninvasive SAR measurement.

3.6.1 Synthetic Data Generation

For the simulation results, we require synthetic data that we would expect from the real problem. In the 2D case, this was generated using a method of moments (MoM) based algorithm [23]. Under this framework, the domain of interest includes the phantom which is discretized into 2D pulse basis functions. The total field inside the domain of interest is obtained by solving a Fredholm integral equation of the second kind. Once these fields are obtained, the external fields (which are used as the synthetic measurement data for the SAR problem) can be obtained from a simple matrix-vector product. The grid spacing for the pulse basis functions was chosen to be $\lambda/50$, where λ is the wavelength inside of the phantom. The total grid side length was 24 (cm) centered at the origin. The results using the synthetic data were verified by also using simulation data from ANSYS HFSS.

3.6.2 2D TM Problem

We start by reintroducing the phantom-DUT system with surfaces Σ, Γ, and $\mathcal{S}$ in Figure 3.6. We have also introduced d, which is the distance separating the surfaces Σ and Γ, and also the surfaces Σ^- and Γ^-, upon which we enforce Love's condition. For the 2D simulation, each the surfaces are in fact circles. The phantom itself is also circular and the DUT being used is an infinite line source with unit current magnitude.

In order to represent a realistic scenario, the phantom relative complex permittivity was chosen to be $\varepsilon_r = 41.5 - j19.4$, which corresponds to the value recommended by the IEEE standard [5] for the frequency $f = 900$ (MHz). The phantom itself has a radius between

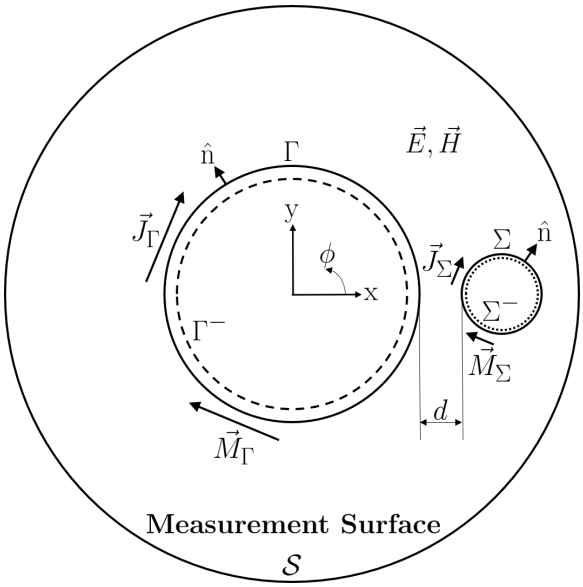


Fig. 3.6: Reconstruction and measurement surfaces for the 2D TM problem.

10.8 and 12 (cm); the maximum extent depends on the pulse basis function discretization size for the synthetic data generation.[8] The choice of 12 (cm) approximately corresponds to the largest dimension of the SAR phantom [5]. The reconstruction surface Γ has a radius of 12.1 (cm). The reconstruction surface Σ has a radius of 1 (cm) and is centered on the DUT which is along the x-axis.

In terms of discretization, there are 360 evenly spaced receivers on the surface $\mathcal{S}$ with a radius of 75 (cm). There are 360 pulse basis functions on Γ and 36 pulses on Σ. The Love's surfaces are concentric to their respective surfaces and are inset by the arc length between pulses on the reconstruction surfaces.

[8] In the MoM solver that was used, the phantom was generated by checking the pulse center position against a chosen radius. This means that large pulses can extend past the chosen radius, which was found to cause numerical artifacts if the pulses also extend past the reconstruction surface. The choice of 10.8 (cm) corresponds to the maximum phantom radius allowed for a grid spacing of $\lambda/10$ and a reconstruction surface radius of 12.1 (m). Although a finer spacing is used in the results, we will show that the SAR distribution for the true fields is approximately the same.

Two choices of d will be used for the results: 16.9 and 0.9 (cm). These choices of d will illustrate the effects of shadowing on the ill-posedness of the problem, and thus the achievable results. A summary of the problem parameters is given in Table 3.1.

Table 3.1: Problem parameters for the 2D TM simulation using MATLAB generated data.

Parameter	**Value**
R_Γ	12.1 (cm)
R_Σ	1 (cm)
R_{Γ^-}	R_Γ - $2\pi R_\Gamma/N_\Gamma$ (cm)
R_{Σ^-}	R_Σ - $2\pi R_\Sigma/N_\Sigma$ (cm)
$R_\mathcal{S}$	75 (cm)
N_Γ	360
N_Σ	36
$N_\mathcal{S}$	360
ε_r	$41.5 - j19.4$
f	900 (MHz)

3.6.3 2D TM Synthetic Results

In order to justify the numerical techniques outlined in this chapter, we will show a progression of results beginning from the fundamental case which does not use any of the numerical techniques from Section 3.3. A spacing of $d = 16.9$ (cm) is chosen in order to reduce shadowing effects and thus isolate the effect of the numerical techniques.

We start by solving the system $Ax = b$ by itself. Figure 3.7 shows the magnitudes of the surface fields on Γ obtained by solving the ill-posed system. These fields are juxtaposed with the true distributions. We immediately see the problems caused by the ill-posedness of the problem, which results in bad field distributions. Note that these fields still result in the observed electric field values in b. Figure 3.8 shows the absolute relative percent error for each entry in $\dfrac{b - Ax^*}{b}$, where x^* are the currents (or fields) obtained from the algorithm. This result helps to illustrate the problems that come with inverse problems; even if the

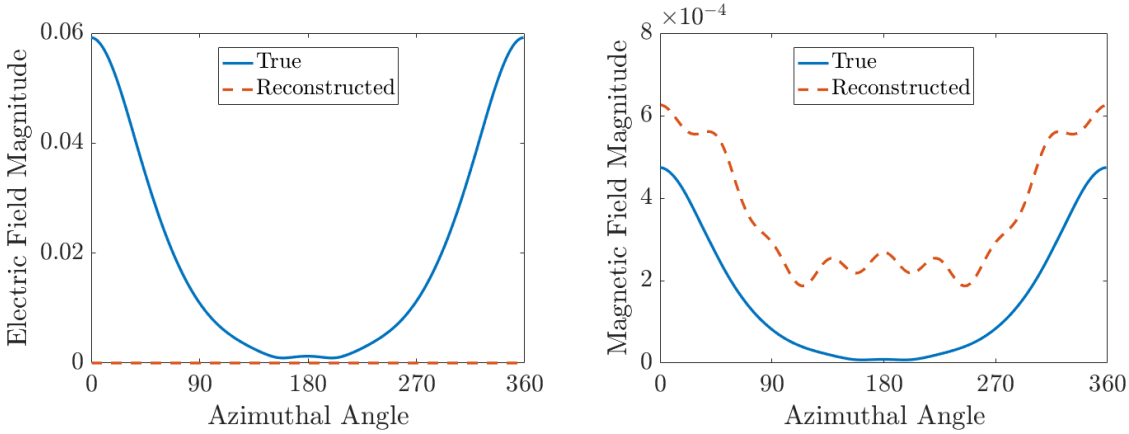


Fig. 3.7: Reconstructed phantom surface a) electric field and b) magnetic field magnitudes for the fundamental problem compared to the true distributions.

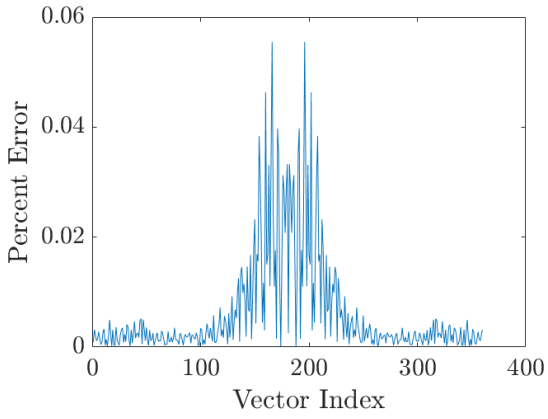


Fig. 3.8: Absolute relative percent error for each entry in $\frac{b - Ax^*}{b}$.

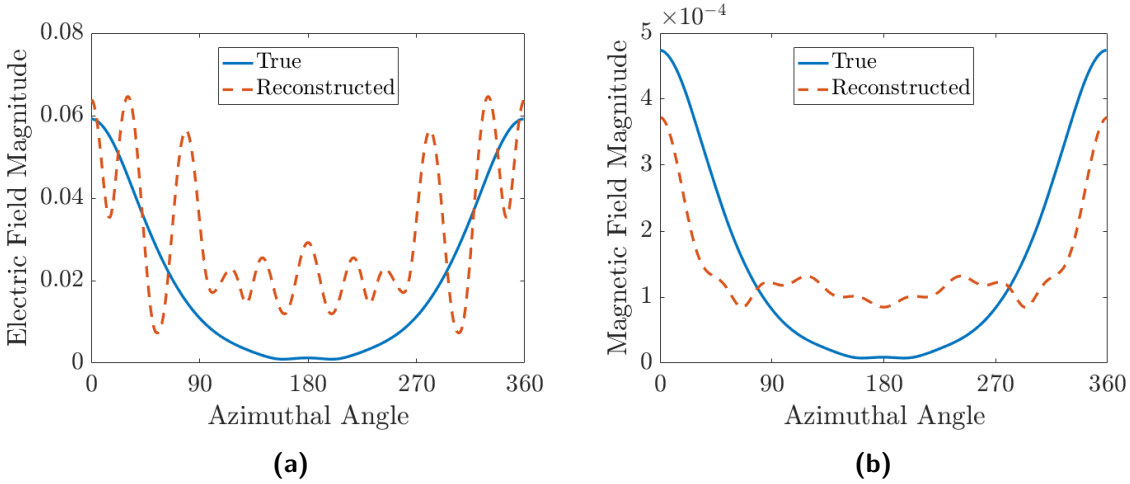


Fig. 3.9: Reconstructed phantom surface a) electric field and b) magnetic field magnitudes using current balancing compared to the true distributions.

data is met, the obtained solution can be completely wrong. As a corollary, we see evidence of the presence of non-radiating fields which greatly affect the results.

One of the more striking aspects of these results is the near-zero electric field. This is due to the lack of current balancing, which results in uneven current accuracy. The next step in our progression is thus to add the current balancing. Figure 3.9 shows the magnitudes of the phantom surface fields when balancing the currents. Although these results aren't very useful, they are much closer to the desired fields.

In order to increase the accuracy of the reconstructed fields, we begin to add more information regarding the true fields (we can also think of this as filtering out the incorrect non-radiating fields). The first thing that we add is the enforcement of Love's condition for the electric field. In order to use this condition effectively, we must add a weight to the equations. Due to finite discretization, the internal fields produced by even the true currents are not quite zero, so we must favor the original equations. Using a weight of $\lambda_B = 1/100$, the phantom surface fields were obtained and are presented in Figure 3.10.

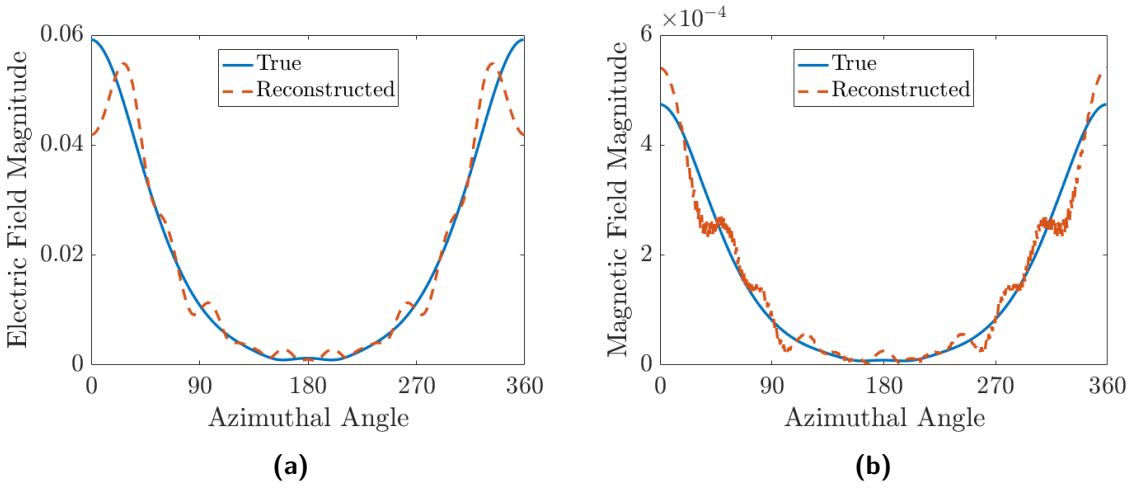


Fig. 3.10: Reconstructed phantom surface a) electric field and b) magnetic field magnitudes when enforcing Love's condition for the electric field compared to the true distributions.

Note that this choice of weight is not necessarily the best choice. Nevertheless, the obtained fields are quite decent although it is unlikely to result in an accurate SAR distribution.

When comparing the two plots in Figure 3.10, we notice that the curve for the magnetic field "looks" different. The dashed line makes it difficult to see but there is a small field component with a very large spatial frequency. We can filter out this component by employing Krylov subspace regularization (i.e., early truncation of the CG algorithm). Truncating at 100 CG iterations, the phantom surface fields were obtained and are presented in Figure 3.11. The early truncation successfully removed the erroneous components, and also slightly increased the peak electric field accuracy.

In order to further improve the accuracy, we also enforce Love's condition for the magnetic field. Since the magnetic field is generally of a small magnitude, we use a larger weight for these equations. Figure 3.12 shows the reconstructed phantom surface fields for the weights $\lambda_B = 1/100$ and $\lambda_C = 1/5$. These fields are now quite close to the true values. The relative normed error for the electric field ($||E_z^{\text{true}} - E_z^{\text{rec}}||_2/||E_z^{\text{true}}||_2$, where "rec" indicates the

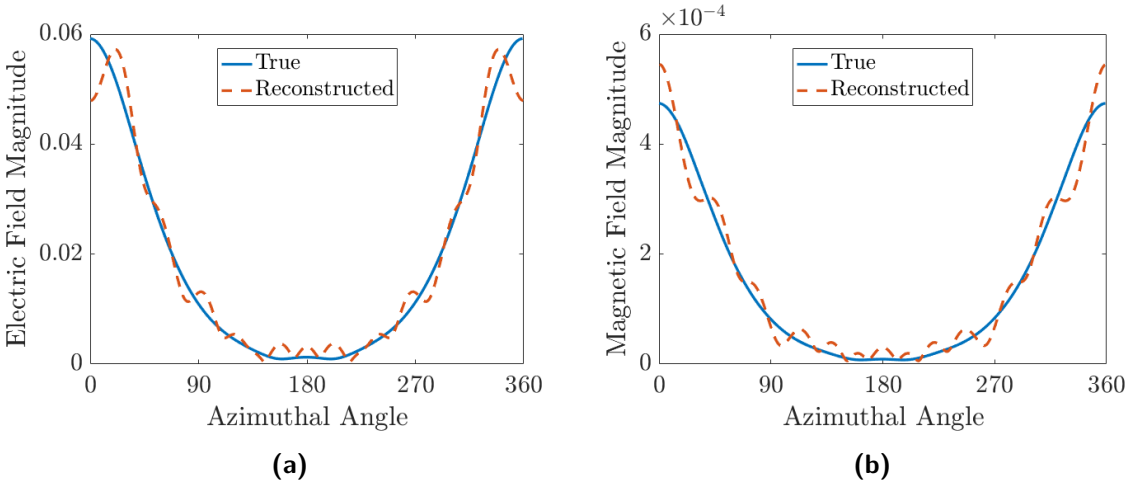


Fig. 3.11: Reconstructed phantom surface a) electric field and b) magnetic field magnitudes using current balancing compared to the true distributions.

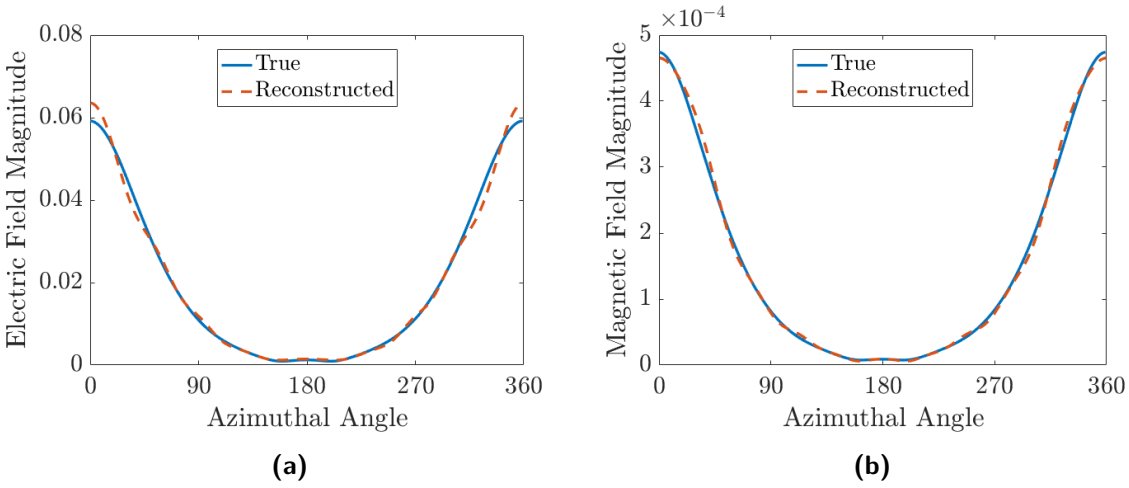


Fig. 3.12: Reconstructed phantom surface a) electric field and b) magnetic field magnitudes when enforcing Love's condition for both fields compared to the true distributions.

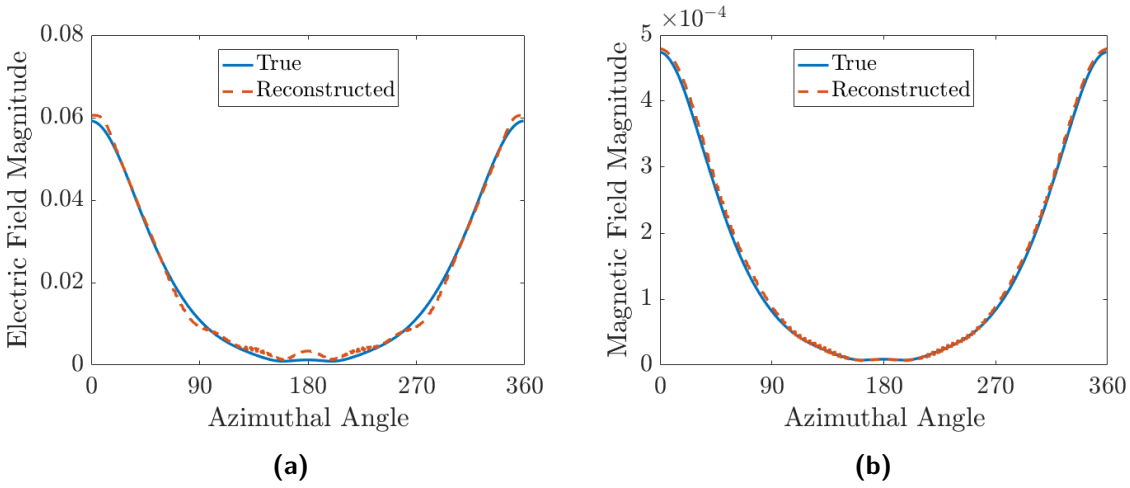


Fig. 3.13: Reconstructed phantom surface a) electric field and b) magnetic field magnitudes when using all techniques compared to the true distributions.

reconstructed field) was 10.4% and the relative normed error for the magnetic field was 4.1%.

There remains one numerical technique to use, which is the unenforcement of Love's condition for the DUT. This provides more freedom on the DUT currents to satisfy Love's condition for the phantom currents.[9] Using this technique, and by using the weights $\lambda_B = 1/100$ and $\lambda_C = 3$, we obtain the results shown in Figure 3.13. The relative normed field errors have now decreased to 5.2% for the electric field and 1.1% for the magnetic field. These are actually quite good and would likely produce a good SAR distribution. However, a spacing of $d = 16.9$ (cm) is not realistic, so the surface current method has not yet been validated.

A much more realistic choice for d is 0.9 (cm) (which results in a DUT position of $x = 14$ (cm)). By using the regularization parameters given in Table 3.2, we obtain the phantom surface fields shown in Figure 3.14. Note that for this particular problem, it was found that the unenforcement of Love's condition was not beneficial, so it is not employed. The

[9]Due to finite discretization, the internal field is never quite zero. By relaxing Love's condition on the DUT, we can get closer to the remaining condition on the phantom.

Table 3.2: Regularization parameters for the $d = 0.9$ (cm) simulation.

Parameter	Value
λ_B	1/20
λ_C	3
CG iterations	30

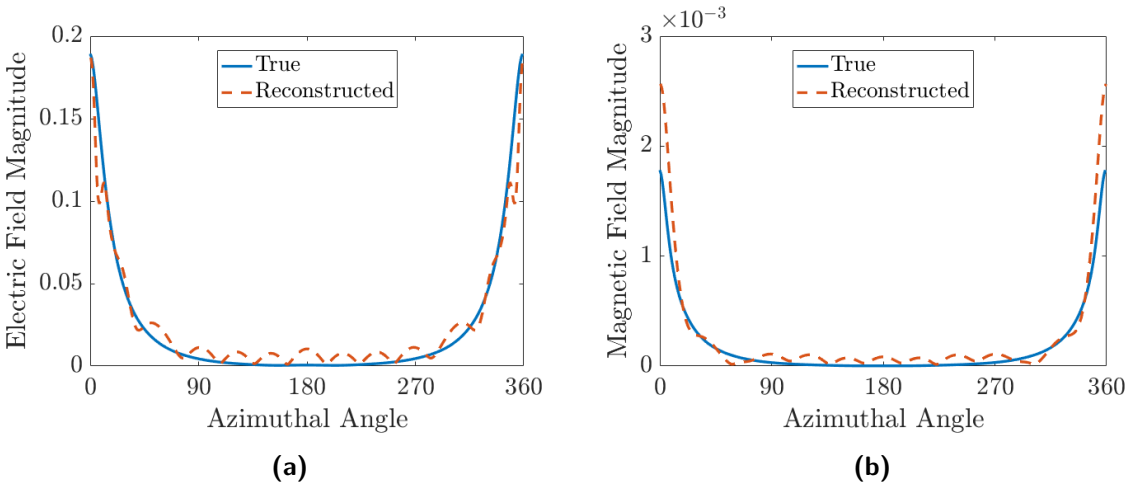


Fig. 3.14: Reconstructed phantom surface a) electric field and b) magnetic field magnitudes for $d = 0.9$ (cm) compared to the true distributions.

reconstructed fields are worse when compared to the previous results, which is due to the increased ill-posedness[10] of the problem. The relative normed error was 9.2% for the electric field and 61.1% for the magnetic field. The large error in the magnetic field is partly due to phase discrepancies. The results are also quite sensitive to the parameters in Table 3.2, which is a major disadvantage of the uncoupled surface field method.

