Randomization for Efficient Nonlinear Parametric Inversion

Selin Sariaydin

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Mathematics

Eric de Sturler, Chair
Serkan Gugercin
Christopher A. Beattie
Misha E. Kilmer
Matthias Chung

April 19, 2018
Blacksburg, Virginia

Keywords: DOT, PaLS, stochastic programming, randomization, inverse problems,
optimization, model order reduction

Randomization for Efficient Nonlinear Parametric Inversion

Selin Sariaydin

(ABSTRACT)

Nonlinear parametric inverse problems appear in many applications in science and engineering. We focus on diffuse optical tomography (DOT) in medical imaging. DOT aims to recover an unknown image of interest, such as the absorption coefficient in tissue to locate tumors in the body. Using a mathematical (forward) model to predict measurements given a parametrization of the tissue, we minimize the misfit between predicted and actual measurements up to a given noise level. The main computational bottleneck in such inverse problems is the repeated evaluation of this large-scale forward model, which corresponds to solving large linear systems for each source and frequency at each optimization step. Moreover, to efficiently compute derivative information, we need to solve, repeatedly, linear systems with the adjoint for each detector and frequency. As rapid advances in technology allow for large numbers of sources and detectors, these problems become computationally prohibitive. In this thesis, we introduce two methods to drastically reduce this cost.

To efficiently implement Newton methods, we extend the use of simultaneous random sources to reduce the number of linear system solves to include simultaneous random detectors. Moreover, we combine simultaneous random sources and detectors with optimized ones that lead to faster convergence and more accurate solutions.

We can use reduced order models (ROM) to drastically reduce the size of the linear systems to be solved in each optimization step while still solving the inverse problem accurately. However, the construction of the ROM bases still incurs a substantial cost. We propose to use randomization to drastically reduce the number of large linear solves needed for constructing the global ROM bases without degrading the accuracy of the solution to the inversion problem.

We demonstrate the efficiency of these approaches with 2-dimensional and 3-dimensional examples from DOT; however, our methods have the potential to be useful for other applications as well.

Randomization for Efficient Nonlinear Parametric Inversion

Selin Sariaydin

(GENERAL AUDIENCE ABSTRACT)

Medical image reconstruction presents huge computational challenges due to the quantity of data generated by modern equipment. Each stage of processing requires the solution of more than a thousand large, three-dimensional problems. Moreover, as rapid advances in technology allow for ever larger numbers of sources and detectors and using multiple frequencies, these problems become computationally prohibitive. In this thesis, we develop two computational methods to drastically reduce this cost and produce good images from measurements.

First, we focus on efficiently estimating the absorption image while we reduce the cost of each optimization step by solving only for a few linear combinations of sources and of detectors.

Second, we can replace the full mathematical model by a reduced mathematical model to drastically reduce the size of the linear systems in each optimization step while still producing good image reconstructions. However, the computation of this reduced model still poses a formidable cost. Hence, we propose to reduce the cost of building the reduced model by sampling the sources and detectors. Using this reduced model for image reconstruction does not degrade the accuracy of the solutions and the quality of the image reconstruction.

We demonstrate the efficiency of these approaches with 2-dimensional and 3-dimensional examples from medical imaging. However, our methods have the potential to be useful for other applications as well.

To Atlas
Forever

Acknowledgments

First and foremost, I would like to express my gratitude towards my advisor, Prof. Eric de Sturler. I am grateful for his guidance, encouragement, and support. His immense knowledge has motivated me throughout my graduate study. He spent so many hours proofreading this thesis, and various academic papers for me that I can not thank him enough.

I would also like to thank Prof. Serkan Gugercin who was a second advisor for me. I owe my passion for reduced order models to his teaching excellence. His impact on my academic life is precious and I am very thankful for his mentorship.

I would like to thank my committee members Christopher Beattie, Misha E. Kilmer and Matthias Chung for their support and helpful directions. Many thanks to the Department of Mathematics staff and our chair, Peter Haskell for various grants and constant support. A very special thanks to Prof. Lizette Zietsman, for her support and kindness. Last but not the least, Eileen Shugart and Dr. Rachel Arnold for helping me become the teacher I am today.

My deepest gratitude is for my family. Especially, I am grateful to my mother, Hava Sariaydin, who always supported me and motivated me to do better. Of course, this list wouldn't be complete without thanking my husband, Serdar Aslan. His love and support helped me to overcome every obstacle. Many thanks to all of my friends for their great friendship and support.

Finally, this dissertation is dedicated to Atlas whom I will love and miss forever.

This material is based upon work supported by the National Science Foundation under Grant No. NSF-DMS 1217156 *and NSF-DMS* 1720305.
Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Inverse problems appear in many applications for identification and localization of anomalous regions, such as finding tumors in the body [5, 16, 21, 40, 61], X-ray techniques for luggage screening [53, 62], and contaminant pools in the earth [56, 63]. In these applications, we aim to recover an unknown image of interest within a given medium using a mathematical model (the forward model). The forward model relates the image of the unknown quantity to the measured data. In this work, we focus on forward models that are large-scale, discretized, partial differential equations (PDEs).

In partial differential equations-based (PDE-based) inverse problems with many measurements, we have to solve many large-scale discretized PDEs for each evaluation of the misfit or objective function. In the nonlinear case, each time the Jacobian is evaluated an additional set of systems must be solved. This leads to a tremendous computational cost, and this is by far the dominant cost for these problems. Hence, we need new computational techniques that characterize the medium quickly and efficiently for such inverse problems.

For the main application discussed in this work, diffuse optical tomography (DOT), the size of a realistic linear system is at least $O(10^6)$. The number of sources and detectors may be a thousand or more, and combined with multiple frequencies, the resulting computational cost is indeed very high. In this thesis, we introduce two new effective and computationally highly efficient methods to reduce the cost of the computing forward models and its derivatives:

- A randomized approach for nonlinear inversion.
- Randomization for the efficient computation of Reduced Order Models (ROMs).

The ideas presented in this dissertation are relevant to many other large-scale applications as well. We include a brief overview of the dissertation outline below.

A Randomized Approach for Nonlinear Inversion.
Several authors have proposed to drastically reduce the number of systems to be solved by exploiting stochastic techniques [32] and posing the problem as a stochastic optimization problem [54]. In this approach, the misfit or objective function is estimated using only a few appropriately chosen random linear combinations of the sources, referred to as *simultaneous random sources* [9, 32, 43, 44].

While some have reported good solution quality at a greatly reduced computational cost using these randomized approaches, for our problem of interest, DOT, the approach often does not lead to sufficiently accurate solutions. Therefore, in Chapter 2, we propose *two innovations* to the previously described technique. First, to efficiently exploit Newton-type methods, we show that using random linear combinations for detectors reduces the number of additional adjoint solves for the detectors. Second, after solving to a modest tolerance, we use a few simultaneous sources and detectors that are optimized to maximize the Frobenius norm of the sampled Jacobian to improve the rate of convergence and obtain more accurate solutions. We complement these optimized sources and detectors by random linear combinations of the sources and detectors constrained to a complementary subspace.

The first innovation is aimed at extending the idea of randomized sources to Newton-type methods, where we need to compute the Jacobian efficiently. In particular, we show that an approach similar to randomized source selection can be used to reduce the number of additional adjoint solves for the detectors. This contribution allows us to develop a Newton-based optimization approach to minimize the residual norm for parametric inversion.

The goal of our second innovation is to combat the observed stagnation in the residual norm decrease of our new stochastic optimization approach. Thus, in our new approach, we first solve with a fixed set of simultaneous random sources and detectors up to an intermediate tolerance on the objective function. To improve convergence, we exploit the (often fairly good) approximate solution after reaching a modest intermediate tolerance to obtain simultaneous sources and detectors that are optimized for convergence of the nonlinear least squares problem; we refer to these as *simultaneous optimized sources and detectors*. In this thesis, we propose two different ways to generate optimized sources and detectors. In Chapter 2, when a chosen intermediate tolerance is reached, we compute the full Jacobian once and compute orthonormal optimized simultaneous sources and detectors. Then, we complement these optimized simultaneous sources and detectors with a new set of random simultaneous sources and detectors constrained to a complementary subspace. We demonstrate the effectiveness of this approach with 2D and 3D examples from DOT.

In Chapter 3, we provide an alternative approach to compute simultaneous optimized sources and detectors using existing simultaneous random sources and detectors. In this approach, when a chosen intermediate tolerance is reached, similar to the previous approach, we compute the full Jacobian once. Then, we *replace* the least effective randomized sources and detectors by simultaneous sources and detectors that are optimized to maximize the Frobenius norm of the sampled Jacobian. There are several ways to do this. Hence, we provide

some implementation strategies and numerical results.

Randomization for the efficient computation of Reduced Order Models
In the past, the use of reduced order models (ROM) as surrogates in place of full order PDE solves has been proposed to drastically reduce the size of the linear systems solved in each optimization step for DOT, while still solving the inverse problem accurately [23]. However, as the number of sources and detectors increases, the construction of the ROM bases still incurs a substantial cost. Interpolatory model reduction requires the solution of large linear systems for all sources and frequencies as well as for all detectors and frequencies for each chosen interpolation point, followed by an expensive rank-revealing factorization to reduce the dimension. The rank-revealing factorization in building the ROM basis reveals that the standard method for ROM construction solves many more linear systems than needed. In Chapter 4, we propose to drastically reduce the number of large linear solves for constructing the global ROM basis using randomization techniques introduced in Chapter 2. In particular, we employ randomization to capture essentially the same subspace at much lower cost via sampling. Moreover, we provide a theoretical justification for exploiting low rank structure in the reduction basis and connect our approach to randomization in computing the interpolatory model reduction bases to tangential interpolation. 2D and 3D experiments justify the effectiveness of our approach.

Due to the sizes of the linear systems and the fact that the problems are 3D, the use of sparse solvers for large linear systems becomes impractical and iterative methods must be employed. The results in [11] suggest that in general high accuracy solves are not necessary to compute interpolatory reduced order models. Therefore, in Chapter 5, we include a numerical study on how sensitive the quality of the reduced order model is to the chosen tolerance.

Appendix A includes some detailed derivations.

Notation

- The matrices are denoted by bold-face capital letters: $\mathbf{A}$ and $\mathbf{\Phi}$
- The vectors are denoted by bold-face lower-case letters: $\mathbf{x}$, $\boldsymbol{\lambda}$
- Unless otherwise stated, $\|\cdot\|$ denotes the Frobenius norm
- We use $\imath$ to denote $\sqrt{-1}$

Chapter 2

Randomized Approach to Nonlinear Inversion Combining Random and Optimized Simultaneous Sources and Detectors

2.1 Introduction

The solution of nonlinear inverse problems requires solving many large-scale discretized PDEs in the evaluation of the forward problem. In parameterized inverse problems, we can compute the response of the system for a particular input by numerically solving the PDE. The forward model used in this thesis, see section 2.2.1, is already regularized using the parametric level set (PaLS) approach [1], and we focus on efficiently solving the nonlinear least squares problem

$$\min_{\mathbf{p}} g(\mathbf{p}) := \min_{\mathbf{p}} \frac{1}{2} \|\mathbb{M}(\mathbf{p}) - \mathbf{d}\|_2^2, \tag{2.1}$$

where $\mathbb{M}(\mathbf{p})$ is the vector of computed measurements obtained using the forward model for the parameter vector $\mathbf{p}$, and $\mathbf{d}$ is the vector of measured data at the detectors.

Each evaluation of $g(\mathbf{p})$ requires the solution of the PDE for all inputs and each frequency. Moreover, to efficiently compute derivative information using the co-state approach [59], we also need to solve linear systems with the adjoint for each detector and each frequency. This leads to an enormous computational bottleneck, as rapid advances in technology allow for large numbers of sources and detectors. Multiply this by the number of frequencies, and the number of linear systems to solve in the solution of (2.1) is very large indeed. For the main application discussed in this thesis, diffuse optical tomography (DOT), the number

of sources and the number of detectors may each be a thousand or more; the number of frequencies used is typically modest (less than ten) [23].

To solve the minimization problem (2.1), we use the Trust region algorithm with Gauss-Newton REGularized model Solution (TREGS) [24] that has proven very effective for parameterized problems of the type we consider in this thesis. In [23], reduced order models have been used to approximate both the function evaluation as well as its derivatives to compute regularized Gauss-Newton steps in TREGS. Here, we explore an alternative approach, following the work by Haber, Chung, and Herrmann [32]. The main idea in their paper was to drastically reduce the number of systems to be solved by exploiting randomization [32], posing the problem as a stochastic optimization problem [54]. In their approach, the misfit or objective function is estimated using only a few, appropriately chosen random linear combinations of the sources, referred to as *random simultaneous sources*, that are kept fixed over many optimization steps. In [54], this approach is referred to as the Sample Average Approximation (SAA) method.

The use of random simultaneous sources has been well-studied in various papers (see [9, 38, 44, 50, 58], and the references therein). While replacing the original objective function by the stochastic optimization problem corresponding to the random simultaneous sources seems to work well for direct current resistivity and seismic tomography [32], for the DOT problem, we find the approach does not lead to accurate recovery of the parameters. Therefore, we propose *two innovations* to the use of random simultaneous sources and detectors. First, the randomized treatment of the detector solves, and second combining random simultaneous sources and detectors with optimized simultaneous sources and detectors to best capture the sensitivity.

The first innovation extends the idea of randomized sources to randomized detectors which are needed for Newton-type methods to compute the Jacobian efficiently. In particular, we show that an approach similar to randomized simultaneous sources can be used to reduce the number of additional adjoint solves for the detectors.

The second innovation avoids stagnation in the residual norm decrease of the stochastic optimization approach. Using random simultaneous sources and detectors provides moderately accurate parameter solution estimates at a drastically reduced number of linear system solves. Thus, in our new approach, we first solve with a fixed set of random simultaneous sources and detectors to an intermediate tolerance. After reaching this intermediate tolerance, we use a few random simultaneous sources and detectors that are optimized to maximize the Frobenius norm of the sampled Jacobian (see Section 2.3). We complement these sources and detectors by random linear combinations of the sources and detectors constrained to a complementary subspace; we refer to these as *optimized simultaneous sources and detectors*. After this update, the optimization converges rapidly to a solution of the same quality as obtained using all sources and detectors. Our use of optimized simultaneous sources and detectors is based on two motivations. First, the regularized model problem solves in the TREGS nonlinear least squares solver [24] used for minimization in our method, and second

the fact that these optimized directions are best informed by the data.

The rest of this chapter is organized as follows. In Section 2.2, we briefly review DOT, PaLS, and TREGS. In Section 2.3, we show that using random linear combinations for detectors reduces the number of additional adjoint solves for the detectors. Further, in Section 2.4, we improve this approach by combining random and optimized simultaneous sources and detectors. We also give an outline of our implementation strategies. In Section 2.5, we demonstrate the effectiveness of combining random and optimized simultaneous sources and detectors using 2D and 3D experiments. Finally, we draw some conclusions and discuss future work in Section 2.6.

2.2 Background

In this section, we briefly review DOT, PaLS, and TREGS.

2.2.1 Diffuse Optical Tomography

DOT is a non-invasive, low cost alternative for breast and brain imaging compared with X-Ray and MRI. In DOT, near infra-red light from an array of sources is transmitted through the medium and measured with an array of detectors, see Figure 2.1 for the basic set-up of the problem. We consider the diffusion model for the photon flux $\phi(\mathbf{x})$ driven by an input

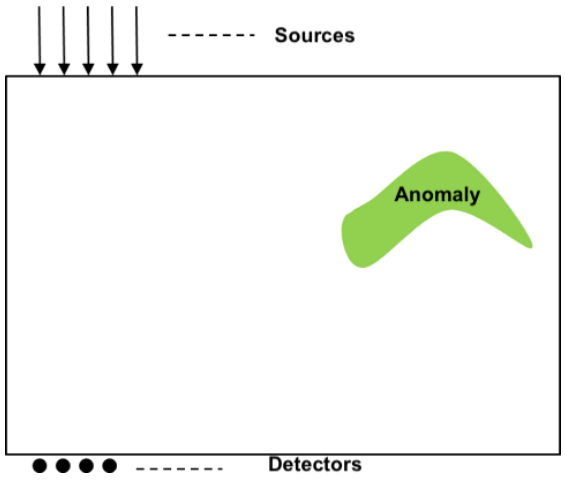


Figure 2.1: Schematic of a simple 2D DOT problem

source $g(\mathbf{x})$. The forward model for DOT, posed in the frequency domain, is given by

$$\begin{aligned} &-\nabla\cdot(D(\mathbf{x})\nabla\phi(\mathbf{x}))+\mu(\mathbf{x})\phi(\mathbf{x})+\frac{\imath\omega}{\nu}\phi(\mathbf{x})=g(\mathbf{x}), \\ &\text{for}\quad \mathbf{x}=(x_1,x_2,x_3)^T \text{ and } -a<x_1<a,\quad -b<x_2<b,\quad 0<x_3<c, \\ &\phi(\mathbf{x})=0 \text{ if } 0\le x_3\le c \text{ and either } x_1=\pm a, \text{ or } x_2=\pm b, \\ &0.25\phi(\mathbf{x})+\frac{D(\mathbf{x})}{2}\frac{\partial\phi(\mathbf{x})}{\partial\xi}=0 \text{ for } x_3=0, \text{ or } x_3=c, \end{aligned} \tag{2.2}$$

where $D(\mathbf{x})$ is the diffusion coefficient, $\mu(\mathbf{x})$ the absorption coefficient, ω is the frequency modulation of light, ν is the speed of light in the medium, and $\boldsymbol{\xi}$ is the outward unit normal [5].

Assuming that the diffusion coefficient is known (a common assumption for breast imaging), we use measurements and the forward model to recover the absorption coefficient of the medium, which can be used to distinguish healthy tissue from tumors [16]. Typical inversion methods would optimize for the desired physical quantity over a collection of grid points/voxels resulting in a parameter vector with at least $O(10^6)$ unknowns. Instead, we assume that the absorption field, $\mu(\mathbf{x})$, is expressible with a modest number of (unknown) parameters, $\mathbf{p}=[p_1,p_2,\ldots,p_{n_p}]^T$ as $\mu(\mathbf{x};\mathbf{p})$ using the PaLS approach. The details of the PaLS approach are given in Section 2.2.2.

Let n_d, n_s, and n_ω denote the number of detectors, sources, and frequencies, respectively. The discretization of (2.2) leads to computed measurements, $\mathbf{m}_i(\omega_j,\mathbf{p})\in\mathbb{C}^{n_d}$, for each source term, $\mathbf{b}_i$,

$$\mathbf{m}_i(\omega_j,\mathbf{p})=\mathbf{C}\left(\frac{\imath\omega_j}{\nu}\mathbf{E}+\mathbf{A}(\mathbf{p})\right)^{-1}\mathbf{b}_i, \tag{2.3}$$

where the rows of $\mathbf{C}$ correspond to the detectors[1]. $\mathbf{A}(\mathbf{p})$ derives from a finite difference discretization of the diffusion and absorption terms in (2.2), and $\mathbf{E}$ derives from the frequency term in (2.2). $\mathbf{E}$ is almost the identity except that it has zero rows for points on the boundary, $x_3=0$, $x_3=c$, in (2.2); so, $\mathbf{E}$ is singular.

For simplicity, we consider the nonlinear residual for a single frequency, $\omega_j=0$. In vector form, the residual is defined as follows

$$\mathbf{f}(\mathbf{p})=\begin{bmatrix}\mathbf{f}_1(\mathbf{p})\\ \vdots\\ \mathbf{f}_{n_s}(\mathbf{p})\end{bmatrix}=\begin{bmatrix}\mathbf{m}_1(\mathbf{p})-\mathbf{d}_1\\ \vdots\\ \mathbf{m}_{n_s}(\mathbf{p})-\mathbf{d}_{n_s}\end{bmatrix}=\begin{bmatrix}\mathbf{C}\mathbf{A}^{-1}(\mathbf{p})\mathbf{b}_1-\mathbf{d}_1\\ \vdots\\ \mathbf{C}\mathbf{A}^{-1}(\mathbf{p})\mathbf{b}_{n_s}-\mathbf{d}_{n_s}\end{bmatrix}, \tag{2.4}$$

where $\mathbf{f}_i\in\mathbb{R}^{n_d}$, $\mathbf{d}_i$ is the data vector with the measurements from the detectors corresponding to source $\mathbf{b}_i$, and the nonlinear least squares problem (2.1) becomes

$$\min_{\mathbf{p}}\frac{1}{2}\|\mathbf{f}(\mathbf{p})\|_2^2. \tag{2.5}$$

[1]In practice, we also split $\mathbf{m}_i$ in its real and imaginary parts.

For Newton-type algorithms, it is also necessary to construct the Jacobian of $\mathbf{f}(\mathbf{p})$,

$$\mathbf{J} = \frac{\partial \mathbf{f}(\mathbf{p})}{\partial \mathbf{p}} = \left[\begin{array}{ccc} \frac{\partial \mathbf{f}(\mathbf{p})}{\partial \mathbf{p}_1} & \cdots & \frac{\partial \mathbf{f}(\mathbf{p})}{\partial \mathbf{p}_{n_p}} \end{array} \right], \tag{2.6}$$

where the components of $\mathbf{J}$ are given by the small vectors

$$\mathbf{J}_{jk}(\mathbf{p}) = \frac{\partial}{\partial \mathbf{p}_k}(\mathbf{C}\mathbf{A}^{-1}(\mathbf{p})\mathbf{b}_j) = -\mathbf{C}\mathbf{A}^{-1}(\mathbf{p})\frac{\partial \mathbf{A}(\mathbf{p})}{\partial \mathbf{p}_k}\mathbf{A}^{-1}(\mathbf{p})\mathbf{b}_j \in \mathbb{R}^{n_d}. \tag{2.7}$$

Evaluating the objective function at $\mathbf{p}$ requires solving $n_s \cdot n_\omega$ large linear systems. Once $\mathbf{f}(\mathbf{p})$ and $\mathbf{A}^{-1}(\mathbf{p})\mathbf{b}_j$ are available, evaluating $\mathbf{J}$ using the co-state approach [59] requires solving an additional $n_d \cdot n_\omega$ adjoint systems for detectors. As a result, standard optimization approaches require $O(10^3 - 10^4)$ large linear system solves at each optimization step. The size of a realistic linear system is at least $O(10^6)$. This leads to an enormous computational bottleneck, and new computational techniques are needed.

We use TREGS [24] to solve the nonlinear least squares problem (2.5). The TREGS algorithm combines a trust region method with a regularized minimization of the Gauss-Newton (GN) model [25]. The local (GN) model at the current parameter vector, $\mathbf{p}_c$, is given by

$$g(\mathbf{p}_c + \boldsymbol{\delta}) \approx m_{GN}(\mathbf{p}_c + \boldsymbol{\delta}) = \frac{1}{2}\mathbf{f}_c{}^T\mathbf{f}_c + \mathbf{f}_c^T\mathbf{J}_c\boldsymbol{\delta} + \frac{1}{2}\boldsymbol{\delta}^T\mathbf{J}_c{}^T\mathbf{J}_c\boldsymbol{\delta}, \tag{2.8}$$

and its minimization is equivalent to the least squares problem problem

$$\min_{\boldsymbol{\delta}} \|\mathbf{J}_c\boldsymbol{\delta} + \mathbf{f}(\mathbf{p}_c)\|_2^2. \tag{2.9}$$

The TREGS algorithm favors updates corresponding to (1) the large singular values and (2) the left singular vectors with large components in $\mathbf{f}$ as determined by a generalized cross validation-like (GCV) criterion. Since the Jacobian tends to be ill-conditioned, the emphasis on large singular values leads to relatively small steps that provide relatively large reductions in the GN model (2.8). We refer the reader to [24] for more details of TREGS.

2.2.2 Parametric Level Set Methods

The Parametric Level Set (PaLS) approach [1] has been used to reduce the dimension of the search space, reconstruct complex geometries, and provide regularization to compensate for the influence of noise and ill-posedness in DOT [1, 6, 23]. We parameterize the absorption field using PaLS, and solve for a modest number of parameters that describe the shape of potential anomalies (tumors), rather than solving for absorption at every grid point. Hence, the parameterized absorption is defined as

$$\mu(\mathbf{x}) = \mu(\mathbf{x}; \mathbf{p}),$$

where $\mathbf{p} \in \mathbb{R}^{n_p}$ denotes the vector of parameters. Using PaLS, we parameterize the absorption $\mu(\mathbf{x}; \mathbf{p})$ as follows.

Let $\varphi : \mathbb{R}^+ \to \mathbb{R}$ be a smooth, compactly supported radial basis function (CSRBF), γ be a positive, small, real number, and $\|\mathbf{x}\|^\dagger := \sqrt{\|\mathbf{x}\|_2^2 + \gamma^2}$ denote the (regularized) Euclidean norm. Then, the PaLS function η with a vector of unknown parameters $\mathbf{p}$ consisting of expansion coefficients α_j, dilation coefficients β_j, and center locations $\boldsymbol{\chi}_j$ is defined as

$$\eta(\mathbf{x}, \mathbf{p}) := \sum_{j=1}^{m_0} \alpha_j \varphi(\|\beta_j(\mathbf{x} - \boldsymbol{\chi}_j)\|^\dagger). \tag{2.10}$$

The PaLS approach uses an approximate Heaviside function $H_\epsilon(r)$, where r is a scalar, to create a differentiable, but sharp transition from anomaly to background. The absorption $\mu(\mathbf{x}, \mathbf{p})$ takes the value $\mu_{in}(\mathbf{x})$ if $\mathbf{x}$ is inside the region and $\mu_{out}(\mathbf{x})$ if $\mathbf{x}$ is outside the region,

$$\mu(\mathbf{x}, \mathbf{p}) = \mu_{in}(\mathbf{x}) H_\epsilon(\eta(\mathbf{x}, \mathbf{p}) - c) + \mu_{out}(\mathbf{x})(1 - H_\epsilon(\eta(\mathbf{x}, \mathbf{p}) - c)), \tag{2.11}$$

where $c \in \mathbb{R}$ is a chosen cut-off parameter for the level set.

Figure 2.2 illustrates how PaLS represents the absorption field. Using PaLS, edges and complex boundaries can be captured with relatively few basis functions. The PaLS representation also regularizes the problem as a function of the number of basis elements used, hence no further regularization is needed. For further discussion of the PaLS parameters for DOT, we refer the reader to [1, 23].

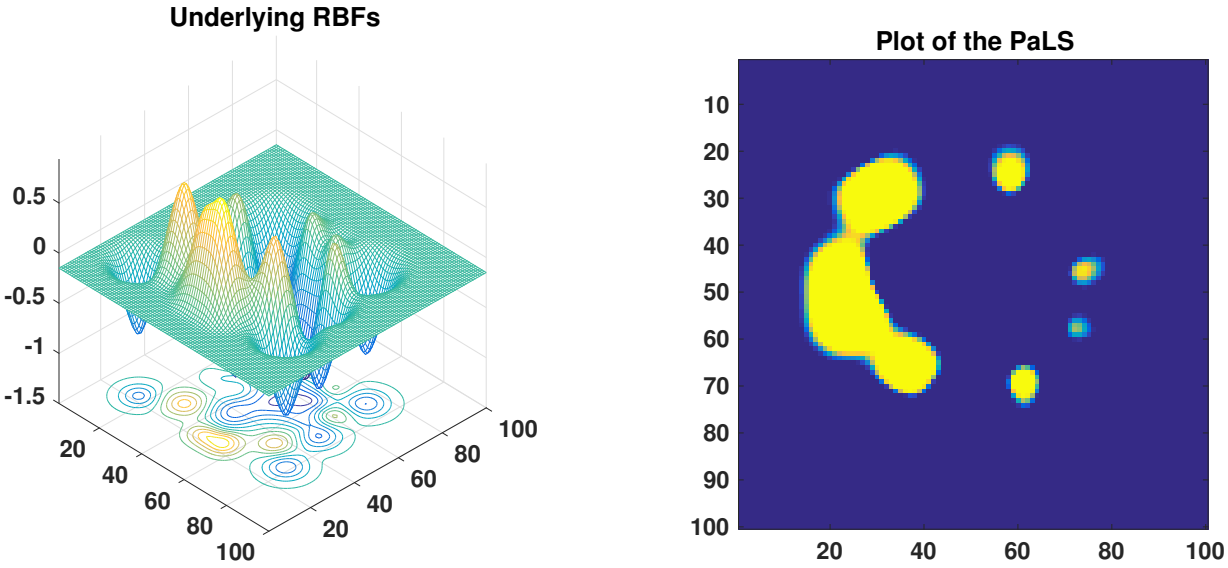
2.3 A Randomized Approach

We recast the nonlinear least squares problem as a stochastic optimization problem using randomization to drastically reduce the number of large linear systems in solving (2.4) and (2.7). The columns of $\mathbf{B} = [\mathbf{b}_1, \cdots, \mathbf{b}_{n_s}]$ are source terms, and we refer to any linear combination of these sources as a simultaneous source. Simultaneous random sources, $\mathbf{Br}$, with $\mathbf{r} \in \mathbb{R}^{n_s}$ a random vector, have been used in several areas [9, 32, 43, 44]. In this chapter, we introduce the concept of *optimized simultaneous sources and detectors* to improve the rate of convergence of the optimization and the quality of the inverse solution.

2.3.1 A Stochastic Optimization Approach

To recast (2.4)–(2.5) as a stochastic optimization problem, we first write the residual in matrix form. For a single frequency, we get

$$\mathbf{F}(\mathbf{p}) = [\mathbf{f}_1(\mathbf{p}) \;\; \mathbf{f}_2(\mathbf{p}) \;\; \cdots \;\; \mathbf{f}_{n_s}(\mathbf{p})] = \mathbf{C}\mathbf{A}^{-1}(\mathbf{p})\mathbf{B} - \mathbf{D}, \tag{2.12}$$



(a) (b)

Figure 2.2: (a) Surface and contour plot of a test anomaly on 100×100 mesh with 25 basis functions where the cut off is at $c = 0.15$. (b) The PaLS function of the test anomaly on the left. If $\eta(\mathbf{x}, \mathbf{p}) \geq 0.15$, then $\mathbf{x}$ is inside the anomaly (light) and if $\eta(\mathbf{x}, \mathbf{p}) < 0.15$, then $\mathbf{x}$ is outside the anomaly (dark).

where the vectors $\mathbf{f}_i \in \mathbb{R}^{n_d}$ are defined in (2.4), and consequently $\mathbf{f}(\mathbf{p}) = \text{vec}(\mathbf{F}(\mathbf{p}))$.[2] The columns of $\mathbf{D} = [\mathbf{d}_1, \cdots, \mathbf{d}_{n_s}]$ are the measurements corresponding to source $\mathbf{b}_i$. We have

$$\min_{\mathbf{p}} \|\mathbf{f}(\mathbf{p})\|_2^2 \quad = \quad \min_{\mathbf{p}} \sum_{j=1}^{n_s} \|\mathbf{C}\mathbf{A}^{-1}(\mathbf{p})\mathbf{b}_j - \mathbf{d}_j\|_2^2 = \min_{\mathbf{p}} \|\mathbf{C}\mathbf{A}^{-1}(\mathbf{p})\mathbf{B} - \mathbf{D}\|^2. \tag{2.13}$$

Each evaluation of the objective function requires solving $n_s \cdot n_\omega$ linear systems. Haber et al. [32] reduce this cost using simultaneous random sources combined with (stochastic) trace estimators, following Hutchinson [36].

Let $\mathbf{r}$ be a random vector with mean $\mathbf{0}$ and identity covariance matrix, and let $\mathbb{E}$ denote the expected value with respect to the random vector $\mathbf{r}$. Then

$$\mathbb{E}\left(\mathbf{r}^T\mathbf{F}(\mathbf{p})^T\mathbf{F}(\mathbf{p})\mathbf{r}\right) = \text{trace}\left(\mathbf{F}(\mathbf{p})^T\mathbf{F}(\mathbf{p})\right) = \|\mathbf{F}(\mathbf{p})\|^2,$$

As a particular choice, we choose $\mathbf{r}$ to be a realization from the Rademacher distribution, where each component of $\mathbf{r}$ is independently and identically distributed (i.i.d.) taking values from $\{-1, +1\}$, each with probability $\frac{1}{2}$. Then, as shown in [36], $\mathbf{r}^T\mathbf{F}(\mathbf{p})^T\mathbf{F}(\mathbf{p})\mathbf{r}$ is a minimum

[2] For multiple frequencies, we need to compute the residual for each frequency, $[\mathbf{F}(\omega_1, \mathbf{p})\ \mathbf{F}(\omega_2, \mathbf{p})\ \cdots] = [\mathbf{C}\mathbf{A}^{-1}(\omega_1, \mathbf{p})\,\mathbf{B} - \mathbf{D}_1 \quad \mathbf{C}\mathbf{A}^{-1}(\omega_2, \mathbf{p})\,\mathbf{B} - \mathbf{D}_2 \quad \cdots]$.

variance and unbiased estimator of the trace of $\mathbf{F}(\mathbf{p})^T\mathbf{F}(\mathbf{p})$. Thus, the nonlinear least squares problem can be written as a stochastic minimization problem

$$\min_{\mathbf{p}} \|\mathbf{F}(\mathbf{p})\|^2 = \min_{\mathbf{p}} \text{trace } \mathbf{F}(\mathbf{p})^T\mathbf{F}(\mathbf{p}) = \min_{\mathbf{p}} \mathbb{E}\left(\mathbf{r}^T\mathbf{F}(\mathbf{p})^T\mathbf{F}(\mathbf{p})\mathbf{r}\right). \tag{2.14}$$

For a random vector $\mathbf{r}$ and simultaneous random source $\mathbf{Br}$, we have

$$\mathbf{F}(\mathbf{p})\mathbf{r} = (\mathbf{C}\mathbf{A}^{-1}(\mathbf{p})\mathbf{B} - \mathbf{D})\mathbf{r} = \mathbf{C}\mathbf{A}^{-1}(\mathbf{p})\mathbf{B}\mathbf{r} - \mathbf{D}\mathbf{r}. \tag{2.15}$$

So, computing $||\mathbf{F}(\mathbf{p})\mathbf{r}||_2^2$ requires a single PDE solve per realization rather than n_s solves, which drastically reduces the cost of a function evaluation.