The last step after obtaining the surface phantom fields is to obtain the internal SAR distribution. As previously mentioned, there is a slight offset between the phantom and the reconstruction surface Γ, which could potentially invalidate the formulation in Section 3.5. We start by comparing the normalized true and reconstructed SAR distributions, shown in

[10]Since the phantom and DUT currents are closer, the columns of A are increasingly near-parallel which results in a problem which is more ill-posed.

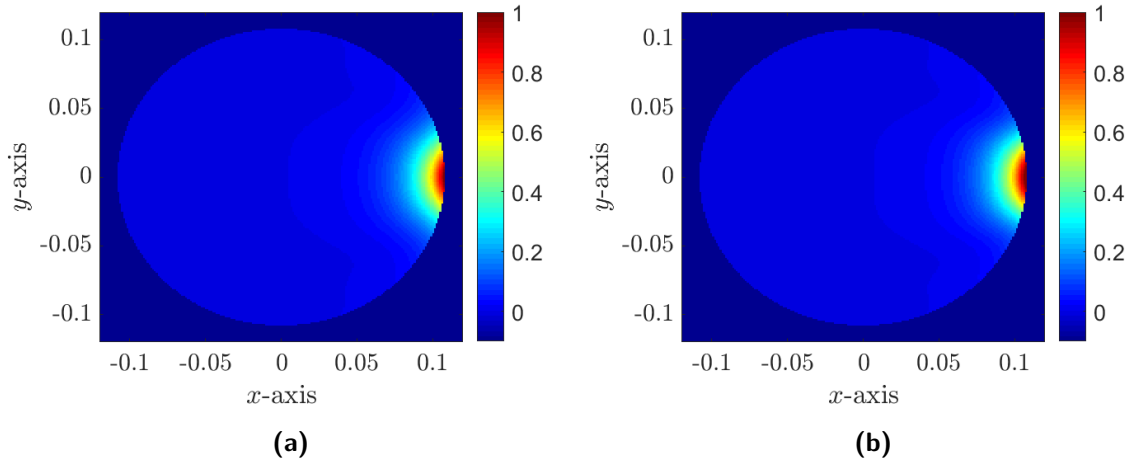


Fig. 3.15: Normalized SAR distributions for the a) true internal electrical field and b) true fields on Γ.

Figure 3.15. The true distribution is obtained by using the internal electric field from the MoM forward solver, and the reconstructed distribution is obtained with the true surface currents on Γ. The figure clearly shows that the offset is not a problem and that with the exact surface fields on Γ, an accurate SAR distribution is obtained. The peak relative SAR error $(\max\{SAR^{\text{rec}}\} - \max\{SAR^{\text{true}}\}/\max\{SAR^{\text{true}}\})$ was 7.2%. In a practical case we would want the reconstruction surface Γ to be as close to the phantom surface as possible, but for now the placement is sufficient.

Using the reconstructed currents in Figure 3.14, the resulting SAR distribution is also fairly good. Figure 3.16 shows the comparison of the normalized true and reconstructed SAR distributions. Since SAR measurements are localized to the location of maximum SAR, we are particularly interested in the accuracy of the distribution near the DUT. To start, the peak relative error was 4.3%, which is quite good, even if the surface fields were not as accurate (the magnetic field in particular). We can also look at the relative percent error for each pixel in the SAR distribution, shown in Figure 3.17a. What this shows is that the SAR distribution is fairly accurate in the area nearest to the DUT, and wildly

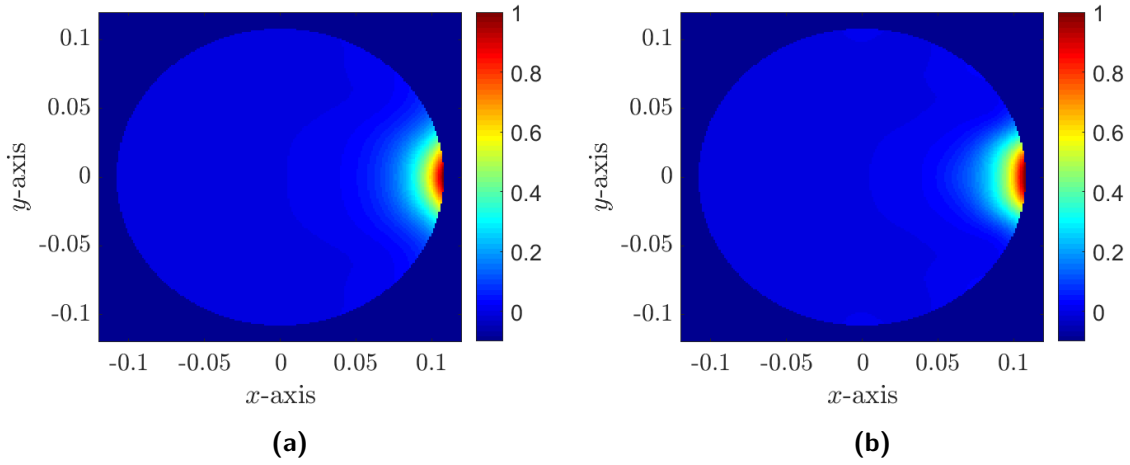


Fig. 3.16: Normalized SAR distributions for the a) true and b) reconstructed fields.

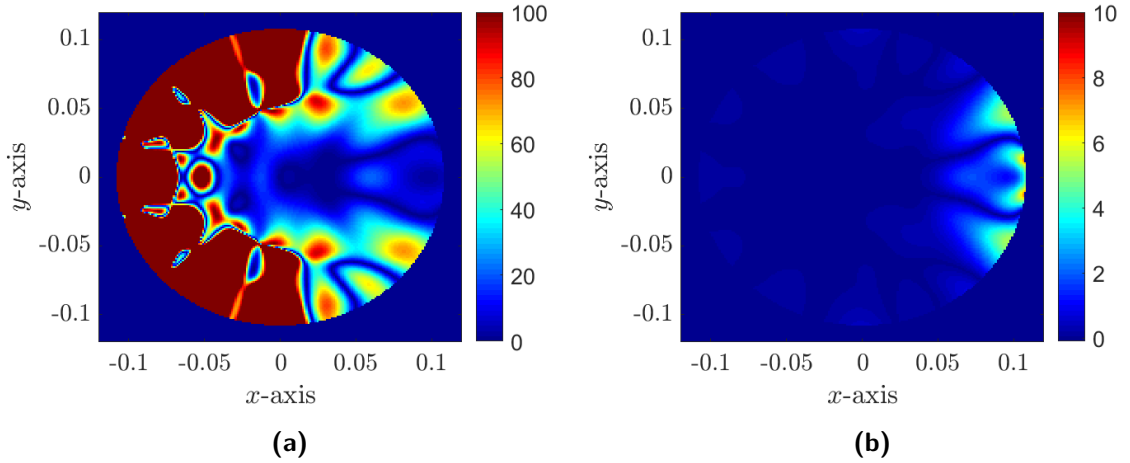


Fig. 3.17: Percent error for each pixel in the SAR distribution relative to a) the true pixel field and b) the maximum true pixel field. Note that the color scales differ in magnitude.

inaccurate in the area furthest away. In practice, this could be completely acceptable, since the inaccurate regions correspond with the areas with minimum SAR values (i.e., we are not concerned with these values). This is emphasized in Figure 3.17b, which shows the pixel percent error relative to the maximum pixel field. (Note that the color scales differ in magnitude for each subplot.)

3.6.4 2D TM HFSS Results

In this section we will present results for the same problem shown previously, but with data generated through ANSYS HFSS in order to validate the synthetic results. The problem parameters are the same as the synthetic case with $d = 0.9$ (cm), with the exception that the phantom radius is actually 12 (cm).

In order to simulate a 2D problem on 3D modelling software, we can employ a clever technique using image theory. To this end, a short phantom cylinder and a coaxially-fed monopole antenna are sandwiched between two large parallel perfect electric conductor (PEC) plates. The PEC plates are chosen to be large in order to approximate infinitely large PEC planes. By image theory, we can remove the PEC planes if we periodically extend the antenna (including currents) and phantom to infinity in the normal directions to the PEC planes. There will be some approximation error due to the finite planes, but overall it is a good approximation. Figures 3.18 and 3.19 show the normal and magnified views of the HFSS model. The rectangular boxes along the x-axis are non-model objects that are used to visualize the fields and ensure that there is an E_z component only. To further validate the approximation, a simulation without the phantom was performed. Electric field values were collected on a ring of radius 1 (cm) centered on the DUT (coinciding with Σ), and on a ring of radius 75 (cm) (coinciding with $\mathcal{S}$). These field values were compared to a *scaled* infinite line source. Figures 3.20 and 3.21 show the scaled comparisons for the two cases described. Note that in Figure 3.21, the DUT is not centered about the observation domain, which results in an asymmetric pattern. Overall, we see a fairly good match, although there is some error due to the approximation (this includes the error occurring from finite discretization and also the HFSS convergence criterion).

Using the regularization parameters in Table 3.3, we are able to obtain the surface phantom

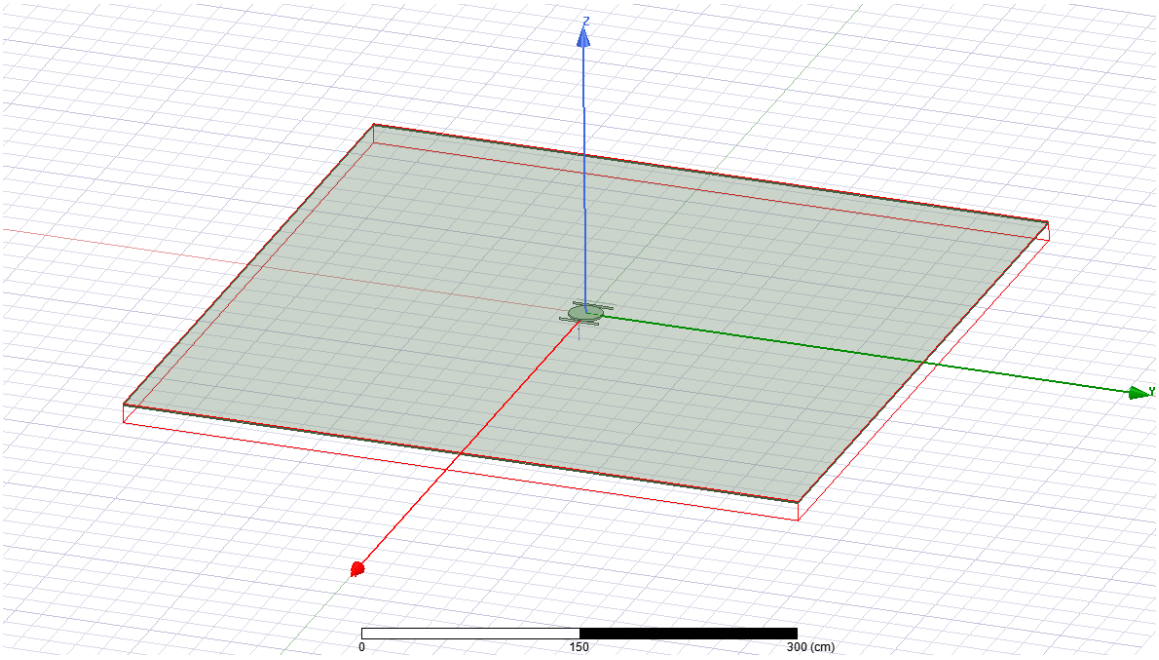


Fig. 3.18: HFSS model for 2D approximation. The z-axis is blue, the y-axis is green, and the x-axis is red. [Image used courtesy of ANSYS, Inc.]

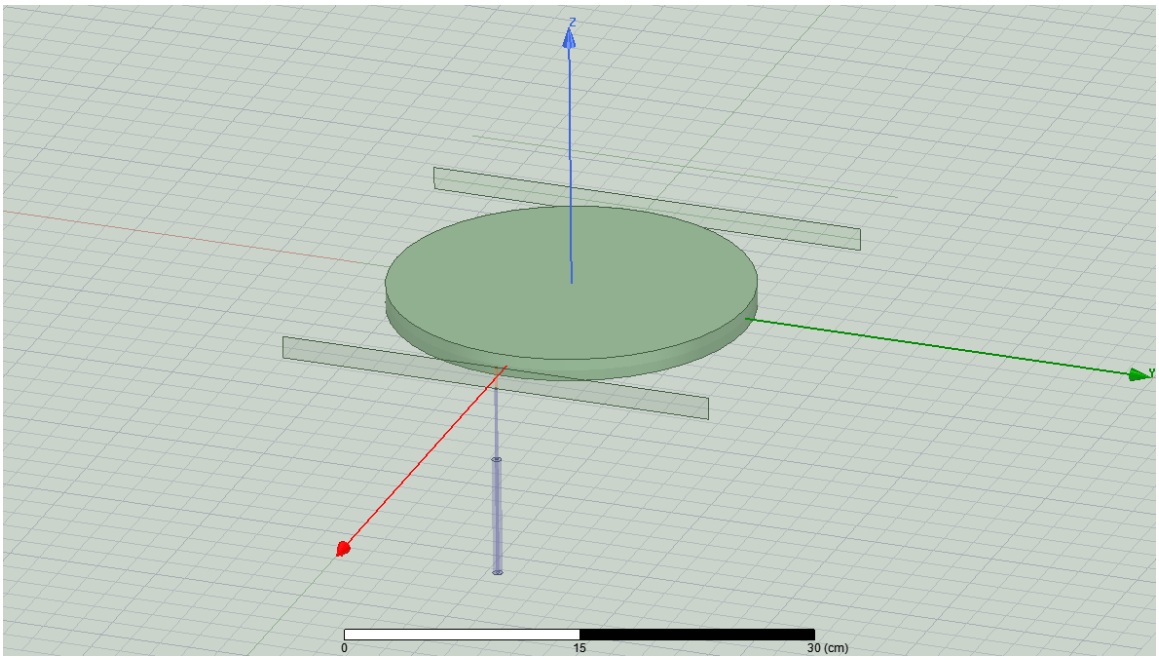


Fig. 3.19: Magnified HFSS model for 2D approximation. The z-axis is blue, the y-axis is green, and the x-axis is red. [Image used courtesy of ANSYS, Inc.]

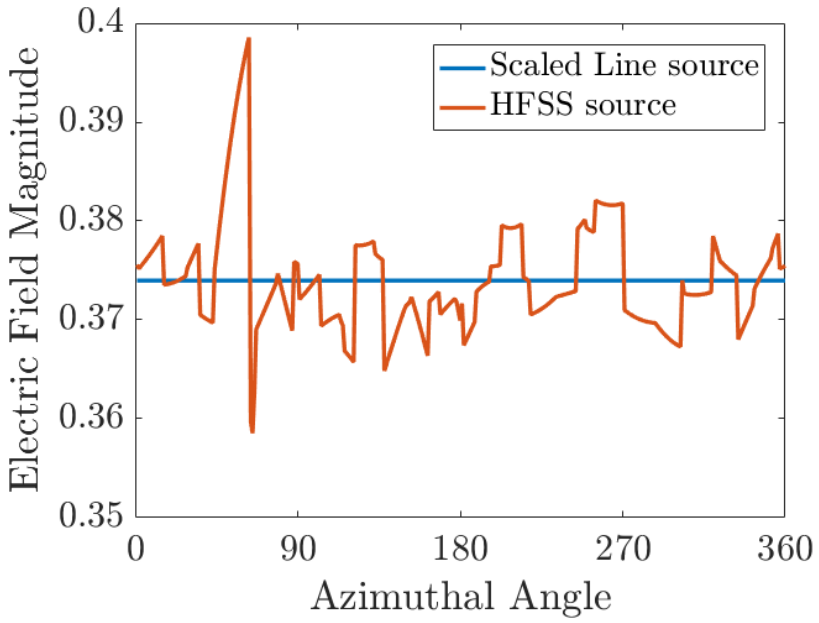


Fig. 3.20: Comparison between the approximate HFSS infinite line source and a scaled ideal infinite line source about the observation domain Σ.

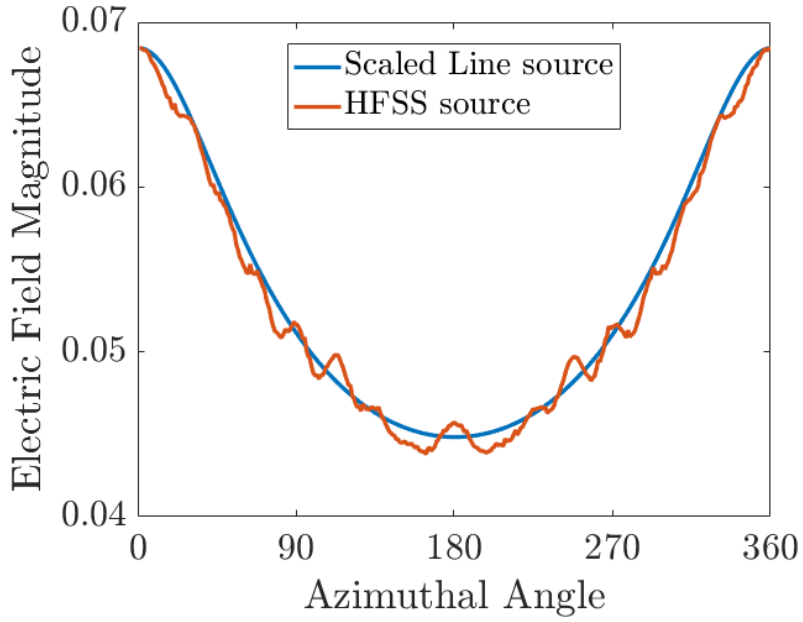


Fig. 3.21: Comparison between the approximate HFSS infinite line source and a scaled ideal infinite line source about the observation domain $\mathcal{S}$.

Table 3.3: Regularization parameters for the HFSS simulation.

Parameter	Value
λ_B	1/50
λ_C	3
CG iterations	55

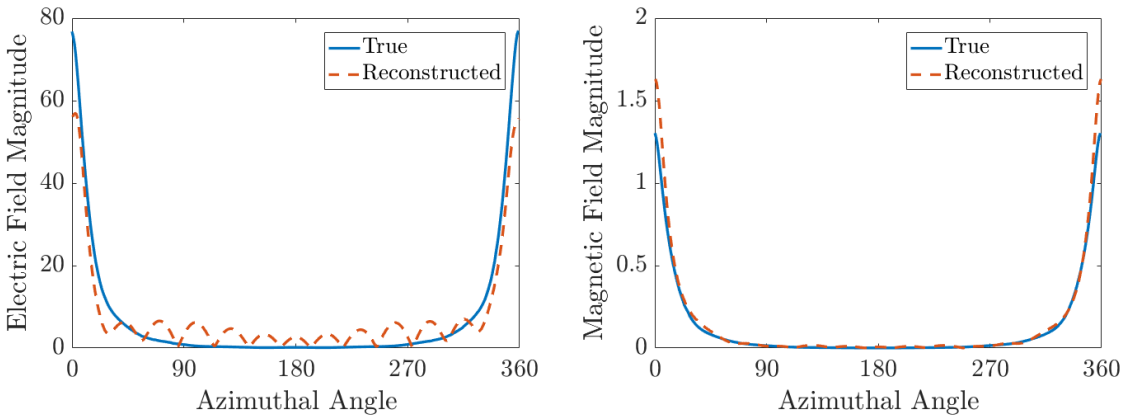


Fig. 3.22: Reconstructed phantom surface a) electric field and b) magnetic field magnitudes using HFSS data compared to the true distributions.

fields shown in Figure 3.22. Compared to the results for synthetic data, these results are much worse. However, this is to be expected since we already know that the 2D approximation is not perfect. In terms of magnitude only, the relative normed errors were 26.0% for the electric field and 25.1% for the magnetic field. In terms of magnitude and phase, the normed errors were 81.0% for the electric field and 78.2% for the magnetic field. This indicates that the reconstructed fields were out of phase, and that the SAR algorithm has difficulty with phase reconstruction. When the observed field is far, phase differences for source currents are greater than magnitude differences so this result makes sense.

Next, we would like to look at the SAR distributions. Figure 3.23 shows the normalized SAR distributions for the true and reconstructed fields. Although the phantom surface fields were quite inaccurate in terms of phase, the SAR distribution, which relies on mag-

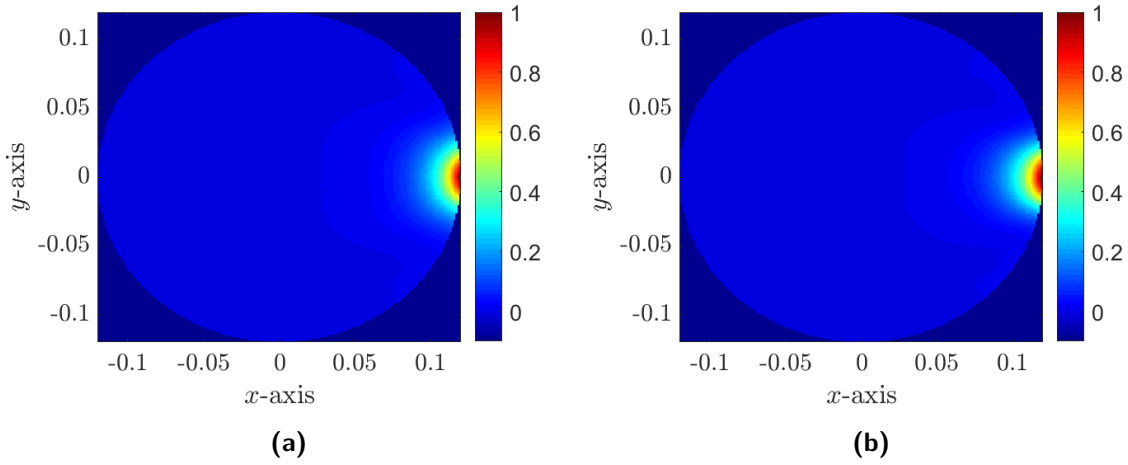


Fig. 3.23: Normalized SAR distributions for the a) true and b) reconstructed fields using the HFSS simulation data.

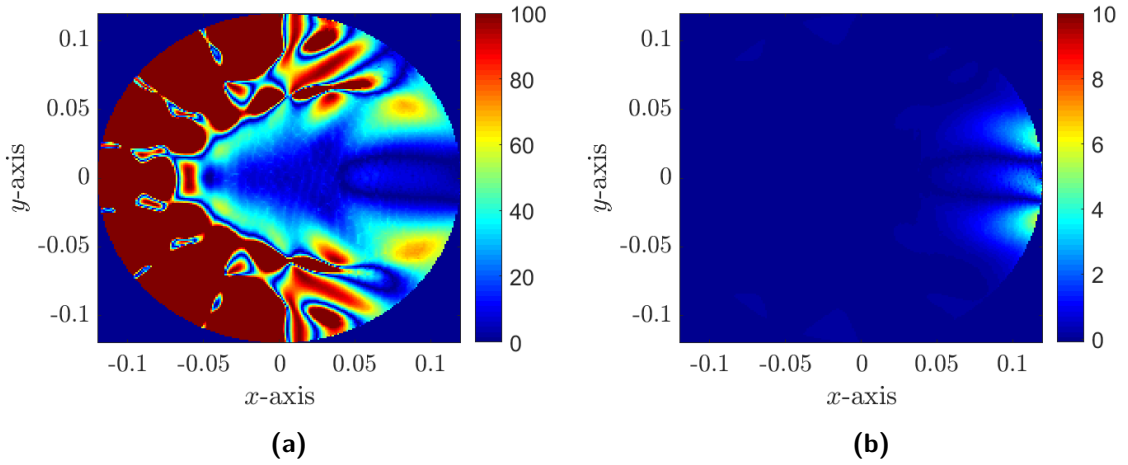


Fig. 3.24: Percent error for each pixel in the SAR distribution relative to a) the true pixel field and b) the maximum true pixel field. Note that the color scales differ in magnitude.

nitude, appears to be fairly good. The peak SAR error for these results was 0.97%, which is very good. Figure 3.24 shows the percent error for each pixel in the SAR distribution. Again, we see a similar kind of behavior as with the synthetic results — the area near the DUT is fairly accurate, and the rest is inaccurate. Due to the finite volume required by standard SAR measurements, this distribution may be accurate enough for a noninvasive

SAR measurement. Note that the accuracy may be coincidental, since it is clear from the magnitudes of each field component that the errors at $\phi = 0$ are effectively cancelled, which results in the correct SAR value.

3.6.5 Preliminary 3D Results

Having completed a 2D simulation with synthetic MATLAB data and HFSS simulated data, the next step is to perform a 3D simulation. A fully vectorial 3D noninvasive SAR algorithm was implemented which only employed the current balancing and the unweighted enforcement of Love's condition on the electric field. Although the full algorithm was not implemented, the results presented in this section are useful to see if there are any added complications from a 3D space and more importantly, a non-trivial DUT.

We re-introduce the 2D problem setup in Figure 3.25, which may be applied to the 3D case by assuming each circular domain is a sphere. Since some of the necessary numerical

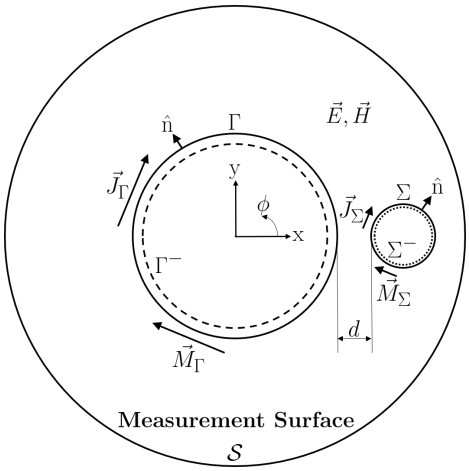


Fig. 3.25: 3D problem setup. All of the circular domains are spheres.

techniques have not been implemented, some of the problem parameters have been modified to account for this; in general, sizes and distances have been increased. A summary of the

Table 3.4: Problem parameters for the 3D simulation using HFSS simulated data.

Parameter	Value
R_Γ	12.1 (cm)
R_Σ	12.1 (cm)
Max Edge Length on Γ, Σ	$\lambda/10$
R_{Γ^-}	R_Γ - $\lambda/10$ (cm)
R_{Σ^-}	R_Σ - $\lambda/10$ (cm)
Max Edge Length on Γ^-, Σ^-	$\lambda/15$
$R_\mathcal{S}$	100 (cm)
$\Delta\phi$ on $\mathcal{S}$	3°
$\Delta\theta$ on $\mathcal{S}$	3°
d	25.8 (cm)
ε_r	41.5 - j19.4

problem parameters can be found in Table 3.4. The DUT reconstruction surface Σ is now the same size as Γ to accommodate the DUT which is now a half-wavelength dipole oriented along the z-axis. The separation distance has been increased to 25.8 (cm) since not all of the necessary numerical techniques have been implemented (this results in a DUT centered at $x = 50$ (cm)). The reconstruction surfaces were meshed with a maximum edge length of $\lambda/10$ (using the open-source software Gmsh [38]). The Love's surfaces were meshed with a maximum edge length of $\lambda/15$. The measured data for this problem were generated using ANSYS HFSS. The electric field was sampled on a sphere of radius 100 (cm) with points spaced by $\Delta\phi = \Delta\theta = 3^\circ$.

The reconstructed surface fields on Γ were compared to the values obtained through HFSS. For this case, only the magnetic field could be reconstructed reasonably well. Without some of the additional numerical techniques, there was a magnitude mismatch but the distribution itself was reconstructed fairly well. Figures 3.26, 3.27, 3.28, and 3.29 show the "front", "right", "back", and "bottom" views of the phantom (referenced from the perspective of the DUT) for both the reconstructed and HFSS phantom surface magnetic fields.

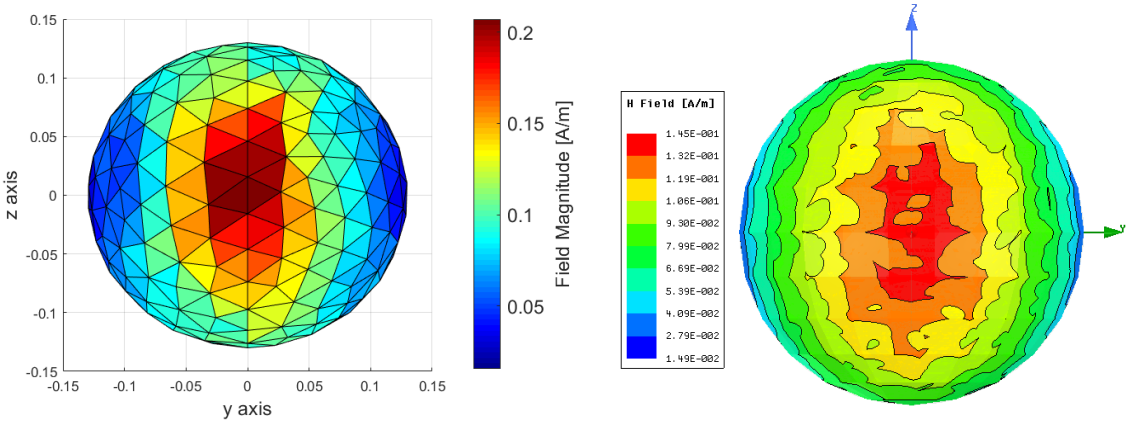


Fig. 3.26: The a) reconstructed and b) simulated phantom surface magnetic field magnitudes for the "front" view. [Image used courtesy of ANSYS, Inc.]

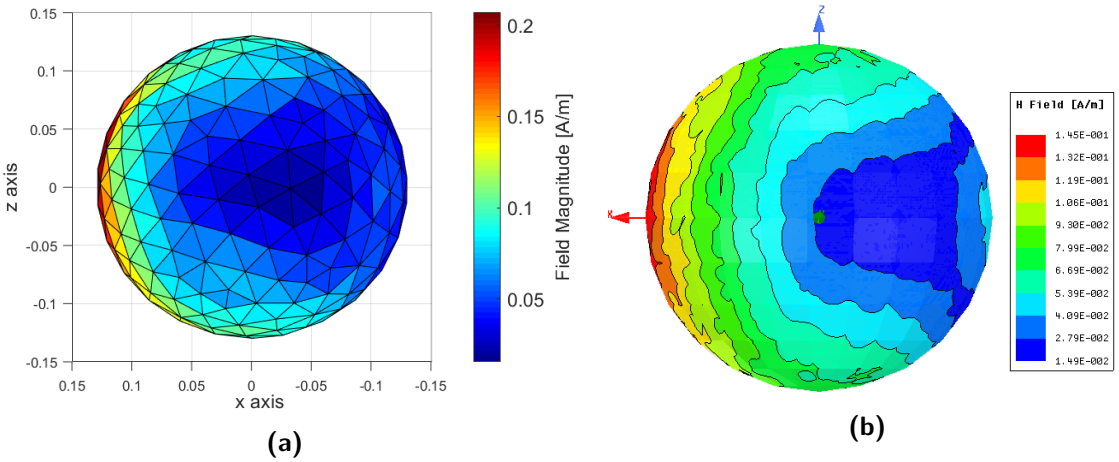


Fig. 3.27: The a) reconstructed and b) simulated phantom surface magnetic field magnitudes for the "right" view. [Image used courtesy of ANSYS, Inc.]

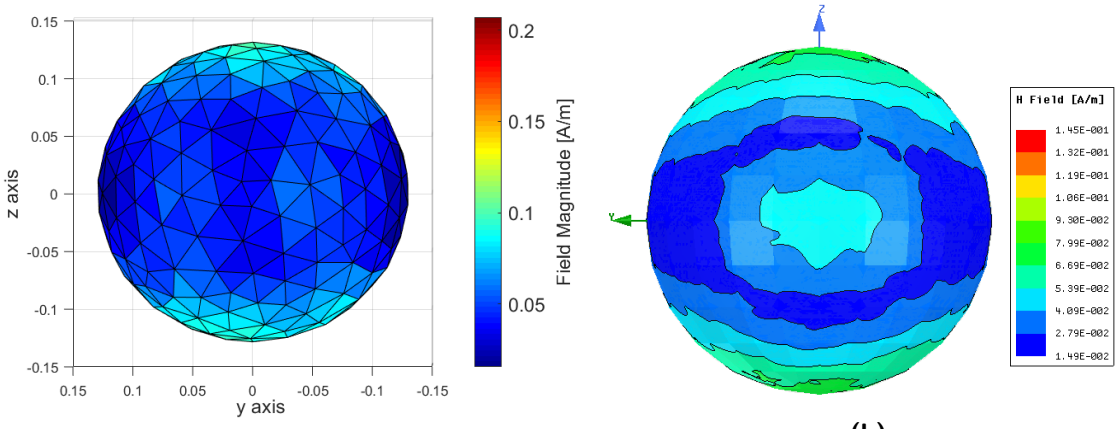


Fig. 3.28: The a) reconstructed and b) simulated phantom surface magnetic field magnitudes for the "back" view. [Image used courtesy of ANSYS, Inc.]

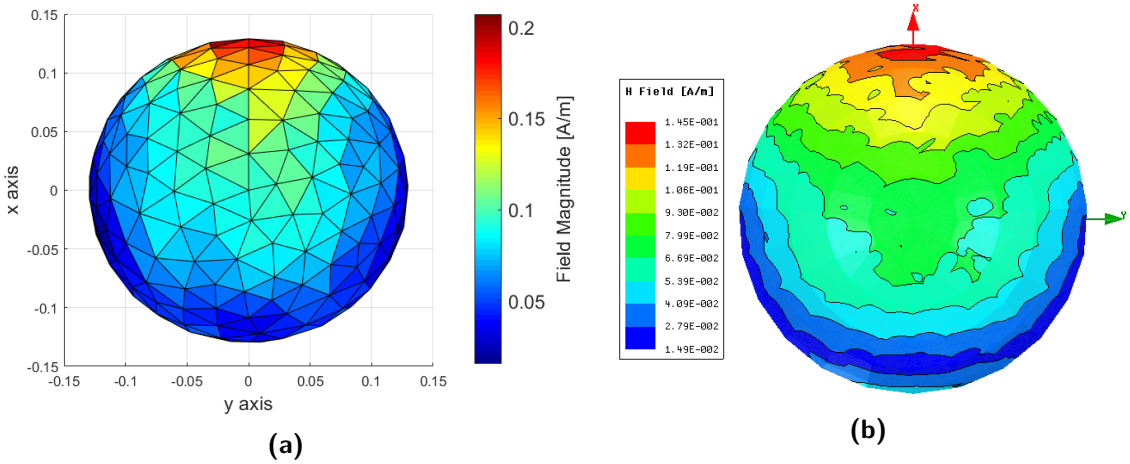


Fig. 3.29: The a) reconstructed and b) simulated phantom surface magnetic field magnitudes for the "bottom" view. [Image used courtesy of ANSYS, Inc.]

It should be noted that the HFSS and MATLAB color scales are slightly different even though they use the same colors; the HFSS scale has more green while the MATLAB scale has more blue. Recognizing this difference, the shape of the reconstructed distribution is quite good, with the only discrepancy being the lack of a local maximum at the back of the phantom. Since there were no hints of major problems for the 3D problem, we speculate that results similar to the 2D case could be achieved if all of the numerical techniques are implemented.

3.7 Discussion

In this chapter, the surface current method for the noninvasive SAR characterization problem was formulated and presented. Results were shown in the 2D case for both synthetic MATLAB and HFSS simulated data and preliminary results were shown in the 3D case for HFSS simulated data. In the 2D case, it was quite difficult to obtain accurate surface fields, especially in terms of phase. In addition, it was found that the tolerance on the regularization parameters was very tight, which is not good for a practical scenario. Despite this, the reconstructed SAR distributions were quite good. It is important to note that none of the examples presented added any measurement noise, which is a practical consideration. Given the tight tolerances on the regularization parameters, it is likely that the surface current method would be too ill-posed to be practical. In order to reduce the ill-posedness, we require more information. This can take the form of extra measurement data, or an initial guess to the CG algorithm. However, it may not be practical to take many measurements or to know a good initial guess in a realistic scenario. Fortunately, there is one other piece of information that has not been used: the link between the phantom and DUT currents. The incorporation of this link into the surface current method forms the basis of the following chapter.

Chapter 4

Regularized Simultaneous Inversion Method

In this chapter, we present the formulation and evaluation of the novel noninvasive SAR measurement method. We will begin by providing a high-level discussion on the benefits of using the current link, and the two options regarding its implementation (one of which is used for the novel algorithm). Following this discussion will be the formulation of the proposed method, including any necessary numerical techniques. Next, we will provide the theoretical framework which motivates choosing this method over alternatives. In particular, we make the argument that the method is particularly resistant to noise. However, we do not make the claim that it is the "best" method and instead provide factors within the theoretical framework which would determine the ideal method in a practical scenario. Finally, we will provide results which form a preliminary verification of the theoretical framework and also investigate the general effect of noisy measurement data.

4.1 Relationship Between the Currents

It has been highlighted many times that the surface current method in the previous chapter did not assume any dependence between the currents, which in reality they do. By not enforcing this condition, we allow equivalent currents which produce the measured fields but do not satisfy the inherent relation between the phantom and DUT surface fields. This results in an inverse problem where it can be difficult to obtain the true physical solution.

The relation between the DUT and phantom currents can be obtained by taking advantage of the field/current equivalence under the equivalence principle (and Love's condition in particular). By considering the tangential phantom surface fields on Γ and the DUT surface currents on Σ, the relation between the currents is the forward relation

$$y = Bx, \tag{4.1}$$

where $y = [H_\Gamma; E_\Gamma] = [J_\Gamma; M_\Gamma]$, $x = [J_\Sigma; M_\Sigma]$, and B is the modified[1] discretized green's function operator which takes into account the scattering from the phantom.[2]

Before we get into the implementation, we would like to briefly discuss the benefits of using this relation with regards to non-uniqueness and instability. Under the general equivalence principle, the DUT currents are related to the external fields $(\vec{E}, \vec{H})$ and the arbitrary internal fields $(\vec{E}_a, \vec{H}_a)$ by

$$\vec{J}_\Sigma = \hat{n} \times (\vec{H} - \vec{H}_a), \tag{4.2}$$

$$\vec{M}_\Sigma = -\hat{n} \times (\vec{E} - \vec{E}_a), \tag{4.3}$$

[1] There is an additional implied operator which extracts the tangential field components.
[2] A semicolon indicates the concatenation of vectors which are assumed to be column vectors.

where $\hat{n}$ is the unit outer normal vector. If we try to reconstruct the DUT currents, we can obtain almost anything due to the arbitrary nature of the internal fields and from the instability of inverse problems. Consider instead the infinite precision case with infinitely discretized phantom and DUT currents where the evanescent fields are perfectly captured by infinite measured data. If we enforce Love's condition on the phantom currents with an infinite number of virtual points at an infinitesimal distance from the phantom reconstruction surface, then the reconstructed phantom fields using the surface current method will necessarily be the true fields by the uniqueness theorem. (The DUT currents remain non-unique due to arbitrary internal fields). Alternatively, we can also reconstruct the surface currents while enforcing the current link. Although the DUT currents can be non-unique, the fields external to the DUT must be unique since the external surface field in the boundary condition does not suffer from instability and is thus unique. This means that the current link enforcement will also produce the true phantom surface fields, i.e., the current link will implicitly enforce Love's condition. Since we only require the phantom surface fields, the two techniques are in fact conceptually equivalent for the noninvasive SAR problem.

However, in the case of finite numerical precision, the previous statements do not hold. In general, non-radiating phantom fields with respect to the measurement domain can be reconstructed. The only thing which may prevent this (i.e., reduce the effective null space) is the enforcement of Love's condition or the current link. Although these produce the same results conceptually, it can be argued that the current link will be better at recovering the true fields. To understand this argument, we must look at the forms of the two techniques. Under the enforcement of Love's condition, the regularization term is $A_E y = 0$, where A_E is the discretized green's function operator for the virtual measurement points. Under the current link, the regularization term is $Bx = y$. The true phantom fields y are enforced

through a forward relation under the current link, but they are enforced through an inverse relation under the enforcement of Love's condition. Thus, there is an additional degree of non-uniqueness or instability when enforcing Love's condition. (Note that the degree of ill-posedness is low since the virtual points are very close to the phantom currents.) We speculate that it is more likely that the current link can obtain the true phantom fields. (In Section 4.5, the simultaneous inversion method using the current link is used to solve the same problem in Chapter 3 for $d = 0.9$ (cm). The results are noticeably better than those in Chapter 3.)

Once the current link B is obtained, we have a choice vis-à-vis its implementation; we can perform a substitution in the original surface current matrix equation in Chapter 3, or we can leave the relation as an explicit regularization term. Let $A = [A_y \; A_x]$ such that $A_y y + A_x x = b$, where b represents the externally measured fields. The procedure for the substitution method can be summarized as follows:

1. Construct the operator A under the surface current formulation in Chapter 3.
2. Construct the operator B which represents the current link.
3. Substitute B into $Ax = b$ such that $A_y Bx + A_x x = b$.
4. Solve the inverse problem for the DUT currents x.
5. Obtain the phantom surface fields by evaluating $y = Bx$, and finally obtain the internal SAR distribution.

The procedure for the simultaneous inversion (SI) method is summarized as follows:

1. Construct the operator A under the surface current formulation in Chapter 3.
2. Construct the operator B which represents the current link.

3. Simultaneous invert the phantom and DUT currents, x and y, while enforcing $y = Bx$ as a regularization term.

4. Use the phantom currents to obtain the internal SAR distribution.

The substitution method is similar to the characterization of a single antenna element, whereas the SI method is similar to the characterization of two dependent antenna elements. The formulation of the SI method will be reviewed in detail in the following section.

4.2 Formulation

For convenience, we redefine the DUT currents $x = [J_\Sigma; M_\Sigma]$, the phantom currents $y = [J_\Gamma; M_\Gamma]$, and introduce the total currents $z = [y; x]$. We then have the fundamental equation

$$Az = b, \tag{4.4}$$

where b represents the externally measured fields, and A represents the discretized free-space green's function operator. We also have the current relation

$$Bx = y, \tag{4.5}$$

where B represents the modified[3] discretized *inhomogeneous* green's function operator. In order to use (4.5) as a standard regularization term, we take the difference equation

$$Bx - y = 0, \tag{4.6}$$

and use the implied matrix

$$C = \begin{bmatrix} -I & B \end{bmatrix}, \tag{4.7}$$

[3]There is an additional implied operator which extracts the tangential field components.

where I is an appropriately sized identity matrix. Since C satisfies

$$Cz = 0, \tag{4.8}$$

we can use this as a regularization term by minimizing the cost functional over the total currents z,

$$\min_z \{||Az - b||_2^2 + \lambda_C^2 ||Cz||_2^2\}, \tag{4.9}$$

where λ_C is the regularization weight.

The matrix B can be obtained by formulating the forward problem for the surface currents and fields, which implicitly requires the green's function in the presence of the phantom. Fortunately, there is also the option to use Mitharwal's method [24, 25] which is based on the PMCHW scattering formulation which can obtain B without the explicit use of the inhomogeneous green's function. Due to limiting ourselves to 2D problems, in this chapter we use the MoM code which generated the synthetic results in order to form the matrix B. Details on this can be found in Section 4.2.2.