In contrast to the approach in [32], we use a Newton-type method, so we also need to reduce the cost of Jacobian evaluations. Therefore, we propose a variation that also drastically reduces the cost of computing $\mathbf{A}^{-T}(\mathbf{p})\mathbf{C}^T$ for the Jacobian.

Let $\boldsymbol{\ell} \in \mathbb{R}^{n_d}$ and $\mathbf{r} \in \mathbb{R}^{n_s}$ with all components i.i.d. uniformly from $\{-1, +1\}$. Then,

$$\begin{aligned}\mathbb{E}\left[\left(\boldsymbol{\ell}^T\mathbf{F}\mathbf{r}\right)^2\right] &= \mathbb{E}\left[\left(\sum_{j=1}^{n_s}\sum_{i=1}^{n_d} \ell_i F_{ij}\, r_j\right)^2\right] = \mathbb{E}\left[\left(\sum_{j=1}^{n_s}\sum_{i=1}^{n_d} \ell_i F_{ij}\, r_j\right)\left(\sum_{m=1}^{n_s}\sum_{k=1}^{n_d} \ell_k F_{km}\, r_m\right)\right] \\ &= \sum_{j,m=1}^{n_s}\sum_{i,k=1}^{n_d} F_{ij}\, F_{km}\, \mathbb{E}\left[\ell_i r_j \ell_k r_m\right]. \end{aligned} \tag{2.16}$$

Since all components of ℓ and r are independent and

$$\mathbb{E}[\ell_i\ell_k] = \begin{cases} 0, & i \neq k \\ 1, & i = k \end{cases} \quad \text{and} \quad \mathbb{E}[r_j r_m] = \begin{cases} 0, & j \neq m \\ 1, & j = m \end{cases}, \tag{2.17}$$

we have

$$\mathbb{E}\left[\left(\boldsymbol{\ell}^T\mathbf{F}\mathbf{r}\right)^2\right] = \sum_{j,m=1}^{n_s}\sum_{i,k=1}^{n_d} F_{ij}\, F_{km}\, \mathbb{E}\left[\ell_i\ell_k\right]\, \mathbb{E}\left[r_j r_m\right] = \sum_{j=1}^{n_s}\sum_{i=1}^{n_d} F_{ij}^2 = ||\mathbf{F}||^2. \tag{2.18}$$

This requires a single additional adjoint solve rather than an additional n_d solves for the Jacobian.

Typically, we need multiple random samples $\mathbf{r}_j$ and $\boldsymbol{\ell}_j$ to make the variance in our stochastic estimates sufficiently small. Hence, we set

$$\mathbf{R} = \frac{1}{\sqrt{\ell_s}}(\mathbf{r}_1\, \mathbf{r}_2\, \cdots\, \mathbf{r}_{\ell_s}) \quad \in \mathbb{R}^{n_s \times \ell_s}, \tag{2.19}$$

where each column vector $\mathbf{r}_i$ is i.i.d. with zero expectation and covariance equal to the identity and $\ell_s \ll n_s$. Similarly, we set

$$\mathbf{L} = \frac{1}{\sqrt{\ell_d}}(\boldsymbol{\ell}_1\, \boldsymbol{\ell}_2\, \cdots\, \boldsymbol{\ell}_{\ell_d}) \quad \in \mathbb{R}^{n_d \times \ell_d}, \tag{2.20}$$

where each column vector $\boldsymbol{\ell}_i$ is i.i.d. with zero expectation and covariance equal to the identity and $\ell_d \ll n_d$. It is easily verified that these choices give

$$\mathbb{E}[\mathbf{R}\mathbf{R}^T] = \mathbf{I}_{n_s} \quad \text{and} \quad \mathbb{E}[\mathbf{L}\mathbf{L}^T] = \mathbf{I}_{n_d} \tag{2.21}$$

Next, we replace the sources $\mathbf{B}$ by simultaneous random sources $\mathbf{BR}$ and the detectors $\mathbf{C}$ by simultaneous random detectors $\mathbf{CL}$. Assume that $\mathbf{R}$ and $\mathbf{L}$ are independent and we compute unbiased estimates for $\|\mathbf{F}(\mathbf{p})\|^2$.

Theorem 2.1. *Let $\mathbf{R} \in \mathbb{R}^{n_s \times \ell_s}$ and $\mathbf{L} \in \mathbb{R}^{n_d \times \ell_d}$ be as given above. Let $\mathbf{F} \in \mathbb{R}^{n_d \times n_s}$. Then*

$$\mathbb{E}\left[\|\mathbf{L}^T\mathbf{F}\mathbf{R}\|^2\right] = \|\mathbf{F}\|^2. \tag{2.22}$$

Proof.

$$\begin{aligned} \mathbb{E}\left[\|\mathbf{L}^T\mathbf{F}\mathbf{R}\|^2\right] &= \mathbb{E}\left[\operatorname{trace}\left(\mathbf{R}^T\mathbf{F}^T\mathbf{L}\mathbf{L}^T\mathbf{F}\mathbf{R}\right)\right] = \mathbb{E}\left[\operatorname{trace}\left(\mathbf{R}\mathbf{R}^T\mathbf{F}^T\mathbf{L}\mathbf{L}^T\mathbf{F}\right)\right] \\ &= \operatorname{trace}\left(\mathbb{E}\left[\mathbf{R}\mathbf{R}^T\right]\mathbf{F}^T\mathbb{E}\left[\mathbf{L}\mathbf{L}^T\right]\mathbf{F}\right) = \operatorname{trace}\left(\mathbf{F}^T\mathbf{F}\right) = \|\mathbf{F}\|^2. \end{aligned} \tag{2.23}$$

□

Since TREGS has proven very effective for the nonlinear least squares problem in DOT with PaLS, we continue to use the TREGS algorithm in the stochastic minimization problem

$$\min_{\mathbf{p}} \mathbb{E}\left[||\mathbf{L}^T\mathbf{F}(\mathbf{p})\mathbf{R}||^2\right] = \min_{\mathbf{p}} ||\mathbf{F}(\mathbf{p})||^2. \tag{2.24}$$

We derive the least squares problem used in TREGS to compute a regularized Gauss-Newton update for the stochastic problem as follows. For any $\mathbf{p}$,

$$\operatorname{vec}\left(\mathbf{L}^T\mathbf{F}(\mathbf{p})\mathbf{R}\right) = \left(\mathbf{R}^T \otimes \mathbf{L}^T\right)\operatorname{vec}(\mathbf{F}(\mathbf{p})) = \left(\mathbf{R}^T \otimes \mathbf{L}^T\right)\mathbf{f}(\mathbf{p}); \tag{2.25}$$

see [35, lemma 4.3.1]. Using a first order approximation to $\mathbf{f}(\mathbf{p}+\boldsymbol{\delta})$ gives

$$\left(\mathbf{R}^T \otimes \mathbf{L}^T\right)\mathbf{f}(\mathbf{p}+\boldsymbol{\delta}) \approx \left(\mathbf{R}^T \otimes \mathbf{L}^T\right)(\mathbf{f}(\mathbf{p}) + \mathbf{J}\boldsymbol{\delta}),$$

which leads to the (sampled) least squares problem

$$\min_{\boldsymbol{\delta}} \|(\mathbf{R}^T \otimes \mathbf{L}^T)\mathbf{J}\boldsymbol{\delta} + (\mathbf{R}^T \otimes \mathbf{L}^T)\mathbf{f}(\mathbf{p})\|_2^2, \tag{2.26}$$

replacing (2.9). Note that setting up the least squares problem (2.26) does not require any computations beyond $\mathbf{A}(\mathbf{p})^{-1}(\mathbf{BR})$ and $\mathbf{A}(\mathbf{p})^{-T}(\mathbf{C}^T\mathbf{L})$. In addition, (2.26) has the following desirable properties for the sampled Jacobian and residual, which follow directly from (2.21) and well-known properties of the Kronecker product.

$$\mathbb{E}\left[\left((\mathbf{R}^T \otimes \mathbf{L})\mathbf{J}\right)^T(\mathbf{R}^T \otimes \mathbf{L})\mathbf{f}\right] = \mathbf{J}^T\mathbb{E}[\mathbf{R}\mathbf{R}^T \otimes \mathbf{L}\mathbf{L}^T]\mathbf{f} = \mathbf{J}^T\mathbf{f}, \tag{2.27}$$

$$\mathbb{E}\left[\left((\mathbf{R}^T \otimes \mathbf{L}^T)\mathbf{J}\right)^T(\mathbf{R}^T \otimes \mathbf{L}^T)\mathbf{J}\right] = \mathbf{J}^T\mathbb{E}\left[\mathbf{R}\mathbf{R}^T \otimes \mathbf{L}\mathbf{L}^T\right]\mathbf{J} = \mathbf{J}^T\mathbf{J}. \tag{2.28}$$

So, the proposed randomization provides unbiased estimates for the gradient and the Gauss-Newton Hessian.

Two approaches to stochastic optimization are commonly used [54]. One approach, stochastic approximation (SA), uses a new random vector (or small set of random vectors) in each optimization step. The other approach, sample average approximation (SAA), uses a fixed set of random vectors over multiple (or many) optimization steps. In this paper, we focus on the SAA approach [54] to solve the stochastic problem (2.24). The SAA approach approximates (2.24) by the sample average problem. At each iteration, this approach requires solving only $\ell_s + \ell_d$ linear systems for each frequency to estimate the objective function and the Jacobian rather than $n_s + n_d$.

We give two representative solutions for our problem using the SAA approach in Figure 2.3. For DOT, the use of random simultaneous sources and detectors initially leads to good

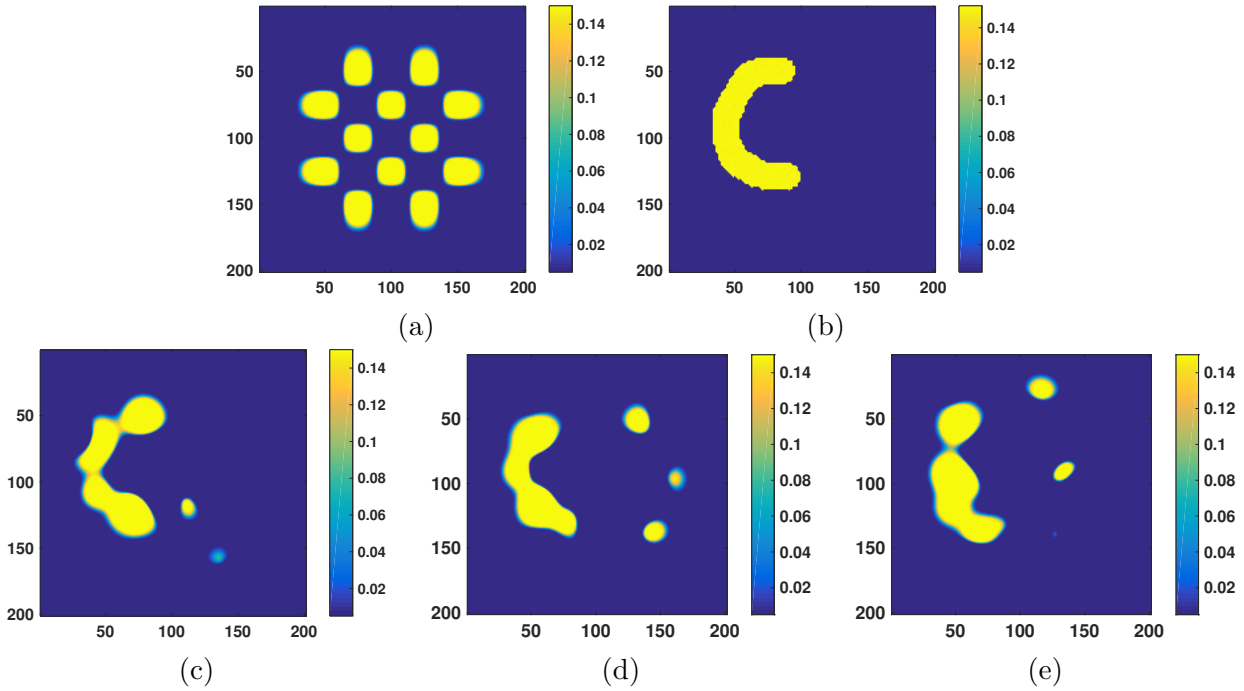


Figure 2.3: Reconstruction of a test anomaly on 201×201 mesh with 32 sources, 32 detectors, using only the zero frequency.
(a) Initial configuration with 25 basis functions arranged in a 5×5 grid where 12 basis functions have positive expansion factors (visible as high absorption regions) and 13 basis functions have negative expansion factors (invisible). (b) True shape of the anomaly. (c) Reconstruction using the full order model. (d) and (e) are two reconstruction results using random simultaneous sources and detectors with $\ell_s = \ell_d = 10$.

progress. However, later in the iteration the convergence slows down, and in many cases, for our problem, it does not lead to sufficiently accurate solutions. With the SAA approach,

optimization (typically) does not reach the noise level. The standard optimization using all sources and all detectors does converge to the noise level. We will demonstrate this in Section 2.5. In the next section, we provide a solution to this problem.

2.4 Improving the Randomized Approach

In the standard SAA approach, when convergence slows down or a minimum is found for the chosen sample (but not for the true problem), we choose a new set of random simultaneous sources and detectors (a new sample) to improve the solution. However, for our problem this approach leads to slow convergence and stagnation. We note that the (worst case) convergence rate for these stochastic methods is typically $O(N^{-1/2})$. Hence, after exploiting the relatively fast initial convergence for our problem, we want to avoid stagnation of convergence in the next phase. One approach is to add additional random simultaneous sources and detectors, that is, increase the sample size, as proposed in [17, 22, 48, 49, 50], with good results. However, this requires progressively more, expensive, solves. Therefore, for efficiency, we choose to keep the number of simultaneous sources and detectors fixed. To improve convergence, we exploit the (often fairly good) approximate solution after reaching a modest intermediate tolerance to obtain simultaneous sources and detectors that are optimized for convergence of the nonlinear least squares problem. We make this precise below.

The nonlinear least squares algorithm TREGS focuses on the dominant singular values of the Jacobian to compute good updates to the parameter vector [24]. The corresponding right singular vectors capture the directions in parameter space of largest sensitivity in the objective function. Hence, we want to update $\mathbf{R}$ and $\mathbf{L}$ so as to capture the largest singular values in $\mathbf{J}$ while respecting the Kronecker product structure in (2.26). This is important for two reasons. First, for the same (fixed) small number of simultaneous sources and detectors, this gives us locally (at the current $\mathbf{p}$) the best approximation to what TREGS would do using all sources and detectors. Second, the directions corresponding to the dominant right singular vectors are best informed by the data. So, when a chosen intermediate tolerance is reached, our method computes the full Jacobian $\mathbf{J}$ once, which requires total of $n_s + n_d$ solves. Then, we compute a small number, q_s respectively q_d, of orthonormal, optimized simultaneous sources ($\widehat{\mathbf{R}}$) and detectors ($\widehat{\mathbf{L}}$). In practice, very small q_s and q_d, 2 or 4, seem to be sufficient. We provide some experimental results regarding the number of optimized directions in Section 2.5. Since $\mathbf{J}$ is typically of rank substantially lower than n_p [1, 24], we expect that computing optimized directions can be done with an approximation to $\mathbf{J}$ that can be computed at substantially lower cost than solving for all sources and detectors (for each frequency).

Ideally, we would maximize

$$\|(\widehat{\mathbf{R}}^T \otimes \widehat{\mathbf{L}}^T)\mathbf{J}\|^2. \tag{2.29}$$

However, the structure of the Kronecker product combined with the constraints that $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{L}}$ be isometric matrices lead to a nonlinear constrained optimization problem. For efficiency, we replace this by an easier problem. We solve two consecutive (separate) optimization problems; one for $\widehat{\mathbf{L}} \in \mathbb{R}^{n_{\mathrm{d}} \times q_{\mathrm{d}}}$ and one for $\widehat{\mathbf{R}} \in \mathbb{R}^{n_{\mathrm{s}} \times q_{\mathrm{s}}}$. To describe this two-step optimization, we need some additional notation. For $\mathbf{y} \in \mathbb{R}^k$, $\mathbf{Y} \in \mathbb{R}^{k \times q}$, and $\mathbf{G} \in \mathbb{R}^{km \times n}$ partitioned as follows, where each $\mathbf{G}_{i,j} \in \mathbb{R}^k$,

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{1,1} & \mathbf{G}_{1,2} & \dots & \mathbf{G}_{1,n} \\ \mathbf{G}_{2,1} & \mathbf{G}_{2,2} & \dots & \mathbf{G}_{1,n} \\ & \vdots & & \\ \mathbf{G}_{m,1} & \mathbf{G}_{m,2} & \dots & \mathbf{G}_{m,n} \end{bmatrix}, \tag{2.30}$$

we define

$$\mathbf{y}^T \star \mathbf{J} = \begin{bmatrix} \mathbf{y}^T\mathbf{G}_{1,1} & \mathbf{y}^T\mathbf{G}_{1,2} & \dots & \mathbf{y}^T\mathbf{G}_{1,n} \\ \mathbf{y}^T\mathbf{G}_{2,1} & \mathbf{y}^T\mathbf{G}_{2,2} & \dots & \mathbf{y}^T\mathbf{G}_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{y}^T\mathbf{G}_{m,1} & \mathbf{y}^T\mathbf{G}_{m,2} & \dots & \mathbf{y}^T\mathbf{G}_{m,n} \end{bmatrix} \in \mathbb{R}^{m \times n} \quad \text{and} \tag{2.31}$$

$$\mathbf{Y}^T \star \mathbf{G} = \begin{bmatrix} \mathbf{y}_1^T \star \mathbf{G} \\ \mathbf{y}_2^T \star \mathbf{G} \\ \vdots \\ \mathbf{y}_q^T \star \mathbf{G} \end{bmatrix} \in \mathbb{R}^{qm \times n}. \tag{2.32}$$

Furthermore, let $S^{k \times \ell} = \{\mathbf{\Theta} \in \mathbb{R}^{k \times \ell} | \ \mathbf{\Theta}^T\mathbf{\Theta} = \mathbf{I}_\ell\}$. We replace maximizing (2.29) by solving consecutively the following problems,

$$\widehat{\mathbf{L}} = \arg \max_{\widetilde{\mathbf{L}} \in S^{n_d \times q_d}} \|\widetilde{\mathbf{L}}^T \star \mathbf{J}\|, \tag{2.33}$$

$$\widehat{\mathbf{R}} = \arg \max_{\widetilde{\mathbf{R}} \in S^{n_s \times q_s}} \|\widetilde{\mathbf{R}}^T \star (\widehat{\mathbf{L}}^T \star \mathbf{J})\|_{\mathrm{F}}, \tag{2.34}$$

where the partitioning of $\mathbf{J}$ is given by (2.7). Note that $\widehat{\mathbf{R}}^T \star (\widehat{\mathbf{L}}^T \star \mathbf{J}) = (\widehat{\mathbf{R}}^T \otimes \widehat{\mathbf{L}}^T)\mathbf{J}$. This process can also be iterated to improve the approximation further.

In addition, since $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{L}}$ are only optimal at the current $\mathbf{p}$, we complement these optimized simultaneous sources and detectors with a new set of random simultaneous sources and detectors constrained to the orthogonal complement of the span of the optimized directions, keeping the total number of columns in $\mathbf{R}$ and $\mathbf{L}$ the same as before. This procedure can be carried out periodically or for a sequence of prescribed tolerances, but in our experiments it never needs to be done more than once.

In the remainder of this section, we first discuss computing the optimized simultaneous detectors ($\widehat{\mathbf{L}}$) and sources ($\widehat{\mathbf{R}}$) and then complementing these with random simultaneous sources constrained to Range$(\widehat{\mathbf{R}})^\perp$ and random simultaneous detectors constrained to Range$(\widehat{\mathbf{L}})^\perp$.

2.4.1 Computing Optimized Simultaneous Sources and Detectors.

To solve the problems (2.33 – 2.34), we need a minor variation of the well known min-max characterization of the singular values.

Lemma 2.2. *Let $\mathbf{X} \in \mathbb{C}^{m\times n}$, let $k \leq \min(m,n)$, and let the singular value decomposition (SVD) of $\mathbf{X}$ be given by $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{Y}^*$ with leading singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k$, $\mathbf{U}_k = [\mathbf{u}_1\, \mathbf{u}_2 \ldots \mathbf{u}_k]$ (the first k columns of $\mathbf{U}$) and $\mathbf{Y}_k = [\mathbf{y}_1\, \mathbf{y}_2 \ldots \mathbf{y}_k]$. Then*

$$\sum_{j=1}^{k} \sigma_j^2 = \max_{\mathbf{Z}\in S^{m\times k}} \|\mathbf{Z}^*\mathbf{X}\|^2, \tag{2.35}$$

$$\mathbf{U}_k = \arg\max_{\mathbf{Z}\in S^{m\times k}} \|\mathbf{Z}^*\mathbf{X}\|^2. \tag{2.36}$$

Furthermore, any $\widetilde{\mathbf{U}}_k \in S^{m\times k}$ such that $\mathrm{Range}(\mathbf{U}_k) = \mathrm{Range}(\widetilde{\mathbf{U}}_k)$ also solves (2.36).

The proof is straightforward and given only for completeness.

Proof. From the unitary invariance of the Frobenius norm and Corollary 3.1.3 in [35] we have for $\mathbf{Z} \in S^{m\times k}$,

$$\|\mathbf{Z}^*\mathbf{X}\|^2 \leq \sum_{j=1}^{k} \sigma_j^2,$$

However, for $\mathbf{Z} = \mathbf{U}_k$ equality is obtained, since $\mathbf{U}_k^*\mathbf{X} = \mathrm{diag}(\sigma_1, \ldots, \sigma_k)\mathbf{Y}_k^*$. □

The following theorem shows how to solve optimization problems like (2.33) and (2.34).

Theorem 2.3. *Let $\mathbf{G} \in \mathbb{R}^{km\times n}$ be partitioned as in (2.30), and let*

$$\widetilde{\mathbf{G}} = \begin{bmatrix} \mathbf{G}_{1,1} & \mathbf{G}_{1,2} & \ldots & \mathbf{G}_{1,n} & \mathbf{G}_{2,1} & \mathbf{G}_{2,2} & \ldots & \mathbf{G}_{m,n} \end{bmatrix} \in \mathbb{R}^{k\times mn}$$

and have the SVD $\widetilde{\mathbf{G}} = \mathbf{U}\mathbf{\Sigma}\mathbf{Y}^T$. Then $\mathbf{U}_q$, the matrix containing the first q columns of $\mathbf{U}$, is a solution to

$$\arg\max_{\widetilde{\mathbf{L}}\in S^{k\times q}} \|\widetilde{\mathbf{L}}^T \star \mathbf{G}\|.$$

Furthermore, any isometric matrix $\widetilde{\mathbf{U}}_q \in \mathbb{R}^{k\times q}$ such that $\mathrm{Range}(\widetilde{\mathbf{U}}_q) = \mathrm{Range}(\mathbf{U}_q)$ is also a solution.

Proof. The proof follows from the property that the Frobenius norm of a matrix depends only on its coefficients. Hence, we have that for $\mathbf{y} \in \mathbb{R}^k$,

$$\|\mathbf{y}^T \star \mathbf{G}\| = \|\mathbf{y}^T\widetilde{\mathbf{G}}\|,$$

and as a result

$$\|\widetilde{\mathbf{L}}^T \star \mathbf{G}\| \quad = \quad \|\widetilde{\mathbf{L}}^T \widetilde{\mathbf{G}}\|.$$

Now, using the results from Lemma 2.2, we get

$$\arg \max_{\widetilde{\mathbf{L}} \in S^{k \times q}} \|\widetilde{\mathbf{L}}^T \star \mathbf{G}\|^2 = \arg \max_{\widetilde{\mathbf{L}} \in S^{k \times q}} \|\widetilde{\mathbf{L}}^T \widetilde{\mathbf{G}}\|^2 \quad = \quad \mathbf{U}_q,$$

and

$$\max_{\widetilde{\mathbf{L}} \in S^{k \times q}} \|\widetilde{\mathbf{L}}^T \star \mathbf{G}\|^2 \quad = \quad \sum_{j=1}^{q} \sigma_j^2.$$

□

Using Theorem 2.3, we solve (2.33) - (2.34) as follows. First, we set

$$\widetilde{\mathbf{J}} \quad = \quad \left[\begin{array}{cccccccc} \mathbf{J}_{1,1} & \mathbf{J}_{1,2} & \dots & \mathbf{J}_{1,n_p} & \mathbf{J}_{2,1} & \mathbf{J}_{2,2} & \dots & \mathbf{J}_{n_s,n_p} \end{array} \right] \in \mathbb{R}^{n_d \times n_s n_p}, \tag{2.37}$$

where each $\mathbf{J}_{i,j} \in \mathbb{R}^{n_d}$, see (2.7), and we compute the SVD of $\widetilde{\mathbf{J}} = \mathbf{U}\mathbf{\Sigma}\mathbf{Y}^T$. We set $\widehat{\mathbf{L}} = \mathbf{U}_{q_d}$ to solve (2.33). Second, we compute

$$\mathbf{K} \quad = \quad \widehat{\mathbf{L}}^T \star \mathbf{J} \in \mathbb{R}^{q_d n_s \times n_p}. \tag{2.38}$$

Note that $\mathbf{K}$ can be partitioned in a similar fashion as $\mathbf{J}$, with vectors $\mathbf{K}_{i,j} \in \mathbb{R}^{n_s}$. Third, we set

$$\widetilde{\mathbf{K}} \quad = \quad \left[\begin{array}{cccc} \mathbf{K}_{1,1} & \mathbf{K}_{1,2} & \dots & \mathbf{K}_{q_d,n_p} \end{array} \right] \in \mathbb{R}^{n_s \times q_d n_p}, \tag{2.39}$$

and compute its SVD, $\widetilde{\mathbf{K}} = \mathbf{\Phi}\mathbf{\Omega}\mathbf{\Psi}^T$. Following Theorem 2.3, the solution for (2.34) is given by $\widehat{\mathbf{R}} = \mathbf{\Phi}_{q_s}$.

While this procedure gives good solutions, in general it does not maximize (2.29). However, we have the following useful result for the case that $\|\mathbf{J}\|$ can be preserved exactly.

Theorem 2.4. *If isometric matrices $\widetilde{\mathbf{R}} \in \mathbb{R}^{n_s \times q_s}$ and $\widetilde{\mathbf{L}} \in \mathbb{R}^{n_d \times q_d}$ exist such that*

$$\|(\widetilde{\mathbf{R}}^T \otimes \widetilde{\mathbf{L}}^T)\mathbf{J}\| \quad = \quad \|\mathbf{J}\|,$$

then solving (2.33) and (2.34) finds isometric matrices $\widehat{\mathbf{R}} \in \mathbb{R}^{n_s \times r}$ and $\widehat{\mathbf{L}} \in \mathbb{R}^{n_s \times s}$ such that

$$\|(\widehat{\mathbf{R}}^T \otimes \widehat{\mathbf{L}}^T)\mathbf{J}\| \quad = \quad \|\mathbf{J}\|,$$

$r \leq q_s$, and $s \leq q_d$.

Proof. First, we note that under the given assumptions $\widetilde{\mathbf{R}} \otimes \widetilde{\mathbf{L}}$ is also isometric and that for any $\mathbf{G} \in \mathbb{R}^{m\times n}$ and isometric matrix $\mathbf{Q} \in \mathbb{R}^{m\times k}$,

$$\|\mathbf{Q}^T\mathbf{G}\| \;=\; \|\mathbf{G}\| \quad\Leftrightarrow\quad \text{Range}(\mathbf{G}) \subseteq \text{Range}(\mathbf{Q}). \tag{2.40}$$

Now, assume that $\widetilde{\mathbf{R}} \in \mathbb{R}^{n_s\times q_s}$ and $\widetilde{\mathbf{L}} \in \mathbb{R}^{n_d\times q_d}$ exist such that $\|(\widetilde{\mathbf{R}}^T \otimes \widetilde{\mathbf{L}}^T)\mathbf{J}\| = \|\mathbf{J}\|$. It follows from (2.40) that $\text{Range}(\mathbf{J}) \subseteq \text{Range}(\widetilde{\mathbf{R}} \otimes \widetilde{\mathbf{L}})$, and therefore

$$\mathbf{J} = (\widetilde{\mathbf{R}} \otimes \widetilde{\mathbf{L}})\mathbf{Z} \quad \text{with} \quad \mathbf{Z} = (\widetilde{\mathbf{R}} \otimes \widetilde{\mathbf{L}})^T\mathbf{J}.$$

This implies that the kth block row of $\mathbf{J}$ satisfies

$$[\mathbf{J}_{k,1}\ \mathbf{J}_{k,2}\ \ldots\ \mathbf{J}_{k,n_p}] = [\widetilde{r}_{k,1}\widetilde{\mathbf{L}} \quad \widetilde{r}_{k,2}\widetilde{\mathbf{L}} \quad \ldots \quad \widetilde{r}_{k,q_s}\widetilde{\mathbf{L}}]\mathbf{Z},$$

which in turn shows that $\text{Range}([\mathbf{J}_{k,1}\ \mathbf{J}_{k,2}\ \ldots\ \mathbf{J}_{k,n_p}]) \subseteq \text{Range}(\widetilde{\mathbf{L}})$. As this holds for every $k = 1, 2, \ldots, n_s$, $\text{Range}(\widetilde{\mathbf{J}}) \subseteq \text{Range}(\widetilde{\mathbf{L}})$ (where $\widetilde{\mathbf{J}} \in \mathbb{R}^{n_d\times n_s n_p}$ is defined in (2.37)), and hence $s = \text{Rank}(\widetilde{\mathbf{J}}) \le q_d$. Let $\widetilde{\mathbf{J}}$ have the SVD $\widetilde{\mathbf{J}} = \mathbf{U}\mathbf{\Sigma}\mathbf{Y}^T$. Since $s = \text{Rank}(\widetilde{\mathbf{J}})$, $\text{Range}(\widetilde{\mathbf{J}}) = \text{Range}(\mathbf{U}_s)$, and we take $\widehat{\mathbf{L}} = \mathbf{U}_s$. Moreover, we have

$$\|\widehat{\mathbf{L}}^T\widetilde{\mathbf{J}}\| = \|\widetilde{\mathbf{J}}\| = \|\mathbf{J}\|.$$

Next, following the proof of Theorem 2.3, we get for $\mathbf{K} = \widehat{\mathbf{L}}^T \star \mathbf{J}$, see (2.38), $\|\mathbf{K}\| = \|\mathbf{J}\|$. For the k$th$ row of $\widehat{\boldsymbol{\ell}}_j^T \star \mathbf{J}$, we have

$$\begin{aligned}\widehat{\boldsymbol{\ell}}_j^T[\mathbf{J}_{k,1} \quad \mathbf{J}_{k,2} \quad \ldots \quad \mathbf{J}_{k,n_p}] \;&=\; \widehat{\boldsymbol{\ell}}_j^T[\widetilde{r}_{k,1}\widetilde{\mathbf{L}} \quad \widetilde{r}_{k,2}\widetilde{\mathbf{L}} \quad \ldots \quad \widetilde{r}_{k,q_s}\widetilde{\mathbf{L}}]\mathbf{Z} \\ &=\; [\widetilde{r}_{k,1}\widehat{\boldsymbol{\ell}}_j^T\widetilde{\mathbf{L}} \quad \widetilde{r}_{k,2}\widehat{\boldsymbol{\ell}}_j^T\widetilde{\mathbf{L}} \quad \ldots \quad \widetilde{r}_{k,q_s}\widehat{\boldsymbol{\ell}}_j^T\widetilde{\mathbf{L}}]\mathbf{Z},\end{aligned}$$

and therefore the jth block row of $\mathbf{K}$ is given by

$$\begin{aligned}\begin{bmatrix}\mathbf{K}_{j,1} & \mathbf{K}_{j,2} & \ldots & \mathbf{K}_{j,n_p}\end{bmatrix} \;&=\; \widehat{\boldsymbol{\ell}}_j^T \star \mathbf{J} \\ &=\; \begin{bmatrix}\widetilde{\mathbf{r}}_1(\widehat{\boldsymbol{\ell}}_j^T\widetilde{\mathbf{L}}) & \widetilde{\mathbf{r}}_2(\widehat{\boldsymbol{\ell}}_j^T\widetilde{\mathbf{L}}) & \ldots & \widetilde{\mathbf{r}}_{q_s}(\widehat{\boldsymbol{\ell}}_j^T\widetilde{\mathbf{L}})\end{bmatrix}\mathbf{Z}.\end{aligned}$$

It follows that $\text{Range}([\mathbf{K}_{j,1}\ \mathbf{K}_{j,2}\ \ldots\ \mathbf{K}_{j,n_p}]) \subseteq \text{Range}(\widetilde{\mathbf{R}})$, and hence $\text{Range}(\widetilde{\mathbf{K}}) \subseteq \text{Range}(\widetilde{\mathbf{R}})$, where $\widetilde{\mathbf{K}}$ is defined in (2.39), and $r = \text{Rank}(\widetilde{\mathbf{K}}) \le q_s$. Following the algorithm above, let the SVD of $\widetilde{\mathbf{K}}$ be given by $\widetilde{\mathbf{K}} = \mathbf{\Phi}\mathbf{\Omega}\mathbf{\Psi}^T$. Then $\text{Range}(\widetilde{\mathbf{K}}) = \text{Range}(\mathbf{\Phi}_r)$, and we take $\widehat{\mathbf{R}} = \mathbf{\Phi}_r$. We have

$$\begin{aligned}\left\|(\widehat{\mathbf{R}} \otimes \widehat{\mathbf{L}})^T\mathbf{J}\right\| \;&=\; \left\|\widehat{\mathbf{R}}^T \star (\widehat{\mathbf{L}}^T \star \mathbf{J})\right\| \;=\; \left\|\widehat{\mathbf{R}}^T \star \mathbf{K}\right\| \\ &=\; \left\|\widehat{\mathbf{R}}^T\widetilde{\mathbf{K}}\right\| \;=\; \left\|\widetilde{\mathbf{K}}\right\| \\ &=\; \|\mathbf{K}\| \;=\; \|\mathbf{J}\|\,.\end{aligned} \tag{2.41}$$

□

Note that following (2.40), the result also shows that

$$\text{Range}(\mathbf{J}) \;\subseteq\; \text{Range}(\widehat{\mathbf{R}} \otimes \widehat{\mathbf{L}}), \tag{2.42}$$

$$\mathbf{J} \;=\; (\widehat{\mathbf{R}} \otimes \widehat{\mathbf{L}})(\widehat{\mathbf{R}}^T \otimes \widehat{\mathbf{L}}^T)\mathbf{J}. \tag{2.43}$$

2.4.2 Computing Complementary Random Simultaneous Sources and Detectors.

We extend the optimized sources and detectors with random simultaneous sources and detectors. Let $\mathbf{R}_f = [\widehat{\mathbf{R}} \;\; \mathbf{R}_c] \in \mathbb{R}^{n_s \times n_s}$ and $\mathbf{L}_f = [\widehat{\mathbf{L}} \;\; \mathbf{L}_c] \in \mathbb{R}^{n_d \times n_d}$ be orthogonal matrices. We have

$$\begin{aligned} \|\mathbf{F}(\mathbf{p})\| &= \|\mathbf{L}_f^T \mathbf{F}(\mathbf{p}) \mathbf{R}_f\| \\ &= \left\| \begin{bmatrix} \widehat{\mathbf{L}}^T \mathbf{F}(\mathbf{p}) \widehat{\mathbf{R}} & \widehat{\mathbf{L}}^T \mathbf{F}(\mathbf{p}) \mathbf{R}_c \\ \mathbf{L}_c^T \mathbf{F}(\mathbf{p}) \widehat{\mathbf{R}} & \mathbf{L}_c^T \mathbf{F}(\mathbf{p}) \mathbf{R}_c \end{bmatrix} \right\| \end{aligned} \tag{2.44}$$

The $(1,1)$-block of this matrix can be computed using the known optimized sources and detectors. We estimate the remaining blocks, proceeding more or less as before. We pick random matrices $\mathbf{Y} = (\ell_s - q_s)^{-1/2}[\mathbf{y}_1 \;\; \mathbf{y}_2 \;\ldots\; \mathbf{y}_{\ell_s - q_s}]$, where each column vector $\mathbf{y}_j \in \mathbb{R}^{\ell_s - q_s}$ is i.i.d. with zero mean and identity covariance, and $\mathbf{Z} = (\ell_d - q_d)^{-1/2}[\mathbf{z}_1 \;\; \mathbf{z}_2 \;\ldots\; \mathbf{z}_{\ell_d - q_d}]$, where each column vector $\mathbf{z}_j \in \mathbb{R}^{n_d - q_d}$ i.i.d. with zero mean and identity covariance. Next, we set the new matrices $\mathbf{R}$ and $\mathbf{L}$ to

$$\mathbf{R} = [\widehat{\mathbf{R}} \;\; (\mathbf{R}_c \mathbf{Y})], \tag{2.45}$$

$$\mathbf{L} = [\widehat{\mathbf{L}} \;\; (\mathbf{L}_c \mathbf{Z})]. \tag{2.46}$$

We have the following results.

Theorem 2.5. *Let $\mathbf{R} \in \mathbb{R}^{n_s \times \ell_s}$ and $\mathbf{L} \in \mathbb{R}^{n_d \times \ell_d}$ be given in (2.45) and (2.46), respectively. Let $\mathbf{F} \in \mathbb{R}^{n_d \times n_s}$. Then,*

$$\mathbb{E}[\mathbf{R}\mathbf{R}^T] = \mathbf{I}_{n_s}, \tag{2.47}$$

$$\mathbb{E}[\mathbf{L}\mathbf{L}^T] = \mathbf{I}_{n_d}, \tag{2.48}$$

$$\mathbb{E}[\|\mathbf{L}^T \mathbf{F}(\mathbf{p}) \mathbf{R}\|^2] = \|\mathbf{F}(\mathbf{p})\|^2. \tag{2.49}$$

Proof. For (2.47), we have

$$\begin{aligned} \mathbb{E}\left[[\widehat{\mathbf{R}} \;\; (\mathbf{R}_c \mathbf{Y})] [\widehat{\mathbf{R}} \;\; (\mathbf{R}_c \mathbf{Y})]^T \right] &= \widehat{\mathbf{R}} \widehat{\mathbf{R}}^T + \mathbb{E}[\mathbf{R}_c \mathbf{Y} \mathbf{Y}^T \mathbf{R}_c^T] \\ &= \widehat{\mathbf{R}} \widehat{\mathbf{R}}^T + \mathbf{R}_c \mathbb{E}[\mathbf{Y}\mathbf{Y}^T] \mathbf{R}_c^T = \mathbf{I}_{n_d}. \end{aligned}$$

An analogous derivation holds for (2.48). The proof for the last equation follows the proof for theorem 2.1, using the results above. □

As a result of Theorem 2.5, the new $\mathbf{R}$ and $\mathbf{L}$ again give for the expectation of the sampled gradient and the expectation of the sampled Gauss-Newton Hessian the true gradient and Gauss-Newton Hessian,

$$\begin{aligned} \mathbb{E}[((\mathbf{R}^T \otimes \mathbf{L})\mathbf{J})^T (\mathbf{R}^T \otimes \mathbf{L})\mathbf{f}] &= \mathbf{J}^T \mathbf{f}, \\ \mathbb{E}[((\mathbf{R}^T \otimes \mathbf{L}^T)\mathbf{J})^T ((\mathbf{R}^T \otimes \mathbf{L}^T)\mathbf{J})] &= \mathbf{J}^T \mathbf{J}. \end{aligned}$$

2.4.3 Implementation

In this section, we outline the efficient computation of the residual and Jacobian. We estimate the norm of the residual using

$$(\mathbf{R}^T \otimes \mathbf{L}^T)\mathbf{f}(\mathbf{p}) = \begin{bmatrix} \mathbf{L}^T\mathbf{C}\mathbf{A}^{-1}(\mathbf{p})\mathbf{B}\mathbf{r}_1 - \mathbf{D}\mathbf{r}_1 \\ \vdots \\ \mathbf{L}^T\mathbf{C}\mathbf{A}^{-1}(\mathbf{p})\mathbf{B}\mathbf{r}_{\ell_s} - \mathbf{D}\mathbf{r}_{\ell_s} \end{bmatrix} = \begin{bmatrix} \mathbf{L}^T\mathbf{C}\mathbf{z}_1 - \mathbf{D}\mathbf{r}_1 \\ \vdots \\ \mathbf{L}^T\mathbf{C}\mathbf{z}_{\ell_s} - \mathbf{D}\mathbf{r}_{\ell_s} \end{bmatrix}, \tag{2.50}$$

where we solve $\mathbf{A}(\mathbf{p})\mathbf{z}_i = \mathbf{B}\mathbf{r}_i$ for $\mathbf{z}_i$, $i = 1 \cdots \ell_s$. This reduces the number of large solves from n_s to ℓ_s per frequency. To compute the Jacobian, we solve the systems, $\mathbf{A}^T(\mathbf{p})\mathbf{y}_j = \mathbf{C}^T\boldsymbol{\ell}_j$ for $\mathbf{y}_j$, $j = 1 \cdots \ell_d$. This reduces the additional number of large solves from n_d to ℓ_d per frequency. We can use iterative solvers or direct sparse solvers depending on the size of the system [37]. To obtain the k-th column of $(\mathbf{R}^T \otimes \mathbf{L}^T)\mathbf{J}$, we compute

$$\left[\mathbf{y}_1^T \frac{\partial \mathbf{A}(\mathbf{p})}{\partial \mathbf{p}_k}\mathbf{z}_1 \;\; \cdots \;\; \mathbf{y}_{\ell_d}^T \frac{\partial \mathbf{A}(\mathbf{p})}{\partial \mathbf{p}_k}\mathbf{z}_1 \;\; \mathbf{y}_1^T \frac{\partial \mathbf{A}(\mathbf{p})}{\partial \mathbf{p}_k}\mathbf{z}_2 \;\; \cdots \;\; \mathbf{y}_{\ell_d}^T \frac{\partial \mathbf{A}(\mathbf{p})}{\partial \mathbf{p}_k}\mathbf{z}_{\ell_s} \right]^T, \tag{2.51}$$

where $\dfrac{\partial \mathbf{A}(\mathbf{p})}{\partial \mathbf{p}_k}$ is a diagonal matrix if we only invert for absorption. If we also invert for diffusion, this matrix has 5 (in 2D) or 7 (in 3D) diagonals. Moreover, after a few iterations, the changes in $\mathbf{A}(\mathbf{p})$ are highly localized, and $\dfrac{\partial \mathbf{A}(\mathbf{p})}{\partial \mathbf{p}_k}$ contains mostly zero coefficients; see [23]. In that case, we first find the few nonzero components of $\dfrac{\partial \mathbf{A}(\mathbf{p})}{\partial \mathbf{p}_k}$ for each k, and the corresponding nonzeros in $\mathbf{z}_i$ and $\mathbf{y}_j$. Then, compute $\mathbf{y}_i^T \dfrac{\partial \mathbf{A}(\mathbf{p})}{\partial \mathbf{p}_k}\mathbf{z}_j$ exploiting the fact that there are few nonzero components in $\dfrac{\partial \mathbf{A}(\mathbf{p})}{\partial \mathbf{p}_k}\mathbf{z}_j$.

2.5 Numerical Experiments

In this section, we provide three numerical experiments, two 2D examples and one 3D example to demonstrate the effectiveness of combining random simultaneous sources and detectors with optimized simultaneous sources and detectors. We show that using optimized simultaneous sources and detectors not only produces reconstruction results that are close to those obtained using all sources and all detectors, but it also reduces the computational cost.

The experimental set up we use is that described in [23], in which model reduction was proposed as an alternative approach to reduce the cost of the inversion process in DOT. For each test case, we construct anomalies in the pixel basis, and we add a small normally distributed random heterogeneity to both the background and to the anomaly to make the

medium inhomogeneous. This ensures a modest mismatch between the exact image and the representation we use to reconstruct the image, so that we avoid the so-called *'inverse crime'*. We use this absorption image to compute the true measured data. We also add $\delta = 0.1\%$ white noise to the measured data, which is the same noise level as in [23]. PaLS [1] and TREGS [24] are used to reconstruct the absorption images.

2D Experiments. Both experiments are carried out on a 201×201 grid, which yields $40,401$ degrees of freedom in the forward model (2.2). The model has 32 sources, 32 detectors, and we use only the zero frequency. Our model has 25 CSRBFs, which leads to 100 parameters (four per 2D basis function) for the nonlinear optimization. We use the same starting guess for each example (see Figure 2.4), with 25 basis functions arranged in a 5×5 grid, where 12 basis functions have a positive expansion coefficient (visible as high absorption regions) and 13 basis functions have a negative expansion coefficient (invisible).

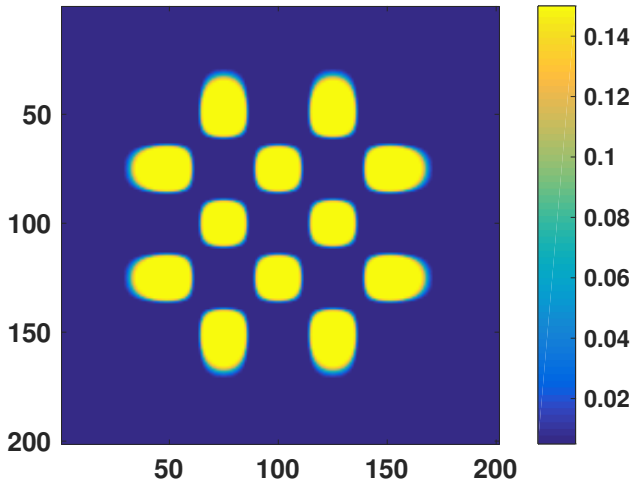


Figure 2.4: Initial configuration with 25 basis functions arranged in a 5×5 grid where 12 basis functions have a positive expansion factors (visible) and 13 basis functions have negative expansion factors (invisible).

We use 10 random simultaneous sources and detectors in each example. As discussed in Section 2.3, we update random simultaneous sources and detectors by optimized simultaneous sources and detectors after a chosen intermediate tolerance has been reached. We find that, in general, the noise level δ is a good choice as the intermediate tolerance ($||\mathbf{f}(\mathbf{p})||_2^2 = \delta$). Since the PaLS representation regularizes the problem, we consider the problem converged when the squared residual norm falls below δ^2, which is called the discrepancy principle (the factor $\frac{1}{2}$ in (2.1) is dropped for convenience). We run both experiments for 50 trials. In each trial, the random simultaneous sources and detectors are chosen independently to get representative reconstruction results.

Example 1. The true absorption image for Example 1 is given in Figure 2.5a. We also include the reconstruction results using all sources and all detectors for comparison (see Figure 2.5b). As can be seen in Figure 2.5c at the intermediate tolerance, SAA gives a

good localization of the anomaly; however, there is no further improvement using SAA (see Figure 2.5d). Figure 2.5e-g show that using optimized simultaneous sources and detectors leads to solutions of the same quality as obtained using all sources and detectors. We report the total number of PDE solves required for each approach in Table 2.1 for a representative result from 50 trials.

While initially the SAA estimate is unbiased, bias arises as we optimize for a specific small set of random simultaneous sources and detectors [54]. The algorithm stops prematurely as the bias, a systematic underestimation of the error/misfit [54, Section 5.1.2], makes it appear as if convergence has been reached. This can make a big difference, since it is usually the case that substantial improvement in the shape of the anomaly occurs towards the end of the iterative process. Figure 2.7a-b demonstrates how poor the reconstructions using only the SAA approach can be at the convergence tolerance when underestimation is severe. To make a fair comparison in terms of the number of large systems solved, we solve the full system on the side to check convergence of the SAA approach. Table 2.3 shows that in terms of the true function evaluation, the SAA approach does not reach the convergence tolerance. Once we use a few optimized sources and detectors, this is no longer an issue (see Table 2.3).

The main purpose of the SAA approach and our modification is to reduce the large number of discretized PDE solves that is necessary for the inversion. In Table 2.1, we give a comparison of the total number of PDE solves for Example 1. Our approach reduces both the computational cost and the number of large-scale linear systems that needs to be solved. Additionally, combining simultaneous random and optimized simultaneous sources and detectors drastically improves the reconstruction results of the SAA approach.

Example 2. The true absorption image for Example 2 is given in Figure 2.6a. We also include the reconstruction result using all sources and all detectors for comparison in Figure 2.6b. In Figure 2.6e-g, we show the reconstruction results for combining random and optimized simultaneous sources and detectors. The results show similar behavior as was observed in Example 1. Note that SAA gives a good localization of the anomaly in Figure 2.6c. No further improvement appears using the SAA approach after the maximum iterations, see Figure 2.6d.

To emphasize the bias issue, we include a comparison of the true objective function $\|\mathbf{f}\|_2^2$ and its SAA estimate relative to the stopping criterion (δ^2) for selected iterations in Table 2.3. Similar to Example 1, using only few optimized sources and detectors overcomes the bias issue.

In Table 2.2, we give the total number of PDE solves required for each approach for Example 2. Our approach improves the rate of convergence of the optimization and reduces the number of large-scale linear systems solves. Moreover, combining random and optimized simultaneous sources and detectors improves the quality of the inverse solution. Figure 2.7c-d demonstrates how poor the reconstructions using only the SAA approach can be at the convergence tolerance when underestimation is severe.

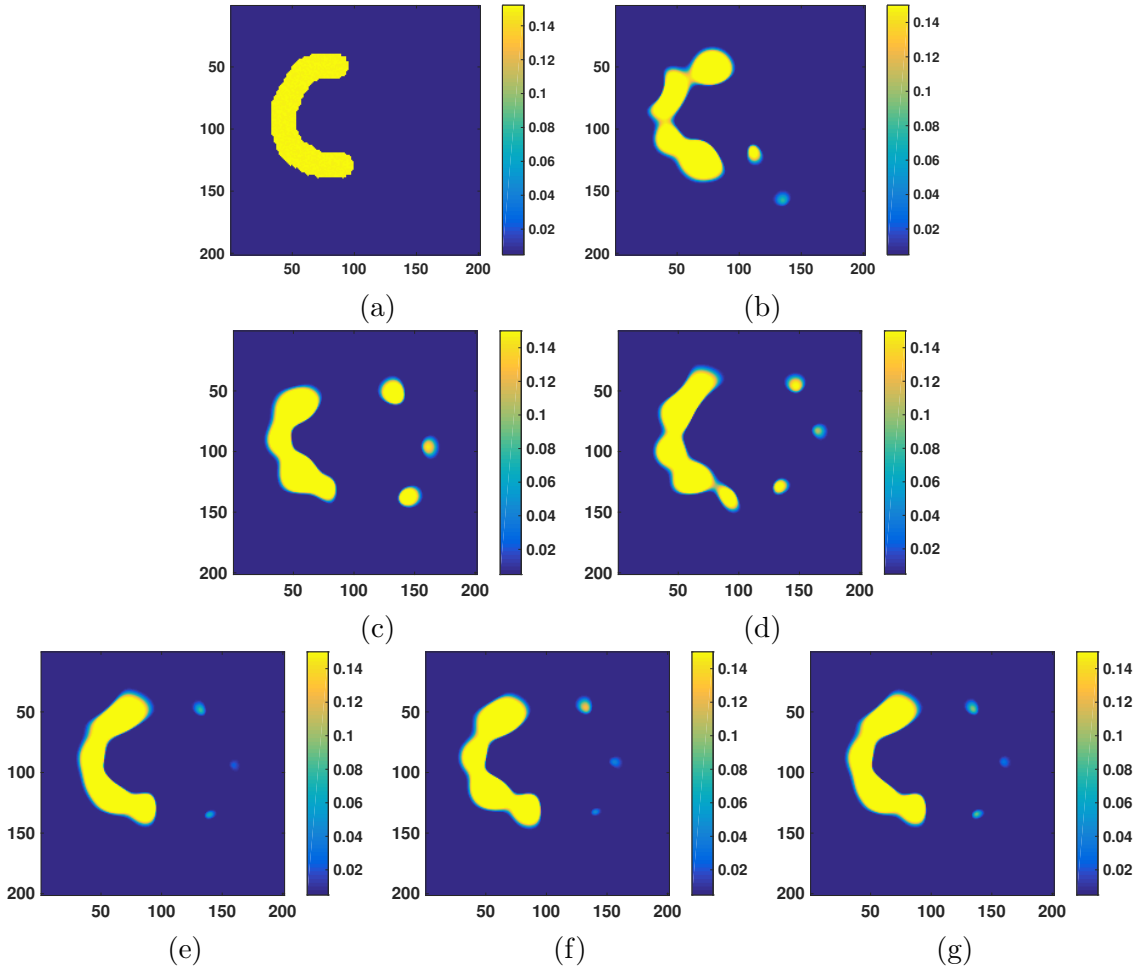


Figure 2.5: Results for Example 1. Reconstruction of a test anomaly on 201×201 mesh with 32 sources and detectors, 25 basis functions, and using only the zero frequency. The SAA approach uses 10 random simultaneous sources and detectors.
(a) True shape of the anomaly. (b) Reconstruction using all sources and all detectors. (c) Reconstruction using the SAA approach at a chosen intermediate tolerance. (d) Reconstruction using the SAA approach after the maximum number of iterations. (e) Reconstruction with SAA and 1 optimized simultaneous source and detector. (f) Reconstruction with SAA and 2 optimized simultaneous sources and detectors. (g) Reconstruction with SAA and 3 optimized simultaneous sources and detectors.

	Iteration Number	Function Evaluations	Jacobian Evaluations	Total PDE Solves	Tol
SAA* (intermediate tol)	10	11	6	170	δ
1 Opt simult src/det	18	19	10	524	δ^2
2 Opt simult srcs/dets	18	19	10	524	δ^2
3 Opt simult srcs/dets	16	17	8	484	δ^2
All srcs/All dets	71	72	47	3808	δ^2
SAA**	32	33	19	520	δ^2
SAA ***	(92)	(93)	(67)	1700	δ^2

Table 2.1: Example 1 Results. The total number of iterations, function evaluations, Jacobian evaluations and PDE solves required on average for 50 trials to reach the stopping criterion, δ^2.
*The first row gives the cost to reach the intermediate tolerance for the SAA approach, δ. ** Since the SAA estimate becomes biased and underestimates the objective function, the algorithm stops prematurely. ***The SAA approach measuring the convergence with the true objective function. Parentheses indicate that the SAA approach does not reach the tolerance.

	Iteration Number	Function Evaluations	Jacobian Evaluations	Total PDE Solves	Tol
SAA* (intermediate tol)	10	11	5	160	δ
1 Optimized src/det	13	14	6	424	δ^2
2 Optimized srcs/dets	13	14	7	434	δ^2
3 Optimized srcs/dets	14	15	7	444	δ^2
All srcs/All dets	25	26	14	1280	δ^2
SAA**	28	29	16	450	δ^2
SAA ***	(90)	(91)	(67)	1580	δ^2

Table 2.2: Example 2 Results. The total number of iterations, function evaluations, Jacobian evaluations and PDE solves required on average for 50 trials to reach the stopping criterion, δ^2.
*The first row gives the costs to reach the intermediate tolerance for the SAA approach, δ. ** Since the SAA estimate becomes biased and underestimates the objective function, the algorithm stops prematurely. ***The SAA approach measuring the convergence with the true objective function. Parentheses indicate that the SAA approach does not reach the tolerance.

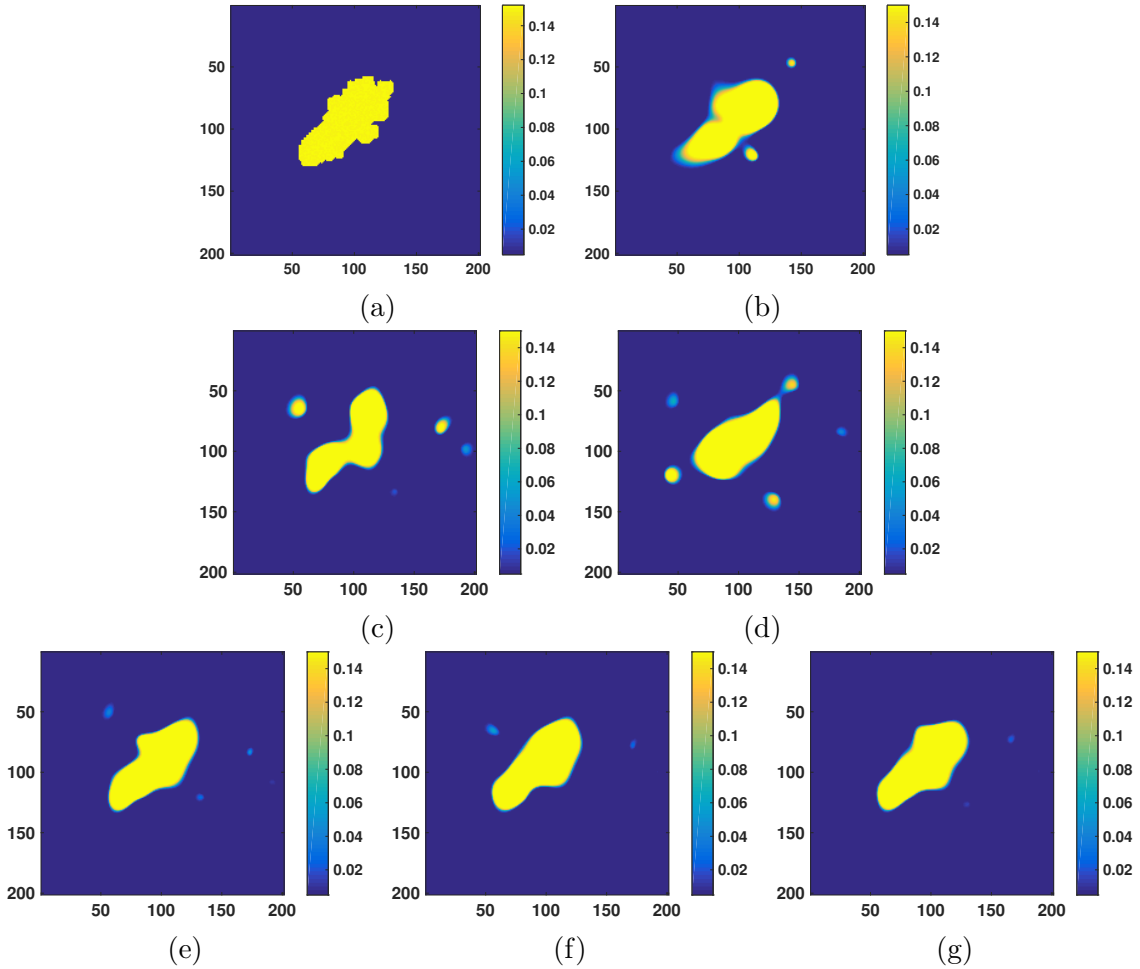
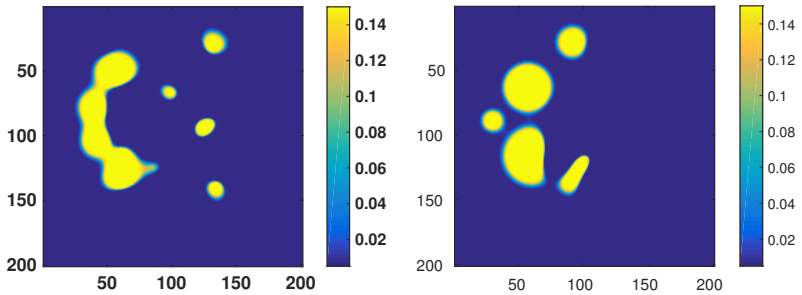


Figure 2.6: Results for Example 2. Reconstruction of a test anomaly on 201×201 mesh with 32 sources and detectors, 25 basis functions, and using only the zero frequency. The SAA approach uses 10 random simultaneous sources and detectors.
(a) True shape of the anomaly. (b) Reconstruction using all sources and all detectors. (c) Reconstruction using SAA approach at intermediate tolerance. (d) Reconstruction using the SAA approach after the maximum iterations. (e) Reconstruction with SAA and 1 simultaneous optimized source and detector. (f) Reconstruction with SAA and 2 optimized simultaneous sources and detectors. (g) Reconstruction with SAA and 3 optimized simultaneous sources and detectors



(a) Example 1 (b) Example 1

(c) Example 2 (d) Example 2

Figure 2.7: Example of poor SAA reconstructions for each test case. Reconstruction of a test anomaly on 201×201 mesh with 32 sources and detectors, 25 basis functions, and using only the zero frequency.

	SAA Approach			Rand & Optimized Simult Src/Det		
	Iter	True $\|\mathbf{f}\|_2^2$ (δ^2)	Estimated $\|\mathbf{f}\|_2^2$ (δ^2)	Iter	True $\|\mathbf{f}\|_2^2$ (δ^2)	Estimated $\|\mathbf{f}\|_2^2$ (δ^2)
Example 1	1	118940	38820	1-5	(SAA)*	(SAA)*
	6	1192.5	391.15	6	1194.9	1197.3
	11	56.550	22.575	13	118.73	118.56
	15	3.748	0.8650	16	19.894	19.929
	(99)	–	–	22	0.8403	0.8389
Example 2	1	90904	21257	1-10	(SAA)*	(SAA)*
	8	2344.5	1629.3	11	847.4	922.22
	12	176.1	97.845	16	160.05	156.7
	17	1.787	0.8599	20	12.53	12.92
	(99)	–	–	28	0.1426	0.1180

Table 2.3: Subset of results for Example 1 and Example 2. The comparison of the true objective function $\|\mathbf{f}\|_2^2$ and its SAA estimate relative to the stopping criterion (δ^2) for selected iterations. For the SAA approach, the estimated residual is obtained with 10 random simultaneous sources and detectors. Parentheses indicate that the SAA approach does not reach the tolerance. The estimated residual for combining random and optimized simultaneous sources and detectors that we report here are those obtained when using 3 optimized simultaneous sources and detectors for Example 1; 2 optimized simultaneous sources and detectors for Example 2. *(SAA) indicates that we initially use the SAA approach at the intermediate tolerance.

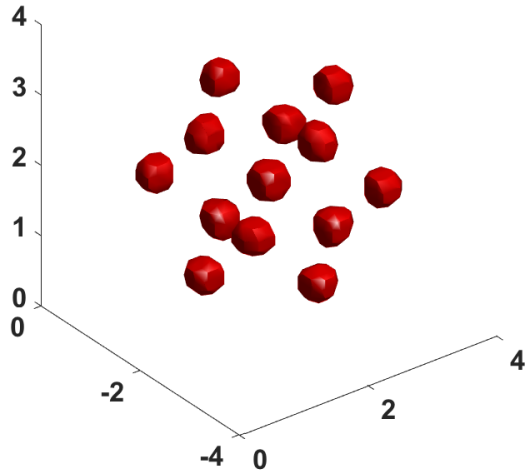


Figure 2.8: Initial configuration with 27 basis functions arranged in a $3 \times 3 \times 3$ grid where 13 basis functions have a positive expansion factors (visible) and 14 basis functions have negative expansion factors (invisible).

3D Experiment. The mesh is $32 \times 32 \times 32$ ($n = 32768$), which gives $32,768$ degrees of freedom in the forward model (2.2). The model has 225 sources at the top and 225 detectors on the bottom, and we use only the zero frequency. In the PaLS approach, we use 27 CSRBFs, which leads to 135 parameters (five per 3D basis function) for the nonlinear optimization. The absorption image using the initial set of parameters is given in Figure 2.8 where 13 basis functions have a positive expansion coefficient (visible as high absorption regions) and 14 basis functions have a negative expansion coefficient (invisible). In our approach, we use 15 random simultaneous sources and detectors.

Example 3. The true absorption image for Example 3 is given in Figure 2.9a. We also include the reconstruction result using all sources and all detectors for comparison in Figure 2.9b. The results show similar behavior as was observed in 2D experiments. Note that SAA gives a good localization of the anomaly in Figure 2.9c. However, no further improvement appears using the SAA approach after the maximum iterations, see Figure 2.9d. In Figure 2.9e-g, we show the reconstruction results for combining random and optimized simultaneous sources and detectors.