After B is obtained, the last step is to determine the regularization weight. With only one parameter to determine, we will utilize the technique known as the *L-curve* method [28] which will allow us to find an appropriate weight for the algorithm. (The surface current method in Chapter 3 required more than one parameter, which is why the weights were chosen in an *ad-hoc* way.)

4.2.1 L-curve Method

Consider a general regularization method with a general regularizing term L, where we seek to minimize the functional over z

$$\min_z \{||Az - b||_2^2 + \lambda^2 ||Lz||_2^2\}. \tag{4.10}$$

The regularization weight λ represents a trade-off between satisfying the data equation $Az = b$ and the regularization term $Lz = 0$. The *L-curve* is a plot of the data norm versus the regularizer norm. Figure 4.1 shows a typical L-curve. For small values of λ, the data

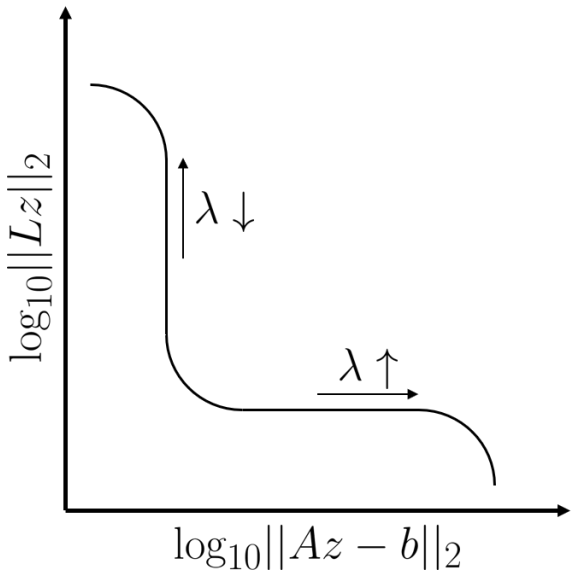


Fig. 4.1: General L-curve plot. For small values of λ, $\|Az - b\|_2$ is primarily satisfied. For large values of λ, $\|Lz\|_2$ is primarily satisfied. The corner or elbow of the curve represents the optimal λ which is the compromise to satisfying both norms.

equation is satisfied well, but the solution is not correct since the regularizer is not satisfied (i.e., we get the under-regularized solution). When λ is large, the opposite happens and the solution is again not correct (i.e., the solution is over-regularized). The corner or elbow of the curve represents the best compromise to satisfying both terms and is generally close to

the optimal solution [28].

The corner of the L-curve can be determined mathematically by finding the point of maximum curvature [28]. For general L-curves which do not necessarily have nice shapes, finding the corner is a very complicated matter which is beyond the scope of this thesis. We will instead find the optimal point by directly comparing the reconstructed phantom surface fields to the true distribution. We will visually confirm that the optimal point occurs at the physical corner of the L-curve so that we may conclude that it would be possible to write an algorithm which can find the corner.

4.2.2 Implementation

In this section, some of the last few details regarding the actual implementation used for the results in this Chapter are discussed.

Current Balancing

In Chapter 3, many additional numerical techniques were incorporated into the surface current method. For the SI method, we retain only one of these, which is the current balancing. This technique was found to be a critical adjustment for the SI method; without it, good results were not possible. Since the currents are not equal in magnitude, in the unbalanced case the electric current error has little to no effect in the term $||Cz||_2^2$ which results in poor currents. However, in the case of the substitution method, it was found that the currents should be unbalanced in order to achieve the best results. It was also found that the unbalanced substitution method was generally more sensitive to added measurement noise. Both the balanced and unbalanced substitution methods will be compared to the SI method in the results.

Obtaining the Current Link

For this thesis, we use the MoM-based forward solver (which was used to obtain the synthetic measurement data) to obtain the current link B. Due to the use of acceleration and compression methods, we did not have access to the explicit matrix. In order to obtain the explicit matrix, we can iteratively form each column of the matrix by running the MoM solver for a corresponding column in the identity matrix. For example, the first column of B corresponds to the phantom fields obtained when using $[1; 0; ...0]$ as the DUT currents. The second column is obtained by using $[0; 1; 0; ...0]$, and so on.

Measurement Spacing

As described in Chapter 3, the maximum spacing allowed for the measurement is likely similar to the case for antenna measurements. Once again, the maximum spacing is not investigated in this chapter. An angular spacing of $\Delta\varphi = 1^\circ = \pi/180$ is maintained for all of the results in this chapter.

Measurement Noise

In the results section, the performance of the SI and substitution methods will be compared when adding white noise to the measurement data. We define the noise level (NL) of the added noise (n) to be the ratio between the maximum noise magnitude added and the minimum synthetic measurement magnitude:

$$\mathrm{NL} = \frac{\max\{|n|\}}{\min\{|E^{\mathrm{meas}}|\}}. \tag{4.11}$$

The random noise itself is a complex quantity which may be expressed as

$$n = \frac{\mathrm{NL} \cdot \min\{|E^{\mathrm{meas}}|\}}{\sqrt{2}}\big(n_R + jn_I\big), \tag{4.12}$$

where n_R and n_I are uniform random numbers between -1 and 1.

CG Algorithm

As in Chapter 3, the CG algorithm is used to minimize the cost function of the problem. We always assume a trivial initial guess of zero for all of the results in this section. When not employing Krylov subspace regularization, the CG algorithm stops at a tolerance of $r_i^H r_i < 10^{-16}$ or at the maximum number of iterations which is set to the spectral size of the matrix.[4]

4.3 Motivation for Simultaneous Inversion

In this section, we present the theoretical framework supporting the choice of the SI method, i.e., the motivation for using this method. We make two main propositions in this section: 1) the SI method is particularly resistant to noise, and 2) the best algorithm (SI or substitution) for a given scenario depends on several factors. Four factors have been identified which can contribute to the relative performance between the SI and substitution methods. While some of these may be related, each of the factors are distinct in that they can be evaluated with their own metric. These factors are:

1. Ill-posedness of B versus the ill-posedness of A,
2. Ill-posedness of B with respect to A,
3. Accuracy of B,
4. Spectral decomposition of the true phantom surface fields.

[4]By spectral size we refer to the number of spectral components in the matrix which is equivalent to the number of singular values. Alternatively, it is equivalent to the maximum dimension of the Krylov subspace for the least-squares problem which is also the maximum number of iterations required for the CG algorithm to converge under infinite precision arithmetic.

These conditions will be described in the following sections and will be verified and evaluated in the results.

4.3.1 Factor 1

The first factor is concerned with the relative ill-posedness of the separate A and B matrices. Consider a general matrix equation which can be decomposed into

$$A_y y + A_x x = b, \tag{4.13}$$

where y and x are unknowns, A_y and A_x are discrete operators, and b is known data. Furthermore, there is a known relation between the two sets of unknowns,

$$Bx = y. \tag{4.14}$$

Suppose that A_x and A_y are ill-posed operators; due to the presence of small singular values, these matrices are effectively low-pass spatial frequency filters. Next suppose that we add a high spatial frequency component to x:

$$x = \overline{x} + \beta s, \tag{4.15}$$

where $\overline{x}$ is the true x, β is a real constant, and s is a high spatial frequency vector corresponding to a small singular value in A_x (i.e., the spatial frequency of s is above the effective cut-off frequency of A_x). If we evaluate the matrix vector product $A_x x$,

$$\begin{aligned} A_x x &= A_x \overline{x} + A_x \beta s, \\ &\approx A_x \overline{x} + 0, \\ &= A_x \overline{x}. \end{aligned} \tag{4.16}$$

If the filtering is sufficiently strong, the added signal βs can be quite large without changing the product. If we try to solve (4.13) by itself, there is nothing preventing the addition of a large βs to the solution (this is just another demonstration of the instability of ill-posed problems).

Suppose that the relation B is actually well-posed, which means that B does not behave like a low-pass filter for spatial frequencies. If we evaluate the matrix vector product Bx for the candidate solution,

$$\begin{aligned} Bx &= B\overline{x} + B\beta s, \\ &= B\overline{x} + \beta s', \end{aligned} \tag{4.17}$$

where s' is a new high spatial frequency signal scaled by the appropriate singular values of B. Unlike in the case with the ill-posed A_x, the effect of s is retained.

If we choose to substitute B into (4.13), we see that despite having a well-posed B, the total system will remain ill-posed. Using the candidate solution in the first term,

$$\begin{aligned} A_y Bx &= A_y B\overline{x} + A_y B\beta s, \\ &= A_y B\overline{x} + A_y \beta s', \\ &\approx A_y B\overline{x} + 0, \\ &= A_y B\overline{x}. \end{aligned} \tag{4.18}$$

We see that the well-posed B does not prevent a high spatial frequency component from being inverted when trying to solve for x. In other words, the two norms

$$||A_x x + A_y y - b||_2^2, \tag{4.19}$$

and

$$||A_x(\overline{x} + \beta s) + A_y B(\overline{x} + \beta s) - b||_2^2, \tag{4.20}$$

can result in practically the same norm which means that the inverted x will suffer from instability. On the other hand, if we choose to leave the relation B as a regularization term (as it is in the SI method), the combined norm becomes

$$\begin{aligned} &||A_x(\overline{x} + \beta s) + A_y y - b||_2^2 + \lambda_C^2 ||B(\overline{x} + \beta s) - y||_2^2 \approx \\ &||A_x \overline{x} + A_y y - b||_2^2 + \lambda_C^2 ||B\overline{x} + \beta s' - y||_2^2. \end{aligned} \tag{4.21}$$

Unlike in the substitution case, the vector s' appears in the regularizer norm. Unless the minimization algorithm modifies y to maintain both norms, an erroneous high spatial frequency component to x will not be inverted. We speculate that it is more likely that the component is suppressed by the iterative solver (i.e., the CG algorithm).

An important fact to remember is that in the SAR characterization problem, we only need the phantom fields y to be accurate. In the substitution case, any instability in x above the cut-off of B is necessarily carried over to y. On the other hand, the SI method formulation does not enforce this. It is this distinction which is likely responsible for the some of the differences between the substitution and SI methods. Not only is x more likely to be unstable under the substitution scheme, but y is also more likely to reflect the instability in x.

In essence, there are two ways that the SI method can suppress instability due to noise under Factor 1. The first way is from the relaxation of the data equation; the more that the data equation is enforced, the more noise that is inverted which results in instability. We speculate that the SI method is less sensitive to noise inversion by relaxing this norm. The second way that the SI method suppresses instability is by the relaxation of the regularizer

norm. Once again, since the substitution method enforces this strictly, any instability in x is carried over to y, provided that B is well-posed. On the other hand, the SI method does not necessarily need to carry over instability into y. In the case where B is also ill-posed and has the same cut-off as A_x, then the two methods tend to be the same since any instability in x will not be present in y. When B is ill-posed but has a larger cut-off frequency than A_x, then there exists a band of spatial frequencies which do not affect the data norm but affect the regularizer norm. We expect that the substitution and SI methods will differ for this case as well.

An simple way of determining which type of behavior to expect is to find the condition number[5] of A and B (it was found that the condition numbers of A_x, A_y, and A where similar for the examples shown in the results). If they are similar, then the two methods may tend to be the same. If B has a significantly smaller condition number than A, then the SI method will be comparatively more resistant to noise.

4.3.2 Factor 2

The second factor which may determine the relative performance between the substitution and SI methods is not as intuitive as the first. To understand the meaning of the relative ill-posedness between matrices, we present a discussion on a fundamental ill-posed problem, which uses a 2-by-2 matrix with subparallel[6] rows. We will investigate the base performance with respect to noise and when using simultaneous inversion versus substitution. This analysis will help increase our understanding of inverse problems in general and provide an understanding of the second factor.

[5] The condition number is indicative of the ill-posedness of the matrix [28].
[6] Almost parallel.

Consider the matrix A with subparallel rows defined as

$$A = \begin{bmatrix} 3.0 & 1.0 \\ 3.1 & 1.1 \end{bmatrix}, \tag{4.22}$$

which satisfies the equation

$$Ax = b, \tag{4.23}$$

where $x = [x_1; x_2]$ is a vector of unknowns and $b = [b_1; b_2]$ is known data. The true value of x is known for this exercise and is

$$x = \begin{bmatrix} 1.0 \\ 3.0 \end{bmatrix}. \tag{4.24}$$

The known data can be obtained by the product

$$b = Ax = \begin{bmatrix} 6.0 \\ 6.4 \end{bmatrix}. \tag{4.25}$$

Since this matrix is small, it is easy to derive an analytical solution for the unknown vector x. Using Cramer's rule,[7]

$$x_1 = \frac{\begin{vmatrix} 6.0 & 1.0 \\ 6.4 & 1.1 \end{vmatrix}}{\begin{vmatrix} 3.0 & 1.0 \\ 3.1 & 1.1 \end{vmatrix}} = 1.0, \tag{4.26}$$

$$x_2 = \frac{\begin{vmatrix} 3.0 & 6.0 \\ 3.1 & 6.4 \end{vmatrix}}{\begin{vmatrix} 3.0 & 1.0 \\ 3.1 & 1.1 \end{vmatrix}} = 3.0. \tag{4.27}$$

For a general right-hand side, the analytical solution can be written as

$$x_1 = 5.5b_1 - 5b_2, \tag{4.28}$$

$$x_2 = 15b_2 - 15.5b_1. \tag{4.29}$$

Due to the small determinant (caused by subparallel rows), any small errors in b are amplified in the resulting numerical solution. This is an example of the more general rule which relates the relative errors in b to the relative errors in x: [8]

$$\frac{||\delta x||_2}{||x||_2} \leq \kappa(A) \frac{||\delta b||_2}{||b||_2}, \tag{4.30}$$

where δ denotes the small error in either x or b, and $\kappa(A)$ is the condition number of the matrix. The condition number for general matrices is the ratio of the maximum and mini-

[7] See http://mathworld.wolfram.com/CramersRule.html.
[8] See https://blogs.mathworks.com/cleve/2017/07/17/what-is-the-condition-number-of-a-matrix.

mum singular values $\kappa(A) = \sigma_{\max}/\sigma_{\min}$, and it is generally large for ill-posed systems which suffer from instability [28].

The above reveals a key subtlety of linear ill-posed problems — if there is no data error, there is no error in x and the ill-posed problem can be solved. It is only with the addition of error that the ill-posed problem becomes difficult to solve. In practical physical problems, there is always some amount of noise which will require some amount of regularization to reduce solution instability. Furthermore, the inherent error that results from discretizing the continuous green's function results in instability even in the synthetic noiseless case.

Returning to the 2-by-2 matrix problem described at the beginning of this section, we would like to look at the performance with respect to noise. If we add noise to the data vector,

$$b \to b + b_e, \quad b_e = \begin{bmatrix} 0.1 \\ -0.1 \end{bmatrix}, \tag{4.31}$$

we can solve the inverse problem to see the effects of instability. Figure 4.2 shows the graphical solution of the problem which is obtained by finding the intersection of the functions $x_1 = \dfrac{b_1 - a_{12}x_2}{a_{11}}$ and $x_1 = \dfrac{b_2 - a_{22}x_2}{a_{21}}$. The true solution is also plotted to show the large change in x for a small change in b. Even with a small data error, the resulting error in x is very large, which is indicative of the instability and ill-posedness of the problem. Figure 4.2 also gives an intuitive visual understanding of ill-posed problems; due to the near parallel lines, the solution can slide up and down the pair of lines very easily from small perturbations.

Now consider that we had access to additional information about the system; in partic-

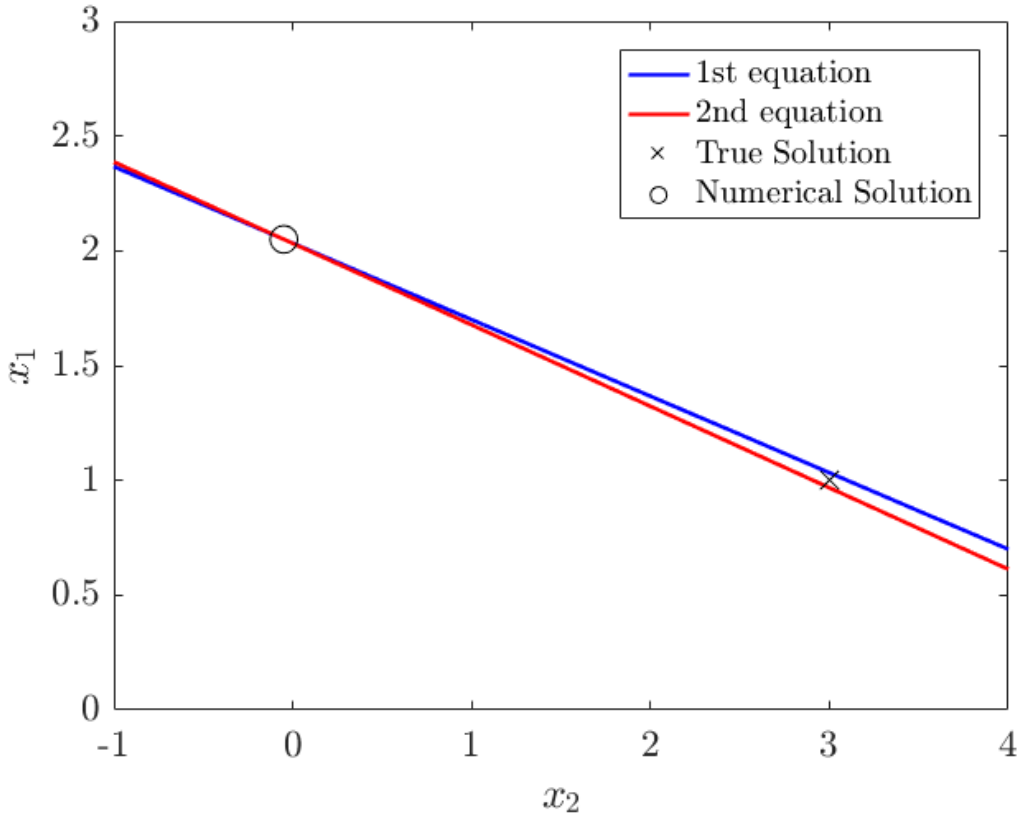


Fig. 4.2: Graphical solution for the 2-by-2 system. The lines represent the two equations of the system whose intersection is the numerical solution. The true solution is also plotted for contrast.

ular, we know how x_1 and x_2 are related. In this case, the relation is just the ratio $B = 1/3$ where

$$x_1 = Bx_2. \tag{4.32}$$

By itself, this condition is useless since there are infinite solutions which exist. However, if this is used in conjunction with the 2-by-2 system, it is actually a very useful condition. Its usefulness comes from the fact that the relation is noiseless (i.e., it does not depend on the measured data b), which means that $Bx_2 = x_1$ *will always cross the true solution.* No matter

how far the intersection between the two equations in $Ax = b$ finds itself, the subparallel lines (which effectively behave like a single line) will cross the condition $Bx_2 = x_1$ precisely at the location of the true solution. Letting $A_1 = [a_{11}; a_{21}]$ and $A_2 = [a_{21}; a_{22}]$, we can incorporate the relation B by substituting this into the system:

$$\begin{aligned} Ax &= b, \\ A_1x_1 + A_2x_2 &= b, \\ A_1Bx_2 + A_2x_2 &= b. \end{aligned} \tag{4.33}$$

By solving the least-squares problem in (4.33), we can obtain the solution which mostly satisfies $Ax = b$ and is subject to $x_1 = Bx_2$. Figure 4.3 shows the same graphical representation of the noisy system which is magnified to show the solution to (4.33). The substitution and true solutions are nearly on top of each other, with the substitution solution from solving (4.33) being $[0.9995; 2.9984]$. Notice that we effectively have a new 2-equation system; the first equation being the pair of subparallel lines, and the second equation being $Bx_2 = x_1$. Since these two effective lines are not subparallel, the system is not ill-posed and we can almost get the exact solution. The condition $Bx_2 = x_1$ can thus be thought as having a regularizing effect. Since $Bx_2 = x_1$ also reveals information on x, we are able to obtain the true solution very easily.

A distinction to be noted is the difference between the mean and deviation of the noise, and its effect on the modified system in (4.33). In general, any noise with a mean of zero (but otherwise arbitrary) will only result in changing the "width" of the 2D cone formed by $Ax = b$, whereas noise with a non-zero mean and no variance will shift the cone itself. (In reality, both types of noise will change the width and placement of the cone, but their influences on each of these is different which is why we make the distinction.) A

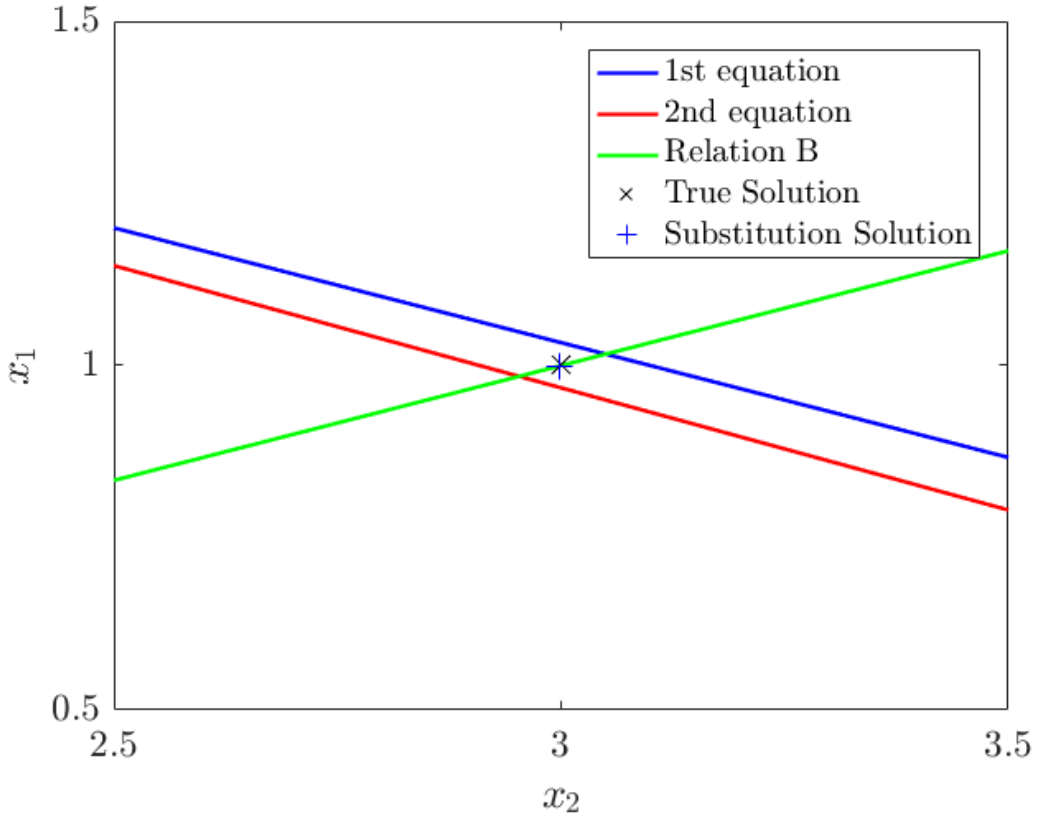


Fig. 4.3: Graphical solution for the 2-by-2 system when enforcing the relation between unknowns. The true solution is also plotted for contrast.

choice of $b_e = [2; -2]$ (large deviation but zero mean) results in the solution $[0.9896; 2.9688]$, whereas a choice of $b_e = [0.2; 0.2]$ (no deviation but a mean of 0.2) results in the solution $[1.0483; 3.1450]$. The effective system is still well-posed, but it is only the mean noise which really has an effect — this is consistent with the idea that the subparallel equations in $Ax = b$ are effectively one equation when regularized by $Bx_2 = x_1$.[9] Note that for large systems for practical problems (e.g., the SAR characterization problem), even if the (typically white) noise has a mean of zero over many measurements, locally there can be a mean

[9] If we take the effective equation to be the mean equation, i.e. $\frac{A_1 + A_2}{2}x = \frac{b_1 + b_2}{2}$, then any noise variance in b is cancelled for this problem.

noise for two subparallel equations.

A critical point that needs to be made is that this method of regularization — substituting $Bx_2 = x_1$ into the system of equations — is not guaranteed to work. Consider the case where the effective equations implied by $Ax = b$ and $Bx_2 = x_1$ are subparallel: any noise mean in the data b will drastically affect the regularized point and the system remains ill-posed. However, using $Bx_2 = x_1$ is still useful since the condition is not subject to any noise; this will reduce but not eliminate the solution instability.

Suppose that instead of using a direct substitution, we use the condition $Bx_2 = x_1$ as an explicit regularization term (as in the SI method). In this case we seek to minimize both the original system $\|Ax - b\|_2^2$ and the additional system $\|Bx_2 - x_1\|_2^2$. An important feature of this method is that when the combined system remains ill-posed, the relaxation of the condition $Bx_2 - x_1 = 0$ (which does not need to be 0) means that the algorithm has the potential to recover the true solution under noisy conditions. Additionally, the regularization can be weighted as necessary in order to find the optimal point. In contrast, the algorithm which uses strict substitution will never be able to recover the true solution since the iterative solver can only search along $Bx_2 = x_1$. This issue can be compounded if there is slight error in B itself; if it represents a continuous operator (which it does in the noninvasive SAR case), then it means the substitution method can suffer from greater error.

To demonstrate these conclusions, the original example needs to be modified to a combined system which is ill-posed. Let

$$A = \begin{bmatrix} 0.90 & -1.00 \\ 0.91 & -1.00 \end{bmatrix}, \quad x = \begin{bmatrix} 1.0 \\ 1.0 \end{bmatrix}, \tag{4.34}$$

and

$$b = Ax = \begin{bmatrix} -0.10 \\ -0.09 \end{bmatrix}. \tag{4.35}$$

To reinforce the ideas that have been presented, we begin by solving the noiseless system, shown in Figure 4.4. There are two new quantities in this plot: the SI solution, which is

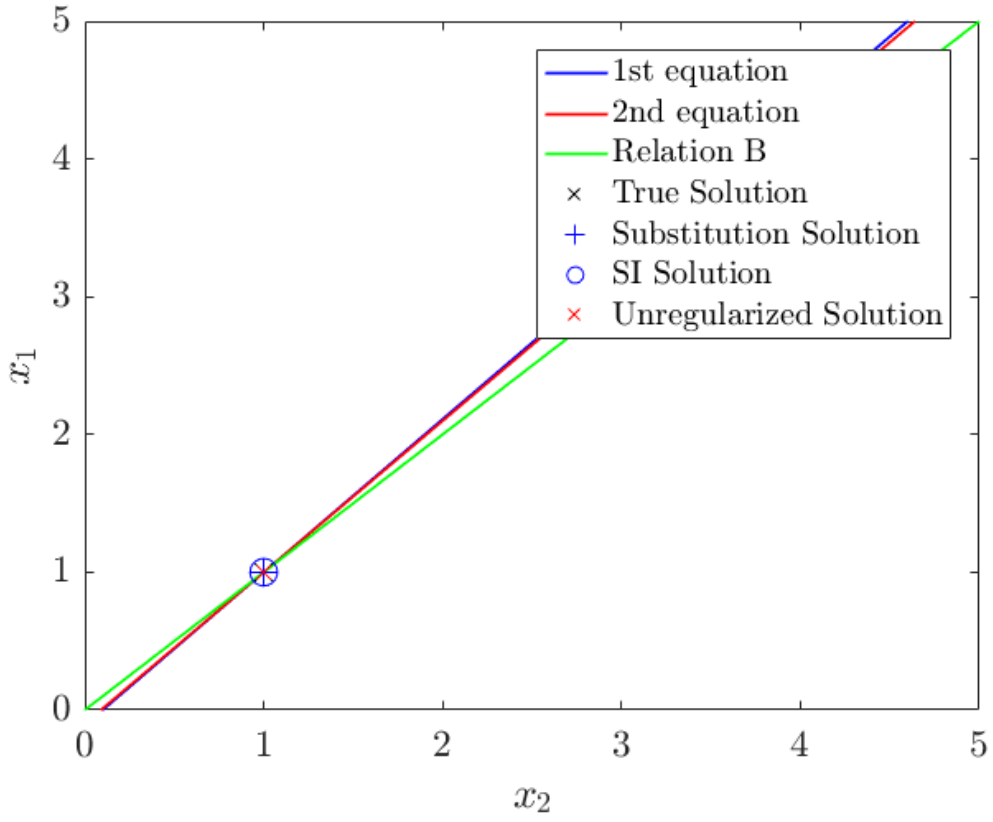


Fig. 4.4: Graphical solution for the noiseless 2-by-2 system. The regularized solution is the solution obtained from using the condition B as an explicit regularization term. The unregularized solution is the base numerical solution which does not consider B.

the solution obtained from using $Bx_2 = x_1$ as a regularization term and simultaneously solving for x_1 and x_2, and the unregularized solution, which is the base solution obtained

from minimizing $||Ax - b||_2^2$ only. As expected, in the truly noiseless system each solution is correct and is perfectly placed over the true solution.

Next, we will look at the behavior of the system when adding noise. Additionally, we will begin weighting the regularization term $Bx_2 = x_1$, such that we seek to minimize the functional over x,

$$\min_x \{||Ax - b||_2^2 + \lambda^2 ||Bx_2 - x_1||_2^2\}, \tag{4.36}$$

where λ is the weight of the regularization term. The choice of λ will decide how well each term in the functional is satisfied. It will be demonstrated shortly that (for reasonable weights) the SI solution will generally be bounded between the base unregularized solution, and the solution obtained through substitution. (For very large weights, the system becomes ill-posed in a different way — if we only need to satisfy the condition $x_1 = Bx_2$, there are infinite solutions and the algorithm becomes unstable with respect to λ). Figure 4.5 shows the noisy system obtained when adding $b_e = [-0.01; 0.01]$. The unregularized solution has drastically moved and is at $x = [3.0; 2.8]$. As expected from the zero-mean noise, the substitution solution is almost correct at $x = [1.0055; 1.0055]$. The regularization weight was scanned from 10^{-5} to 10^5 and the best solution was chosen for the plot, which was at $x = [1.0055; 1.0055]$. Figure 4.6 shows the same plot but magnified. In addition, there is a dashed line which represents the "path" that the SI solution takes, which starts at the unregularized solution and ends at the substitution solution. We again see how a zero-mean noise in this example does not significantly affect the substitution solution.

When the noise does have a mean, we begin to see the advantages of the SI strategy. Figure 4.7 shows the noisy system when adding $b_e = [0.01; 0.01]$. Since the effective system formed when including $Bx_2 = x_1$ is ill-posed, a non-zero mean noise results in a significant shift for

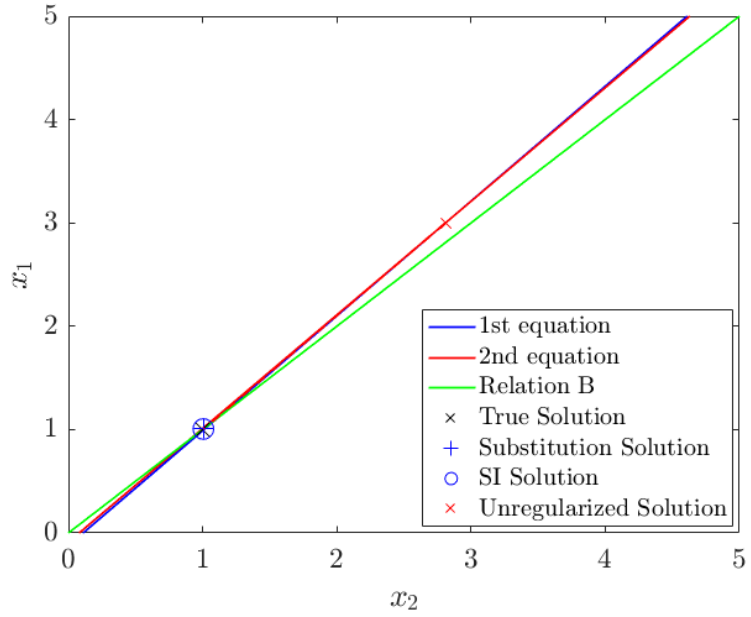


Fig. 4.5: Graphical solution for the noisy 2-by-2 system with data error $b_e = [-0.01; 0.01]$.

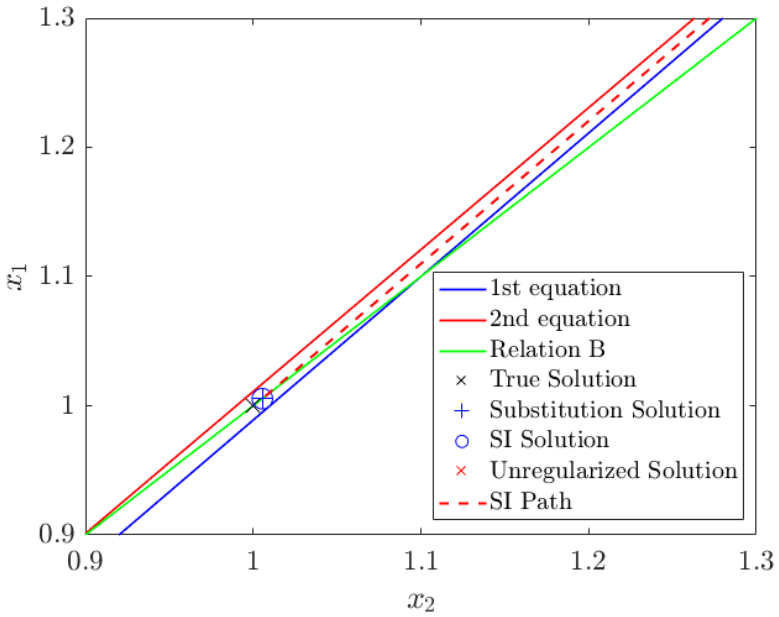


Fig. 4.6: Magnified graphical solution for the noisy 2-by-2 system with data error $b_e = [-0.01; 0.01]$.

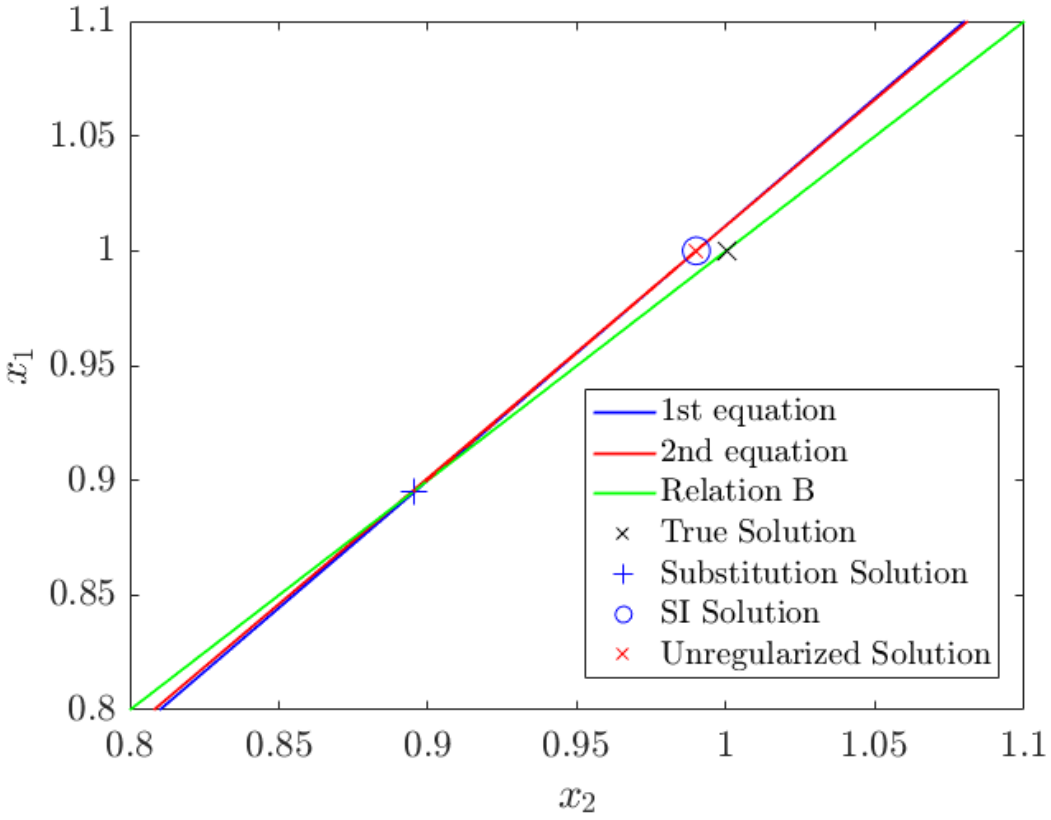


Fig. 4.7: Magnified graphical solution for the noisy 2-by-2 system with data error $b_e = [0.01; 0.01]$.

the substitution solution. In contrast, the unregularized solution does not move drastically since it is not sliding up and down the effective line drawn by $Ax = b$. Consequently, a small choice of λ for the SI solution is best which results in a solution ($[1.00; 0.99]$) which is much closer to the true solution ($[1.00; 1.00]$) than in the substitution case ($[0.90; 0.90]$).

The past two examples have demonstrated some critical qualities of the noisy system. In the case where the effective system created by $Ax = b$ and $Bx_2 = x_1$ is ill-posed, the base unregularized solution and the substitution solution are each resistant and sensitive to one of each type of noise. The unregularized solution is sensitive to noise variance in b and

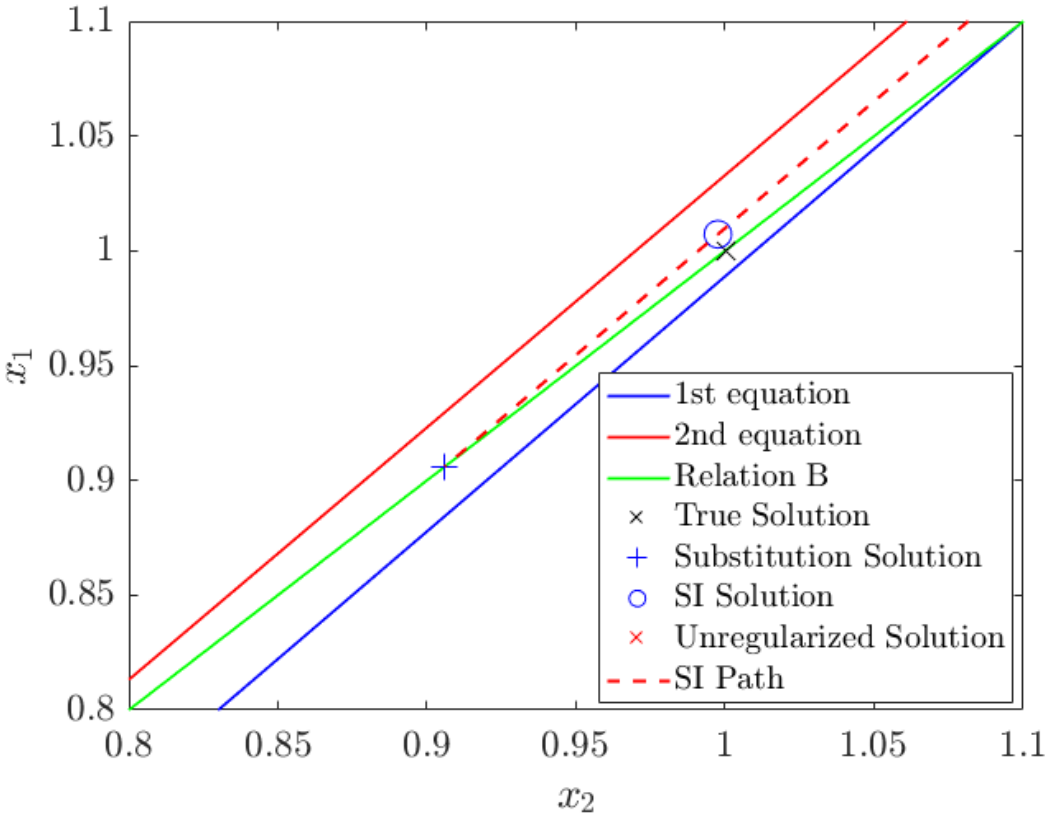


Fig. 4.8: Magnified graphical solution for the noisy 2-by-2 system with data error $b_e = [-0.01; 0.03]$.

is relatively insensitive to a mean noise. Conversely, the substitution solution is relatively insensitive to noise variance, but is sensitive to a mean noise.

In general, the noise in the data will be a combination of the two types. In this case, both the unregularized and substitution solutions will perform poorly. However, the SI solution has a chance to get close to the true solution. Figure 4.8 show the zoomed-in noisy system when adding $b_e = [-0.01; 0.03]$. The base solution is not on the plot and is located at $x = [5.0; 4.6]$. Due to the noise mean, the substitution solution is again incorrect and is $x = [0.9; 0.9]$. In contrast, a choice of $\lambda = 0.47$ gives a regularized SI solution of

$x = [1.007; 0.998]$, which is very close to the true solution. By considering both the unregularized and substitution solutions, the SI algorithm can take advantage of the specific noise insensitivity of each solution in order to obtain a good result. It should be noted, however, that the results in Figure 4.8 represents the ideal case where the true solution is between the unregularized and substitution solutions. When this is not the case, the best SI solution is whichever other solution that is closest. In any case, the SI solution will always be at worst equally accurate to one of the two other solutions, which means that the method will be the most resistant to noise in general.

The only problem with the SI approach is choosing λ. In the examples provided above, prior knowledge of the true solution was needed; in practical problems, this may not always be available. If there was a mechanism which allowed the algorithm to choose an appropriate λ, then the simultaneous inversion of x_1 and x_2 would be very effective. Ideally, there would be some quantity that could be calculated which exhibits a particular response in the vicinity of the true solution. When x is unknown, there are still two pieces of information available to the algorithm: the differences $Ax - b$ and $Bx_2 - x_1$. Figure 4.9 shows the plot of the norms of these differences as λ is varied from 10^{-5} to 10^5 for the problem shown in Figure 4.8. Unfortunately, this plot does not help in finding the optimal λ. As expected, the plot is bounded by the unregularized and substitution solutions which are linked by a straight line (as was seen in the previous plots).