The straightforward inversion using all sources and detectors requires $9,225$ large linear solves for the reconstruction. Table 2.4 shows that our approach reduces the number of large linear solves by *about a factor 10* compared to using all sources and detectors, while approximating the original shape well. Clearly, there is a large improvement to be gained

	Iteration Number	Function Evaluations	Jacobian Evaluations	Total PDE Solves	Tol
SAA* (intermediate tol)	5	6	4	150	δ
2 Opt simult srcs/dets	12	13	6	885	δ^2
4 Opt simult srcs/dets	9	10	3	795	δ^2
All srcs/All dets	25	26	15	9225	δ^2
SAA **	(99)	(100)	(67)	2505	δ^2

Table 2.4: Example 3 Results. The total number of iterations, function evaluations, Jacobian evaluations and PDE solves required to reach the stopping criterion, δ^2.
*The first row gives the costs to reach the intermediate tolerance for the SAA approach, δ.
**The SAA approach measuring the convergence with the true objective function. Parentheses indicate that the SAA approach does not reach the tolerance.

by using only a small number of optimized simultaneous sources and detectors. For larger problems with many sources and detectors and using multiple frequencies, we expect much larger gains.

Overall, our approach improves the rate of convergence of the optimization and reduces the number of large-scale linear systems solves. Moreover, combining random and optimized simultaneous sources and detectors improves the quality of the inverse solution.

2.6 Conclusions

We use the SAA approach to estimate the objective function and the Jacobian using only a few random simultaneous sources and detectors in DOT problems. While this approach is reasonably effective for the application in [32], it does not work quite that well for DOT. Since convergence to the noise level slows down for later iterations, and the SAA approach regularly does not converge to the noise level, we propose using optimized simultaneous sources and detectors. With the addition of optimized directions, we observe faster convergence, good quality reconstructions, and robustness. This technique could be quite useful in other applications as well. Although the approach has proved successful experimentally, we aim to understand the underlying theory better. In the future, we plan to analyze, more fundamentally, what are the most effective simultaneous sources and detectors for fast convergence of the inverse problem: randomized, optimized (and in what sense), and their combination.

We intend to update the TREGS algorithm and study how small we can make the number of simultaneous sources and detectors (random and optimized) and still obtain good solutions and fast convergence. Moreover, finding more appropriate stopping criteria for the randomized approach may also improve our results.

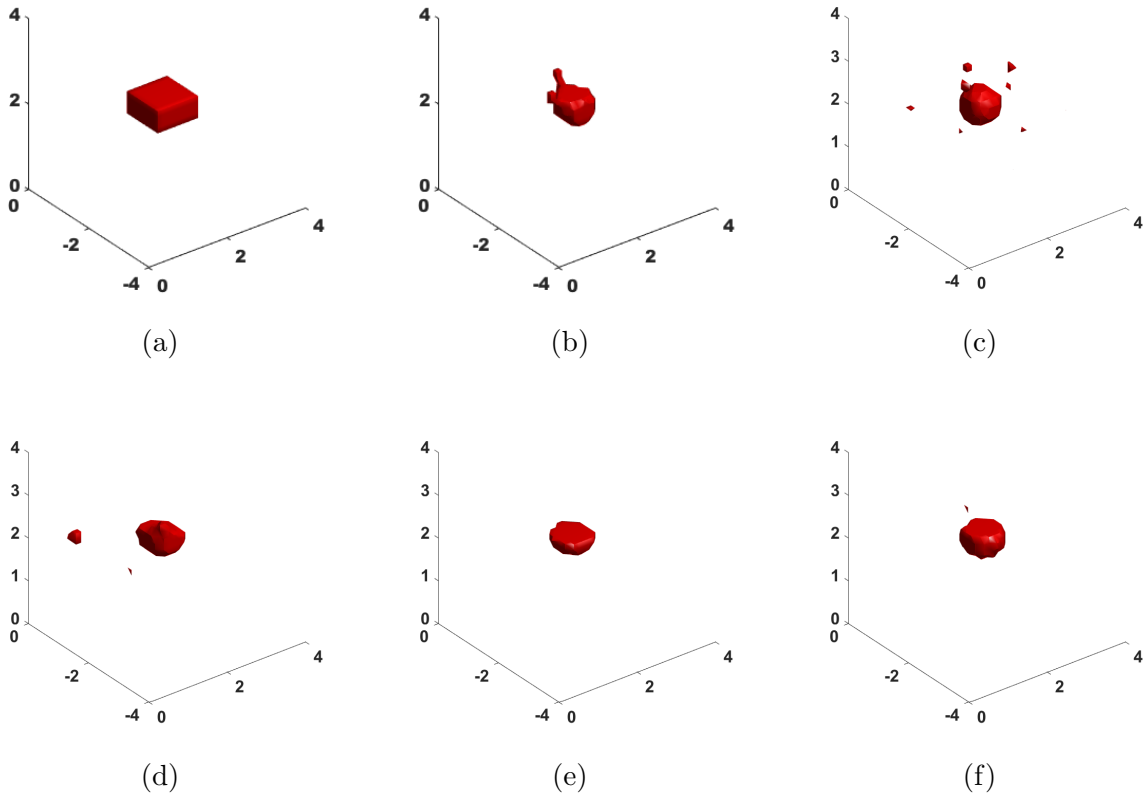


Figure 2.9: Results for Example 3. Reconstruction of a test anomaly on $32 \times 32 \times 32$ mesh with 225 sources and detectors, 27 basis functions, and using only the zero frequency. The SAA approach uses 15 random simultaneous sources and detectors.
(a) True shape of the anomaly. (b) Reconstruction using all sources and all detectors. (c) Reconstruction using SAA approach at intermediate tolerance. (d) Reconstruction using the SAA approach after the maximum iterations. (e) Reconstruction with SAA and 2 optimized simultaneous sources and detectors. (f) Reconstruction with SAA and 4 optimized simultaneous sources and detectors.

Chapter 3

An Alternative Way to Compute Optimized Sources and Detectors

In this chapter, we introduce an alternative approach to compute optimized simultaneous sources and detectors. The goal of our innovation is to combat the observed stagnation in the residual norm decrease of our new stochastic optimization approach, see Section 2.3.1. Specifically, using random simultaneous sources and detectors provides moderately accurate parameter solution estimates at a drastically reduced number of linear system solves. However, for our problem of interest, diffuse optical tomography, the approach often does not lead to sufficiently accurate solutions. Thus, in our alternative approach, we first solve with a fixed set of simultaneous random sources and detectors up to an intermediate tolerance on the objective function. After reaching this intermediate tolerance, we *replace* a few of those in the original set by simultaneous sources and detectors that are optimized to maximize the Frobenius norm of the sampled Jacobian; we refer to these as *optimized simultaneous sources and detectors*. We make this precise below.

The nonlinear least squares algorithm TREGS focuses on the dominant singular values of the Jacobian to compute good updates to the parameter vector [24]. The corresponding right singular vectors capture the directions in parameter space of largest sensitivity in the objective function. Hence, we want to update $\mathbf{R}$ and $\mathbf{L}$ so as to capture the largest singular values in $\mathbf{J}$ while respecting the Kronecker product structure in (2.26). This is important for two reasons. (1) For the same (fixed) small number of simultaneous sources and detectors, this gives us locally (at the current $\mathbf{p}$) the best approximation to what TREGS would do using all sources and detectors. (2) The directions corresponding to the dominant right singular vectors are best informed by the data. So, when a chosen intermediate tolerance is reached, our method computes the full Jacobian $\mathbf{J}$ once and *replaces* the least effective directions in the Range($\mathbf{R}$) and Range($\mathbf{L}$) by directions that maximize

$$\|(\widehat{\mathbf{R}}^T \otimes \widehat{\mathbf{L}}^T)\mathbf{J}\|_F^2, \tag{3.1}$$

for the updated $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{L}}$ (but with the same number of columns). This procedure can be carried out periodically or for a sequence of prescribed tolerances, but in our experiments it never needs to be done more than once.

For simplicity, the following discussion assumes that $\mathbf{R} \in \mathbb{R}^{n_s \times \ell_s}$ and $\mathbf{L} \in \mathbb{R}^{n_d \times \ell_d}$ have orthonormal columns. In our implementation, the length of the columns is $n_s^{1/2}$ for $\mathbf{R}$ and $n_d^{1/2}$ for $\mathbf{L}$, but as the columns in each matrix have equal norm, this has no effect on the analysis (and the issue could be remedied trivially by scaling all columns first). In addition, the columns of $\mathbf{R}$, and $\mathbf{L}$ are only orthogonal in expectation; $\mathbb{E}(\mathbf{r}_i^T \mathbf{r}_j) = 0$. However, for $\ell_s \ll n_s$ and $\ell_d \ll n_d$ the columns are close to orthogonal, and it appears that in practice the assumption (simplification) does not impact the effectiveness of our approach.

We want to update $\mathbf{R}$ and $\mathbf{L}$ to obtain $\widehat{\mathbf{R}} \in \mathbb{R}^{n_s \times \ell_s}$ and $\widehat{\mathbf{L}} \in \mathbb{R}^{n_d \times \ell_d}$ such that (3.1) is (approximately) maximized. There are several ways to do this. Since $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{L}}$ would only be optimal at the current $\mathbf{p}$, we choose to replace only a few components in the spaces spanned by the columns of $\mathbf{R}$ and $\mathbf{L}$. After this update in random simultaneous sources and detectors, convergence of the optimization algorithm to a solution of the same quality as obtained using all sources and detectors is rapid.

The details of updating $\mathbf{R}$ and $\mathbf{L}$ are given in Section 3.1 and 3.2. We also give an outline of our implementation strategies in Section 3.3. In practice, removing/adding only 1 or 2 sources and detectors seem to be sufficient. We provide some experimental results regarding the number of updated directions in Section 3.4.

3.1 Removing Random Simultaneous Sources and Detectors.

We first consider truncating $\mathbf{L}^{n_d \times \ell_d}$ to $\widetilde{\mathbf{L}}^{n_d \times (\ell_d - s)}$ such that $\text{Range}(\widetilde{\mathbf{L}}) \subset \text{Range}(\mathbf{L})$ and $\|(\mathbf{R}^T \otimes \widetilde{\mathbf{L}}^T)\mathbf{J}\|_F^2$ is maximum. This is equivalent to the following optimization problem. Let $S = \{\mathbf{\Theta} \in \mathbb{R}^{\ell_d \times (\ell_d - s)} |\ \mathbf{\Theta}^T \mathbf{\Theta} = \mathbf{I}\}$ and $\widetilde{\mathbf{L}} = \mathbf{L}\mathbf{\Gamma}$. We want to find $\mathbf{\Gamma}$ such that

$$\mathbf{\Gamma} \quad = \quad \arg\max_{\widetilde{\mathbf{\Gamma}} \in S} \|(\mathbf{R}^T \otimes (\mathbf{L}\widetilde{\mathbf{\Gamma}})^T)\mathbf{J}\|_F^2. \tag{3.2}$$

To solve (3.2) we need a variation of the min-max characterization of singular values.

Lemma 3.1. *Let $\mathbf{X} \in \mathbb{C}^{m \times n}$, let $k \leq \min(m, n)$, and let $T = \{\mathbf{\Theta} \in \mathbb{C}^{m \times k} | \mathbf{\Theta}^* \mathbf{\Theta} = \mathbf{I}\}$. Furthermore, let the SVD of $\mathbf{X}$ be given by $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{Y}^*$ with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k$, $\mathbf{U}_k = [\mathbf{u}_1\, \mathbf{u}_2\, \ldots\, \mathbf{u}_k]$ (the first k columns of $\mathbf{U}$) and $\mathbf{Y}_k = [\mathbf{y}_1\, \mathbf{y}_2\, \ldots\, \mathbf{y}_k]$. Then*

$$\sum_{j=1}^{k} \sigma_j^2 \quad = \quad \max_{\mathbf{Z} \in T} \|\mathbf{Z}^* \mathbf{X}\|_F^2, \tag{3.3}$$

$$\mathbf{U}_k \quad = \quad \arg\max_{\mathbf{Z} \in T} \|\mathbf{Z}^* \mathbf{X}\|_F^2. \tag{3.4}$$

Furthermore, any $\widetilde{\mathbf{U}}_k \in T$ *with orthonormal columns such that* $\text{Range}(\mathbf{U}_k) = \text{Range}(\widetilde{\mathbf{U}}_k)$ *also solves (3.4).*

Proof. See Lemma 2.2 in Chapter 2. □

The only complication in (3.2) now is the Kronecker structure of $(\mathbf{R}^T \otimes (\mathbf{L}\widetilde{\mathbf{\Gamma}})^T)$, which we can remove by combining $\mathbf{R}$ and $\mathbf{J}$. For ease of exposition, we first consider a single simultaneous source $\mathbf{r}$ and ℓ_d simultaneous detectors. With $\mathbf{J}_{jk} \in \mathbb{R}^{n_d}$ given by (2.7), and w_i the ith component of $\mathbf{r}$, we have

$$\begin{aligned}\left[\mathbf{r}^T \otimes \widetilde{\mathbf{L}}^T\right]\mathbf{J} &= \begin{bmatrix} r_1\widetilde{\boldsymbol{\ell}}_1^T & r_2\widetilde{\boldsymbol{\ell}}_1^T & \dots & r_{n_s}\widetilde{\boldsymbol{\ell}}_1^T \\ \vdots & \vdots & \vdots & \vdots \\ r_1\widetilde{\boldsymbol{\ell}}_{\ell_d-s}^T & r_2\widetilde{\boldsymbol{\ell}}_{\ell_d-s}^T & \dots & r_{n_s}\widetilde{\boldsymbol{\ell}}_{\ell_d-s}^T \end{bmatrix} \begin{bmatrix} \mathbf{J}_{1,1} & \mathbf{J}_{1,2} & \dots & \mathbf{J}_{1,n_p} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{J}_{n_s,1} & \mathbf{J}_{n_s,2} & \dots & \mathbf{J}_{n_s,n_p} \end{bmatrix} \\ &= \begin{bmatrix} \widetilde{\boldsymbol{\ell}}_1^T \\ \vdots \\ \widetilde{\boldsymbol{\ell}}_{\ell_d-s}^T \end{bmatrix} \begin{bmatrix} \mathbf{r}\star\mathbf{J}_1 & \mathbf{r}\star\mathbf{J}_2 & \dots & \mathbf{r}\star\mathbf{J}_{n_p} \end{bmatrix}, \end{aligned} \tag{3.5}$$

where, for the vector $\mathbf{r}$ and column $\mathbf{J}_k$, we define

$$\mathbf{r}\star\mathbf{J}_k = r_1\mathbf{J}_{1,k} + r_2\mathbf{J}_{2,k} + \cdots + r_{n_s}\mathbf{J}_{n_s,k} \quad \text{for } k = 1, 2, \dots, n_p. \tag{3.6}$$

Furthermore, we define

$$\widehat{\mathbf{J}} = \mathbf{r}\star\mathbf{J} = \begin{bmatrix} \mathbf{r}\star\mathbf{J}_1 & \mathbf{r}\star\mathbf{J}_2 & \dots & \mathbf{r}\star\mathbf{J}_{n_p} \end{bmatrix}. \tag{3.7}$$

Using (3.7) and $\widetilde{\mathbf{L}} = \mathbf{L}\mathbf{\Gamma}$, we can simplify (3.5),

$$\left[\mathbf{r}^T \otimes \widetilde{\mathbf{L}}^T\right]\mathbf{J} = \widetilde{\mathbf{L}}^T\widehat{\mathbf{J}} = \mathbf{\Gamma}^T\mathbf{L}^T\widehat{\mathbf{J}},$$

and as a result (3.2) reduces to

$$\mathbf{\Gamma} = \arg\max_{\widetilde{\mathbf{\Gamma}}\in S} \|\widetilde{\mathbf{\Gamma}}^T\mathbf{L}^T\widehat{\mathbf{J}}\|_F^2. \tag{3.8}$$

The solution of (3.8) follows directly from Lemma 3.1. Hence we have the following theorem.

Theorem 3.2. *Let* $\mathbf{J}$ *and* $\mathbf{J}_{j,k}$ *be defined by (2.6) and (2.7) respectively, let* $\widehat{\mathbf{J}}$ *be defined as in (3.7) for a single simultaneous source* $\mathbf{r}$*, and let* $\mathbf{L} \in \mathbb{R}^{n_d\times\ell_d}$ *have orthonormal columns. Furthermore, let*

$$\mathbf{L}^T\widehat{\mathbf{J}} = \mathbf{\Phi}\mathbf{\Omega}\mathbf{\Psi}^T. \tag{3.9}$$

Then the solution to (3.2) for a single simultaneous source is given by

$$\mathbf{\Gamma} = [\boldsymbol{\varphi}_1\, \boldsymbol{\varphi}_2 \dots \boldsymbol{\varphi}_{\ell_d-s}]. \tag{3.10}$$

Proof. The proof follows from (3.5), (3.7), and Lemma 3.1. □

Next we consider the general case of (3.2), that is, for multiple simultaneous sources. Using (3.5), we have

$$\left[\mathbf{R}^T \otimes \widetilde{\mathbf{L}}^T\right] \mathbf{J} \;=\; \begin{bmatrix} \mathbf{r}_1^T \otimes \widetilde{\mathbf{L}}^T \\ \mathbf{r}_2^T \otimes \widetilde{\mathbf{L}}^T \\ \vdots \\ \mathbf{r}_{\ell_s}^T \otimes \widetilde{\mathbf{L}}^T \end{bmatrix} \mathbf{J} \;=\; \begin{bmatrix} \widetilde{\mathbf{L}}^T \widehat{\mathbf{J}}_1 \\ \widetilde{\mathbf{L}}^T \widehat{\mathbf{J}}_2 \\ \vdots \\ \widetilde{\mathbf{L}}^T \widehat{\mathbf{J}}_{\ell_s} \end{bmatrix}, \tag{3.11}$$

where the matrices $\widehat{\mathbf{J}}_i$ are defined as

$$\widehat{\mathbf{J}}_i = \mathbf{r}_i \star \mathbf{J} \quad \text{for } i = 1, \ldots, \ell_s. \qquad \text{(cf. (3.7))} \tag{3.12}$$

Theorem 3.3. *Let* $\mathbf{J}$ *be defined by (2.6), let* $\widehat{\mathbf{J}}_k$ *be defined as in (3.12), and let* $\mathbf{R} \in \mathbb{R}^{n_s \times \ell_s}$ *and* $\mathbf{L} \in \mathbb{R}^{n_d \times \ell_d}$ *have orthonormal columns. Furthermore, define the SVD*

$$\mathbf{L}^T \left[\widehat{\mathbf{J}}_1 \; \widehat{\mathbf{J}}_2 \; \ldots \; \widehat{\mathbf{J}}_{\ell_s}\right] \;=\; \mathbf{\Phi \Omega \Psi}^T. \tag{3.13}$$

Then the solution to (3.2) is given by

$$\mathbf{\Gamma} \;=\; [\boldsymbol{\varphi}_1 \; \boldsymbol{\varphi}_2 \; \ldots \; \boldsymbol{\varphi}_{\ell_d - s}]. \tag{3.14}$$

Proof. The proof mostly follows the proof of Theorem 3.2.

$$\begin{aligned} \|(\mathbf{R}^T \otimes \widetilde{\mathbf{L}}^T)\mathbf{J}\|_F^2 \;&=\; \left\| \begin{bmatrix} \widetilde{\mathbf{L}}^T \widehat{\mathbf{J}}_1 \\ \widetilde{\mathbf{L}}^T \widehat{\mathbf{J}}_2 \\ \vdots \\ \widetilde{\mathbf{L}}^T \widehat{\mathbf{J}}_{\ell_s} \end{bmatrix} \right\|_F^2 = \left\| \widetilde{\mathbf{L}}^T \left[\widehat{\mathbf{J}}_1 \; \widehat{\mathbf{J}}_2 \; \ldots \; \widehat{\mathbf{J}}_{\ell_s}\right] \right\|_F^2 \\ &=\; \left\| \mathbf{\Gamma}^T \mathbf{L}^T \left[\widehat{\mathbf{J}}_1 \; \widehat{\mathbf{J}}_2 \; \ldots \; \widehat{\mathbf{J}}_{\ell_s}\right] \right\|_F^2, \end{aligned} \tag{3.15}$$

which puts the problem in the form of Lemma 3.1. Hence the solution is given by (3.14). □

Next, consider truncating $\mathbf{R}^{n_s \times \ell_s}$ to $\widetilde{\mathbf{R}}^{n_s \times (\ell_s - s)}$ such that $\text{Range}(\widetilde{\mathbf{R}}) \subset \text{Range}(\mathbf{R})$ and $\|(\widetilde{\mathbf{R}}^T \otimes \mathbf{L}^T)\mathbf{J}\|_F^2$ is maximized. This is equivalent to the following optimization problem. Let $S = \{\mathbf{\Theta} \in \mathbb{R}^{\ell_s \times (\ell_s - s)} \,|\, \mathbf{\Theta}^T \mathbf{\Theta} = \mathbf{I}\}$ and $\widetilde{\mathbf{R}} = \mathbf{R}\mathbf{\Gamma}$. We want to find $\mathbf{\Gamma}$ such that

$$\mathbf{\Gamma} \;=\; \arg\max_{\widetilde{\mathbf{\Gamma}} \in S} \|((\mathbf{R}\widetilde{\mathbf{\Gamma}})^T \otimes \mathbf{L})\mathbf{J}\|_F^2. \tag{3.16}$$

We combine $\mathbf{L}$ and $\mathbf{J}$ to remove the Kronecker structure. For ease of exposition, we first consider ℓ_s simultaneous sources and a single simultaneous detector $\boldsymbol{\ell}$.

$$\left[\widetilde{\mathbf{R}}^T \otimes \boldsymbol{\ell}^T\right]\mathbf{J} \;=\; \begin{bmatrix} (\widetilde{\mathbf{r}}_1^T \otimes \boldsymbol{\ell}^T)\mathbf{J} \\ (\widetilde{\mathbf{r}}_2^T \otimes \boldsymbol{\ell}^T)\mathbf{J} \\ \vdots \\ (\widetilde{\mathbf{r}}_{\ell_s-s}^T \otimes \boldsymbol{\ell}^T)\mathbf{J} \end{bmatrix} = \widetilde{\mathbf{R}}^T \begin{bmatrix} \boldsymbol{\ell}^T\mathbf{J}_{1,1} & \boldsymbol{\ell}^T\mathbf{J}_{1,2} & \dots & \boldsymbol{\ell}^T\mathbf{J}_{1,n_p} \\ \boldsymbol{\ell}^T\mathbf{J}_{2,1} & \boldsymbol{\ell}^T\mathbf{J}_{2,2} & \dots & \boldsymbol{\ell}^T\mathbf{J}_{2,n_p} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\ell}^T\mathbf{J}_{n_s,1} & \boldsymbol{\ell}^T\mathbf{J}_{n_s,2} & \dots & \boldsymbol{\ell}^T\mathbf{J}_{n_s,n_p} \end{bmatrix} = \widetilde{\mathbf{R}}^T\widetilde{\mathbf{J}}, \tag{3.17}$$

where we define, for convenience and later use, the notation

$$\widetilde{\mathbf{J}} = \boldsymbol{\ell} \circledast \mathbf{J} = \begin{bmatrix} \boldsymbol{\ell}^T\mathbf{J}_{1,1} & \boldsymbol{\ell}^T\mathbf{J}_{1,2} & \dots & \boldsymbol{\ell}^T\mathbf{J}_{1,n_p} \\ \boldsymbol{\ell}^T\mathbf{J}_{2,1} & \boldsymbol{\ell}^T\mathbf{J}_{2,2} & \dots & \boldsymbol{\ell}^T\mathbf{J}_{2,n_p} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\ell}^T\mathbf{J}_{n_s,1} & \boldsymbol{\ell}^T\mathbf{J}_{n_s,2} & \dots & \boldsymbol{\ell}^T\mathbf{J}_{n_s,n_p} \end{bmatrix}. \tag{3.18}$$

We now define for multiple detectors

$$\widetilde{\mathbf{J}}_k = \boldsymbol{\ell}_k \circledast \mathbf{J} \quad \text{for } k = 1, \dots, \ell_d. \tag{3.19}$$

Hence, we obtain

$$\left[\widetilde{\mathbf{R}}^T \otimes \mathbf{L}^T\right]\mathbf{J} \;=\; \begin{bmatrix} \widetilde{\mathbf{R}}^T\widetilde{\mathbf{J}}_1 \\ \widetilde{\mathbf{R}}^T\widetilde{\mathbf{J}}_2 \\ \vdots \\ \widetilde{\mathbf{R}}^T\widetilde{\mathbf{J}}_{\ell_d} \end{bmatrix}.$$

Theorem 3.4. *Let $\mathbf{J}$ be defined by (2.6), let $\widetilde{\mathbf{J}}_k$ be defined as in (3.18) and (3.19), and let $\mathbf{R} \in \mathbb{R}^{n_s \times \ell_s}$ and $\mathbf{L} \in \mathbb{R}^{n_d \times \ell_d}$ have orthonormal columns. Furthermore, define the SVD*

$$\mathbf{R}^T \begin{bmatrix} \widetilde{\mathbf{J}}_1 & \widetilde{\mathbf{J}}_2 & \dots & \widetilde{\mathbf{J}}_{\ell_d} \end{bmatrix} \;=\; \boldsymbol{\Phi}\boldsymbol{\Omega}\boldsymbol{\Psi}^T. \tag{3.20}$$

Then the solution to (3.16) is given by

$$\boldsymbol{\Gamma} \;=\; [\boldsymbol{\varphi}_1 \; \boldsymbol{\varphi}_2 \; \dots \; \boldsymbol{\varphi}_{\ell_s-s}]. \tag{3.21}$$

Proof. We use the definition of the Frobenius norm to put the problem in the form of Lemma 3.1.

$$\|(\widetilde{\mathbf{R}}^T \otimes \mathbf{L}^T)\mathbf{J}\|_F^2 \;=\; \left\| \begin{bmatrix} \widetilde{\mathbf{R}}^T\widetilde{\mathbf{J}}_1 \\ \widetilde{\mathbf{R}}^T\widetilde{\mathbf{J}}_2 \\ \vdots \\ \widetilde{\mathbf{R}}^T\widetilde{\mathbf{J}}_{\ell_d} \end{bmatrix} \right\|_F^2 = \left\| \widetilde{\mathbf{R}}^T \begin{bmatrix} \widetilde{\mathbf{J}}_1 & \widetilde{\mathbf{J}}_2 & \dots & \widetilde{\mathbf{J}}_{\ell_d} \end{bmatrix} \right\|_F^2 = \left\| \boldsymbol{\Gamma}_2^T\mathbf{R}^T \begin{bmatrix} \widetilde{\mathbf{J}}_1 & \widetilde{\mathbf{J}}_2 & \dots & \widetilde{\mathbf{J}}_{\ell_d} \end{bmatrix} \right\|_F^2.$$

Therefore, the solution is given by (3.21). □

3.2 Adding Optimized Simultaneous Sources and Detectors.

Next, we consider extending $\mathbf{L}^{n_d\times\ell}$ to $\widehat{\mathbf{L}}^{n_d\times(\ell+s)}$ in directions that orthogonal to the current space $\mathbf{L}$ such that $\|(\mathbf{R}^T\otimes\widehat{\mathbf{L}}^T)\mathbf{J}\|_F^2$ is maximized. This is equivalent to the following optimization problem. Let $S=\{\boldsymbol{\Theta}\in\mathbb{R}^{n_d\times s}|\,\boldsymbol{\Theta}^T\boldsymbol{\Theta}=\mathbf{I}\}$. $\mathbf{Q}_s=[\mathbf{q}_1\ \mathbf{q}_2\cdots\mathbf{q}_s]$ such that $\mathbf{Q}_s=\mathbf{L}_c\boldsymbol{\Gamma}$, and $[\mathbf{L}\ \mathbf{L}_c]$ is orthogonal. We want to find $\boldsymbol{\Gamma}\in S$ such that

$$\boldsymbol{\Gamma} \;=\; \arg\max_{\widetilde{\boldsymbol{\Gamma}}\in S}\|(\mathbf{R}^T\otimes[\mathbf{L}\quad\mathbf{L}_c\widetilde{\boldsymbol{\Gamma}}]^T)\mathbf{J}\|_F^2. \tag{3.22}$$

Combining $\mathbf{R}$ and $\mathbf{J}$ using $\widehat{\mathbf{J}}_i=\mathbf{r}_i\star\mathbf{J}$, see (3.12), we get

$$\begin{aligned}\left\|\begin{bmatrix}\mathbf{r}_1^T\otimes\mathbf{L}^T\\ \mathbf{r}_1^T\otimes\mathbf{q}_1^T\\ \vdots\\ \mathbf{r}_1^T\otimes\mathbf{q}_s^T\\ \mathbf{r}_2^T\otimes\mathbf{L}^T\\ \vdots\\ \mathbf{r}_{\ell_s}^T\otimes\mathbf{L}^T\\ \mathbf{r}_{\ell_s}^T\otimes\mathbf{q}_1^T\\ \vdots\\ \mathbf{r}_{\ell_s}^T\otimes\mathbf{q}_s^T\end{bmatrix}\mathbf{J}\right\|_F^2 &= \left\|\begin{bmatrix}\mathbf{L}^T\\ \mathbf{q}_1^T\\ \mathbf{q}_2^T\\ \vdots\\ \mathbf{q}_s^T\end{bmatrix}\begin{bmatrix}\mathbf{r}_1\star\mathbf{J} & \mathbf{r}_2\star\mathbf{J} & \dots & \mathbf{r}_{\ell_s}\star\mathbf{J}\end{bmatrix}\right\|_F^2\\ &= \left\|\mathbf{L}^T\begin{bmatrix}\widehat{\mathbf{J}}_1\ \widehat{\mathbf{J}}_2\ \dots\ \widehat{\mathbf{J}}_{\ell_s}\end{bmatrix}\right\|_F^2+\left\|\mathbf{Q}_s^T\begin{bmatrix}\widehat{\mathbf{J}}_1\ \widehat{\mathbf{J}}_2\ \dots\ \widehat{\mathbf{J}}_{\ell_s}\end{bmatrix}\right\|_F^2.\end{aligned} \tag{3.23}$$

Since the first term is fixed, we maximize the second term in (3.23). Taking $\mathbf{Q}_s=\mathbf{L}_c\boldsymbol{\Gamma}$, we get

$$\boldsymbol{\Gamma} \;=\; \arg\max_{\widetilde{\boldsymbol{\Gamma}}\in S}\left\|\widetilde{\boldsymbol{\Gamma}}^T\mathbf{L}_c^T\begin{bmatrix}\widehat{\mathbf{J}}_1\ \widehat{\mathbf{J}}_2\ \dots\ \widehat{\mathbf{J}}_{\ell_s}\end{bmatrix}\right\|_F^2. \tag{3.24}$$

The solution of (3.24) now follows from Lemma 3.1.