Suppose that x_1 and x_2 were only part of a greater system and there was another pair of variables $x_{II} = [x_3; x_4]$ which were unknown. (For simplicity, we will assume that the

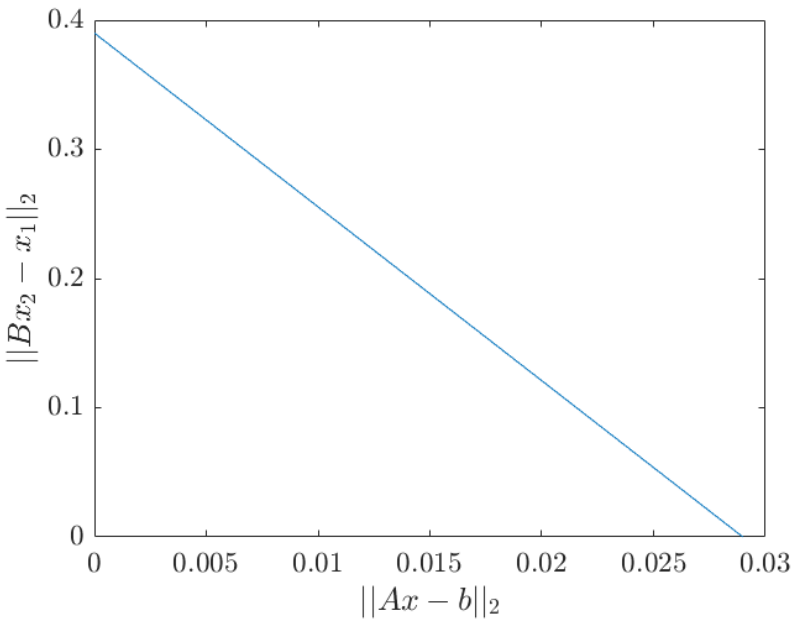


Fig. 4.9: Plot of $||Bx_2 - x_1||_2$ versus $||Ax - b||_2$ for the previous noisy system.

variables are independent). Let

$$A_{II} = \begin{bmatrix} 0.49 & -1.00 \\ 0.48 & -1.00 \end{bmatrix}, \quad x_{II} = \begin{bmatrix} -3.0 \\ 5.0 \end{bmatrix}, \tag{4.37}$$

and

$$b_{II} = A_{II}x_{II} = \begin{bmatrix} -6.47 \\ -6.44 \end{bmatrix}. \tag{4.38}$$

The relation between the variables is $B_{II} = x_3/x_4 = -0.6$. With this new system we can again create a plot of $||Bx_2 - x_1||_2$ versus $||Ax - b||_2$, which will again result in a straight line. Although the two sets of variables are separate, assume that we would like to solve both systems at the same time subject to the same λ; i.e., we seek to minimize over x and

x_{II},

$$\min_{x,x_{II}}\{||Ax-b||_2^2+||A_{II}x_{II}-b_{II}||_2^2+\lambda^2||Bx_2-x_1||_2^2+\lambda^2||B_{II}x_4-x_3||_2^2\}. \tag{4.39}$$

The terms are separated since x and x_{II} are independent. If we take the individual (non-squared) norms and then add them together, the combined regularizer versus data norm produces a plot shown in Figure 4.10. Unlike the individual cases, the plot is not a straight

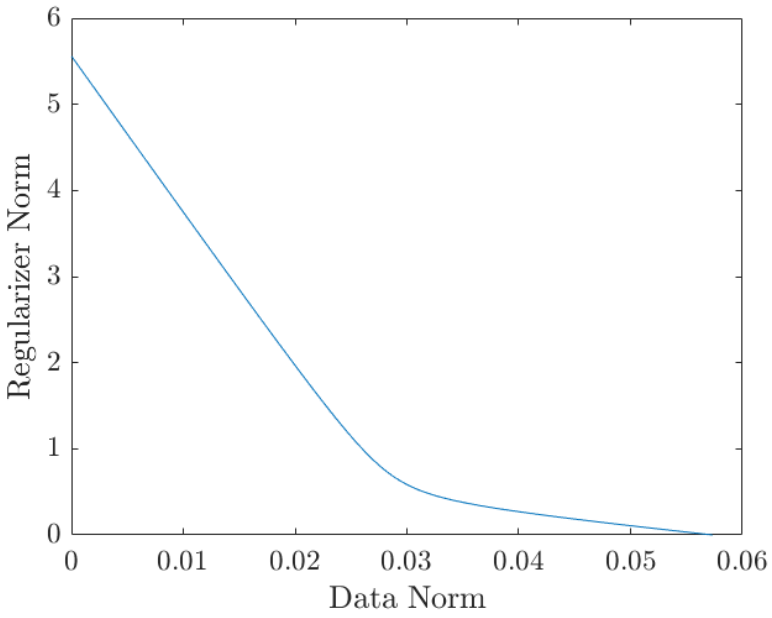


Fig. 4.10: Plot of regularizer norm sum versus data norm sum for the larger system.

line and is rather a curve in the shape of an "L". Additionally, it is intuitive to realize that the optimal point — the value of λ which best satisfies the two systems — is at the bend, or the corner of the L plot. At this point, each of the four norms are evenly satisfied (relative to their range over λ). To understand why a bend appears, recall that the individual norm plots naturally have different slopes and ranges due to the fact that noise will have different effects on the unregularized and substitution solutions. By consequence, the combined plot will not also be linear, and the corner of the curve indicates the best compromise. In fact,

the plot in Figure 4.10 is a rudimentary example of the previously mentioned *L-curve*. For general regularization strategies, the plot of the regularizer norm versus the data norm typically forms an L whose corner represents the optimal solution and choice of λ.

Summary of Discussion

The aim of this section was to provide a discussion to understand the meaning of Factor 2, which is the ill-posedness of B with respect to A. Simple problems for a 2-by-2 matrix system were presented and its performance with respect to noise was analyzed. Three types of solutions were examined: an unregularized solution, a solution obtained by substituting the noiseless condition $Bx_2 = x_1$ into $Ax = b$, and a solution obtained from using $Bx_2 = x_1$ as an explicit regularization term (i.e., simultaneous inversion). When the combined system was ill-posed, very particular behaviors were noted for the 2-by-2 system. In general, the unregularized solution was sensitive to noise variance, but insensitive to mean noise. The substitution solution had the opposite behavior; it was insensitive to noise variance, but sensitive to mean noise. The SI solution was a combination of the two previous solutions, where λ could be tuned in order to favor one or the other. In many cases, a good choice of λ can result in a very accurate solution. Finally, when many variables are considered, the combined norm plot produces an L-curve which can be used to find the optimal λ.

When dealing with an arbitrarily-sized system such as the noninvasive SAR problem, the current link $Bx = y$ can be considered to be ill-posed with respect to $Az = b$ if there exists a 2D projection (over two unknowns) where two equations of $Az = b$ ($z = [y; x]$) are subparallel to an equation in $Bx = y$. This will necessarily happen if a pair of the multi-dimensional rows in A are subparallel to another row in $C = [-I\ B]$. Note that we do not need to consider b, which will only determine the exact intersection point. It is for this reason that we refer to Factor 2 as "the ill-posedness of B with respect to A". If Factor 2 exists in the

noninvasive SAR problem, then we can predict that the three solutions (SI, substitution, and unregularized) will react to the two types of noise in a similar way to the 2-by-2 system.

Factor 2 will be evaluated with two different methods. The first method is to evaluate the ranks of A, C, and $[A;C]$. If the rank of the combined system is less than the sum of the individual ranks, then there is redundant information in B with respect to A. The second method is to verify that the response of the three solutions (SI, substitution, unregularized) to the different types of noise is similar to the simplified case. In the ideal case, it would be sufficient to add a *local* mean noise or noise variance; i.e., we add the noise to two nearby data points (hence the term local) in b corresponding to two subparallel equations in $Az = b$. Unfortunately, the green's function is a continuous operator, so there will not only be two equations which form one effective equation; in general, there will be many subparallel partners to a single equation. Therefore, we will also look at the performance when adding a *global* mean noise or noise variance; i.e., we add the noise to all data points (hence the term global). Regardless of the exact performance, we at least want to see that the SI solution is more resistant to noise than the other methods.

4.3.3 Factor 3

The third factor in the SI versus substitution discussion is the accuracy of B. In the previous discussion, B was considered perfectly accurate and void of noise. For practical problems, this is not necessarily the case. Since the relation $Bx = y$ passes through the origin, any error in B is represented by a shift in the slope. It is intuitive to see that this can cause a large shift in the substitution solution in the case that Factor 2 is present. Since the relation B is the discretized green's function in the noninvasive SAR problem, it will necessarily suffer from some discretization error. The discretization size is thus very influential on this factor. The accuracy of B can be examined by evaluating the condition

$\frac{\|Bx-y\|_2}{\|y\|_2}$ for the true surface currents.

4.3.4 Factor 4

The fourth and arguably most important aspect of the SI versus substitution discussion is the spectral decomposition of the true solution. Consider the case where the condition number of B is high but less than A (Factor 1) and x has a flat spatial frequency spectrum.[10] Both the substitution and SI methods will require additional regularization to filter out all spatial frequencies which are susceptible to instability. However, due to the relaxation of the data norm, the SI method may require much less extra regularization. If this is true, the SI solution will be generally preferable since we capture more information regarding the solution x. Next consider the case where the true solution has a baseband spectral decomposition, i.e., x only contains low spatial frequency components. If the largest significant spatial frequency is below the cut-off of A, then additional regularization can be used without losing any significant information regarding x. Provided that B is accurate, the substitution method will produce a good solution. On the other hand, the SI method may suffer from inaccuracies due to relatively fewer measurement data (more unknowns), among other things. We will show that this is indeed the case for the 2D TM SAR problem.

4.3.5 Factor Influences

The four factors outlined in this chapter are similar but distinct qualities which can be evaluated to predict the differences between the substitution and SI methods for a given problem. The discussions were limited to the end problem, e.g., a fixed condition number or rank. However, there are minor factors that can affect the major factors outlined in this chapter. Firstly, there are the physical parameters, e.g., phantom size and properties, DUT type and position, etc. There are also the chosen discretization strategies (number and type

[10] Each possible spatial frequency component is equally represented in x.

of basis functions) and the number and spacing of measured data. Unfortunately, the effect of these minor factors has not been fully investigated for this thesis. Only a preliminary investigation into the effect of current discretization is included in this thesis.

4.4 2D TM Implementation

The primary goal of the results presented in this chapter is to demonstrate the noise resistance of the SI method and validate the four major factors outlined in this chapter. A preliminary investigation into the effect of discretization is made by looking at three different cases of discretization.

4.4.1 Physical Parameters

The general 2D TM simulation parameters for each case are the same as in Chapter 3. Figure 4.11 shows the simulation geometry and highlights important parameters shared by each case. The values of these parameters are summarized in Table 4.1. The phantom

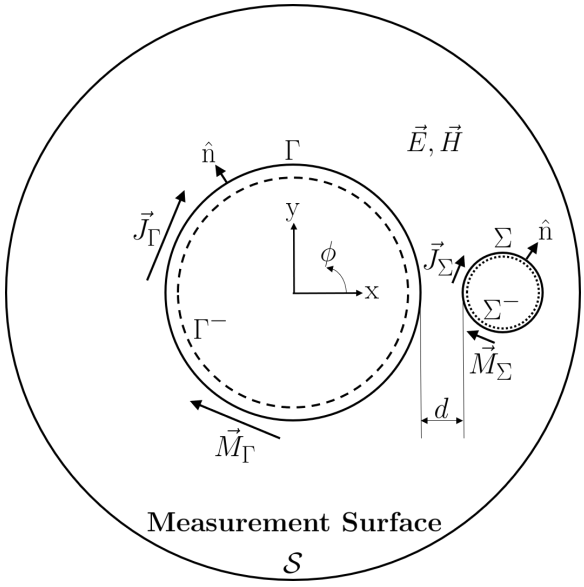


Fig. 4.11: 2D TM problem setup for results.

Table 4.1: Chapter 4 problem parameters for the 2D TM simulation using MATLAB generated data.

Parameter	Value
R_Γ	12.1 (cm)
R_Σ	1 (cm)
d	0.9 (cm)
f	900 (MHz)
ε_r	$41.5 - j19.4$

size and its properties are the same as in Chapter 3; the phantom radius is approximately 10.8 (cm) and the relative complex permittivity is $41.5 - j19.4$ at a frequency of 900 (MHz). The DUT is once again an infinite line source with unit current magnitude.

4.4.2 Generation of the Current Link

As previously mentioned, the current link B is generated by using the MoM-based forward solver which was also used to generate the synthetic data. In order to avoid any possible inverse crime, the discretized phantom domains are different. For the synthetic data, a 250×250 square grid with a total side length of 24 (cm) centered on the phantom was used. For the current link, a 200×200 square grid with the same total side length was used.

4.4.3 Error Metrics

For both the SI and substitution methods, it is only the phantom fields nearest to the DUT that are difficult to recover. This location also coincides with the peak field magnitude and peak SAR. Due to this, we are particularly concerned with the error on the peak field. We formally define the relative peak field error as

$$\text{Relative Peak Field Error} = \frac{|F^{\text{rec}} - F^{\text{true}}_{\text{max}}|}{|F^{\text{true}}_{\text{max}}|}, \tag{4.40}$$

where F^{rec} is the reconstructed field (E or H) and $F^{\text{true}}_{\text{max}}$ refers to the true field, both at the location of the largest true field magnitude.

4.5 Case 1 2D TM Results

The first case of discretization that will be investigated will be the same that was used in Chapter 3 for the surface current method. Table 4.2 summarizes the choices for the number of basis functions and measured data points. The verification of the four factors from the motivation section will be particularly thorough for this first case.

Table 4.2: Case 1 problem parameters for the 2D TM simulation using MATLAB generated data.

N_Γ	360
N_Σ	36
$N_\mathcal{S}$	360

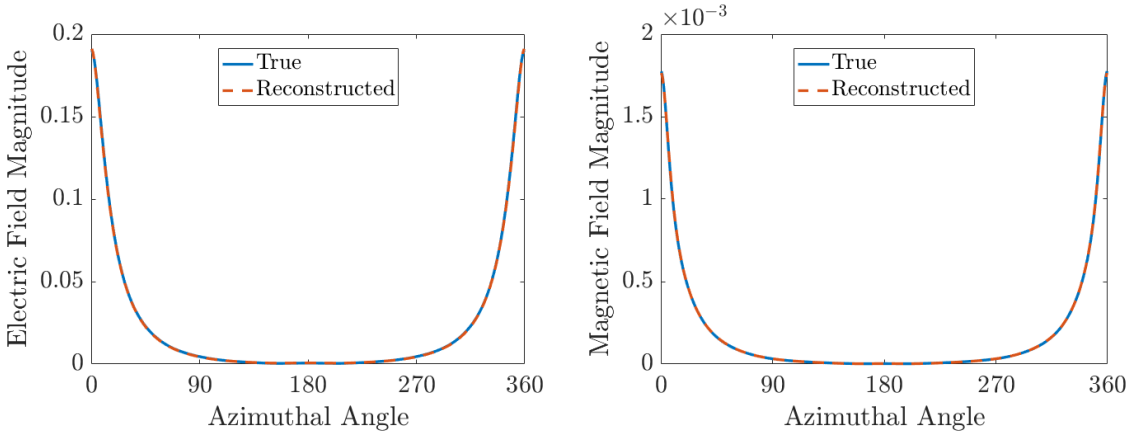
4.5.1 Noiseless Results

We first begin by looking at the performance of the substitution and SI methods in the noiseless case.[11] An important issue which must be noted is that for many cases, the substitution method performed better with unbalanced currents. This can be understood by evaluating Factor 3; it will be shown that the condition B is more accurate in the unbalanced case. All of the results in this chapter will look at the SI[12], balanced substitution, and unbalanced substitution methods.

Figure 4.12 shows the reconstructed phantom surface fields using the SI method. The relative peak field error was 3.00% for the electric field and 0.721% for the magnetic field. For

[11] Although there is no added noise, the operator discretization error will result in an effective noise which is why there can be error in the solution.

[12] It should be noted that the SI method will always use balanced currents.



(a) a) **(b)** b)

Fig. 4.12: Reconstructed phantom surface a) electric field and b) magnetic field magnitudes in using the SI method compared to the true distributions. The relative peak field error was 3.00% for the electric field and 0.721% for the magnetic field. The relative peak field error occurs at $\varphi = 0^\circ = 360^\circ$ which corresponds to the location nearest to the DUT.

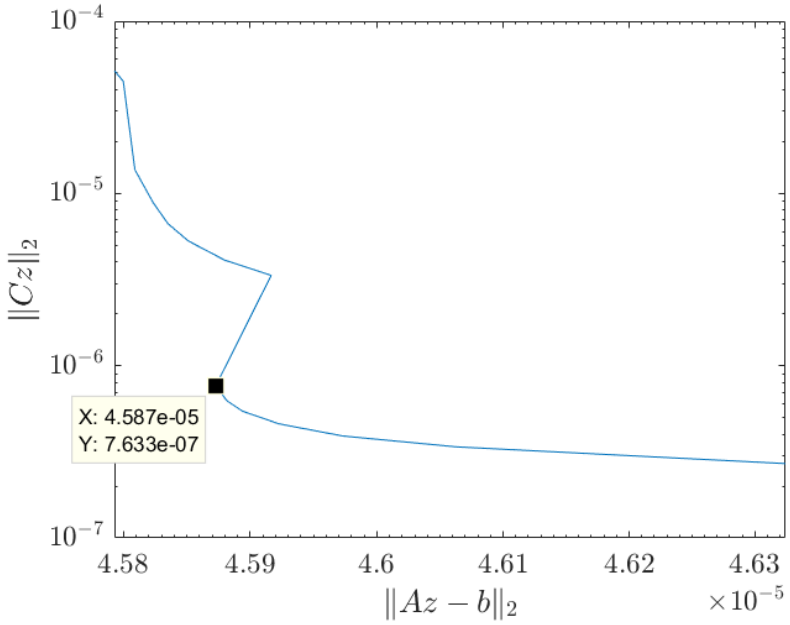


Fig. 4.13: L-curve plot for Case 1 using the range $1/30 < \lambda_C < 30$ with the corner highlighted.

this case of discretization, the SI method produces very good phantom fields. The weight λ_C was 1.63 and was chosen by examining the L-curve shown in Figure 4.13. We would like to highlight that these results are noticeably better than those in Chapter 3, despite using the same problem parameters. This agrees with the idea that the current link is generally better than the enforcement of Love's condition for obtaining the true phantom fields.

Moving on, Figure 4.14 shows the phantom surface fields using the balanced substitution method. The relative peak field error was 25.5% for the electric field and 24.7% for the

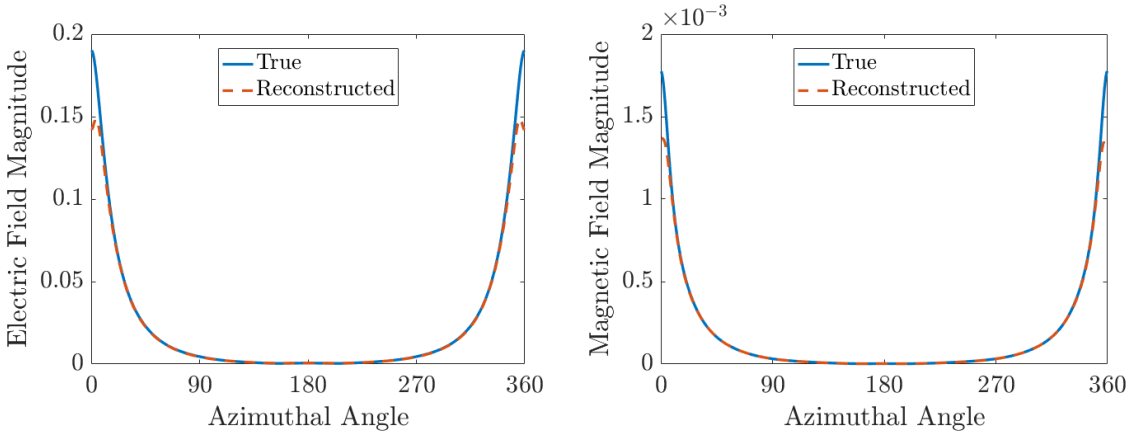
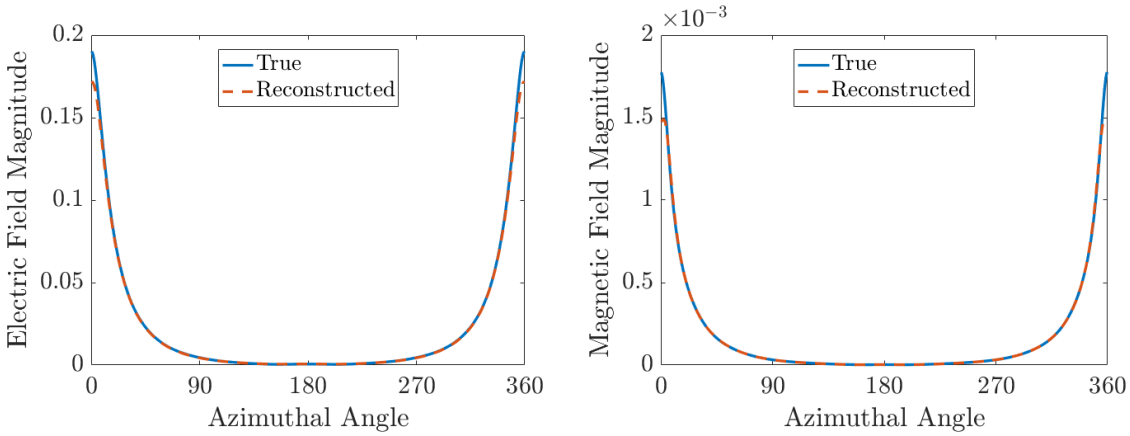


Fig. 4.14: Reconstructed phantom surface a) electric field and b) magnetic field magnitudes in using the balanced substitution method compared to the true distributions. The relative peak field error was 25.5% for the electric field and 24.7% for the magnetic field. The relative peak field error occurs at $\varphi = 0^\circ = 360^\circ$ which corresponds to the location nearest to the DUT.

magnetic field. Unlike the SI method, the substitution method produces fairly significant errors in the reconstructed fields.[13] To understand this difference, we need to examine the major factors presented earlier in this chapter. The examination and verification of the major factors will be done in the following section. Finally, Figure 4.15 shows the phantom surface fields using the unbalanced substitution method. The relative peak field error was

[13]We would like to emphasize that there is an effective added noise in the "noiseless" case due to finite discretization error.



(a) a) **(b)** b)

Fig. 4.15: Reconstructed phantom surface a) electric field and b) magnetic field magnitudes in using the unbalanced substitution method compared to the true distributions. The relative peak field error was 13.0% for the electric field and 18.9% for the magnetic field. The relative peak field error occurs at $\varphi = 0^\circ = 360^\circ$ which corresponds to the location nearest to the DUT.

13.0% for the electric field and 18.9% for the magnetic field. This is a clear improvement over the balanced scheme; to understand why these differences occur, we need to examine the major factors outlined in this chapter.

4.5.2 Verification of Section 4.3

In the motivation section, there were four major factors that were argued to be a reflection of the relative performance of the substitution and SI methods, particularly in the case of noisy data. These four factors were:

1. the ill-posedness of B versus the ill-posedness of A,
2. the ill-posedness of B with respect to A,
3. the accuracy of B,
4. the spectral decomposition of true phantom surface fields.

The first factor can be evaluated directly through a comparison of matrix condition numbers. In addition, we can look at reconstructed DUT fields and verify that the relaxation of the data equation results in more stable currents. The second factor was proposed to be evaluated by comparing the ranks of the individual and combined matrices. In order to make a thorough verification, we will also attempt to replicate the behaviors which were observed in the 2-by-2 system which will indicate that B is ill-posed with respect to A. The third factor will be examined by evaluating the current link with the true fields and currents. The fourth factor is not the simplest to evaluate thoroughly; we will defer to the noisy results[14] to demonstrate this factor.