Theorem 3.5. *Let* $\mathbf{J}$ *be defined by (2.6) and (2.7), let* $\widehat{\mathbf{J}}_k$ *be defined as in (3.12) and let* $\mathbf{R}\in\mathbb{R}^{n_s\times\ell_s}$ *and* $\mathbf{L}\in\mathbb{R}^{n_d\times\ell_d}$ *have orthonormal columns and* $[\mathbf{L}\ \mathbf{L}_c]$ *be orthogonal. Furthermore, let*

$$\mathbf{L}_c^T\begin{bmatrix}\widehat{\mathbf{J}}_1\ \widehat{\mathbf{J}}_2\ \dots\ \widehat{\mathbf{J}}_{\ell_s}\end{bmatrix}=\boldsymbol{\Phi}\boldsymbol{\Omega}\boldsymbol{\Psi}^T. \tag{3.25}$$

Then the solution to (3.22) is given by

$$\boldsymbol{\Gamma} \;=\; [\boldsymbol{\varphi}_1\,\boldsymbol{\varphi}_2\,\dots\,\boldsymbol{\varphi}_s]. \tag{3.26}$$

Proof. The proof follows directly from (3.23) and (3.24), and Lemma 3.1. □

The final derivation we discuss is to add optimized simultaneous sources. Consider extending $\mathbf{R}^{n_s \times \ell}$ to $\widehat{\mathbf{R}}^{n_s \times (\ell+s)}$ so that $(\widehat{\mathbf{R}}^T \otimes \mathbf{L}^T)\mathbf{J}\|_F^2$ is maximized. This is equivalent to the following optimization problem. Let $S = \{\boldsymbol{\Theta} \in \mathbb{R}^{n_s \times s} | \ \boldsymbol{\Theta}^T\boldsymbol{\Theta} = \mathbf{I}\}$ and $\widetilde{\mathbf{Q}}_s = [\widetilde{\mathbf{q}}_1 \ \ \widetilde{\mathbf{q}}_2 \cdots \widetilde{\mathbf{q}}_s]$ such that $\widetilde{\mathbf{Q}}_s = \mathbf{R}_c\boldsymbol{\Gamma}$ where $[\mathbf{R} \ \ \mathbf{R}_c]$ is orthogonal. We want to find $\boldsymbol{\Gamma} \in S$ such that

$$\boldsymbol{\Gamma} \ = \ \arg\max_{\widetilde{\boldsymbol{\Gamma}} \in S} \|([\mathbf{R} \quad \mathbf{R}_c\widetilde{\boldsymbol{\Gamma}}]^T \otimes \mathbf{L}^T)\mathbf{J}\|_F^2. \tag{3.27}$$

Combining $\mathbf{L}$ and $\mathbf{J}$ using $\widetilde{\mathbf{J}}_k = \boldsymbol{\ell}_k \circledast \mathbf{J}$, see (3.19), we get

$$\begin{aligned}
\left\| \begin{bmatrix} \mathbf{R}^T \otimes \mathbf{L}^T \\ \widetilde{\mathbf{q}}_1^T \otimes \ \mathbf{L}^T \\ \vdots \\ \widetilde{\mathbf{q}}_{\ell_s}^T \otimes \ \mathbf{L}^T \end{bmatrix} \mathbf{J} \right\|_F^2 &= \left\| \begin{bmatrix} \mathbf{R}^T \\ \widetilde{\mathbf{q}}_1^T \\ \vdots \\ \widetilde{\mathbf{q}}_s^T \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{J}}_1 \ \widetilde{\mathbf{J}}_2 \ \ldots \ \widetilde{\mathbf{J}}_{\ell_d} \end{bmatrix} \right\|_F^2 \\
&= \left\| \mathbf{R}^T \begin{bmatrix} \widetilde{\mathbf{J}}_1 \ \widetilde{\mathbf{J}}_2 \ \ldots \ \widetilde{\mathbf{J}}_{\ell_d} \end{bmatrix} \right\|_F^2 + \left\| \widetilde{\mathbf{Q}}_s^T \begin{bmatrix} \widetilde{\mathbf{J}}_1 \ \widetilde{\mathbf{J}}_2 \ \ldots \ \widetilde{\mathbf{J}}_{\ell_d} \end{bmatrix} \right\|_F^2.
\end{aligned} \tag{3.28}$$

Since the first term is fixed, we maximize the second term in (3.28). Taking $\widetilde{\mathbf{Q}}_s = \mathbf{R}_c\boldsymbol{\Gamma}$, we get

$$\boldsymbol{\Gamma} \ = \ \arg\max_{\widetilde{\boldsymbol{\Gamma}} \in S} \left\| \widetilde{\boldsymbol{\Gamma}}^T \mathbf{R}_c^T \begin{bmatrix} \widetilde{\mathbf{J}}_1 \ \widetilde{\mathbf{J}}_2 \ \ldots \ \widetilde{\mathbf{J}}_{\ell_d} \end{bmatrix} \right\|_F^2. \tag{3.29}$$

The solution of (3.29) follows from Lemma 3.1.

Theorem 3.6. *Let* $\mathbf{J}$ *be defined by (2.6) and (2.7), let* $\widetilde{\mathbf{J}}_k$ *be defined as in (3.18) and (3.19), and let* $\mathbf{R} \in \mathbb{R}^{n_s \times \ell_s}$ *and* $\mathbf{L} \in \mathbb{R}^{n_d \times \ell_d}$ *have orthonormal columns and* $[\mathbf{R} \ \ \mathbf{R}_c]$ *be orthogonal. Furthermore, let*

$$\mathbf{R}_c^T \begin{bmatrix} \widetilde{\mathbf{J}}_1 \ \widetilde{\mathbf{J}}_2 \ \ldots \ \widetilde{\mathbf{J}}_{\ell_d} \end{bmatrix} = \boldsymbol{\Phi}\boldsymbol{\Omega}\boldsymbol{\Psi}^T. \tag{3.30}$$

Then the solution to (3.27) is given by

$$\boldsymbol{\Gamma} \ = \ [\boldsymbol{\varphi}_1 \, \boldsymbol{\varphi}_2 \, \ldots \, \boldsymbol{\varphi}_s]. \tag{3.31}$$

Proof. The proof follows directly from (3.28), (3.29) and Lemma 3.1. □

3.3 Implementation

Efficient computation of the residual and Jacobian is discussed in Section 2.4.3. In this section, we outline the efficient computation of some critical parts of the algorithm.

Given the general discussion in Section 3.1 and Section 3.2, replacing random simultaneous sources and detectors by optimized simultaneous sources and detectors can be done in several ways. Experiments suggest that using a two-phase alternating approach gives good reconstruction results. In phase one, we alternately remove one source, then one detector, and so on. In phase two, we alternately add one source, then one detector, and so on. This choice requires the SVD of four small matrices. The sizes of these matrices are given in Table 3.1. The cost of these computations is negligible compared with the solution of many large PDEs. Therefore, the main cost of our algorithm is the number of PDE solves for the function and Jacobian evaluations. Section 3.4 includes a detailed discussion of the computational cost.

Method	Sizes for SVD
Removing detectors	$\ell_d \times (\ell_s n_p)$
Removing sources	$\ell_s \times (\ell_d n_p)$
Adding detectors	$(n_d - (\ell_s - s)) \times (\ell_s - s)n_p$
Adding sources	$(n_s - (\ell_d - s)) \times (\ell_d - s)n_p$

Table 3.1: Sizes of matrices for SVD computations to replace random simultaneous sources and detectors by optimized simultaneous sources and detectors.

3.4 Numerical Experiments

We discuss four examples, including three 2D and one 3D, to demonstrate the effectiveness of combining random and optimized simultaneous sources and detectors. We show that using optimized simultaneous sources and detectors not only produces reconstruction results that are close to those obtained using all sources and all detectors, but it also reduces the computational cost. The experimental set up we use is that described in Chapter 2.

All experiments are carried out on a 201×201 grid, which yields $40,401$ degrees of freedom in the forward model (2.2). The model has 32 sources, 32 detectors, and we use only the zero frequency. For each test case, we construct anomalies in the pixel basis, and we add a small normally distributed random heterogeneity to both the background and to the anomaly to make the medium inhomogeneous. This ensures a modest mismatch between the exact image and the representation we use to reconstruct the image, so that we avoid the so-called *'inverse crime'*. We use this absorption image to compute the true measured data. We also add $\delta = 0.1\%$ white noise to the measured data which is the same noise level as in [23]. PaLS [1] and TREGS [24] are used to reconstruct the absorption images. Our model has 25 CSRBFs,

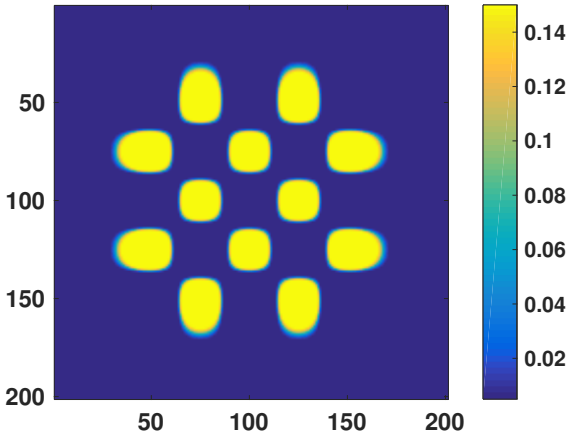


Figure 3.1: Initial configuration with 25 basis functions arranged in a 5×5 grid where 12 basis functions have a positive expansion factors (visible) and 13 basis functions have negative expansion factors (invisible).

which leads to 100 parameters (four per 2D basis function) for the nonlinear optimization. We use the same starting guess for each example (see Figure 3.1), with 25 basis functions arranged in a 5×5 grid, where 12 basis functions have a positive expansion coefficient (visible as high absorption regions) and 13 basis functions have a negative expansion coefficient (invisible).

We use 10 random simultaneous sources and detectors in each example. As discussed in Section 3.1 and Section 3.2, we replace random simultaneous sources and detectors by optimized simultaneous sources and detectors after a chosen intermediate tolerance has been reached. We find that, in general, the noise level δ is a good choice as the intermediate tolerance ($||\mathbf{f}(\mathbf{p})||_2^2 = \delta$). Since the PaLS representation regularizes the problem, we consider the problem converged when the squared residual norm falls below δ^2 (the factor $\frac{1}{2}$ in (2.1) is dropped for convenience). We run each experiment for 50 trials. In each trial, the random simultaneous sources and detectors are chosen independently to get representative reconstruction results.

Example 1. The true absorption image for Example 1 is given in Figure 3.2a. We report the reconstruction results using random simultaneous sources and detectors and combined with optimized simultaneous sources and detectors in Tables 3.2-3.4. We also include the reconstruction results using all sources and all detectors for comparison (see Figure 3.2b). As can be seen in Figure 3.2c at the intermediate tolerance, SAA gives a good localization of the anomaly; however, there is no further improvement using SAA (see Figure 3.2d).

While initially the SAA estimate is unbiased, bias arises as we optimize for a specific small set of random simultaneous sources and detectors [54]. The algorithm stops prematurely as the bias, a systematic underestimation of the error/misfit [54, Section 5.1.2], makes it appear as if convergence has been reached. This can make a big difference since it is usually the

case that substantial improvement in the shape of the anomaly occurs towards the end of the iterative process. To make a fair comparison in terms of the number of large systems solved, we solve the full system on the side to check convergence of the SAA approach. Table 3.2 shows that in terms of the true function evaluation, the SAA approach does not reach to the convergence tolerance. Once we replace a few sources and detectors, this is no longer an issue (see Table 3.2). In Table 3.3, we give the average ratio of the number of times when the estimated residual underestimates the true residual to the total number of iterations using 50 trials for each example. Clearly, there is a large improvement to be gained by replacing a very small number of random simultaneous sources and detectors.

The main purpose of the SAA approach and our modification is to reduce the large number of discretized PDE solves that is necessary for the inversion. In Table 3.4, we give a comparison of the total number of PDE solves for Example 1. Our approach reduces both the computational cost and the number of large-scale linear systems that needs to be solved.

Additionally, combining random and optimized simultaneous sources and detectors drastically improves the reconstruction results of the SAA approach. Figure 3.2 also shows that replacing a few random sources and detectors by optimized sources and detectors leads to solutions of the same quality as obtained using all sources and detectors.

Example 2 and Example 3. In Example 2, we discuss reconstruction results for an image with multiple anomalies. In Figure 3.3, we show the reconstruction results for combining simultaneous random and optimized simultaneous sources and detectors. The true absorption image is given in Figure 3.3a. We also include the reconstruction result using all sources and all detectors for comparison. The results show similar behavior as was observed in Example 1 for a fixed set of random simultaneous sources and detectors. In Table 3.5, we give the total number of PDE solves required for each approach for Example 2.

The true shape of the absorption and the reconstruction results for Example 3 are given in Figure 3.4. We give the total number of PDE solves for each approach in Table 3.6.

3D Experiment. The mesh is $32 \times 32 \times 32$, which gives $32,768$ degrees of freedom in the forward model (2.2). The model has 225 sources at the top and 225 detectors on the bottom, and we use only the zero frequency. In the PaLS approach, we use 27 CSRBFs, which leads to 135 parameters (five per 3D basis function) for the nonlinear optimization. The absorption image using the initial set of parameters is given in Figure 3.5 where 13 basis functions have a positive expansion coefficient (visible as high absorption regions) and 14 basis functions have a negative expansion coefficient (invisible). In our approach, we use 12 random simultaneous sources and detectors.

Example 4. The true absorption image for Example 4 is given in Figure 3.6a. We also include the reconstruction result using all sources and all detectors for comparison in Figure 3.6b. In Figure 3.6e-g, we show the reconstruction results for combining random and optimized simultaneous sources and detectors. The results show similar behavior as was

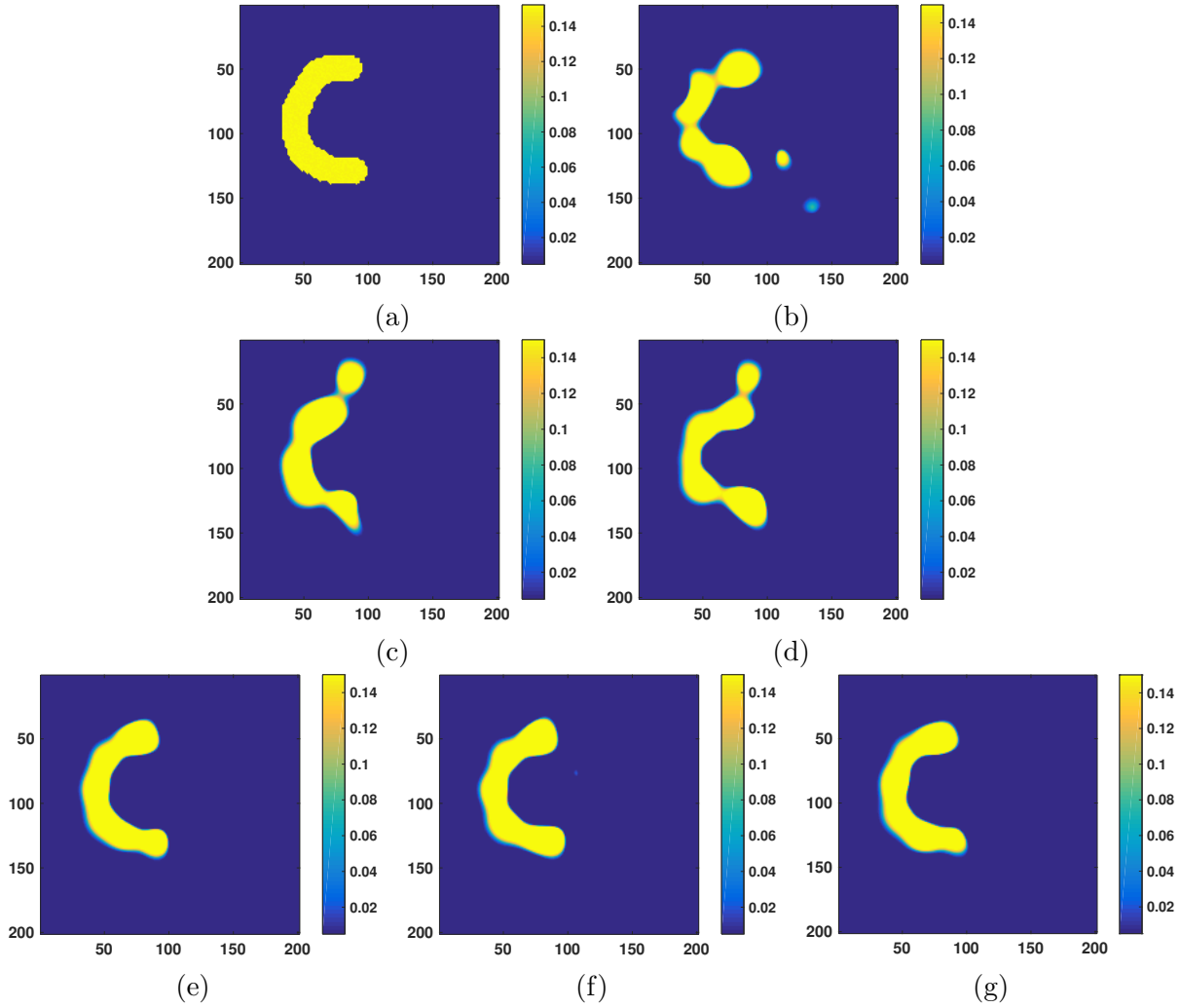


Figure 3.2: Results for Example 1. Reconstruction of a test anomaly on 201×201 mesh with 32 sources and detectors, 25 basis functions, and using only the zero frequency. The SAA approach uses 10 random simultaneous sources and detectors.
(a) True shape of the anomaly. (b) Reconstruction using all sources and all detectors. (c) Reconstruction using SAA approach at intermediate tolerance. (d) Reconstruction using the SAA approach after the maximum iterations. (e) Reconstruction with SAA and replacing 1 random simultaneous source and detector by a optimized simultaneous source and detector. (f) Reconstruction with SAA and replacing 2 random simultaneous sources and detectors by optimized simultaneous sources and detectors. (g) Reconstruction with SAA and replacing 3 random simultaneous sources and detectors by optimized simultaneous sources and detectors.

	SAA Approach			Simult Rand & Optimized Src/Det		
	Iter	True $\|\mathbf{f}\|_2^2$ (δ^2)	Estimated $\|\mathbf{f}\|_2^2$ (δ^2)	Iter	True $\|\mathbf{f}\|_2^2$ (δ^2)	Estimated $\|\mathbf{f}\|_2^2$ (δ^2)
Example 1	1	118980	60504	1-9	(SAA)*	(SAA)*
	7	4508.8	2943	10	298.46	313.03
	13	78.698	42.345	14	166.06	156.42
	19	1.607	1.069	16	5.951	6.251
	20	1.735	0.845	19	2.392	2.576
	(99)	–	–	20	0.738	0.806
Example 2	1	89777	119030	1-8	(SAA)*	(SAA)*
	6	26144	31707	9	278.26	554.33
	11	278.69	432.4	13	46.7	105.78
	17	1.9412	1.869	18	3.58	7.642
	18	1.838	0.977	21	2.838	2.999
	(99)	–	–	22	0.428	0.960
Example 3	1	90913	30028	1-8	(SAA)*	(SAA)*
	7	9017.3	2708.3	9	1188.7	1084.7
	14	116.22	33.279	16	292.4	264.67
	20	2.4558	1.1767	21	17.268	35.567
	21	1.2917	0.9255	26	12.522	13.947
	(99)	–	–	27	0.128	0.242

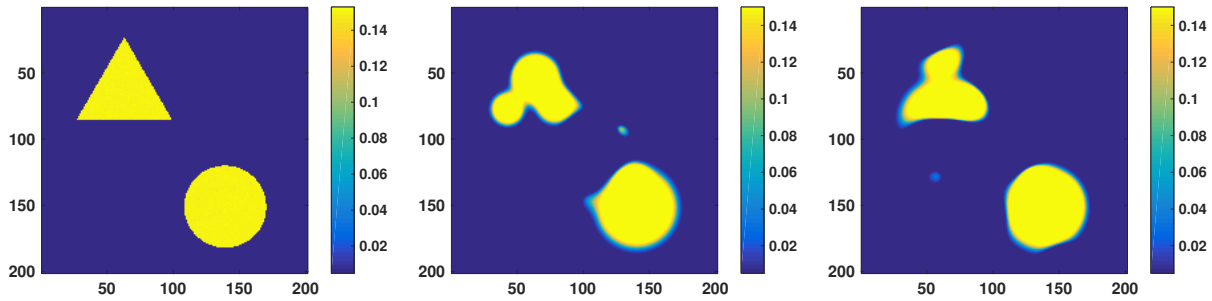
Table 3.2: Subset of results for Example 1,2 and 3. The comparison of the true objective function $\|\mathbf{f}\|_2^2$ and its SAA estimate relative to the stopping criterion (δ^2) for selected iterations. For the SAA approach, the estimated residual is obtained with 10 random simultaneous sources and detectors. Parentheses indicate that the SAA approach does not reach the tolerance. The estimated residual for combining simultaneous random and optimized sources and detectors that we report here are those obtained when replacing 2 random sources and 2 detectors by optimized sources and detectors for Example 1; replacing 1 random source and 1 detector by an optimized source and detector for Example 2; and replacing 3 random sources and 3 detectors by optimized sources and detectors for Example 3. *(SAA) indicates that we initially use the SAA approach at the intermediate tolerance.

	Example 1 (m/n)	Example 2 (m/n)	Example 3 (m/n)
Replacing 1 src/1 det	9 /14	9/14	7/12
Replacing 2 srcs/2 dets	7/14	6/14	5/13
Replacing 3 srcs/3 dets	4/16	4/15	2/13
SAA	24/28	24/28	23/28

Table 3.3: The average number of times (m) that the estimated residual underestimates the true residual out of the total number of iterations (n) on average for 50 trials to reach to the stopping criterion, δ^2.

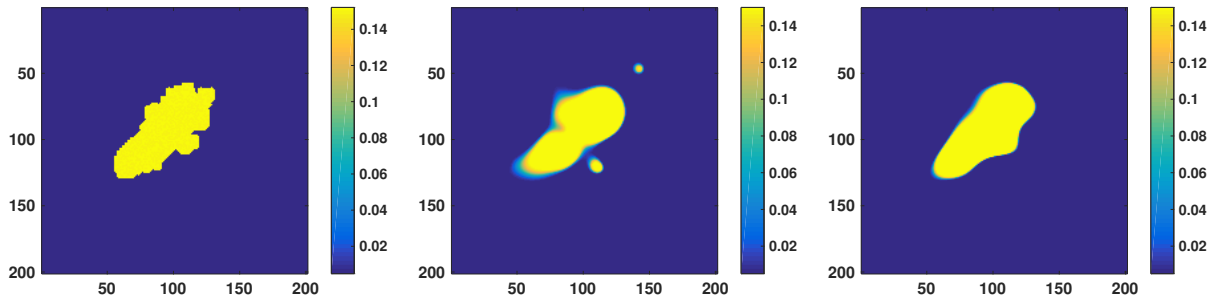
	Iteration Number	Function Evaluations	Jacobian Evaluations	Total PDE Solves	Tol
SAA* (intermediate tol)	9	10	5	150	δ
Replacing 1 src/1 det	23	25	12	434	δ^2
Replacing 2 srcs/2 dets	23	25	12	434	δ^2
Replacing 3 srcs/3 dets	25	27	13	464	δ^2
All srcs/All dets	71	72	47	3808	δ^2
SAA**	28	29	18	470	δ^2
SAA ***	(92)	(93)	(67)	1700	δ^2

Table 3.4: Example 1 Results. The total number of iterations, function evaluations, Jacobian evaluations and PDE solves required on average for 50 trials to reach the stopping criterion, δ^2.
*The first row gives the costs to reach the intermediate tolerance for the SAA approach, δ. ** Since the SAA estimate becomes biased and underestimates the objective function, the algorithm stops prematurely. ***The SAA approach measuring the convergence with the true objective function. Parentheses indicate that the SAA approach does not reach the tolerance.



(a) (b) (c)

Figure 3.3: Results for Example 2. Reconstruction of a test anomaly on 201×201 mesh with 32 sources and detectors, 25 basis functions, and using only the zero frequency.
(a) True shape of the anomaly. (b) Reconstruction using all sources and all detectors. (c) Reconstruction with SAA and replacing 1 simultaneous random source and detector by optimized source and detector.



(a) (b) (c)

Figure 3.4: Results for Example 3. Reconstruction of a test anomaly on 201×201 mesh with 32 sources and detectors, 25 basis functions, and using only the zero frequency.
(a) True shape of the anomaly. (b) Reconstruction using all sources and all detectors. (c) Reconstruction with SAA and replacing 3 random simultaneous sources and detectors by optimized sources and detectors.

	Iteration Number	Function Evaluations	Jacobian Evaluations	Total PDE Solves	Tol
SAA (intermediate)	9	10	5	150	δ
Replacing 1 src/1 det	23	25	12	434	δ^2
Replacing 2 srcs/2 dets	23	25	12	434	δ^2
Replacing 3 srcs/3 dets	24	26	13	454	δ^2
All srcs/All dets	97	98	69	5344	δ^2
SAA*	28	29	16	450	δ^2
SAA**	(92)	(93)	(65)	1580	δ^2

Table 3.5: Example 2 Results. The total number of iterations, function evaluations, Jacobian evaluations and PDE solves required on average for 50 trials to reach the stopping criterion, δ^2.
*The first row gives the costs to reach the intermediate tolerance for the SAA approach δ. ** Since the SAA estimate becomes biased and underestimates the objective function, the algorithm stops prematurely. ***The SAA approach measuring the convergence with the true objective function. Parentheses indicate that the SAA approach does not reach the tolerance.

	Iteration Number	Function Evaluations	Jacobian Evaluations	Total PDE Solves	Tol
SAA (intermediate)	10	11	5	160	δ
Replacing 1 src/1 det	22	24	11	414	δ^2
Replacing 2 srcs/2 dets	23	25	11	424	δ^2
Replacing 3 srcs/3 dets	23	25	12	434	δ^2
All srcs/All dets	25	26	14	1280	δ^2
SAA*	28	29	16	450	δ^2
SAA**	(96)	(97)	(68)	1650	δ^2

Table 3.6: Example 3 Results. The total number of iterations, function evaluations, Jacobian evaluations and PDE solves required on average for 50 trials to reach the stopping criterion, δ^2.
*The first row gives the costs to reach the intermediate tolerance for the SAA approach, δ. ** Since the SAA estimate becomes biased and underestimates the objective function, the algorithm stops prematurely. ***The SAA approach measuring the convergence with the true objective function. Parentheses indicate that the SAA approach does not reach the tolerance.

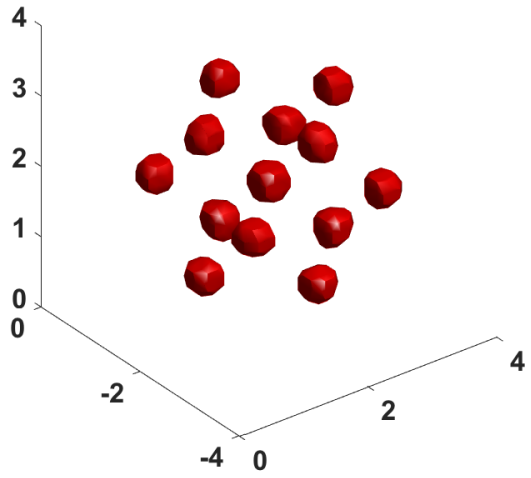


Figure 3.5: Initial configuration with 27 basis functions arranged in a $3 \times 3 \times 3$ grid where 13 basis functions have a positive expansion factors (visible) and 14 basis functions have negative expansion factors (invisible).

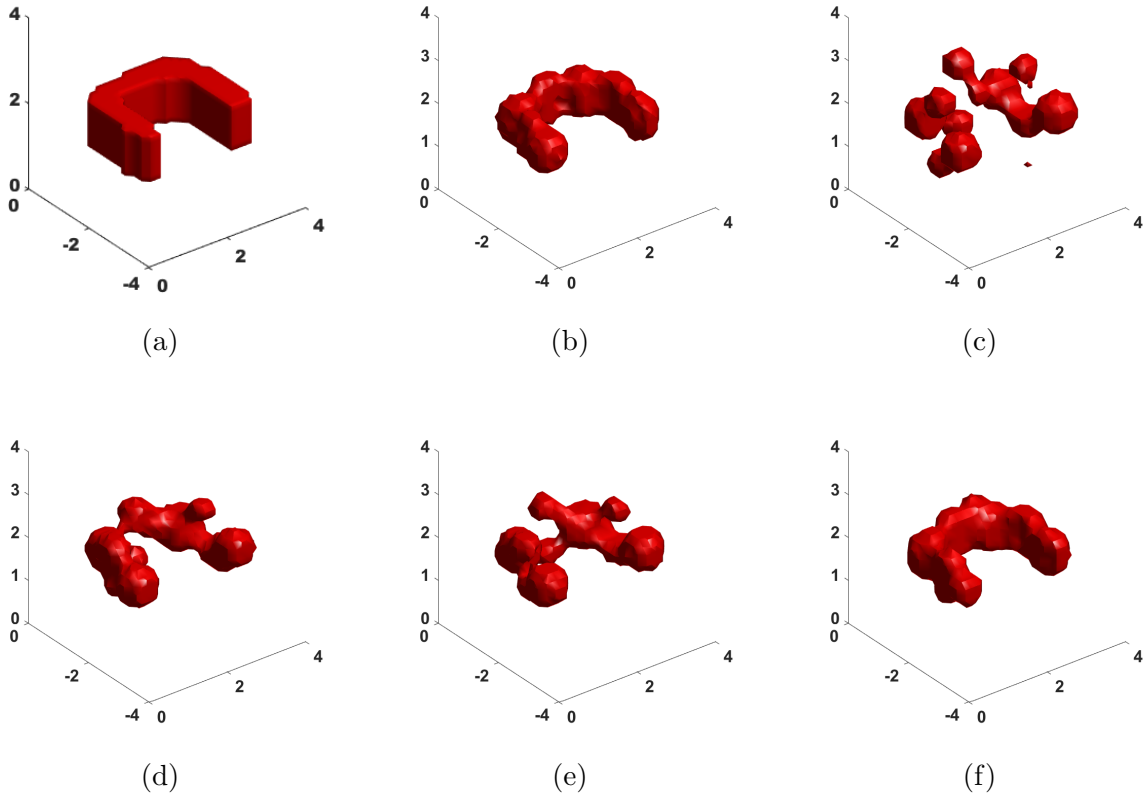
	Iteration Number	Function Evaluations	Jacobian Evaluations	Total PDE Solves	Tol
SAA* (intermediate tol)	5	6	4	100	δ
Replacing 1 src/1 det	8	9	4	706	δ^2
Replacing 2 srcs/2 dets	17	18	8	862	δ^2
All srcs/All dets	16	17	9	5850	δ^2
SAA **	(99)	(100)	(57)	1884	δ^2

Table 3.7: Example 4 Results. The total number of iterations, function evaluations, Jacobian evaluations and PDE solves required to reach the stopping criterion, δ^2.
*The first row gives the costs to reach the intermediate tolerance for the SAA approach, δ. ** The SAA approach measuring the convergence with the true objective function. Parentheses indicate that the SAA approach does not reach the tolerance.

observed in 2D experiments. Note that SAA gives a good localization of the anomaly in Figure 3.6c. However, no further improvement appears using the SAA approach after the maximum iterations, see Figure 3.6d.

The straightforward inversion using all sources and detectors requires 5850 large linear solves for the reconstruction. Table 3.7 shows that our approach reduces the number of large linear solves by *about a factor 7* compared to the all sources and detectors, while approximating the original shape well. For larger problems with many sources and detectors and using multiple frequencies, we expect much larger gains.

Overall, our approach improves the rate of convergence of the optimization and reduces the number of large-scale linear systems solves. Moreover, combining random and optimized simultaneous sources and detectors improves the quality of the inverse solution.



(a) (b) (c)

(d) (e) (f)

Figure 3.6: Results for Example 4. Reconstruction of a test anomaly on $32 \times 32 \times 32$ mesh with 225 sources and detectors, 27 basis functions, and using only the zero frequency. The SAA approach uses 15 random simultaneous sources and detectors.
(a) True shape of the anomaly. (b) Reconstruction using all sources and all detectors. (c) Reconstruction using SAA approach at intermediate tolerance. (d) Reconstruction using the SAA approach after the maximum iterations. (e) Reconstruction with SAA and replacing 1 random simultaneous source and detector by an optimized source and detector. (f) Reconstruction with SAA and replacing 2 random simultaneous sources and detectors by optimized sources and detectors.

Chapter 4

Randomization for the Efficient Computation of Reduced Order Models

In DOT, the forward model is described by a large-scale discretized partial differential equation (PDE). The main computational bottleneck is the repeated evaluation of this forward model and its derivatives to recover the desired image. The number of sources and detectors may be a thousand or more in 3D, and combined with multiple frequencies the resulting computational cost can be prohibitively expensive. Hence, we need new computational techniques that characterize the medium quickly and efficiently for such inverse problems.

The use of interpolatory parametric reduced models as surrogates for the full forward model to reduce the size of the linear systems solved in each optimization step has been considered in DOT. While approximating both the cost functional and the Jacobian accurately [23]. This approach significantly reduces the cost of the inversion process by drastically reducing the computational cost of solving the forward problems. However, to compute the bases used in the ROM, many large linear systems need to be solved, followed by an expensive rank-revealing factorization.

This rank-revealing factorization in building the ROM basis reveals that the standard method for ROM construction solves many more linear systems than needed [23, 45]. So, using ROMs reduces the size of the linear systems solved in each step of the optimization, but it still requires the solution of many large systems to build the ROM basis. In Chapter 2, we show that randomization can reduce the number of large linear system solves [6]. Hence, in this chapter, we combine these two approaches to obtain an effective and computationally highly efficient approach for nonlinear parameter inversion. We employ randomization to capture essentially the same ROM space at much lower cost via sampling.

In Section 4.1, we briefly review DOT. In Section 4.2, we give the system-theoretic notation

for DOT and a brief overview of interpolatory model reduction. In Section 4.3, we introduce a new randomized approach to generate ROMs efficiently. In Section 4.4, we provide a theoretical justification for exploiting low rank structure in the reduction basis, and we connect our approach, using randomization to compute the interpolatory model reduction bases, to tangential interpolation. Numerical results are given in Section 4.5 for 2D and 3D problems. Conclusions and future work are outlined in Section 4.6.

For large 3D problems, we use iterative solvers to generate the candidate ROM basis. In chapter 5, we explore ways to incorporate low accuracy iterative solves in interpolatory model reduction.