Factor 1

Using the parameters of Case 1, we first look at the condition numbers of the important matrices. Table 4.3 summarizes the condition numbers of A and B for the balanced and unbalanced cases. Both matrices have large condition numbers, but the current link B is

Table 4.3: Placeholder.

	Matrix	Condition Number	Ratio
Balanced	A	9.4×10^{16}	$\approx 10^6$
	B	2.5×10^{10}	
Unbalanced	A	1.7×10^{17}	$\approx 10^4$
	B	1.3×10^{13}	

less ill-conditioned. Under the low-pass filter analogy, the substitution method will be much more susceptible to inverting erroneous high frequency components. We also see that in the unbalanced case, the A matrix is more poorly conditioned, and the ratio of condition numbers decreases. This indicates that the unbalanced strategy may be more susceptible to noise. However, it will be shown that the current link is more accurate in the unbalanced

[14]The noisy results are the reconstructed solutions using measurement data with added synthetic white noise.

case which results in a higher achievable accuracy.

To see the increased susceptibility to noise, we can look at the reconstructed DUT currents. Figure 4.16 shows the plots of the reconstructed DUT currents for the balanced substitution and SI methods. It is clear from this figure that the substitution method has

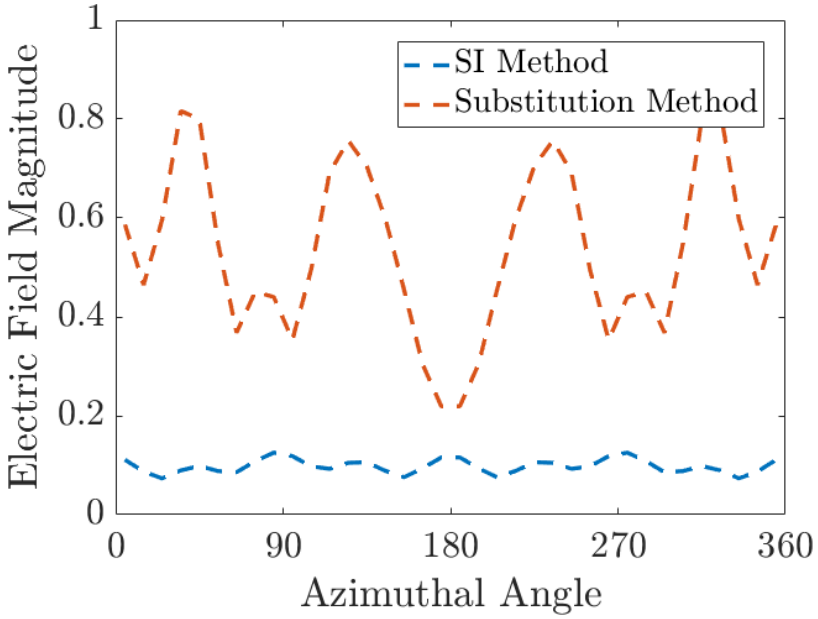


Fig. 4.16: Reconstructed DUT currents for the balanced substitution and SI methods.

reconstructed large spatial frequencies that the SI method has suppressed. This agrees with our speculations in Section 4.3.1. Most of these components are dampened when we evaluate $y = Bx$ to obtain the phantom surface fields. However, any erroneous components in x which are below the cut-off of B are carried over to y in the substitution case, which explains the increased error for Case 1.

Table 4.4: Comparison of matrix ranks for A, C, and concatenated $[A;C]$ matrices.

	Matrix	Rank	Rows Lost
Balanced	A	37	
	C	720	
	$[A;C]$	745	12
Unbalanced	A	35	
	C	720	
	$[A;C]$	741	14

Factor 2

Next, we would like to verify the discussion regarding the second factor: the ill-posedness of B with respect to A. We start by looking at the rank[15] of the individual and combined matrices. Table 4.4 summarizes the rank of A, C, and the concatenated $[A;C]$ matrices. Note that we must use $C = [\text{-}I\ B]$ in order to properly compare the ranks. Since the number of columns for these matrices is 792 (total number of unknowns), we would expect the rank of the concatenated matrix to be the addition of the individual ranks.[16] What we see instead is a loss of 12 and 14 rows when combining the systems. This indicates that there is some redundant[17] information in C and suggests that at least one of the possible 2D projections for the system would demonstrate the combined ill-posed system shown in the motivation section.

In the discussion of the second factor, a few properties of the combined A-B system were noted. These properties, which presuppose that B is ill-posed with respect to A, were

1. The substitution solution is sensitive to a local noise mean, and insensitive to a local noise variance.

[15]The rank was evaluated using the built-in MATLAB function.

[16]If the addition exceeds the number of columns (which it does not in this case), the maximum rank of the concatenation is the number of columns. Note also that this only applies if the number of columns is smaller than the number of rows for the concatenated matrix (which it does in this case).

[17]That is, some of the rows in C are linearly dependent (to a certain tolerance) on the rows in A.

2. The unregularized solution is insensitive to a local noise mean, and sensitive to a local noise variance.

Since the results in Table 4.4 suggest that this should be the case, we next examine how the results change for a local mean noise and a local noise variance. As mentioned in the discussion for Factor 2, it may not be possible to properly replicate the local noise effects so we will also investigate adding a global mean noise and global noise variance to all measured data in b.

Starting with a global mean noise, we can predict that the substitution solution significantly changes, while the unregularized solution does not. By consequence, an appropriate choice of λ_C should give a reasonable solution for the simultaneous inversion. Table 4.5 summarizes the relative peak field errors for the phantom fields when adding the global noise $b_e = 10^{-3}$ to each data point in b. As predicted, a mean noise does not affect the

Table 4.5: Relative peak error for the simultaneous inversion (si), substitution (su), and unregularized solutions (un) when adding a global mean noise.

		Error	**Noiseless Error**	**Change**
Balanced	E_{si}	28.9%	2.9%	+26.0%
	H_{si}	10.2%	0.7%	+9.5%
	E_{su}	3533.0%	25.5%	+3507.5%
	H_{su}	1673.3%	24.7%	+1648.6%
	E_{un}	47.0%	46.9%	−0.1%
	H_{un}	119.7%	119.7%	0.0%
Unbalanced	E_{su}	3379.8%	13.0%	+3366.8%
	H_{su}	17410.1%	18.9%	+17391.2%

unregularized solution, but significantly affects the substitution solution. Due to the self-correcting L-curve, a choice of $\lambda_C = 6 \times 10^{-3}$, which favors the unregularized solution, results in a relatively stable solution for the simultaneous inversion method. In addition, we see that the unbalanced substitution method is more susceptible to a global mean noise.

Since white noise has a mean of zero, it is more likely that there is a local mean noise. Table 4.6 summarizes the relative peak errors for the phantom fields when adding the local noise $b_e = 10^{-3}$ to the first two measured data in b. Again, the balanced substitution so-

Table 4.6: Relative peak error for the simultaneous inversion (si), substitution (su), and unregularized solutions (un) when adding a local mean noise.

		Error	**Noiseless Error**	**Change**
Balanced	E_{si}	106.3%	2.9%	+103.4%
	H_{si}	53.7%	0.7%	+53.0%
	E_{su}	441.4%	25.5%	+415.9%
	H_{su}	1617.7%	24.7%	+1593.0%
	E_{un}	2938.1%	46.9%	+2891.2%
	H_{un}	534.1%	119.7%	+414.4%
Unbalanced	E_{su}	255.8%	13.0%	+242.8%
	H_{su}	404.7%	18.9%	+385.8%

lution changes significantly. The unregularized solution also changes significantly, but this makes sense since there is a noise variance with respect to nearby equations which are also subparallel.[18] Nonetheless, with a choice of $\lambda_C = 2.6 \times 10^{-2}$, the simultaneous inversion solution is relatively stable. The unbalanced substitution solution is also relatively good (but still worse than the SI method).

Following the previous cases, we add a global noise variance to the measured data. Table 4.7 summarizes the errors when adding $b_e = [10^{-3}; -10^{-3}]$ to each pair of measurements in b. As predicted, the substitution and regularized solutions essentially do not change. Surprisingly, the unregularized solution is unchanged as well; a possible explanation for this is that since each equation in $Ax = b$ generally has at least two other equations which are

[18] If we consider the second and third equations of $Ax = b$, these are also subparallel and there is both a local mean noise and local noise variance.

Table 4.7: Relative peak error for the simultaneous inversion (si), substitution (su), and unregularized solutions (un) when adding a global noise variance.

		Error	**Noiseless Error**	**Change**
Balanced	E_{si}	3.1%	2.9%	+0.2%
	H_{si}	0.7%	0.7%	+0.0%
	E_{su}	25.5%	25.5%	+0.0%
	H_{su}	24.7%	24.7%	+0.0%
	E_{un}	47.0%	46.9%	+0.1%
	H_{un}	119.7%	119.7%	+0.0%
Unbalanced	E_{su}	13.0%	13.0%	+0.0%
	H_{su}	18.9%	18.9%	+0.0%

subparallel, the noise variance is effectively cancelled out.[19]

If instead we add the same noise variance b_e locally to a pair of points in b, we may see the behavior that was predicted. Table 4.8 summarizes the results obtained from this case. Unfortunately, we do not see the behavior that we predicted; that is, the unregularized

Table 4.8: Relative peak error for the simultaneous inversion (si), substitution (su), and unregularized solutions (un) when adding a local noise variance.

		Error	**Noiseless Error**	**Change**
Balanced	E_{si}	3.1%	2.9%	+0.2%
	H_{si}	0.4%	0.7%	−0.3%
	E_{su}	25.3%	25.5%	−0.2%
	H_{su}	24.5%	24.7%	−0.2%
	E_{un}	48.6%	46.9%	+1.7%
	H_{un}	120.1%	119.7%	+0.4%
Unbalanced	E_{su}	12.3%	13.0%	−0.7%
	H_{su}	17.8%	18.9%	−1.1%

solution does not see a large error. However, when looking at the reconstructed phantom fields for the unregularized solution, we saw that the local noise variance introduces high

[19] If we consider the first four equations in $Ax = b$, we can separate these into pairs of two. The effective lines drawn by each pair in a 2D projection is almost equivalent to the mean equation, which means that there will effectively be zero noise in the problem.

spatial frequency components which affected the accuracy of the distribution but not the fields nearest to the DUT. If we look at the relative normed error for the unregularized y_{un},[20] it was 116% for the noiseless case and 154% for the local noise variance. The unregularized solution is indeed susceptible to noise variance, but perhaps not as much as the substitution methods are sensitive to mean noise.

Although there were a few unpredicted results, overall there is good agreement with the properties predicted by the first two factors in Section 4.3. When also considering the discussion on the matrix ranks, we can conclude that the ill-posedness of B with respect to A contributes to the greater noise sensitivity of the substitution method.

Factor 3

To check the accuracy of B, we evaluate the relative norm

$$\frac{||Bx_T - y_T||_2}{||y_T||_2}, \tag{4.41}$$

where the subscript T denotes the true currents. For the single case of discretization, we can only compare the accuracies of the balanced and unbalanced schemes. These norms were found to be 2.04% and 0.33% for the balanced and unbalanced cases respectively. We clearly see that the unbalanced scheme results in a more accurate B, which can explain the greater accuracy of the unbalanced substitution method for the noiseless case.

Factor 4

In the plots for the noiseless results, it is clear that the phantom surface fields are quite smooth for the 2D TM problem. It would make sense that the substitution methods could

[20]The relative normed error of the unregularized y_{un} is $\|y_{un} - y_T\|_2 / \|y_T\|_2$, where T denotes the true currents or fields.

ultimately produce good results since we can afford to regularize the solution without hurting the accuracy too much. In the following section, we will see that this is indeed the case.

4.5.3 Noisy Results

In this section, the performance of both the substitution and SI methods will be evaluated for noisy data. The problem parameters are the same as the noiseless case, but now there is random white noise added to the synthetic measurement data.

Using random uniform noise with NL = 0.5 (see Section 4.2.2), the phantom fields were reconstructed using the substitution and SI methods. As mentioned previously, there is sufficient noise to require additional regularization, which was implemented with Krylov subspace regularization. However, we start by looking at the solutions with no Krylov subspace regularization, which will henceforth be referred to as the unregularized SI or unregularized substitution solutions. The solutions which use the Krylov subspace regularization will be referred to as the regularized SI or regularized substitution solutions.

Table 4.9 summarizes the mean (E and H) relative peak field error for the unregularized algorithms.[21] As expected, the substitution methods are very sensitive to added measurement

Table 4.9: Case 1 unregularized mean relative peak field errors for the SI, balanced substitution, and unbalanced substitution methods for a noise level of NL=0.5.

Method	Mean Relative Peak Field Error
SI	17.8%
Balanced Substitution	4.67×10^3%
Unbalanced Substitution	1.3×10^4%

[21] The mean relative peak field error is the mean of the individual relative peak field errors for each field component.

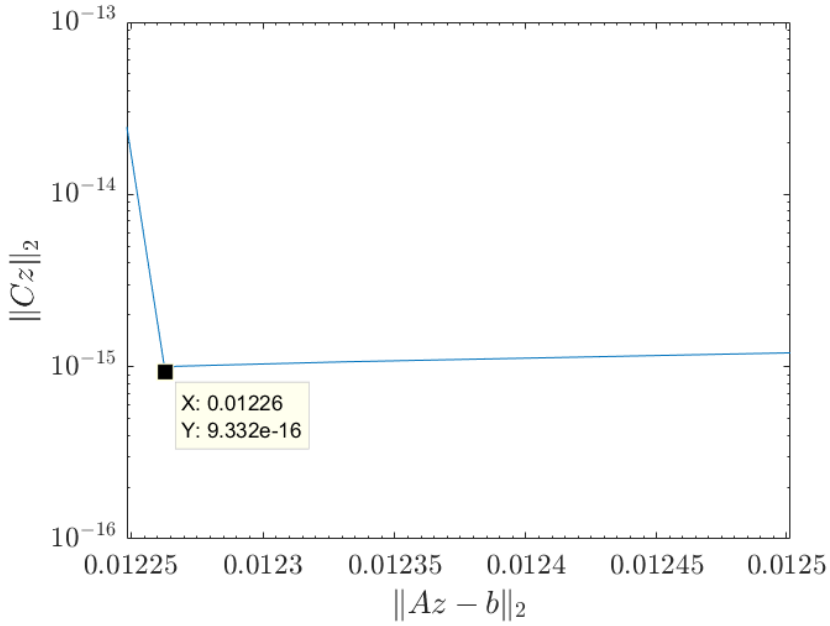


Fig. 4.17: L-curve for the unregularized SI method using the noisy data for Case 1 with the corner highlighted.

noise. In contrast, with an appropriate λ_C, the SI method is reasonable. The regularization weight was found to be 3.2×10^9 which corresponds to the corner of the L-curve shown in Figure 4.17. Although this seems to be a very large value for the regularization weight, the corner is still clearly recognizable, and it corresponds to this weight.[22] With such a large weight, the influence of noise in the data equation $Az = b$ is minimized and a reasonable solution can be obtained.

If we allow additional Krylov subspace regularization, we see that the substitution methods can obtain good results. Table 4.10 summarizes the mean relative peak field error for the three methods. In addition, the iteration at which the CG algorithm is truncated is

[22]It is useful to recall that the current link is supposed to be satisfied by the true solution. A large regularization weight does not necessarily prevent the recovery of the true solution. The concept of over-regularization does not apply to the current link in the same way that is does for unbiased or general regularizers.

Table 4.10: Case 1 regularized mean relative peak field errors for the SI, balanced substitution, and unbalanced substitution methods for a noise level of NL=0.5.

Method	Error	Truncation Iteration
SI	11.2%	200
Balanced Substitution	7.2%	7
Unbalanced Substitution	2.0%	5

presented. With sufficient truncation, the substitution methods are able to get very good results which are better than the SI method. Evidently the true phantom fields are sufficiently smooth (for the 2D TM case) that we can afford to severely truncate the CG algorithm. Note that the substitution methods, in addition to requiring more truncation, were also more sensitive to the choice of truncation. We will examine the sensitivity to truncation in the following section.

Sensitivity to Truncation

We begin by comparing the balanced and unbalanced substitution methods for a few choices of truncation. Table 4.11 summarizes the mean relative peak field errors for these choices. These results show that the substitution method is indeed sensitive to the choice of truncation. In addition, we see that the unbalanced method is even more sensitive to truncation, which follows from the increased noise sensitivity. If it is reasonable to truncate early then there will be no problem, but if more information is needed, there are only a few iterations

Table 4.11: Mean relative peak field errors for the substitution methods for varying choices of truncation.

CG Iterations	Balanced Error	Unbalanced Error
2	11.8%	6.6%
5	7.2%	2.0%
7	7.4%	46.8%
10	15.1%	94.1%
13	81.9%	184.5%

before the solution becomes unstable.

Comparatively, the SI method is much less sensitive to the choice of truncation. Table 4.12 summarizes the mean relative peak field errors for a few choices of truncation. These results again highlight the strength of the SI method; the self-correcting nature of the L-curve allows for practically any choice of truncation.

Table 4.12: Mean relative peak field errors for the SI method for varying choices of truncation.

CG Iterations	λ_C	**Error**
50	1.5×10^4	11.2%
100	10^9	11.2%
200	10^9	11.2%
500	10^{11}	11.1%

Although the substitution methods can produce better results, they are much more sensitive to the choice of truncation. On the other hand, the SI method is quite insensitive; the consistency of the SI method makes it an attractive option for a practical measurement, where precision is generally more important than accuracy.[23]

Sensitivity to λ_C

We would like to quickly discuss the sensitivity of the SI method with respect to the regularization weight λ_C. In a practical scenario, it may be beneficial to use a known regularization weight, perhaps if the algorithm to find the L-curve corner is inconsistent or requires a lot of time. The SI method would be more useful if the choice of λ_C is not critical. Fortunately, this does appear to be the case. When truncating at 200 CG iterations, any choice of λ_C in the range $10^8 < \lambda_C < 10^{11}$ results in a mean relative peak field error under 12%. This

[23] All physical measurements suffer from imprecision due to many factors. A method which can guarantee a 10% error is often preferable to a method which can produce anything from 5% to 50%.

is a very large range, which is good for the SI method. Note that in general, this range decreases as the degree of truncation is also reduced.

Sensitivity to Noise Randomness

In all of the previous results, the same random seed for the noise was used. The performance of the three methods was briefly examined for varying seeds of random noise to ensure that the previous results were not due to a "lucky" seed. Fortunately, the overall performance, and more importantly the conclusions, did not change from seed to seed. The SI method was consistently insensitive to measurement noise and the regularized substitution methods produced the best results (when truncating the CG algorithm).

4.6 Case 2 2D TM Results

When considering Factor 1, a case of interest was the scenario where the condition numbers of A and B are similar. To this end, we reduce the number of phantom current basis functions to 36, and also reduce the number of measurements to 72[24]. We begin by examining the major factors for this problem in order to predict the performance.

4.6.1 Factor 1

The condition numbers of A and B for the balanced and unbalanced cases are summarized in Table 4.13. Since the condition numbers are relatively similar, we can expect similar results for the balanced substitution and SI methods based on the discussion presented in Section 4.3.1. The unbalanced substitution method may differ depending on Factor 3.

[24]These measurements are spaced by $\Delta\varphi = \pi/180 = 1^\circ$ beginning at $\varphi = 0$.

Table 4.13: Case 2 balanced and unbalanced condition numbers for Factor 1.

	Matrix	Condition Number	Ratio
Balanced	A	1.1×10^{17}	$\approx 10^2$
	B	3.1×10^{15}	
Unbalanced	A	1.9×10^{18}	≈ 1
	B	7.4×10^{18}	

4.6.2 Factor 2

To examine Factor 2, we again look at the ranks of A, C, and the concatenated $[A;C]$ matrix. Table 4.14 summarizes these ranks and indicates any redundancy between the matrices. In both the unbalanced and balanced cases, there is redundant information in C

Table 4.14: Comparison of matrix ranks for A, C, and concatenated A-C matrices.

	Matrix	Rank	Rows Lost
Balanced	A	18	
	C	72	
	$[A;C]$	87	3
Unbalanced	A	18	
	C	72	
	$[A;C]$	85	5

which indicates that B is ill-posed with respect to A. We can expect that the substitution methods are very sensitive to noise.

4.6.3 Factor 3

To examine Factor 3, we again look at how well the current link is satisfied with the true currents, i.e., we evaluate the relative norm $\frac{||Bx_T - y_T||_2}{||y_T||_2}$. Similarly to Case 1, the unbalanced norm (0.33%) was smaller than the balanced norm (1.49%), so we can expect the unbalanced substitution method to be better than the balanced scheme.

4.6.4 Factor 4

Since the true phantom fields have not changed, we can expect that the substitution methods can produce reasonable results using noisy data if we sufficiently truncate the CG algorithm.

4.6.5 Case 2 Results

We start by looking at the mean relative peak field error for the noiseless[25] measurement data. Table 4.15 summarizes these results for the three methods. As predicted by Factor 1, the SI and balanced substitution methods produce very similar results. As predicted by Factor 3, the balanced substitution method produces better results.

Table 4.15: Case 1 mean relative peak field errors for the SI, balanced substitution, and unbalanced substitution methods for noiseless data.

Method	Mean Relative Peak Field Error
SI	5.4%
Balanced Substitution	5.2%
Unbalanced Substitution	3.4%

We would like to bring attention to the fact that the substitution methods produced better results than in Case 1 for noiseless data, despite using the same number of unknowns and reducing the number of measured data. Under conventional numerical wisdom, fewer data for the same number of unknowns should produce worse results. However, by carefully examining Factor 1, we can explain this difference. Recall that the fields that we really want are the phantom fields y. Since the cut-off frequencies of A and B are similar, any instability in x which results from the data equation is subsequently filtered by the operation $y = Bx$. The DUT currents are free to change as necessary to satisfy the data norm, all without significantly affecting the phantom currents y. In fact, if we look at the reconstructed DUT currents in Figure 4.18, we see that the currents are the same for both methods and that

[25] We would like to emphasize that there is an effective added noise due to finite discretization error.