4.1 Introduction

DOT is a non-invasive, low cost alternative for breast and brain imaging compared with X-Ray and MRI. In DOT, near infra-red light from an array of sources is transmitted through the medium and measured with an array of detectors. Here, we assume that the diffusion coefficient is known, and we use measurements and the forward model to recover the absorption coefficient of the medium. The absorption coefficient can be used to distinguish healthy tissue from tumors [5].

We consider the diffusion model, posed in the frequency domain, for the photon flux $\phi(\mathbf{x})$ driven by an input source $g(\mathbf{x})$. The forward model for DOT is given by

$$
\begin{aligned}
&-\nabla\cdot(D(\mathbf{x})\nabla\phi(\mathbf{x}))+\mu(\mathbf{x})\phi(\mathbf{x})+\frac{\imath\omega}{\nu}\phi(\mathbf{x})=g(\mathbf{x}), \qquad (4.1)\\
&\text{for}\quad \mathbf{x}=(x_1,x_2,x_3)^T \text{ and } -a<x_1<a,\quad -b<x_2<b,\quad 0<x_3<c,\\
&\phi(\mathbf{x})=0 \text{ if } 0\le x_3\le c \text{ and either } x_1=\pm a, \text{ or } x_2=\pm b,\\
&0.25\phi(\mathbf{x})+\frac{D(\mathbf{x})}{2}\frac{\partial\phi(\mathbf{x})}{\partial\xi}=0 \text{ for } x_3=0, \text{ or } x_3=c,
\end{aligned}
$$

where $D(\mathbf{x})$ is the diffusion coefficient, $\mu(\mathbf{x})$ the absorption coefficient, ω is the frequency modulation of light, ν is the speed of light in the medium, and $\boldsymbol{\xi}$ is the outward unit normal on the boundary [5].

The Parametric Level Set (PaLS) approach [1] has been used to reconstruct complex geometries, and provide regularization to compensate for the influence of noise and ill-posedness of DOT [1, 6, 23]. We parameterize the absorption field $\mu(\mathbf{x};\mathbf{p})$ using PaLS and reduce the dimension of the parameter space. Hence, we only solve for a modest number of parameters, $\mathbf{p}\in\mathbb{R}^{n_p}$, that describe the shape of potential anomalies (tumors), rather than solving for absorption at every grid point.

Let n_s be the number of sources on the top, n_d be the number of detectors on the bottom, and n_ω be the number of frequencies to generate the DOT data. The discretization of (4.1)

can be done by finite element or finite difference techniques. Let the rows of $\mathbf{C} \in \mathbb{R}^{n_d \times n}$ correspond to the detectors, the columns of $\mathbf{B} \in \mathbb{R}^{n \times n_s}$ correspond to the sources, and let $\widehat{\mathbf{m}}(\omega; \mathbf{p})$ be the vector of detector outputs. After discretization of (4.1), measurements at the detector locations satisfy

$$\widehat{\mathbf{m}}(\omega; \mathbf{p}) = \boldsymbol{\Psi}(\omega; \mathbf{p})\ \hat{\mathbf{u}}(\omega) \quad \text{where} \quad \boldsymbol{\Psi}(\omega; \mathbf{p}) = \mathbf{C}\left(\frac{\imath\omega}{\nu}\,\mathbf{E} + \mathbf{A}(\mathbf{p})\right)^{-1}\mathbf{B}, \tag{4.2}$$

and $\mathbf{A}(\mathbf{p}) \in \mathbb{R}^{n \times n}$ results from a finite difference discretization of the diffusion and absorption terms, and $\mathbf{E} \in \mathbb{R}^{n \times n}$ is the identity except for zero rows corresponding to points on the boundaries $x_3 = 0$ and $x_3 = c$ in (4.1). Thus, $\mathbf{E}$ is singular. For a given frequency ω, and parameter vector $\mathbf{p}$, $\boldsymbol{\Psi}(\omega; \mathbf{p})$ gives the map from inputs $\hat{\mathbf{u}}(\omega)$ to outputs $\widehat{\mathbf{m}}(\omega; \mathbf{p})$; this is known as the transfer function. Combining all the predicted observation vectors, we obtain the complex matrix

$$\mathbb{M}(\mathbf{p}) = [\widehat{\mathbf{m}}_1(\omega_1; \mathbf{p}),\ \ldots,\ \widehat{\mathbf{m}}_{n_s}(\omega_1; \mathbf{p}),\ \widehat{\mathbf{m}}_1(\omega_2; \mathbf{p}), \ldots, \widehat{\mathbf{m}}_{n_s}(\omega_{n_\omega}; \mathbf{p})], \tag{4.3}$$

of dimension $n_d \times (n_s \cdot n_\omega)$. Given the empirical data $\mathbb{D}$, we solve the nonlinear least squares problem

$$\hat{\mathbf{p}} := \arg\ \min_{\mathbf{p}} \|\mathbb{M}(\mathbf{p}) - \mathbb{D}\|_F^2. \tag{4.4}$$

The solution $\hat{\mathbf{p}}$ identifies the absorption field and thus the anomaly. The PaLS representation regularizes the problem, so no further regularization is needed [1]. Evaluating the objective function at $\mathbf{p}$ requires solving the systems

$$\left(\frac{\imath\omega_j}{\nu}\,\mathbf{E} + \mathbf{A}(\mathbf{p})\right)\hat{\mathbf{Y}} = \mathbf{B}, \tag{4.5}$$

for each frequency, leading to $n_s \cdot n_\omega$ large linear systems. For Newton-type algorithms, it is also necessary to construct the Jacobian of $\boldsymbol{\Psi}(\mathbf{p})$,

$$\mathbf{J}(\mathbf{p}) = -\mathbf{C}\left(\frac{\imath\omega_j}{\nu}\mathbf{E} + \mathbf{A}(\mathbf{p})\right)^{-1}\frac{\partial}{\partial \mathbf{p}}\mathbf{A}(\mathbf{p})\left(\frac{\imath\omega_j}{\nu}\mathbf{E} + \mathbf{A}(\mathbf{p})\right)^{-1}\mathbf{B}, \tag{4.6}$$

where $\frac{\partial}{\partial \mathbf{p}_k}\mathbf{A}(\mathbf{p})$ is diagonal and inexpensive to compute. Evaluating $\mathbf{J}$ using the co-state approach [59] requires solving the systems

$$\left(\frac{\imath\omega_j}{\nu}\,\mathbf{E} + \mathbf{A}(\mathbf{p})\right)^T\widehat{\mathbf{Z}} = \mathbf{C}^T, \tag{4.7}$$

for each frequency, leading to an additional $n_d \cdot n_\omega$ adjoint systems for the detectors. As a result, standard optimization approaches require $O(10^3 - 10^4)$ large linear system solves at each optimization step. The size of a realistic linear system is at least $O(10^6)$. Repeated evaluations of (4.5) and (4.7) lead to a computational bottleneck for solving the inverse problem.

For the solution to the minimization problem (4.4), we use the Trust region algorithm with REGularized model Solution (TREGS) [24] that has proven very effective for problems of this type.

4.2 Interpolatory Model Reduction

Projection-based parametric model reduction has been widely used to replace a dynamical system with high dimension, n, by one with smaller dimension, r, to allow rapid yet accurate simulation over the range of parameters [4, 10, 12, 13, 34]. To explain the ROM construction, we consider the time domain representation of the frequency domain equation (4.2),

$$\frac{1}{\nu}\mathbf{E}\,\dot{\mathbf{y}}(t;\mathbf{p}) = -\mathbf{A}(\mathbf{p})\mathbf{y}(t;\mathbf{p}) + \mathbf{B}\mathbf{u}(t) \quad \text{with} \quad \mathbf{m}(t;\mathbf{p}) = \mathbf{C}\mathbf{y}(t;\mathbf{p}). \tag{4.8}$$

We seek to replace this high dimensional dynamical system by a reduced order model

$$\frac{1}{\nu}\mathbf{E}_r\,\dot{\mathbf{y}}_r(t;\mathbf{p}) = -\mathbf{A}_r(\mathbf{p})\mathbf{y}_r(t;\mathbf{p}) + \mathbf{B}_r\mathbf{u}(t) \quad \text{with} \quad \mathbf{m}_r(t;\mathbf{p}) = \mathbf{C}_r\mathbf{y}_r(t;\mathbf{p}), \tag{4.9}$$

with the associated reduced transfer function

$$\mathbf{\Psi}_r(\omega;\mathbf{p}) = \mathbf{C}_r\left(\frac{\imath\omega}{\nu}\mathbf{E}_r + \mathbf{A}_r(\mathbf{p})\right)^{-1}\mathbf{B}_r \in \mathbb{C}^{n_d\times n_s}, \tag{4.10}$$

where

$$\begin{aligned}\mathbf{E}_r = \mathbf{W}_r^T\mathbf{E}\mathbf{V}_r, \quad \mathbf{B}_r &= \mathbf{W}_r^T\mathbf{B}, \quad \mathbf{C}_r = \mathbf{C}\mathbf{V}_r,\\ \mathbf{A}_r(\mathbf{p}) &= \mathbf{W}_r^T\mathbf{A}(\mathbf{p})\mathbf{V}_r.\end{aligned}$$

and $\mathbf{V}_r \in \mathbb{C}^{n\times r}$ and $\mathbf{W}_r \in \mathbb{C}^{n\times r}$ are the computed projection basis matrices. To obtain a high-fidelity approximation to $\mathbf{\Psi}(\omega;\mathbf{p})$ over all parameters and frequencies of interest, we need to construct appropriate matrices $\mathbf{W}_r$ and $\mathbf{V}_r$. Several parametric model reduction methods exist to select $\mathbf{W}_r$ and $\mathbf{V}_r$; see, e.g., [12, 13, 14, 19, 20, 34]. In DOT, the function and Jacobian evaluations correspond to transfer function evaluations and their derivatives at parameter points that arise in the minimization (4.4). Therefore, a natural choice is to use interpolatory model reduction to construct $\mathbf{W}_r$ and $\mathbf{V}_r$ [8, 18, 26, 31]. Following well-known theorems regarding interpolatory parametric model reduction in [8, Theorem 4.1-4.2], generating the ROM using Algorithm 1 (below) for selected interpolation points in frequency and parameter space, $(\omega_j, \boldsymbol{\pi}_i)$ for $j = 1, \cdots, n_\omega$ and $i = 1, \cdots, n_k$, ensures that the ROM and its derivatives match the full order model and its derivatives *exactly* at the selected points, that is,

$$\begin{aligned}\mathbf{\Psi}(\omega_j;\mathbf{p}_i) &= \mathbf{\Psi}_r(\omega_j;\mathbf{p}_i),\\ \nabla_{\mathbf{p}}\mathbf{\Psi}(\omega_j;\mathbf{p}_i) &= \nabla_{\mathbf{p}}\mathbf{\Psi}_r(\omega_j;\mathbf{p}_i),\\ \mathbf{\Psi}'(\omega_j;\mathbf{p}_i) &= \mathbf{\Psi}_r'(\omega_j;\mathbf{p}_i),\end{aligned} \tag{4.11}$$

where $'$ denotes the derivative with respect to ω. In other words, the optimization algorithm does not see the difference between the reduced order model and the full order model at the selected points.

Algorithm 1 [23]

Given parameter interpolation points, $\boldsymbol{\pi}_1, \cdots, \boldsymbol{\pi}_{n_k}$, and frequency interpolation points, $\omega_1, \cdots, \omega_{n_\omega}$

(1) Generate candidate bases for $i = 1, \cdots n_k$
$$\mathbf{V}^{(i)} = \left[\left(\tfrac{\imath\omega_1}{\nu}\mathbf{E} - \mathbf{A}(\boldsymbol{\pi}_i)\right)^{-1} \mathbf{B}, \cdots, \left(\tfrac{\imath\omega_{n_\omega}}{\nu}\mathbf{E} - \mathbf{A}(\boldsymbol{\pi}_i)\right)^{-1} \mathbf{B} \right]$$
$$\mathbf{W}^{(i)} = \left[\left(\tfrac{\imath\omega_1}{\nu}\mathbf{E} - \mathbf{A}(\boldsymbol{\pi}_i)\right)^{-T} \mathbf{C}^T, \cdots, \left(\tfrac{\imath\omega_{n_\omega}}{\nu}\mathbf{E} - \mathbf{A}(\boldsymbol{\pi}_i)\right)^{-T} \mathbf{C}^T \right]$$

(2) Use an SVD or rank-revealing QR factorization to compute ROM bases $[\mathbf{V}^{(1)}, \mathbf{V}^{(2)}, \cdots, \mathbf{V}^{(n_k)}] \to \mathbf{V}^{(r)}$ and $[\mathbf{W}^{(1)}, \mathbf{W}^{(2)}, \cdots, \mathbf{W}^{(n_k)}] \to \mathbf{W}^{(r)}$. Then, employ one-sided projection $[\mathbf{V}^{(r)}, \ \mathbf{W}^{(r)}] \to \mathbf{V}_r$ to obtain the global basis $\mathbf{V}_r$ $(= \mathbf{W}_r)$.

After the global basis $\mathbf{V}_r$ has been computed using Algorithm 1, we can compute the reduced order model for every parameter $\mathbf{p}$ as follows

$$\begin{aligned} \mathbf{E}_r = \mathbf{V}_r^T \mathbf{E} \mathbf{V}_r, \quad \mathbf{B}_r &= \mathbf{V}_r^T \mathbf{B}, \quad \mathbf{C}_r = \mathbf{C}\mathbf{V}_r, \\ \mathbf{A}_r(\mathbf{p}) &= \mathbf{V}_r^T \mathbf{A}(\mathbf{p}) \mathbf{V}_r. \end{aligned} \tag{4.12}$$

In our DOT setting, $\mathbf{A}(\mathbf{p}) = \mathbf{A}_0 + \mathbf{A}_1(\mathbf{p})$ where $\mathbf{A}_0$ is constant and $\mathbf{A}_1(\mathbf{p})$ carries the parametric dependency. Since $\mathbf{A}(\mathbf{p}) \in \mathbb{R}^{n \times n}$ results from a finite difference discretization of the diffusion and absorption terms, $\mathbf{A}_1(\mathbf{p})$ is diagonal. Details on efficient computation of $\mathbf{A}_r(\mathbf{p})$ is discussed in [23].

Algorithm 1 is used to construct the projection spaces for given parameter and frequency interpolation points [13, 23]. The selection of interpolation points for our test problems is discussed in Section 4.5. Some of the results in [23] suggest that it may be possible to compute the ROM basis in an offline phase.

Following Algorithm 1, we need to solve systems (4.5) and (4.7) for every parameter and frequency interpolation point to generate candidate basis, leading to $n_k n_\omega n_s + n_k n_\omega n_d$ large linear systems. If the number of sources and detectors is large, the construction of the ROM basis in Algorithm 1 still incurs a substantial cost at the start of the inversion. In fact, for DOT using parametric model order reduction (PMOR) [23], this is currently the dominant cost. On the other hand, after generating a candidate basis, Algorithm 1 uses a rank-revealing reduction to compute the ROM basis. Experiements show that the candidate basis has many more vectors than the final projection basis [23].

In Section 4.3, we employ randomization theory to capture essentially the same ROM space at much lower cost via sampling. Hence, we avoid first solving each linear system in (4.5) and (4.7).

4.3 Randomization and Reduced Order Modeling

As discussed in the previous section, many large linear systems are solved to generate the candidate basis in Algorithm 1. This step is followed by an expensive rank-revealing factorization to obtain a much smaller number of vectors for the global basis [23]. This reflects the fact that the ROM bases generated in Algorithm 1 is very close to a matrix of much lower rank. In this section, we provide a theoretical justification for exploiting low rank structure in the candidate bases computed by Algorithm 1. Moreover, we propose to significantly reduce the number of large linear solves for constructing the global ROM basis via sampling. Then, we link randomized sampling for computing the ROM basis with tangential interpolation.

Interpolatory model reduction is effective for nonlinear parameter inversion when $r \ll n$. However, enforcing full interpolation by solving for all detectors and sources is computationally expensive. To give some perspective on the problem, consider a 3D test problem with 225 sources and detectors with 3 parameter sample points and 4 frequency interpolation points; see Section 4.5.2. The standard method for computing the global ROM basis (Algorithm 1) uses $\mathbf{B}$ and $\mathbf{C}$ to match the full objective function and Jacobian, see (4.11). Algorithm 1 initially generates 5400 directions and therefore solves 5400 large linear systems. However, after a rank-revealing factorization only 1228 of those directions are used for the global ROM basis; see Section 4.5.2. This shows that we have computed a lot of redundant information. An approach based on ideas from Krylov subspace recycling [37, 47] to compute the global basis more efficiently for DOT was recently explored in [45].

Since candidate ROM bases are (nearly) low rank, an alternative approach to drastically reduce the number of vectors in the candidate bases would be to multiply the candidate bases by random matrices to get the low rank structure via sampling [33]. To actually reduce the number of large linear solves, we implement this approach by forming a few linear combination of the right-hand sides into a few stochastic sources and detectors and solve only for the resulting modest numbers of right hand sides,

$$\left(\frac{\imath\omega_j}{\nu}\,\mathbf{E} + \mathbf{A}(\mathbf{p})\right)\widetilde{\mathbf{Y}} = \mathbf{B}\mathbf{R}, \text{ and } \left(\frac{\imath\omega_j}{\nu}\,\mathbf{E} + \mathbf{A}(\mathbf{p})\right)^T\widetilde{\mathbf{Z}} = \mathbf{C}^T\mathbf{L}, \tag{4.13}$$

where $\mathbf{R} = [\mathbf{r}_1 \cdots \mathbf{r}_{\ell_s}]$ and $\mathbf{r}_i \in \mathbb{R}^{n_s}$, for $i = 1 \cdots \ell_s$, with $\ell_s \ll n_s$, and $\mathbf{L} = [\boldsymbol{\ell}_1 \cdots \boldsymbol{\ell}_{\ell_d}]$ and $\boldsymbol{\ell}_j \in \mathbb{R}^{n_d}$, for $j = 1 \cdots \ell_d$, with $\ell_d \ll n_d$ [6, 32]. We use the Rademacher distribution to generate $\mathbf{R}$ and $\mathbf{L}$. Other distributions can be used as well. This approach drastically reduces the number of large linear system solves and adds only $(\ell_s + \ell_d)$ directions to the ROM basis per frequency and interpolation point as opposed to $(n_s + n_d)$, drastically reducing the cost of the rank revealing factorization as well.

Our approach can also be seen as an efficient way of finding a good set of tangential interpolation directions. Therefore, we rephrase the following theorem, from [8], in the context of DOT to motivate the use of stochastic sources and detectors in the tangential interpolation setting for DOT.

Theorem 4.1. *Let* $\mathbf{A}(\mathbf{p})$ *be continuously differentiable in a neighborhood of* $\tilde{\mathbf{p}}$. *Let* $\hat{\omega} \in \mathbb{R}$, *and let* $\left(\frac{\imath\hat{\omega}}{\nu}\mathbf{E} + \mathbf{A}(\tilde{\mathbf{p}})\right)$ *and* $\left(\frac{\imath\hat{\omega}}{\nu}\widetilde{\mathbf{E}}_r + \widetilde{\mathbf{A}}_r(\tilde{\mathbf{p}})\right)$ *be invertible. Suppose the columns of* $\mathbf{L} \in \mathbb{R}^{n_d \times \ell_d}$ *and* $\mathbf{R} \in \mathbb{R}^{n_s \times \ell_s}$ *are linearly independent vectors with* $\ell_s \ll n_s$ *and* $\ell_d \ll n_d$. *Then, both* $\mathbf{L}^T\mathbf{\Psi}(\hat{\omega};\tilde{\mathbf{p}})\mathbf{R}$ *and* $\mathbf{L}^T\widetilde{\mathbf{\Psi}}_r(\hat{\omega};\tilde{\mathbf{p}})\mathbf{R}$ *are differentiable with respect to* $\mathbf{p}$ *in a neighborhood of* $\tilde{\mathbf{p}}$. *If* $\text{Range}\left(\left(\frac{\imath\hat{\omega}}{\nu}\mathbf{E} + \mathbf{A}(\tilde{\mathbf{p}})\right)^{-1}\mathbf{BR}\right) \subset \text{Range}(\widetilde{\mathbf{V}})$ *and* $\text{Range}\left(\left(\frac{\imath\hat{\omega}}{\nu}\mathbf{E} + \mathbf{A}(\tilde{\mathbf{p}})\right)^{-T}\mathbf{C}^T\mathbf{L}\right) \subset \text{Range}(\widetilde{\mathbf{W}})$, *then*

$$\mathbf{L}^T\mathbf{\Psi}(\hat{\omega};\tilde{\mathbf{p}}) = \mathbf{L}^T\widetilde{\mathbf{\Psi}}_r(\hat{\omega};\tilde{\mathbf{p}}) \quad \textit{and} \quad \mathbf{\Psi}_r(\hat{\omega};\tilde{\mathbf{p}})\mathbf{R} = \widetilde{\mathbf{\Psi}}_r(\hat{\omega};\tilde{\mathbf{p}})\mathbf{R}, \tag{4.14}$$

$$\nabla_{\mathbf{p}}\left(\mathbf{L}^T\mathbf{\Psi}(\hat{\omega};\tilde{\mathbf{p}})\mathbf{R}\right) = \nabla_{\mathbf{p}}\left(\mathbf{L}^T\widetilde{\mathbf{\Psi}}_r(\hat{\omega};\tilde{\mathbf{p}})\mathbf{R}\right), \tag{4.15}$$

$$\mathbf{L}^T\mathbf{\Psi}'(\hat{\omega};\tilde{\mathbf{p}})\mathbf{R} = \mathbf{L}^T\widetilde{\mathbf{\Psi}}_r'(\hat{\omega};\tilde{\mathbf{p}})\mathbf{R}, \tag{4.16}$$

where $'$ *denotes the derivative with respect to* ω.

For general parametrized systems, optimal conditions are not known, except for special cases [8]. For a fixed set of parameter samples, the iterative rational Krylov algorithm (IRKA) [29] can be used for parametric systems to compute locally optimal ROMs. However, IRKA requires repeated evaluations of linear systems to choose the tangential directions in an optimal way. For efficiency, we prefer to choose the tangential directions, $\mathbf{R}$ and $\mathbf{L}$, without an iterative process. Therefore, we propose to use stochastic sources and detectors to compute right and left tangential interpolation directions; see Algorithm 2.

Algorithm 2

Given parameter sample points $\boldsymbol{\pi}_i$, frequency interpolation points ω_j, and random matrices $\mathbf{L}_{i_j} \in \mathbb{R}^{n_d \times \ell_d}$ and $\mathbf{R}_{i_j} \in \mathbb{R}^{n_s \times \ell_s}$ for $i = 1, \cdots, n_k$ and $j = 1, \cdots, n_\omega$

(1) Generate candidate bases for $i = 1, \cdots n_k$

$$\widetilde{\mathbf{V}}^{(i)} = \left[\left(\tfrac{\imath\omega_1}{\nu}\mathbf{E} - \mathbf{A}(\boldsymbol{\pi}_i)\right)^{-1}\mathbf{BR}_{i_1}, \cdots, \left(\tfrac{\imath\omega_{n_\omega}}{\nu}\mathbf{E} - \mathbf{A}(\boldsymbol{\pi}_i)\right)^{-1}\mathbf{BR}_{i_{n_\omega}}\right]$$

$$\widetilde{\mathbf{W}}^{(i)} = \left[\left(\tfrac{\imath\omega_1}{\nu}\mathbf{E} - \mathbf{A}(\boldsymbol{\pi}_i)\right)^{-T}\mathbf{C}^T\mathbf{L}_{i_1}, \cdots, \left(\tfrac{\imath\omega_{n_\omega}}{\nu}\mathbf{E} - \mathbf{A}(\boldsymbol{\pi}_i)\right)^{-T}\mathbf{C}^T\mathbf{L}_{i_{n_\omega}}\right]$$

(2) Use an SVD or rank-revealing QR factorization to compute ROM bases $[\widetilde{\mathbf{V}}^{(1)}, \widetilde{\mathbf{V}}^{(2)}, \cdots, \widetilde{\mathbf{V}}^{(n_k)}] \to \widetilde{\mathbf{V}}^{(r)}$ and $[\widetilde{\mathbf{W}}^{(1)}, \widetilde{\mathbf{W}}^{(2)}, \cdots, \widetilde{\mathbf{W}}^{(n_k)}] \to \mathbf{W}^{(r)}$. Then, employ one-sided projection $[\widetilde{\mathbf{V}}^{(r)}, \widetilde{\mathbf{W}}^{(r)}] \to \widetilde{\mathbf{V}}_r$ to obtain the global basis $\widetilde{\mathbf{V}}_r$ $(= \widetilde{\mathbf{W}}_r)$.

Our approach can be combined with the approach introduced in [45] to compute only necessary extensions of the global basis.

After the global basis $\widetilde{\mathbf{V}}_r$ has been computed using Algorithm 2, we can compute the reduced order model for every parameter $\mathbf{p}$ as follows

$$\widetilde{\mathbf{E}}_r = \widetilde{\mathbf{V}}_r^T\mathbf{E}\widetilde{\mathbf{V}}_r, \quad \widetilde{\mathbf{B}}_r = \widetilde{\mathbf{V}}_r^T\mathbf{B}, \quad \widetilde{\mathbf{C}}_r = \mathbf{C}\widetilde{\mathbf{V}}_r$$
$$\widetilde{\mathbf{A}}_r(\mathbf{p}) = \widetilde{\mathbf{V}}_r^T\mathbf{A}(\mathbf{p})\widetilde{\mathbf{V}}_r. \tag{4.17}$$

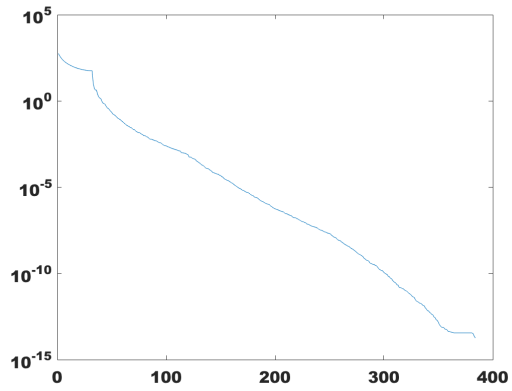


Figure 4.1: Singular values of the candidate basis $\mathbf{V}$ before computing global basis.

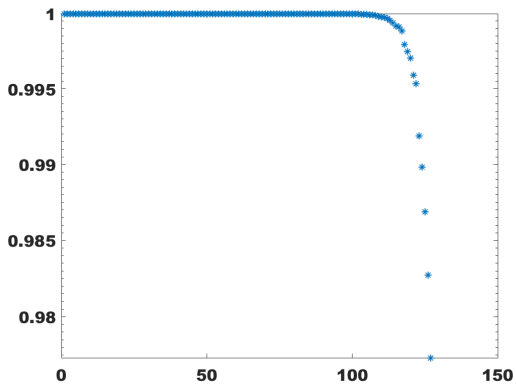


Figure 4.2: The cosine of the canonical angles between Range($\mathbf{V}_r$) and Range($\widetilde{\mathbf{V}}_r$).

Similar to $\mathbf{A}_r(\mathbf{p})$, the computation of $\widetilde{\mathbf{A}}_r(\mathbf{p})$ can be done efficiently [23].

Our approach constructs a global basis $\widetilde{\mathbf{V}}_r$ such that the reduced transfer function $\widetilde{\mathbf{\Psi}}_r(\omega;\mathbf{p})$ tangentially interpolates $\mathbf{\Psi}_r(\omega;\mathbf{p})$ at the sample points along $\mathbf{R} = [\mathbf{r}_1 \cdots \mathbf{r}_{\ell_s}]$ and $\mathbf{L} = [\boldsymbol{\ell}_1 \cdots \boldsymbol{\ell}_{\ell_d}]$; see (4.14)-(4.16). To demonstrate the effective low rank of the candidate basis $\mathbf{V}$ from Algorithm 1, we consider a 2D experiment with 32 sources and 32 detectors. We use 4 parameter sample points and 2 frequencies to construct $\mathbf{V}$. Figure 4.1 shows the fast decay of the singular values of the candidate basis. We also give the cosines of the canonical angles between Range($\mathbf{V}_r$), the global basis using all sources and detectors, and Range($\widetilde{\mathbf{V}}_r$), the global basis using stochastic sources and detectors, for the same experiment in Figure 4.2. The details of this experiment are discussed later in Section 4.5.1. Clearly, the right projection space obtained with Algorithm 2 is very close to the right projection space obtained with Algorithm 1. So, in the DOT setting, the proposed model reduction basis computed using tangential interpolation via stochastic directions spans almost the same subspace as the full interpolation basis (ignoring directions corresponding to tiny singular values). Algorithm 2 greatly reduces the cost of building ROMs.

In the next section, we provide some further theoretical motivation for our approach.

4.4 Analysis of Combining Randomization and ROM

In the previous section, we observe that the right projection space obtained using Algorithm 2 is very close to the the right projection space obtained using Algorithm 1. In the following subsections, we provide some theoretical motivation for the low-rank structure of the global basis computed using Algorithm 1.

We first use the system properties to rewrite the transfer function in terms of a symmetric positive definite (SPD) matrix as in [37, 45]. Then, we analyze the relative change in the global basis computed using Algorithm 1 from parameter to parameter to show the effectiveness of the global basis computed using Algorithm 2. Moreover, we examine the relative changes in the Gramians to justify our claims.

4.4.1 Rewriting the Transfer Function

We first show that by eliminating the boundary conditions and solving for the Schur complement (for $\omega = 0$), the full order transfer function $\mathbf{\Psi}(\omega, \mathbf{p})$ in (4.2) can be *equivalently* written in terms of an SPD matrix.

We follow the same discretization scheme as in [37, 45]. Then, the matrix $\left(\frac{\imath\omega_j}{\nu}\mathbf{E} + \mathbf{A}(\mathbf{p})\right)$ has the following block structure for the two-dimensional case,

$$\begin{bmatrix} \mathbf{G} & \mathbf{D}_1 \\ \mathbf{D}_2 & \mathbf{F}(\mathbf{p}) + \frac{\imath\omega h^2}{\nu}\mathbf{I} \end{bmatrix}, \tag{4.18}$$

where we have ordered the $N_x N_y$ unknowns such that the unknowns on the top and the bottom boundary appear first, followed by lexicographical ordering of internal points. The structure of the blocks are defined as follows

- $\mathbf{G}$ is an invertible diagonal matrix,
- $\mathbf{D}_1$ has at most one nonzero per row, and these occur only in the first $2N_x$ and the last $2N_x$ columns, and
- $\mathbf{D}_2$, although it has different entries, has the same sparsity pattern as $\mathbf{D}_1^T$.

In addition, the matrices $\mathbf{C}$ and $\mathbf{B}$ contain columns from an $N_x N_y \times N_x N_y$ identity matrix and $\mathbf{B}$ is scaled by $\dfrac{1}{h^2}$. Hence, we partition $\mathbf{C}$ and $\mathbf{B}$ conforming with (4.18) as

$$\begin{bmatrix} \mathbf{C}_1 & \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{0} \\ \mathbf{B}_1 \end{bmatrix}. \tag{4.19}$$

In Figure 4.3, we give the sparsity pattern of $\mathbf{A}(\mathbf{p})$ to help visualize the structure of our matrix. For ease of exposition, we first consider $\omega = 0$ in (4.18). We also consider the following block structure for $\mathbf{A}(\mathbf{p})^{-1}$

$$\begin{bmatrix} \mathbf{H} & \mathbf{S}_1 \\ \mathbf{S}_2 & \mathbf{N} \end{bmatrix}. \tag{4.20}$$

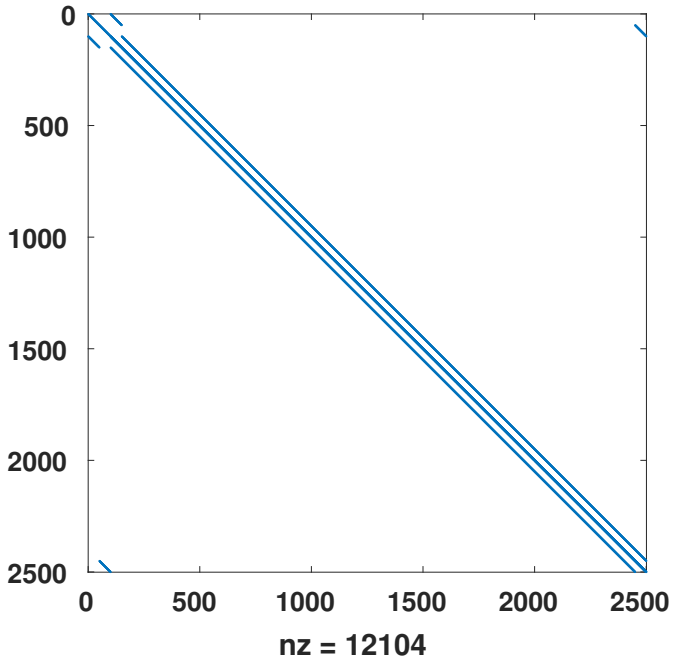


Figure 4.3: Sparsity plot of $\mathbf{A}(\mathbf{p})$.