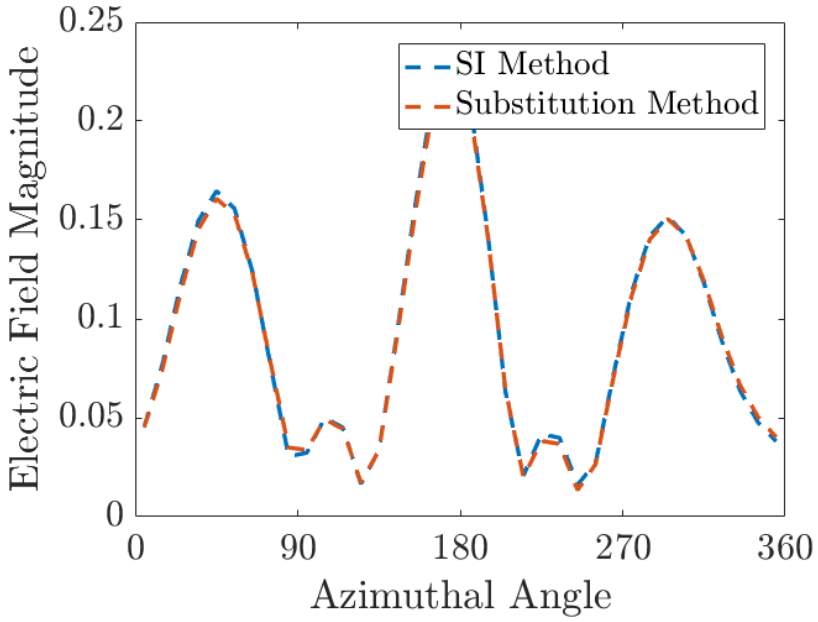


Fig. 4.18: Case 2 reconstructed DUT currents for the balanced substitution and SI methods.

there is a large spatial frequency component. Due to the smoothing properties of B, this is not an issue and the reconstructed phantom currents are accurate.

Next, we look at the noisy results, again using a noise level of NL=0.5. Table 4.16 summarizes the mean relative peak field errors for the unregularized and regularized methods, where "its." denotes the truncation iteration for the CG algorithm. Similarly to Case 1,

Table 4.16: Case 2 unregularized and regularized mean relative peak field errors for the SI, balanced substitution, and unbalanced substitution methods for a noise level of NL=0.5.

Method	Unregularized Error	Regularized Error	Its.
SI	10.51%	10.49%	45
Balanced Sub.	5.8×10^4%	10.0%	5
Unbalanced Sub.	1.1×10^3%	5.1%	2

the SI method can produce a reasonable solution in the unregularized case, and the sub-

stitution methods are very sensitive to noise. Unlike Case 1, the unbalanced substitution method is less sensitive to noise. This can be understood by observing the relative condition numbers of A and B; since they are even closer in the unbalanced scheme[26], any instability in the DUT currents x is filtered when evaluating $y = Bx$. However, since Factor 2 remains satisfied, the unbalanced substitution method is still very sensitive to measurement noise. Finally, due to the increased accuracy of B in the unbalanced scheme, the unbalanced substitution method produces the best results in the regularized case.

4.7 Case 3 2D TM Results

A final case of interest is the scenario where there is a large discrepancy between the condition numbers of A and B, i.e., B has a small condition number. To this end, we reduce the number of phantom current basis functions to 30, and we reduce the number of DUT current basis functions to 3.[27] In addition, we reduce the number of measurements to 30, such that the ratio of measurements to unknowns is the same as in Case 1.[28] We begin by examining the major factors for this problem in order to predict the performance.

4.7.1 Factor 1

The condition numbers of A and B for the balanced and unbalanced cases are summarized in Table 4.17. Since B is effectively well-conditioned, we can expect drastic differences for the SI and substitution methods. Any instability in x will necessarily be represented in y under the substitution scheme, which means that we can expect relatively poor results.

[26] The condition number ratio in the balanced scheme was $\approx 10^2$, and it was ≈ 1 in the unbalanced scheme.
[27] Note that this choice of discretization still satisfies the general rule of thumb for the maximum element size which is $\lambda/10$.
[28] These measurements are spaced by $\Delta\varphi = \pi/180 = 1^\circ$ beginning at $\varphi = 0$.

Table 4.17: Case 3 balanced and unbalanced matrix condition numbers for Factor 1.

	Matrix	Condition Number	Ratio
Balanced	A	1.2×10^{16}	$\approx 10^{16}$
	B	32	
Unbalanced	A	4.5×10^{16}	$\approx 10^{12}$
	B	7.7×10^{4}	

4.7.2 Factor 2

Unlike the previous two cases, we cannot compare the ranks of the matrices, as the sum rank(A)+rank(C) (13+60) exceeds the actual maximum rank of the concatenated matrix $[A; C]$ (66). Indeed, if we evaluate the rank of $[A; C]$, we get the maximum rank. In order to determine if there is any redundancy, we can utilize the observations that were made regarding the sensitivity of the substitution method to a global mean noise. If we add $b_e = 10^{-3}$ to each measured data, the mean relative peak field error is 42.7%; this is not at all close to the errors that were seen for Case 1. We can conclude that the degree of ill-posedness for B relative to A is low. Although Factor 1 predicts poor results for the substitution method, Factor 2 indicates that they will instead be fairly good. We will continue this discussion after showing some results.

4.7.3 Factor 3

To examine Factor 3, we again look at how well the current link is satisfied with the true currents, i.e., we evaluate the relative norm $\frac{||Bx_T - y_T||_2}{||y_T||_2}$. Unlike the previous cases, the unbalanced norm (5.07%) was larger than the balanced norm (3.44%), so we can expect the balanced substitution to be better than the unbalanced scheme.

4.7.4 Factor 4

Since the true phantom fields have not changed, we can expect that the substitution methods can produce reasonable results using noisy data if we sufficiently truncate the CG algorithm.

4.7.5 Case 3 Results

We start by looking at the mean relative peak field error for the noiseless[29] measurement data. Table 4.18 summarizes these results for the three methods. With these results, we

Table 4.18: Case 3 mean relative peak field errors for the SI, balanced substitution, and unbalanced substitution methods for noiseless data.

Method	Mean Relative Peak Field Error
SI	10.7%
Balanced Substitution	8.3%
Unbalanced Substitution	25.5%

begin to see the strength of analyzing the major factors proposed in this chapter. Although Factor 1 indicates that there may be a large discrepancy between the substitution and SI methods, Factor 2 indicates that there will be less instability for the substitution method. In essence, Factor 1 indicates the potential difference between the methods, whereas Factor 2 indicates how large this difference may be. In other words, we need to consider both of these to understand the results. Finally, we note that since B was inaccurate in the unbalanced scheme, the unbalanced substitution method had the worst performance.

Next, we look at the noisy results, again using a noise level of NL=0.5. Table 4.19 summarizes the mean relative peak field errors for the unregularized and regularized methods. We start by examining the unregularized results. Similarly to the previous cases, the SI method is able to produce a reasonable result when using an appropriate weight λ_C (7×10^6). Somewhat surprisingly, the balanced substitution method sees a large error in the unregularized

[29] We would like to emphasize that there is an effective added noise due to finite discretization error.

Table 4.19: Case 3 unregularized and regularized mean relative peak field errors for the SI, balanced substitution, and unbalanced substitution methods for a noise level of NL=0.5.

Method	Unregularized Error	Regularized Error	Its.
SI	20.6%	15.6%	55
Balanced Sub.	7.1×10^4%	43.3%	3
Unbalanced Sub.	337.4%	8.6%	2

case. Perhaps it is in the noisy case where Factor 1 begins to dominate over Factor 2. On the other hand, the unbalanced substitution method is relatively stable, which is likely due to the increased condition number of B which can filter erroneous components in x.

We next look at the regularized results. The most interesting result comes from the balanced substitution method. The reason for the large error is due to a large sensitivity to the truncation iteration. The error at 3 iterations is positive, whereas the error at 2 iterations is negative; this means that a close solution lies in between the two choices of truncation. A different regularization technique may be able to obtain a reasonable solution, but it is evident that the tolerance on the regularization weight would be very small. In contrast, the SI method was once again relatively insensitive to the choice of truncation. Finally, the unbalanced substitution method was the most accurate (likely due to Factor 4), although it was also sensitive to the choice of truncation.

4.8 Discussion

In the motivation section, four factors were identified which could potentially explain the differences between the SI and substitution methods. For each of the three cases of discretization, the noiseless and noisy results were analyzed using these factors. It was found that differences in the noiseless case could be reasonably explained by considering the first three factors. The first factor, which is concerned with the relative condition numbers of

A and B, was useful for judging the potential, but not necessary, differences between the methods. In particular, similar condition numbers indicated similar results for the substitution and SI cases, whereas a condition number mismatch indicated a potential difference. In order to ascertain the actual difference, the second factor needed to be analyzed. The second factor is concerned with any redundancy in B with respect to A; this is represented as a loss of rank from the concatenation of the A and C matrices. When there is redundancy, a situation similar to the 2-by-2 system presented in Section 4.3 occurs which results in instability for the substitution method. In Case 3, we saw that although Factor 1 indicated a large potential difference, a weak Factor 2 determined that this difference would be small in the noiseless case. In essence, Factor 1 explains why there may be a difference, and Factor 2 explains how this difference is felt. Finally, the third factor, which is concerned with the accuracy of B, helped explain the differences between the unbalanced and balanced substitution methods.

In the noisy case, the common behavior was that the SI method was insensitive to noise and the substitution methods were very sensitive to noise. However, since the phantom surface fields are smooth in the 2D TM case, the substitution methods generally performed better than the SI method since the additional Krylov subspace regularization did not affect the solution significantly. In this sense, Factor 4, which is concerned with the smoothness of the phantom currents, is perhaps the most important and deciding factor in the substitution versus SI discussion. If the phantom surface field distribution is not too smooth in the practical case, then the SI method may be preferable. If the fields are smooth, then the substitution method may instead be preferable. Nonetheless, it is clear that both methods would need to be investigated for a practical 3D case.

Chapter 5

Conclusion

In this thesis, the noninvasive SAR characterization problem was approached within the framework of electromagnetic inversion. In order to fully understand and appreciate the problem, we began with an overview of existing invasive SAR measurement methods. Contemporary SAR measurements are made by directly measuring the electric field internal to the homogeneous liquid human phantom, from which the SAR can be calculated. These invasive measurements can be time-consuming and thus there has been a push towards alternative methods which reduce the measurement time. One possible option for acceleration is to perform a noninvasive measurement; that is, we measure the EM fields external to the phantom and then infer the electric field internal to the phantom. The noninvasive measurement would also allow for the use of inhomogeneous solid phantoms and pre-existing field measurement hardware. These advantages come with increased software complexity. In particular, noninvasive SAR measurements within the framework of electromagnetic inversion are inverse problems, which suffer from solution non-uniqueness and instability.

Due to the electromagnetic equivalence and uniqueness theorems, there are many different approaches to the noninvasive SAR characterization problem. These approaches can

be classified by how the scattering of the phantom is modelled. First, there is the inhomogeneous background approach, which uses the inhomogeneous green's function to embed the information of the phantom. Although this is conceptually simple, obtaining the inhomogeneous green's function is not a trivial task. This leads to another group of methods, which are homogeneous (free-space) background approaches. There is the contrast source approach, which replaces the phantom with equivalent volumetric sources which radiate in free space. An important assumption in this formulation is the absence of DUT antenna loading, which invalidates the method in most cases. Another free-space background method is the surface field approach, which replaces the phantom with surfacic equivalent currents which radiate in free space. This method is equivalent to a dual-antenna measurement of the equivalent phantom and DUT currents, so existing near-field antenna measurement algorithms can be used for this approach. Since the antennas are not independent, there are hybrid approaches which incorporate this dependency into the free-space formulation. If the dependency is substituted into the surface field method, the method is similar to previous noninvasive approaches. Alternatively, the dependency can be used as a regularization term, which is the key feature of the simultaneous inversion (SI) method proposed for the first time in this thesis.

The first method which was examined was the surface field approach. The general algorithm and accompanying numerical techniques were largely based on existing algorithms and techniques used for near-field antenna measurements. Two novel numerical techniques were proposed in order to tackle the particularly severe ill-posedness of the noninvasive SAR problem. The surface field algorithm was fully implemented for 2D TM problems and partially implemented for fully-vectorial 3D problems. The 2D algorithm was evaluated using synthetic MATLAB data and also simulated ANSYS HFSS data. Although the SAR distribution could be reconstructed fairly well, there was difficulty in reconstructing the phase

of the surface phantom fields. The 3D algorithm was also evaluated using simulated HFSS data. The algorithm worked reasonably well for a partial implementation. It is reasonable to assume that the fully-implemented 3D algorithm could produce similar results to the 2D case. In the end, it was concluded that the dependency between the phantom and DUT surface fields is too important to ignore, which leads to the novel method proposed by this thesis.

The simultaneous inversion method was presented and proposed to be particularly resistant to measurement noise in comparison to the alternative substitution method. The theoretical framework motivating this proposition was presented. In particular, four major factors were identified which contribute to potential differences in performance between the SI and substitution methods. The main proposition of noise resistance and the four motivating factors were verified for 2D TM problems. The general observation was that with noisy measurement data, the SI method produced a much more stable solution than the substitution method without any additional regularization. When additional regularization was used, the substitution method was generally better, but the tolerance on the regularization weight (which was equivalent to the truncation iteration of the CG algorithm) was found to be very small. For all cases, the SI method and its parameters were insensitive to measurement noise, which supports the proposition that it is particularly resistant to noise. In addition, the four major factors were useful for understanding and predicting the differences between the SI and substitution methods. It was concluded that the optimal algorithm for a practical 3D case would depend on these factors.

References

[1] United Nations, *Sources and Effects of Ionizing Radiation.* New York: United Nations Publications, 2011.

[2] J. Walker, *Fundamentals of Physics*, 9th ed. Hoboken, NJ: Wiley, 2011.

[3] A. V. Vorst, A. Rosen, and Y. Kotsuka, *RF/Microwave Interaction Mechanisms in Biological Materials.* IEEE, 2006. [Online]. Available: https://ieeexplore-ieee-org.uml.idm.oclc.org/document/5237918

[4] Consumer and Clinical Radiation Protection Bureau, Health Canada, "Limits of human exposure to radiofrequency electromagnetic energy in the frequency range from 3 kHz to 300 GHz," 2015.

[5] "IEEE recommended practice for determining the peak spatial-average specific absorption rate (SAR) in the human head from wireless communications devices: Measurement techniques," *IEEE Std 1528-2013 (Revision of IEEE Std 1528-2003)*, pp. 1–246, Sep. 2013.

[6] "Measurement procedure for the assessment of specific absorption rate of human exposure to radio frequency fields from hand-held and body-mounted wireless communication devices - part 1: Devices used next to the ear (frequency range of 300 MHz to 6 GHz)," *IEC 62209-1:2016*, pp. 1–471, July 2016.

[7] A. Faraone, D. O. McCoy, C. M. Chou, and Q. Balzano, "Characterization of miniaturized E-field probes for SAR measurements," in *IEEE International Symposium on Electromagnetic Compatibility. Symposium Record (Cat. No.00CH37016)*, vol. 2, Aug 2000, pp. 749–754 vol.2.

[8] M. Douglas, N. Kuster, and K. Pokovic, "Past, present, and future of SAR evaluations," in *2016 URSI Asia-Pacific Radio Science Conference (URSI AP-RASC)*, Aug 2016, pp. 426–428.

[9] V. Hombach, K. Meier, M. Burkhardt, E. Kuhn, and N. Kuster, "The dependence of EM energy absorption upon human head modeling at 900 MHz," *IEEE Transactions on Microwave Theory and Techniques*, vol. 44, no. 10, pp. 1865–1873, Oct 1996.

[10] A. Christ, N. Chavannes, K. Popovich, H.-U. Gerber, and N. Kuster, "Numerical and experimental comparison of human head models for SAR assessment," pp. 234–240, Oct 2000.

[11] C. Li, M. Douglas, E. Ofli, B. Derat, S. Gabriel, N. Chavannes, and N. Kuster, "Influence of the hand on the specific absorption rate in the head," *IEEE Transactions on Antennas and Propagation*, vol. 60, no. 2, pp. 1066–1074, Feb 2012.

[12] A. Kyriacou, A. Christ, E. Neufeld, E. Cabot, S. Kuhn, M. Douglas, and N. Kuster, "Numerical and experimental comparison of human head models for SAR assessment," in *Bioelectromagnetics Annual Meeting*, June 2010.

[13] H. Virtanen, J. Keshvari, and R. Lappalainen, "Interaction of radio frequency electromagnetic fields and passive metallic implants — a brief review," *Bioelectromagnetics*, vol. 27, pp. 431–439, 2006.

[14] M. Y. Kanda, M. G. Douglas, E. D. Mendivil, M. Ballen, A. V. Gessner, and Chung-Kwang Chou, "Faster determination of mass-averaged SAR from 2-D area scans," *IEEE Transactions on Microwave Theory and Techniques*, vol. 52, no. 8, pp. 2013–2020, Aug 2004.

[15] K. Kiminami, T. Iyama, T. Onishi, and S. Uebayashi, "Novel specific absorption rate (SAR) estimation method based on 2-D scanned electric fields," *IEEE Transactions on Electromagnetic Compatibility*, vol. 50, no. 4, pp. 828–836, Nov 2008.

[16] B. Derat, L. Aberbour, and A. Cozza, "Near-field and vector signal analysis techniques applied to specific absorption rate measurement," in *2015 IEEE MTT-S 2015 International Microwave Workshop Series on RF and Wireless Technologies for Biomedical and Healthcare Applications (IMWS-BIO)*, Sep. 2015, pp. 34–35.

[17] B. Derat, L. Aberbour, K. Quelever, T. Julien, M. Dembinski, S. Pannetrat, and A. Cozza, "A novel technology for fast and accurate specific absorption rate measurement," in *Int. Workshop on Antenna Technology (iWAT)*, Karlsruhe,Germany, March 2013.

[18] P. Mojabi, N. Firoozy, N. Bayat, T. Brown, C. Narendra, P. Mojabi, C. Niu, T. Tiede, T. Neusitzer, X. Li, I. Jeffrey, J. LoVetri, and D. Barber, "Electromagnetic inversion for biomedical imaging, antenna characterization, and sea ice remote sensing applications," in *2016 URSI Asia-Pacific Radio Science Conference (URSI AP-RASC)*, Aug 2016, pp. 586–589.

[19] C. Gilmore, A. Abubakar, W. Hu, T. M. Habashy, and P. M. van den Berg, "Microwave biomedical data inversion using the finite-difference contrast source inversion method," *IEEE Transactions on Antennas and Propagation*, vol. 57, no. 5, pp. 1528–1538, May 2009.

[20] N. Amani, A. Jafargholi, and R. Pazoki, "A broadband VHF/UHF-loaded dipole antenna in the vicinity of a human body," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 10, pp. 5577–5582, Oct 2017.

[21] M. Koohestani, A. A. Moreira, and A. K. Skrivervik, "Feeding structure influence on performance of two UWB antennas near a human arm," in *The 8th European Conference on Antennas and Propagation (EuCAP 2014)*, April 2014, pp. 834–836.

[22] R. F. Harrington, *Time-Harmonic Electromagnetic Fields*. IEEE-Press, 2001. [Online]. Available: http://dx.doi.org/10.1109/9780470546710

[23] J. Richmond, "Scattering by a dielectric cylinder of arbitrary cross section shape," *IEEE Transactions on Antennas and Propagation*, vol. 13, no. 3, pp. 334–341, May 1965.

[24] R. Mitharwal and F. Andriulli, "A regularized boundary element formulation for contactless SAR evaluations within homogeneous and inhomogeneous head phantoms," *Comptes Rendus Physique*, vol. 16, 09 2015.

[25] S. Omi, T. Uno, T. Arima, and J. Wiart, "Reconstruction of internal field of dielectric objects for noninvasive SAR measurement using boundary integral equation," *IEEE Transactions on Electromagnetic Compatibility*, vol. 61, no. 1, pp. 48–56, Feb 2019.

[26] P. C. Hansen, *Rank-deficient and discrete ill-posed problems : numerical aspects of linear inversion*. Philadelphia, Pa: SIAM, 1998.

[27] T. Brown, C. Narendra, Y. Vahabzadeh, C. Caloz, and P. Mojabi, "On the use of electromagnetic inversion for metasurface design," *IEEE Transactions on Antennas and Propagation (early access)*, 2019.

[28] P. C. Hansen, "Numerical tools for analysis and solution of Fredholm integral equations of the first kind," *Inverse Probl.*, vol. 8, pp. 849–872, 1992.

[29] C. A. Balanis, *Antenna Theory: Analysis and Design*. Wiley-Interscience, 2005.

[30] C. Narendra, I. Jeffrey, and P. Mojabi, "Using the source reconstruction method to model incident fields in microwave tomography," *IEEE Antennas and Wireless Propagation Letters*, vol. 16, pp. 46–49, 2017.

[31] J. R. Shewchuk, "An introduction to the conjugate gradient method without the agonizing pain," 1994.

[32] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," nov 2012.

[33] P. Mojabi and J. LoVetri, "Enhancement of the Krylov subspace regularization for microwave biomedical imaging," *IEEE Transactions on Medical Imaging*, vol. 28, no. 12, pp. 2015–2019, Dec 2009.

[34] J. L. A. Quijano and G. Vecchi, "Field and source equivalence in source reconstruction on 3D surfaces," *Progress In Electromagnetics Research*, vol. 103, pp. 67–100, 2010.

[35] ——, "Improved-accuracy source reconstruction on arbitrary 3-D surfaces," *IEEE Antennas and Wireless Propagation Letters*, vol. 8, pp. 1046–1049, 2009.

[36] S. Rao, D. Wilton, and A. Glisson, "Electromagnetic scattering by surfaces of arbitrary shape," *IEEE Transactions on Antennas and Propagation*, vol. 30, no. 3, pp. 409–418, May 1982.

[37] Parini et al., *Theory and Practice of Modern Antenna Range Measurements.* Institution of Engineering and Technology, 2015.

[38] C. Geuzaine and J.-F. Remacle, "Gmsh: A 3-D finite element mesh generator with built-in pre- and post-processing facilities," *International Journal for Numerical Methods in Engineering*, vol. 79, no. 11, pp. 1309–1331, 2009.