Then, using $\mathbf{A}(\mathbf{p})\mathbf{A}(\mathbf{p})^{-1} = \mathbf{A}(\mathbf{p})^{-1}\mathbf{A}(\mathbf{p}) = \mathbf{I}$, we obtain

$$\begin{aligned} \mathbf{H} &= \left[\mathbf{G} - \mathbf{D}_1\mathbf{F}^{-1}\mathbf{D}_2\right]^{-1}, && (4.21)\\ \mathbf{N} &= \left[\mathbf{F} - \mathbf{D}_2\mathbf{G}^{-1}\mathbf{D}_1\right], && (4.22)\\ \mathbf{S}_1 &= -\mathbf{G}^{-1}\mathbf{D}_1\left[\mathbf{F} - \mathbf{D}_2\mathbf{G}^{-1}\mathbf{D}_1\right]^{-1}, && (4.23)\\ \mathbf{S}_2 &= -\left[\mathbf{F} - \mathbf{D}_2\mathbf{G}^{-1}\mathbf{D}_1\right]^{-1}\mathbf{D}_2\mathbf{G}^{-1}. && (4.24) \end{aligned}$$

Using (4.19) and (4.23), we rewrite the transfer function at $\omega = 0$ as

$$\mathbf{\Psi}(0,\mathbf{p}) = \mathbf{C}\mathbf{A}(\mathbf{p})^{-1}\mathbf{B} = \mathbf{C}_1\mathbf{S}_1\mathbf{B}_1 = -\mathbf{C}_1\mathbf{G}^{-1}\mathbf{D}_1\left[\mathbf{F} - \mathbf{D}_2\mathbf{G}^{-1}\mathbf{D}_1\right]^{-1}\mathbf{B}_1. \tag{4.25}$$

Setting

$$\widetilde{\mathbf{C}} = \mathbf{C}_1\mathbf{G}^{-1}\mathbf{D}_1, \quad \widetilde{\mathbf{A}}(\mathbf{p}) = -(\mathbf{F} - \mathbf{D}_2\mathbf{G}^{-1}\mathbf{D}_1), \quad \text{and} \quad \widetilde{\mathbf{B}} = \mathbf{B}_1, \tag{4.26}$$

we define a reduced transfer function for $\omega \neq 0$

$$\widehat{\mathbf{\Psi}}(\omega,\mathbf{p}) = \widetilde{\mathbf{C}}\left(\frac{\imath\omega}{\nu}\mathbf{I} + \widetilde{\mathbf{A}}(\mathbf{p})\right)^{-1}\widetilde{\mathbf{B}}, \tag{4.27}$$

where $\widetilde{\mathbf{A}}(\mathbf{p})$ is an SPD matrix. Next, we revisit the differential-algebraic system (4.8) (scaled by $1/\nu$ for convenience) and the system-theoretic interpretation for the ROM

$$\mathbf{E}\dot{\mathbf{y}}(t;\mathbf{p}) = -\mathbf{A}(\mathbf{p})\mathbf{y}(t;\mathbf{p}) + \mathbf{B}\,\mathbf{u}(t) \quad \text{with} \quad \mathbf{m}(t;\mathbf{p}) = \mathbf{C}\,\mathbf{y}(t;\mathbf{p}), \tag{4.28}$$

where $\mathbf{y}$ denotes the discretized photon flux, and $\mathbf{m}$ is the vector of detector outputs. Partitioning the matrix $\mathbf{E}$ conforming with (4.18) gives

$$\mathbf{E} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}. \tag{4.29}$$

Substituting (4.18), (4.19), and (4.29) into (4.28) gives

$$\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \dot{\mathbf{y}}_1 \\ \dot{\mathbf{y}}_2 \end{bmatrix} = - \begin{bmatrix} \mathbf{G} & \mathbf{D}_1 \\ \mathbf{D}_2 & \mathbf{F}(\mathbf{p}) \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{B}_1 \end{bmatrix} \mathbf{u}, \tag{4.30}$$

and we obtain

$$\mathbf{0} = -\mathbf{G}\mathbf{y}_1 - \mathbf{D}_1\mathbf{y}_2, \tag{4.31}$$

$$\dot{\mathbf{y}}_2 = -\mathbf{D}_2\mathbf{y}_1 - \mathbf{F}\mathbf{y}_2 + \mathbf{B}_1\mathbf{u}. \tag{4.32}$$

Substituting $\mathbf{y}_1 = -\mathbf{G}^{-1}\mathbf{D}_1\mathbf{y}_2$ from (4.31) into (4.32) yields

$$\dot{\mathbf{y}}_2 = -\left[\mathbf{F} - \mathbf{D}_2\mathbf{G}^{-1}\mathbf{D}_1\right]\mathbf{y}_2 + \mathbf{B}_1\mathbf{u} = \widetilde{\mathbf{A}}(\mathbf{p})\mathbf{y}_2 + \widetilde{\mathbf{B}}\mathbf{u}. \tag{4.33}$$

It is straightforward to show that

$$\mathbf{m}(t;\mathbf{p}) = \mathbf{C}\,\mathbf{y}(t;\mathbf{p}) = \mathbf{C}_1\mathbf{G}^{-1}\mathbf{D}_1\mathbf{y}_2 = \widetilde{\mathbf{C}}\mathbf{y}_2. \tag{4.34}$$

Next, we rewrite the transfer function by eliminating the singular matrix $\mathbf{E}$

$$\widehat{\mathbf{\Psi}}(\omega,\mathbf{p}) = \widetilde{\mathbf{C}}\left(\frac{\imath\omega}{\nu}\mathbf{I} + \widetilde{\mathbf{A}}(\mathbf{p})\right)^{-1}\widetilde{\mathbf{B}}. \tag{4.35}$$

Hence, from a system theoretic perspective, we converted systems of differential-algebraic equations (DAE) to systems of ordinary differential equations (ODE). For details on interpolatory model order reduction of DAEs, we refer the reader to [30]. From now on, we assume that we have converted systems of DAE to systems of ODE.

4.4.2 Perturbations in Candidate Basis

The following discussion assumes that the tissue has a relatively small anomaly. In this section, we show that candidate basis is low rank if $\mathbf{p}$ corresponds to a small anomaly and the matrix corresponding to each anomaly is a small perturbation of the discretized Laplacian (that is the matrix for no anomaly). In the following, we consider perturbations from the discretized Laplacian matrix in order to bound the perturbations in the candidate basis. So, the norm of the difference between any two small anomaly solutions is bounded by twice the norm of the difference for small anomaly solution to no anomaly solution.

We anticipate that due to the low-rank structure of the candidate basis computed using Algorithm 1, tangential interpolation using stochastic sources and detectors essentially captures the whole space. We justify this below in a slightly different setting.

In the previous section, we show that the transfer function can be equivalently written as

$$\widehat{\mathbf{\Psi}}(\omega, \mathbf{p}) = \widetilde{\mathbf{C}} \left(\frac{\imath\omega}{\nu} \mathbf{I} + \widetilde{\mathbf{A}}(\mathbf{p}) \right)^{-1} \widetilde{\mathbf{B}}, \tag{4.36}$$

where $\widetilde{\mathbf{A}}(\mathbf{p})$ is an SPD matrix that is a small perturbation of the discretized Laplacian.

Next, for a slightly simplified problem (only for the discretized Laplacian), we show that relative changes in the candidate basis, computed using Algorithm 1, remain small from parameter to parameter. For simplicity, we consider $\omega = 0$ in the following derivations.

Based on the finite difference discretization, we note that, for changes in $\mathbf{p}$, the changes in $\widetilde{\mathbf{A}}(\mathbf{p})$ are quite small relative to the magnitude of the matrix coefficients and occur only on the diagonal; see [23]. Since the optimization constraints the updates in $\mathbf{p}$, after few iterations, the changes in $\widetilde{\mathbf{A}}(\mathbf{p})$ are highly localized. Hence, the perturbation in the discretized matrices,

$$\widetilde{\Delta\mathbf{A}} = \widetilde{\mathbf{A}}(\mathbf{p}_{i+1}) - \widetilde{\mathbf{A}}(\mathbf{p}_i), \tag{4.37}$$

is constraint into a few diagonals with mostly zero entries. Under this assumption, there are only few nonzero entries in $\widetilde{\Delta\mathbf{A}}$ and the nonzero diagonal coefficients are $O(h^2)$. Therefore, we rewrite $\widetilde{\Delta\mathbf{A}}$ as follows

$$\widetilde{\Delta\mathbf{A}} = \widetilde{\mathbf{U}} \Xi \widetilde{\mathbf{U}}^T, \tag{4.38}$$

where Ξ is an $k \times k$ diagonal matrix with entries of magnitude $O(h^2)$. $\widetilde{\mathbf{U}}$ is given by

$$\widetilde{\mathbf{U}} = \left[\mathbf{e}_{j_1} \;\; \mathbf{e}_{j_2} \;\; \cdots \;\; \mathbf{e}_{j_k} \right], \tag{4.39}$$

where $\mathbf{e}_j$ denotes the j^{th} Cartesian unit vector and the indices j_k correspond to pixels where absorption has changed.

Next, we theoretically justify that the updates in the global basis are indeed low rank, using the special structure of $\widetilde{\Delta\mathbf{A}}$ in (4.38), and that the updates remain small as $\mathbf{p}$ changes.

Lemma 4.2. *Let $\Delta\mathbf{V} = \mathbf{V}^{(i+1)} - \mathbf{V}^{(i)}$ denote the perturbation in a candidate basis, where $\mathbf{V}^{(i)}$ and $\mathbf{V}^{(i+1)}$ are defined as in Algorithm 1 for $\omega = 0$. Let $\widetilde{\mathbf{A}}(\mathbf{p}_i)$ and $\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})$ correspond to the discretized matrices for parameter vectors $\mathbf{p}_i$ and $\mathbf{p}_{i+1}$, and define the perturbation $\widetilde{\Delta\mathbf{A}} = \widetilde{\mathbf{A}}(\mathbf{p}_{i+1}) - \widetilde{\mathbf{A}}(\mathbf{p}_i)$ as in (4.38). Then,*

$$\frac{||\Delta\mathbf{V}||}{||\mathbf{V}^{(i)}||} \leq ||\widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\mathbf{U}}||\; ||(\mathbf{I}_k + \Xi\widetilde{\mathbf{U}}^T\widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\mathbf{U}})^{-1}||\; ||\Xi\widetilde{\mathbf{U}}^T||, \tag{4.40}$$

where $\mathbf{I}_k$ is the identity matrix of size $k \times k$.

Proof. From Algorithm 1, we get

$$\Delta \mathbf{V} \;=\; \mathbf{V}^{(i+1)} - \mathbf{V}^{(i)} = \widetilde{\mathbf{A}}(\mathbf{p}_{i+1})^{-1}\mathbf{B} - \widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\mathbf{B}. \tag{4.41}$$

Rewriting $\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})$ as

$$\widetilde{\mathbf{A}}(\mathbf{p}_{i+1}) \;=\; \widetilde{\mathbf{A}}(\mathbf{p}_i)(\mathbf{I}_n + \widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\Delta\mathbf{A}}),$$

we obtain

$$\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})^{-1}\mathbf{B} \;=\; (\mathbf{I}_n + \widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\Delta\mathbf{A}})^{-1}\widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\mathbf{B}, \tag{4.42}$$

where $\mathbf{I}_n + \widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\Delta\mathbf{A}})$ is invertible. Substituting (4.38) into (4.42) yields

$$\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})^{-1}\mathbf{B} \;=\; (\mathbf{I}_n + \widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\mathbf{U}}\Xi\widetilde{\mathbf{U}}^T)^{-1}\widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\mathbf{B}. \tag{4.43}$$

Next, using the Sherman-Morrison-Woodbury formula, we get

$$(\mathbf{I}_n + \widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\mathbf{U}}\Xi\widetilde{\mathbf{U}}^T)^{-1} = \mathbf{I}_n - \widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\mathbf{U}}(\mathbf{I}_k + \Xi\widetilde{\mathbf{U}}^T\widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\mathbf{U}})^{-1}\Xi\widetilde{\mathbf{U}}^T. \tag{4.44}$$

Substitute (4.44) into (4.43) to get

$$\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})^{-1}\mathbf{B} = \left[\mathbf{I}_n - \widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\mathbf{U}}(\mathbf{I}_k + \Xi\widetilde{\mathbf{U}}^T\widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\mathbf{U}})^{-1}\Xi\widetilde{\mathbf{U}}^T\right]\widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\mathbf{B}. \tag{4.45}$$

From (4.41) and (4.45), we derive

$$\begin{aligned}\mathbf{V}^{(i+1)} &= \left[\mathbf{I}_n - \widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\mathbf{U}}(\mathbf{I}_k + \Xi\widetilde{\mathbf{U}}^T\widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\mathbf{U}})^{-1}\Xi\widetilde{\mathbf{U}}^T\right]\mathbf{V}^{(i)},\\ &= \mathbf{V}^{(i)} - \left[\widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\mathbf{U}}(\mathbf{I}_k + \Xi\widetilde{\mathbf{U}}^T\widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\mathbf{U}})^{-1}\Xi\widetilde{\mathbf{U}}^T\right]\mathbf{V}^{(i)},\end{aligned}$$

and hence

$$\Delta \mathbf{V} \;=\; -\left[\widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\mathbf{U}}(\mathbf{I}_k + \Xi\widetilde{\mathbf{U}}^T\widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\mathbf{U}})^{-1}\Xi\widetilde{\mathbf{U}}^T\right]\mathbf{V}^{(i)}, \tag{4.46}$$

where $\widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\mathbf{U}} \in \mathbb{R}^{n\times k}$. Taking norms and using the submultiplicative property in (4.46), we obtain

$$||\Delta\mathbf{V}|| \;\leq\; ||\widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\mathbf{U}}(\mathbf{I}_k + \Xi\widetilde{\mathbf{U}}^T\widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\mathbf{U}})^{-1}\Xi\widetilde{\mathbf{U}}^T||\,||\mathbf{V}^{(i)}||.$$

Then, the relative change in $\mathbf{V}^{(i)}$ is bounded by

$$\frac{||\Delta\mathbf{V}||}{||\mathbf{V}^{(i)}||} \;\leq\; ||\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})^{-1}\widetilde{\mathbf{U}}(\mathbf{I}_k + \Xi\widetilde{\mathbf{U}}^T\widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\mathbf{U}})^{-1}\Xi\widetilde{\mathbf{U}}^T||,$$

which gives the desired inequality

$$\frac{||\Delta\mathbf{V}||}{||\mathbf{V}^{(i)}||} \;\leq\; ||\widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\mathbf{U}}||\,||(\mathbf{I}_k + \Xi\widetilde{\mathbf{U}}^T\widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\mathbf{U}})^{-1}||\,||\Xi\widetilde{\mathbf{U}}^T||. \tag{4.47}$$

□

Remark 4.3. Equation (4.46) shows that the changes in candidate basis lie in the space Range$(\widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\mathbf{U}})$. This confirms our earlier claim that the updates in the global basis are indeed low rank.

Next, we give a bound on (4.46) for a fixed k.

Lemma 4.4. *Let $\Delta\mathbf{V} = \mathbf{V}^{(i+1)} - \mathbf{V}^{(i)}$ denote the perturbation in the candidate basis, where $\mathbf{V}^{(i+1)}$ and $\mathbf{V}^{(i)}$ are defined as in Algorithm 1. Then,*

$$\frac{\|\Delta\mathbf{V}\|}{\|\mathbf{V}^{(i)}\|} \leq Ch, \tag{4.48}$$

where C is a constant and h is the mesh width in each direction.

Proof. To bound $\dfrac{||\Delta\mathbf{V}||}{||\mathbf{V}^{(i)}||}$, we examine the right-hand side of the inequality in (4.47),

$$\frac{||\Delta\mathbf{V}||}{||\mathbf{V}^{(i)}||} \leq ||\widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\mathbf{U}}||\,||(\mathbf{I}_k + \Xi\widetilde{\mathbf{U}}^T\widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\mathbf{U}})^{-1}||\,||\Xi\widetilde{\mathbf{U}}^T||. \tag{4.49}$$

We start with bounding $\|\widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\mathbf{U}}\|$, where $\widetilde{\mathbf{U}}$ is defined in (4.39). In the DOT setting, k is a modest number, and we have

$$\|\widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\mathbf{U}}\| \leq k \max_{e_j} \|\widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\mathbf{e}_j\|. \tag{4.50}$$

We assume $k = 1$ for simplicity and focus on $||\widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\mathbf{e}_j||$. Since $\widetilde{\mathbf{A}}(\mathbf{p}_i)$ is SPD, we express $\widetilde{\mathbf{A}}(\mathbf{p}_i)$ in terms of its eigendecomposition,

$$\widetilde{\mathbf{A}}(\mathbf{p}_i) = \widetilde{\mathbf{\Phi}}\mathbf{\Lambda}\widetilde{\mathbf{\Phi}}^T. \tag{4.51}$$

In the following, we need the eigenvalues and eigenvectors of the 2D discretized Laplacian, which can be found in several textbooks [39]. In our application, we assume the mesh width h is the same in x and y direction. Let k_x and k_y be the wave numbers in the x and y direction. Labeling the eigenvectors by wave number, the components of the (k_x, k_y) eigenvector are given by

$$\phi^{k_xk_y}_{i_xi_y} = \sin(i_xk_x\pi h)\sin(i_yk_y\pi h),\ \ i_x,\ i_y = 1, \cdots, K-1, \tag{4.52}$$

and the components of the normalized eigenvector are given by

$$\widetilde{\phi}^{k_xk_y}_{i_xi_y} = \nu^{k_xk_y}\sin(i_xk_x\pi h)\sin(i_yk_y\pi h),\ \ i_x,\ i_y = 1, \cdots, K-1, \tag{4.53}$$

where $\nu^{k_xk_y}$ is the scaling factor so that $\widetilde{\mathbf{\Phi}}$ is unitary. The corresponding eigenvalues are

$$\lambda^{k_x,k_y} = 4\left[\sin^2\left(\frac{k_x\pi h}{2}\right) + \sin^2\left(\frac{k_y\pi h}{2}\right)\right]. \tag{4.54}$$

Using the eigendecomposition of $\widetilde{\mathbf{A}}(\mathbf{p}_i)$, we get

$$||\widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\mathbf{e}_j|| = ||\widetilde{\mathbf{\Phi}}\mathbf{\Lambda}^{-1}\widetilde{\mathbf{\Phi}}^T\mathbf{e}_j||. \tag{4.55}$$

Since $\widetilde{\mathbf{\Phi}}$ is unitary,

$$||\widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\mathbf{e}_j|| = ||\mathbf{\Lambda}^{-1}\widetilde{\mathbf{\Phi}}^T\mathbf{e}_j||. \tag{4.56}$$

Using (4.52), (4.54) and (4.56) gives

$$||\mathbf{\Lambda}^{-1}\widetilde{\mathbf{\Phi}}^T\mathbf{e}_j||^2 = \sum_{k_y=1}^{K-1}\sum_{k_x=1}^{K-1}\left[\frac{\nu^{k_x,k_y}\sin(jk_x\pi h)\sin(jk_y\pi h)}{4\left[\sin^2\left(\frac{k_x\pi h}{2}\right)+\sin^2\left(\frac{k_y\pi h}{2}\right)\right]}\right]^2. \tag{4.57}$$

Using the fact that the value of the scaling factor, ν^{k_x,k_y}, is $2h$, see [28, pg. 400], we write

$$\begin{aligned}||\mathbf{\Lambda}^{-1}\widetilde{\mathbf{\Phi}}^T\mathbf{e}_j||^2 &= h^2\sum_{k_y=1}^{K-1}\sum_{k_x=1}^{K-1}\left[\frac{\sin(jk_x\pi h)\sin(jk_y\pi h)}{\sin^2\left(\frac{k_x\pi h}{2}\right)+\sin^2\left(\frac{k_y\pi h}{2}\right)}\right]^2 \\ &\leq h^2\sum_{k_y=1}^{K-1}\sum_{k_x=1}^{K-1}\left[\frac{\frac{\pi^2}{4}}{\frac{\pi^2}{4}\left(\sin^2\left(\frac{k_x\pi h}{2}\right)+\sin^2\left(\frac{k_y\pi h}{2}\right)\right)}\right]^2.\end{aligned} \tag{4.58}$$

Since $\beta^2 \leq \frac{\pi^2}{4}\sin^2(\beta)$ for $0 \leq \beta \leq \frac{\pi}{2}$ in (4.58), we get

$$\begin{aligned}||\mathbf{\Lambda}^{-1}\widetilde{\mathbf{\Phi}}^T\mathbf{e}_j||^2 &\leq h^2\sum_{k_y=1}^{K-1}\sum_{k_x=1}^{K-1}\left[\frac{\frac{\pi^2}{4}}{\frac{\pi^2h^2}{4}\left(k_x^2+k_y^2\right)}\right]^2 \\ &= \frac{1}{h^2}\sum_{k_y=1}^{K-1}\sum_{k_x=1}^{K-1}\left[\frac{1}{\left(k_x^2+k_y^2\right)}\right]^2,\end{aligned} \tag{4.59}$$

where the double sum term is $O(1)$. Then, we obtain the desired bound for (4.56)

$$||\mathbf{\Lambda}^{-1}\widetilde{\mathbf{\Phi}}^T\mathbf{e}_j|| \;\leq\; c_1\frac{1}{h}, \tag{4.60}$$

for a modest constant c_1. Since $\|\Xi\widetilde{\mathbf{U}}^T\widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\mathbf{U}}\|$ is $O(h)$, it is now straightforward to show that

$$||(\mathbf{I}_k + \Xi\widetilde{\mathbf{U}}^T\widetilde{\mathbf{A}}(\mathbf{p}_i)^{-1}\widetilde{\mathbf{U}})^{-1}|| \leq c_2, \tag{4.61}$$

where c_2 is a modest constant. Finally, we have

$$|||\Xi\widetilde{\mathbf{U}}^T|| \leq c_3h^2 \tag{4.62}$$

from the definition of Ξ (4.38). Combining (4.60), (4.62), and (4.62), we obtain the desired bound

$$\frac{||\Delta \mathbf{V}||}{||\mathbf{V}^{(i)}||} \leq Ch, \tag{4.63}$$

where $C = c_1 c_2 c_3$. □

Next, we consider two fundamental quantitites describing the behavior of dynamical systems as in (4.36) are the so-called reachability and observability Gramians; see, e.g.,[3].

4.4.3 Perturbations in Gramians

Given the system (4.36) and stable $-\widetilde{\mathbf{A}}(\mathbf{p})$, then $\mathbf{P}$ is the unique positive semi-definite solution of the Lyapunov equation

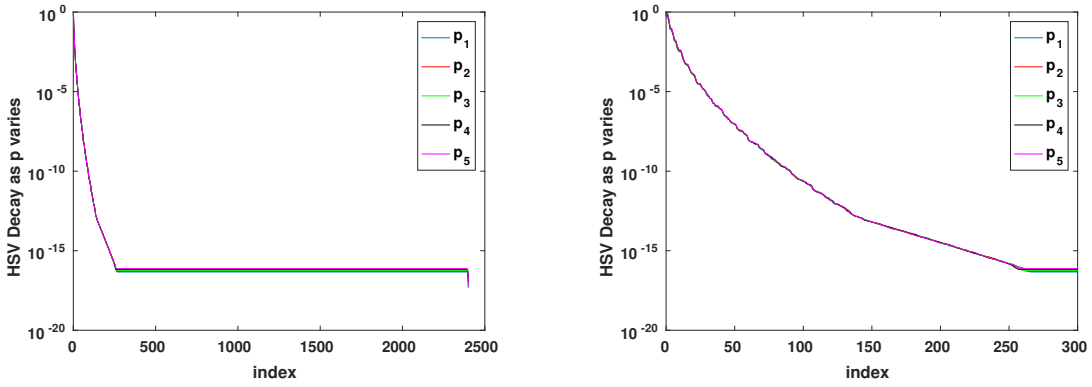
$$\widetilde{\mathbf{A}}(\mathbf{p})\mathbf{P} + \mathbf{P}\widetilde{\mathbf{A}}(\mathbf{p}) = -\mathbf{B}\mathbf{B}^T. \tag{4.64}$$

$\mathbf{P}$ is called the reachability Gramian of (4.36) and provides a measure of which states in the discretized model are unimportant (in terms of reachability) and can be truncated. The dominant eigenspace of $\mathbf{P}$ determines which degrees of freedom to keep in the reduced model. The observability Gramian $\mathbf{T}$ is defined similarly [3]

$$\widetilde{\mathbf{A}}(\mathbf{p})\mathbf{T} + \mathbf{T}\widetilde{\mathbf{A}}(\mathbf{p}) = -\mathbf{C}^T\mathbf{C}. \tag{4.65}$$

The square roots of the eigenvalues of the product of Gramians $\mathbf{PT}$ are called the Hankel singular values. The singular values determine the required order of the ROM. Hence, we show the fast decay of Hankel singular values for a small 2D test problem in Figure 4.4.

In the following, we focus on the reachability Gramian; the observability Gramians can be treated analogously. If the reachability Gramian $\mathbf{P}$ changes only slightly from one parameter point to another, then the underlying dynamics do not vary drastically from parameter to parameter. Note that in our application, the parameter vectors remain in a relatively small range. Therefore, the global interpolatory model reduction subspaces, constructed in Algorithm 2 with an initial sampling (see Section 4.5), will remain good/effective model reduction spaces. In Figure 4.5, we give the cosines of the canonical angles. As can be seen, the angles remain small between the global basis $\widetilde{\mathbf{V}}$ and new reduction spaces $\mathbf{A}(\mathbf{p}_s)^{-1}\mathbf{B}$ over the course of optimization. Thus, even though Theorem 4.1 only guarantees matching the cost function and the Jacobian at the sample points, the reduced parametric transfer function is expected to interpolate the full transfer function approximately, yet accurately enough, even for the other parameter points that the optimization algorithm visits during the inversion process. Next, for a slightly simplified problem (only for the discretized Laplacian), we theoretically justify that changes in the reachability Gramian from parameter to parameter remain small.



(a) (b)

Figure 4.4: Decay of Hankel Singular Values for a simple 2D test problem. The anomaly is on 50×50 with 15 sources and 15 detectors. We use 25 basis functions and only the zero frequency. (a) Decay of Hankel singular values of the system over the first five distinct $\mathbf{p}$ vectors. (b) Decay of truncated Hankel singular values of the system in (a).

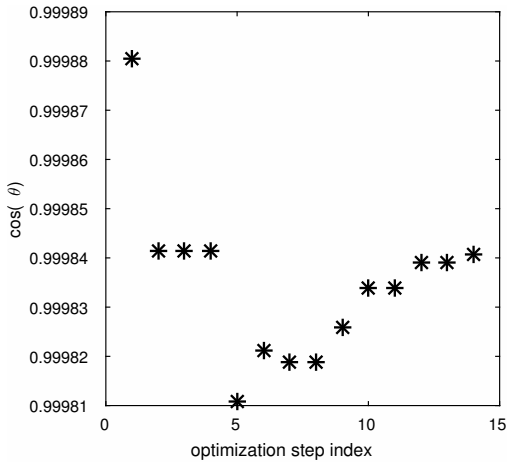


Figure 4.5: Evolution of the subspace gap between the global basis, $\widetilde{\mathbf{V}}$ and new reduction spaces $\mathbf{A}(\mathbf{p}_s)^{-1}\mathbf{B}$ over the course of optimization. The anomaly is on 201×201 mesh with 32 sources and 32 detectors. We use 25 basis functions and only the zero frequency.

Lemma 4.5. *Let $\mathbf{P}_i$ be the solution to the Lyapunov equation*

$$\widetilde{\mathbf{A}}(\mathbf{p}_i)\mathbf{P}_i + \mathbf{P}_i\widetilde{\mathbf{A}}(\mathbf{p}_i) = -\mathbf{B}\mathbf{B}^T, \tag{4.66}$$

and let $\mathbf{P}_{i+1}$ be the solution to the Lyapunov equation

$$\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})\mathbf{P}_{i+1} + \mathbf{P}_{i+1}\widetilde{\mathbf{A}}(\mathbf{p}_{i+1}) = -\mathbf{B}\mathbf{B}^T, \tag{4.67}$$

where $\mathbf{P}_i$ and $\mathbf{P}_{i+1}$ correspond to the Lyapunov solutions for parameter values $\mathbf{p}_i$ and $\mathbf{p}_{i+1}$, respectively. Define the perturbation in the solution to the Lyapunov equation as $\Delta\mathbf{P} = \mathbf{P}_{i+1} - \mathbf{P}_i$. Then,

$$\frac{\|\Delta\mathbf{P}\|}{\|\mathbf{P}_i\|} \leq 2\|\mathbf{Q}\|, \tag{4.68}$$

where $\mathbf{Q}$ is the unique positive semi-definite solution of the Lyapunov equation,

$$\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})\mathbf{Q} + \mathbf{Q}\widetilde{\mathbf{A}}(\mathbf{p}_{i+1}) = \widetilde{\Delta\mathbf{A}}. \tag{4.69}$$

where $\widetilde{\Delta\mathbf{A}}$ is defined as in (4.37).

Proof. Substituting $\mathbf{P}_{i+1} = \mathbf{P}_i + \Delta\mathbf{P}$ in (4.67), and subtracting (4.67) from (4.66), we obtain that $\Delta\mathbf{P}$ solves

$$\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})\Delta\mathbf{P} + \Delta\mathbf{P}\widetilde{\mathbf{A}}(\mathbf{p}_{i+1}) + \widetilde{\Delta\mathbf{A}}\mathbf{P}_i + \mathbf{P}_i\widetilde{\Delta\mathbf{A}} = 0, \tag{4.70}$$

where $\widetilde{\Delta\mathbf{A}}$ is defined as in (4.37). Since $-\widetilde{\mathbf{A}}(\mathbf{p}_i)$ is a stable matrix, we can express the solution, $\Delta\mathbf{P}$, as an integral over matrix exponentials,

$$\Delta\mathbf{P} = \int_0^\infty e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t}\left(\widetilde{\Delta\mathbf{A}}\mathbf{P}_i + \mathbf{P}_i\widetilde{\Delta\mathbf{A}}\right)e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t}\,dt. \tag{4.71}$$

Next, we find an upper bound for $\|\Delta\mathbf{P}\|$

$$\begin{aligned}\|\Delta\mathbf{P}\|^2 =& \text{trace}\left(\Delta\mathbf{P}^T\Delta\mathbf{P}\right)\\ =& \text{trace}\Big(\int_0^\infty\int_0^\infty e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t}\left(\widetilde{\Delta\mathbf{A}}\mathbf{P}_i + \mathbf{P}_i\widetilde{\Delta\mathbf{A}}\right)e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t}e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s}\cdot\\ &\left(\widetilde{\Delta\mathbf{A}}\mathbf{P}_i + \mathbf{P}_i\widetilde{\Delta\mathbf{A}}\right)e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s}\,ds\,dt\Big),\end{aligned} \tag{4.72}$$

which can be split into the following sum

$$\begin{aligned}
\|\Delta\mathbf{P}\|^2 &= \int_0^\infty\int_0^\infty \text{trace}\left(e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t}\widetilde{\Delta\mathbf{A}}\mathbf{P}_i e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t}e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s}\widetilde{\Delta\mathbf{A}}\mathbf{P}_i e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s}\right) ds\, dt \\
&+ \int_0^\infty\int_0^\infty \text{trace}\left(e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t}\widetilde{\Delta\mathbf{A}}\mathbf{P}_i e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t}e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s}\mathbf{P}_i\widetilde{\Delta\mathbf{A}} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s}\right) ds\, dt \\
&+ \int_0^\infty\int_0^\infty \text{trace}\left(e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t}\mathbf{P}_i\widetilde{\Delta\mathbf{A}} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t}e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s}\widetilde{\Delta\mathbf{A}}\mathbf{P}_i e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s}\right) ds\, dt \\
&+ \int_0^\infty\int_0^\infty \text{trace}\left(e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t}\mathbf{P}_i\widetilde{\Delta\mathbf{A}} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t}e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s}\mathbf{P}_i\widetilde{\Delta\mathbf{A}} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s}\right) ds\, dt \\
&= \mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3 + \mathcal{I}_4, \qquad (4.73)
\end{aligned}$$

where we use the fact that the trace and integral operators commute together with the linearity of the trace. To bound each integral, we repeatedly use the cyclic property of the trace and the fact that $\text{trace}(\mathbf{KL}) \leq \|\mathbf{L}\|\ \text{trace}(\mathbf{K})$, where $\mathbf{K}$ and $\mathbf{L}$ are two conforming matrices.

In the following, we bound the first integral as follows

$$
\begin{aligned}
\mathcal{I}_1 &= \int_0^\infty \int_0^\infty \operatorname{trace}\left(e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} \widetilde{\Delta\mathbf{A}} \mathbf{P}_i e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s} \widetilde{\Delta\mathbf{A}} \mathbf{P}_i e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s} \right) ds\, dt \\
&= \int_0^\infty \int_0^\infty \operatorname{trace}\left(e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s} \widetilde{\Delta\mathbf{A}} \mathbf{P}_i e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} \widetilde{\Delta\mathbf{A}} \mathbf{P}_i \right) ds\, dt \\
&\leq \|\mathbf{P}_i\| \int_0^\infty \int_0^\infty \operatorname{trace}\left(e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s} \widetilde{\Delta\mathbf{A}} \mathbf{P}_i e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} \widetilde{\Delta\mathbf{A}} \right) ds\, dt \\
&= \|\mathbf{P}_i\| \int_0^\infty \int_0^\infty \operatorname{trace}\left(e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} \widetilde{\Delta\mathbf{A}} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s} \widetilde{\Delta\mathbf{A}} \mathbf{P}_i \right) ds\, dt \\
&\leq \|\mathbf{P}_i\|^2 \int_0^\infty \int_0^\infty \operatorname{trace}\left(e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} \widetilde{\Delta\mathbf{A}} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s} \widetilde{\Delta\mathbf{A}} \right) ds\, dt \\
&= \|\mathbf{P}_i\|^2 \int_0^\infty \int_0^\infty \operatorname{trace}\left(e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} \widetilde{\Delta\mathbf{A}} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s} \widetilde{\Delta\mathbf{A}} \right) ds\, dt \\
&= \|\mathbf{P}_i\|^2 \int_0^\infty \int_0^\infty \operatorname{trace}\left(e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} \widetilde{\Delta\mathbf{A}} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s} \widetilde{\Delta\mathbf{A}} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s} \right) ds\, dt \\
&= \|\mathbf{P}_i\|^2 \operatorname{trace}\left(\underbrace{\int_0^\infty e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} \widetilde{\Delta\mathbf{A}} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} dt}_{\mathbf{Q}} \underbrace{\int_0^\infty e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s} \widetilde{\Delta\mathbf{A}} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s} ds}_{\mathbf{Q}} \right).
\end{aligned}
\tag{4.74}
$$

Hence, we obtain

$$
\mathcal{I}_1 \leq \|\mathbf{P}_i\|^2 \|\mathbf{Q}\|^2, \tag{4.75}
$$

where $\mathbf{Q}$ is the unique positive semi-definite solution to the Lyapunov equation (4.69) and the integral representation of $\mathbf{Q}$ is given by

$$
\mathbf{Q} = \int_0^\infty e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} \, \widetilde{\Delta\mathbf{A}} \, e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} \, dt. \tag{4.76}
$$

Analogous proceeding for the other terms, we obtain

$$
\|\Delta\mathbf{P}\|^2 \leq 4 \|\mathbf{Q}\|^2 \|\mathbf{P}_i\|^2. \tag{4.77}
$$

From which we obtain

$$\frac{\|\Delta \mathbf{P}\|}{\|\mathbf{P}_i\|} \leq 2\|\mathbf{Q}\|. \tag{4.78}$$

□

The details on bounding other integrals are given in Section A.1.

Next, we show that this bound remains small, justifying the effectiveness of the global basis $\widetilde{\mathbf{V}}_r$, even for the other parameter vectors that the optimization algorithm visits during the inversion process.

Lemma 4.6. *Let* $\mathbf{Q}$ *be the unique positive semi-definite solution of the Lyapunov equation*

$$\widetilde{\mathbf{A}}(\mathbf{p}_{s+1})\mathbf{Q} + \mathbf{Q}\widetilde{\mathbf{A}}(\mathbf{p}_{s+1}) = \widetilde{\Delta\mathbf{A}}, \tag{4.79}$$

where $\widetilde{\mathbf{A}}(\mathbf{p}_{s+1})$ *and* $\widetilde{\Delta\mathbf{A}}$ *are defined as above. Then,*

$$\|\mathbf{Q}\| \leq Ch^2. \tag{4.80}$$

Proof. Let $\widetilde{\mathbf{A}}(\mathbf{p}_{s+1}) = \widetilde{\mathbf{\Phi}}\mathbf{\Lambda}\widetilde{\mathbf{\Phi}}$ such that $\widetilde{\mathbf{\Phi}}$ is a symmetric unitary matrix whose i^{th} column is the eigenvector $\widetilde{\boldsymbol{\phi}}_i$ defined in (4.53), and $\mathbf{\Lambda}$ is the diagonal matrix whose diagonal elements are the corresponding eigenvalues of $\widetilde{\mathbf{A}}(\mathbf{p}_{s+1})$. We define $\mathbf{\Gamma} = \widetilde{\mathbf{\Phi}}\mathbf{Q}\widetilde{\mathbf{\Phi}}$. Since $\widetilde{\mathbf{\Phi}}$ is unitary, we have

$$\|\mathbf{Q}\| = \|\widetilde{\mathbf{\Phi}}\mathbf{\Gamma}\widetilde{\mathbf{\Phi}}\| = \|\mathbf{\Gamma}\| = \left(\sum_{i=1}^{K-1}\sum_{j=1}^{K-1}\gamma_{ij}^2\right)^{\frac{1}{2}}. \tag{4.81}$$

Expanding $\mathbf{Q}$ and $\widetilde{\Delta\mathbf{A}}$ in the eigenvector basis of Lyapunov operator gives

$$\mathbf{Q} = \sum_{i=1}^{K-1}\sum_{j=1}^{K-1}\gamma_{ij}\widetilde{\boldsymbol{\phi}}_i\widetilde{\boldsymbol{\phi}}_j^T, \tag{4.82}$$

and

$$\widetilde{\Delta\mathbf{A}} = \sum_{i=1}^{K-1}\sum_{j=1}^{K-1}\alpha_{ij}\widetilde{\boldsymbol{\phi}}_i\widetilde{\boldsymbol{\phi}}_j^T. \tag{4.83}$$

Then, we can expand (4.79) in the eigenvector basis of the Lyapunov operator

$$\sum_{i=1}^{K-1}\sum_{j=1}^{K-1}\gamma_{ij}(\lambda_i+\lambda_j)\widetilde{\boldsymbol{\phi}}_i\widetilde{\boldsymbol{\phi}}_j^T = \sum_{i=1}^{K-1}\sum_{j=1}^{K-1}\alpha_{ij}\widetilde{\boldsymbol{\phi}}_i\widetilde{\boldsymbol{\phi}}_j^T, \tag{4.84}$$

where λ_i and λ_j are the eigenvalues of $\widetilde{\mathbf{A}}(\mathbf{p}_{s+1})$ corresponding to $\widetilde{\boldsymbol{\phi}}_i$ and $\widetilde{\boldsymbol{\phi}}_j$, respectively. Then, $(\lambda_i + \lambda_j)$ are the corresponding eigenvalues of the Lyapunov operator. Matching terms, we get for γ_{ij}

$$\gamma_{ij} = \frac{\alpha_{ij}}{\lambda_i + \lambda_j}. \tag{4.85}$$

Here, we focus on α_{ij}. Recall that $\widetilde{\Delta\mathbf{A}}$ can be factored as

$$\widetilde{\Delta\mathbf{A}} = \widetilde{\mathbf{U}}\Xi\widetilde{\mathbf{U}}^T, \tag{4.86}$$

where $\Xi \in \mathbb{R}^{k\times k}$ is a diagonal matrix with entries of magnitude $O(h^2)$ and $\widetilde{\mathbf{U}} \in \mathbb{R}^{n\times k}$ is defined as in (4.39). Hence, we expand $\widetilde{\Delta\mathbf{A}}$ in (4.83) in the eigenvector basis as follows

$$\widetilde{\Delta\mathbf{A}} = \sum_{m=1}^{k} \xi_{\ell_m}\, \mathbf{e}_{\ell_m}\mathbf{e}_{\ell_m}^T. \tag{4.87}$$

Since there are only few localized changes in $\widetilde{\Delta\mathbf{A}}$, we focus on $k = 1$ and $\widetilde{\mathbf{U}} = \mathbf{e}_\ell$ for some ℓ, as the case before. We rewrite $\mathbf{e}_\ell$ in terms of $\widetilde{\mathbf{\Phi}}$ as follows

$$\mathbf{e}_\ell = \widetilde{\mathbf{\Phi}}\mathbf{y}_\ell, \tag{4.88}$$

where $\mathbf{y}_\ell = \widetilde{\mathbf{\Phi}}^T\mathbf{e}_\ell$. Then, we obtain

$$\mathbf{e}_\ell\mathbf{e}_\ell^T = \widetilde{\mathbf{\Phi}}\mathbf{y}_\ell\mathbf{y}_\ell^T\widetilde{\mathbf{\Phi}}^T. \tag{4.89}$$

Substituting (4.89) into (4.87) for $k = 1$, we obtain

$$\widetilde{\Delta\mathbf{A}} = \sum_{i=1}^{K-1}\sum_{j=1}^{K-1} \alpha_{ij}^{(1)}\widetilde{\boldsymbol{\phi}}_i\widetilde{\boldsymbol{\phi}}_j^T = \sum_{i=1}^{K-1}\sum_{j=1}^{K-1} \xi_\ell\, \mathbf{y}_{i\ell}\mathbf{y}_{j\ell}\, \widetilde{\boldsymbol{\phi}}_i\widetilde{\boldsymbol{\phi}}_j^T, \tag{4.90}$$

which leads to

$$\alpha_{ij}^{(1)} = \xi_\ell\, \mathbf{y}_{i\ell}\mathbf{y}_{j\ell}, \tag{4.91}$$

where $\mathbf{y}_{i\ell} = \widetilde{\boldsymbol{\phi}}_i^T\mathbf{e}_\ell$ and $\mathbf{y}_{j\ell} = \widetilde{\boldsymbol{\phi}}_j^T\mathbf{e}_\ell$. In the subsequent discussion, we need to distinguish the individual eigenvectors. Hence, we identify the linear index i with the wave number (k_x, k_y), and the linear index j with the wave number (m_x, m_y). Using (4.53) and (4.54), we can rewrite (4.85) for the 2D Laplacian as follows

$$\begin{aligned}
\gamma_{k_x,k_y,m_x,m_y}^{(1)} &= \frac{\alpha_{k_x,k_y,m_x,m_y}^{(1)}}{\lambda^{k_x,k_y} + \lambda^{m_x,m_y}} \\
&= \frac{(\nu^{k_xk_y})^2\xi_\ell \sin(\ell k_x\pi h)\sin(\ell k_y\pi h)\sin(\ell m_x\pi h)\sin(\ell m_y\pi h)}{4\left[\sin^2\left(\frac{k_x\pi h}{2}\right) + \sin^2\left(\frac{k_y\pi h}{2}\right)\right] + 4\left[\sin^2\left(\frac{m_x\pi h}{2}\right) + \sin^2\left(\frac{m_y\pi h}{2}\right)\right]},
\end{aligned} \tag{4.92}$$

where the value of the scaling factor, ν^{k_x,k_y}, is $2h$, see [28, pg. 400]. Substituting (4.92) into (4.81), we obtain

$$\begin{aligned}\|\mathbf{Q}\|^2 &= \sum_{k_x=1}^{K-1}\sum_{k_y=1}^{K-1}\sum_{m_x=1}^{K-1}\sum_{m_y=1}^{K-1}\left(\frac{(2h)^2\xi_\ell \sin(\ell k_x\pi h)\sin(\ell k_y\pi h)\sin(\ell m_x\pi h)\sin(\ell m_y\pi h)}{4\left[\sin^2\left(\frac{k_x\pi h}{2}\right)+\sin^2\left(\frac{k_y\pi h}{2}\right)\right]+4\left[\sin^2\left(\frac{m_x\pi h}{2}\right)+\sin^2\left(\frac{m_y\pi h}{2}\right)\right]}\right)^2 \\ &= h^4\xi_\ell^2\sum_{k_x=1}^{K-1}\sum_{k_y=1}^{K-1}\sum_{m_x=1}^{K-1}\sum_{m_y=1}^{K-1}\left(\frac{\sin(\ell k_x\pi h)\sin(\ell k_y\pi h)\sin(\ell m_x\pi h)\sin(\ell m_y\pi h)}{\sin^2\left(\frac{k_x\pi h}{2}\right)+\sin^2\left(\frac{k_y\pi h}{2}\right)+\sin^2\left(\frac{m_x\pi h}{2}\right)+\sin^2\left(\frac{m_y\pi h}{2}\right)}\right)^2 \\ &\le h^4\xi_\ell^2\sum_{k_x=1}^{K-1}\sum_{k_y=1}^{K-1}\sum_{m_x=1}^{K-1}\sum_{m_y=1}^{K-1}\left(\frac{\frac{\pi^2}{4}}{\frac{\pi^2}{4}\left(\sin^2\left(\frac{k_x\pi h}{2}\right)+\sin^2\left(\frac{k_y\pi h}{2}\right)+\sin^2\left(\frac{m_x\pi h}{2}\right)+\sin^2\left(\frac{m_y\pi h}{2}\right)\right)}\right)^2\end{aligned} \tag{4.93}$$

Using $\beta^2 \le \frac{\pi^2}{4}\sin^2(\beta)$ for $0 \le \beta \le \frac{\pi}{2}$ in (4.93), we get

$$\begin{aligned}\|\mathbf{Q}\|^2 &\le h^4\xi_\ell^2\sum_{k_x=1}^{K-1}\sum_{k_y=1}^{K-1}\sum_{m_x=1}^{K-1}\sum_{m_y=1}^{K-1}\left(\frac{\frac{\pi^2}{4}}{\frac{\pi^2h^2}{4}\left(k_x^2+k_y^2+m_x^2+m_y^2\right)}\right)^2 \\ &= \xi_\ell^2\sum_{k_x=1}^{K-1}\sum_{k_y=1}^{K-1}\sum_{m_x=1}^{K-1}\sum_{m_y=1}^{K-1}\left(\frac{1}{\left(k_x^2+k_y^2+m_x^2+m_y^2\right)}\right)^2,\end{aligned} \tag{4.94}$$

where ξ_ℓ is $O(h^2)$ and the quadruple sum term is $O(1)$. Then, we obtain the desired bound

$$\|\mathbf{Q}\| \le Ch^2, \tag{4.95}$$

□

4.5 Numerical Results

We present three proof-of-concept experiments for 2D and 3D DOT inversion, inverting only for absorption and using multiple frequencies. For comparison, the experimental set up we use is that described in [23], in which model reduction was proposed as an alternative approach to reduce the cost of the inversion process in DOT. For each test case, we construct anomalies in the pixel basis, and we add a small normally distributed random heterogeneity to both the background and to the anomaly to make the medium inhomogeneous. This ensures a modest mismatch between the exact image and the representation we use to reconstruct the image. We also add $\delta = 0.1\%$ white noise to the simulated data in each experiment.

PaLS [1] and TREGS [24] are used to reconstruct the absorption images. We use the first n_k iterations of the optimization using the exact objective function to obtain parameter sample points. We run the FOM and the standard ROM using all sources and detectors using

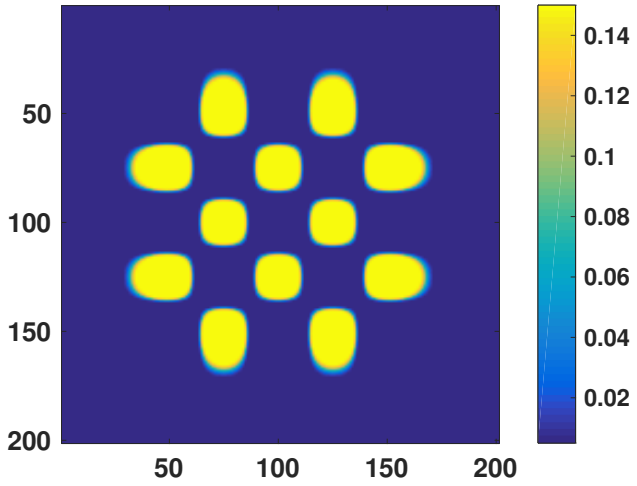


Figure 4.6: Initial configuration with 25 basis functions arranged in a 5×5 grid where 12 basis functions have a positive expansion factors (visible) and 13 basis functions have negative expansion factors (invisible).

Algorithm 1 for comparison. We remove small nonzero singular values of the candidate basis below a modest tolerance to obtain the global basis. For nonzero frequencies, we split the real and complex solutions to work in real arithmetic.

We use ℓ_s number of simultaneous random sources and ℓ_d number of simultaneous random detectors where $\ell_s = \ell_d$. As discussed in Section 4.3, we use these stochastic sources and detectors as the right and left tangential interpolation directions to compute the global basis, see Algorithm 2. We use the same tolerance to remove small nonzero singular values of the candidate basis. We stop the optimization when the residual norm falls below 1.1 times the noise level.

4.5.1 2D Experiment

Example 1. The mesh is 201×201 with 32 sources at the top and 32 detectors on the bottom. We use 4 parameter sample points to build the ROM. We use 2 frequencies with 25 compactly supported radial basis functions (CSRBFs) to reconstruct the anomaly, leading to 100 parameters. The absorption image for the initial set of parameters is given in Figure 4.6 where 12 basis functions have a positive expansion coefficient (visible as high absorption regions) and 13 basis functions have a negative expansion coefficient (invisible).

The true absorption image for Example 2 is given in Figure 4.7a. We also include the reconstruction result using the FOM in Figure 4.7b. Moreover, the reconstruction using the standard ROM with all sources and detectors is given in Figure 4.7c. In Table 4.1, we give the number of large linear systems to be solved in each method for comparison. Our approach reduces the large solver cost by about *a factor 10*, while obtaining similar quality

	FOM	ROM all srcs/dets	ROM stoch. srcs/dets
Experiment 1	1856 large	512 large 1792 small (r=242)	192 large 2112 small (r=212)
Experiment 2	47700 large	5400 large 16875 small (r=1368)	1200 large 17550 small (r=1294)
Experiment 3	47700 large	5400 large 21600 small (r=1228)	1200 large 27000 small (r=1184)

Table 4.1: Number of Large Linear Solves for 2D and 3D Experiments

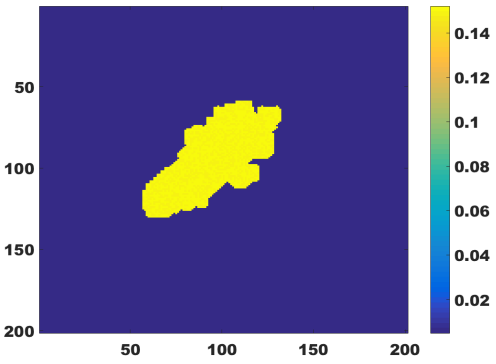
reconstruction results, see Figure 4.7d. For larger numbers of sources and detectors and using multiple frequencies, the computational savings will be significant, see, e.g. the 3D experiment in Section 4.5.2.

4.5.2 3D Experiments

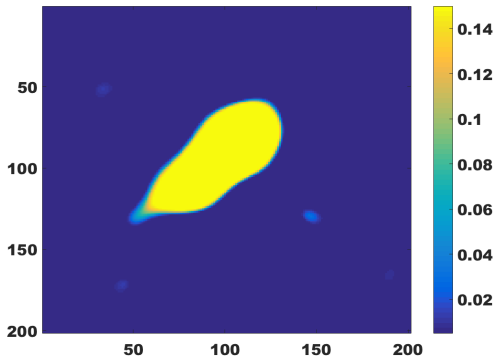
Example 2. The mesh is $32 \times 32 \times 32$, which gives $32,768$ degrees of freedom in the forward model (2.2). The model has 225 sources at the top and 225 detectors on the bottom, and we use 3 frequencies. We use 4 parameter sample points to construct the ROM. We 27 CSRBFs to reconstruct the anomaly, leading to 135 parameters. The initial absorption image is given in Figure 4.8 where 13 basis functions have a positive expansion coefficient (visible as high absorption regions) and 14 basis functions have a negative expansion coefficient (invisible). In our approach, we use 50 stochastic sources and detectors to construct ROM basis, $\ell_s = \ell_d = 50$.

The true absorption image for Example 2 is given in Figure 4.9a. Using the FOM, the optimization algorithm solves $18,900$ linear systems of dimension $32,768$ to reconstruct the absorption image. The reconstruction result using the FOM is given in Figure 4.9b. The optimization using the reduced order model using all sources and detectors requires $5,400$ linear systems of dimension $32,768$ to construct ROM basis and $16,875$ linear systems of dimension $1,368$ to reconstruct the absorption image. The reconstruction result using the standard ROM with all sources and detectors is given in Figure 4.10c. Meanwhile, our approach only requires $1,200$ linear systems of dimension $32,768$ to construct ROM basis and $17,550$ linear systems of dimension $1,294$ to reconstruct the absorption image. Our approach reduces the large solver cost by about *a factor 16*, while obtaining similar quality reconstruction results; see Figure 4.10d. In Table 4.1, we give the number of linear systems to be solved for each method for comparison.

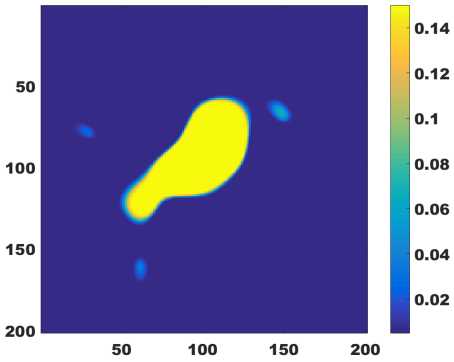
Example 3. We use the same model as in Example 2. In this experiment, we use 3 parameter sample points and 4 frequencies to construct the ROM basis. We also use the same initial absorption image, see Figure 4.8. Recall that the FOM requires the solution



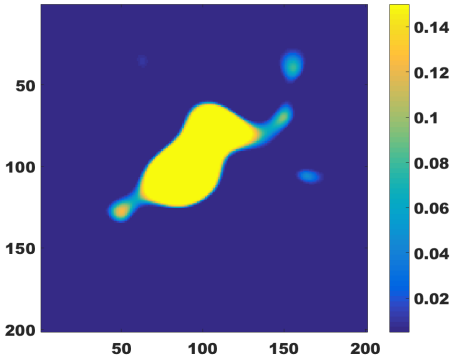
(a) True shape of the anomaly.



(b) Reconstruction using the FOM.



(c) Reconstruction using the standard ROM with all sources and detectors.



(d) Reconstruction using the ROM with stochastic sources and detectors, $\ell_s = \ell_d = 12$.

Figure 4.7: Reconstruction of a simple 2D test anomaly.

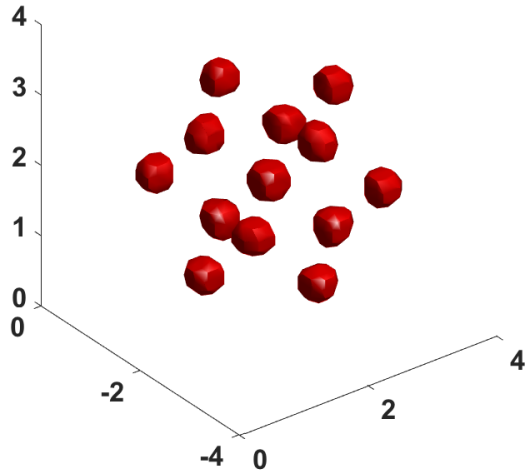
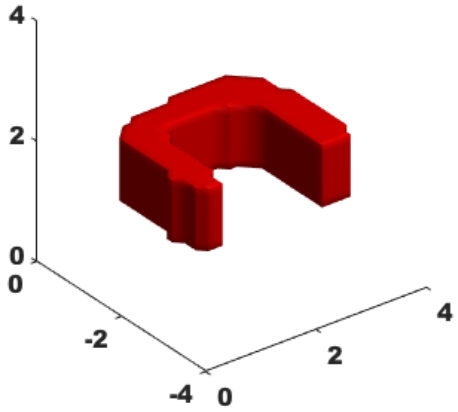


Figure 4.8: Initial configuration with 27 basis functions arranged in a $3 \times 3 \times 3$ grid where 13 basis functions have a positive expansion factors (visible) and 14 basis functions have negative expansion factors (invisible).

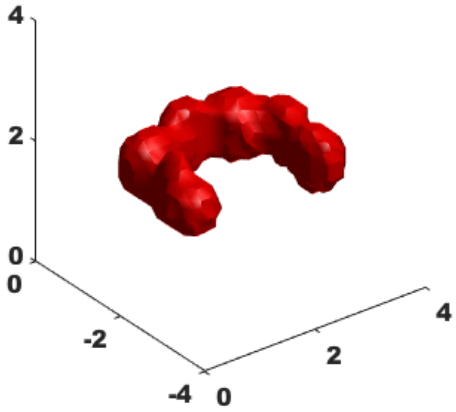
of $47,700$ linear systems of dimension $32,768$. The standard ROM using all sources and detectors requires the solution of $5,400$ linear systems of dimension $1,228$, and reduces the large solver cost by about *a factor 9*. However, our approach only requires the solution of $1,200$ linear systems of dimension $1,184$ and drastically reduces the large solver cost, by about *a factor 40*. The number of linear systems required for each method is given in Table 4.1. The original absorption image and the results of the reconstruction are given in Figure 4.10. For larger problems with many sources and detectors and using multiple frequencies, we expect much larger gains.

4.6 Conclusions

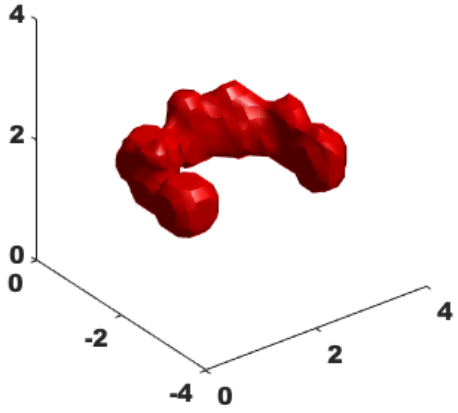
As argued before, the computation of a global basis poses a formidable cost. We have shown that using stochastic sources and detectors to build the global basis can substantially reduce the cost. Our experiments show that even for a small 3D problem, the number of large linear solves can substantially be reduced, while obtaining similar quality reconstructions. For larger problems with many more sources and detectors and multiple frequencies, we expect much larger gains.



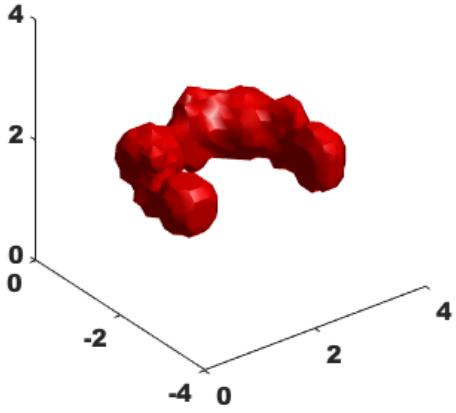
(a) True shape of the anomaly.



(b) Reconstruction using the FOM.

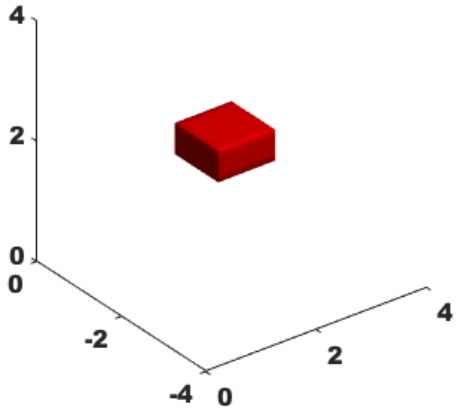


(c) Reconstruction using the standard ROM with all sources and detectors.

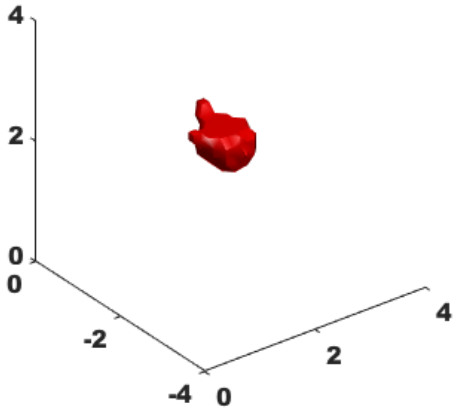


(d) Reconstruction using the ROM with stochastic sources and detectors, $\ell_s = 50$.

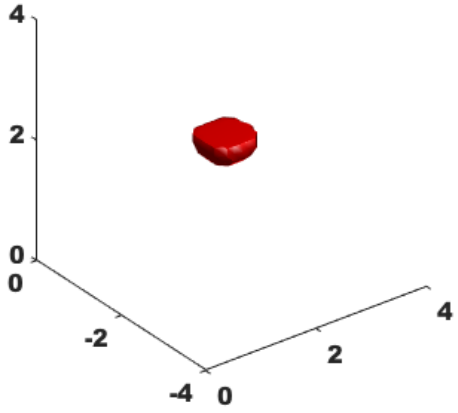
Figure 4.9: Results for Example 2. Reconstruction of a test anomaly on $32 \times 32 \times 32$ mesh with 225 sources and detectors, 27 basis functions, and using three frequencies.



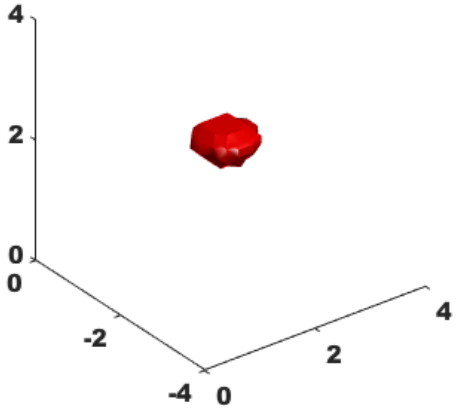
(a) True shape of the anomaly.



(b) Reconstruction using the FOM.



(c) Reconstruction using the standard ROM with all sources and detectors.



(d) Reconstruction using the ROM with stochastic sources and detectors, $\ell_s = 50$.

Figure 4.10: Results for Example 3. Reconstruction of a test anomaly on $32 \times 32 \times 32$ mesh with 225 sources and detectors, 27 basis functions, and using four frequencies.

Chapter 5

Iterative Solution and Tuning Accuracy

5.1 Introduction

In section 4.3, we propose a new approach to substantially reduce the number of large linear solves for constructing the global ROM basis, see Algorithm 2. In addition, we show that by eliminating the boundary conditions and solving for the Schur complement, the transfer function and derivatives can be written in terms of a SPD matrix for the zero frequency case (complex symmetric if ω is nonzero). Let $\mathbf{R} = [\mathbf{r}_1 \cdots \mathbf{r}_{\ell_s}]$ with $\ell_s \ll n_s$, and $\mathbf{L} = [\boldsymbol{\ell}_1 \cdots \boldsymbol{\ell}_{\ell_d}]$ with $\ell_d \ll n_d$ be the stochastic tangential directions, then for each frequency ω_j, we need to solve the linear systems

$$\left(\frac{\imath\omega_j}{\nu}\mathbf{I} + \widetilde{\mathbf{A}}(\mathbf{p})\right)\widetilde{\mathbf{Y}} = \mathbf{B}\mathbf{R}, \text{ and } \left(\frac{\imath\omega_j}{\nu}\mathbf{I} + \widetilde{\mathbf{A}}(\mathbf{p})\right)^T\widetilde{\mathbf{Z}} = \mathbf{C}^T\mathbf{L}, \tag{5.1}$$

where $\mathbf{B} \in \mathbb{R}^{n\times n_s}$, $\mathbf{C} \in \mathbb{R}^{n_d\times n}$. Here, $\widetilde{\mathbf{A}}(\mathbf{p}) \in \mathbb{R}^{n\times n}$ is an SPD matrix. Since the size of a realistic linear system is at least $O(10^6)$, the use of direct sparse solvers for (5.1) becomes impractical, especially for the 3D problems. Therefore, we use iterative methods to handle the large linear systems for interpolatory model reduction [11].

In this chapter, we first explore the use of well known iterative methods to solve the linear systems in (5.1) and discuss implementation details. This is important for two reasons. First, the results in [11] suggest that in general high accuracy solves are not necessary to compute interpolatory reduced order models. Second, since we slightly oversample the reduction basis using Algorithm 2, again high accuracy is unlikely to be important. Substantial computational savings are possible by reducing the solver tolerance. Therefore, we provide a numerical study on how sensitive the quality of the reduced order model is to the chosen tolerance. Many iterative methods have been developed to solve (5.1) [51, 57]. In particular we focus on MINRES [7, 46] for the zero frequency and GMRES [52] for the nonzero

frequencies.

For the zero frequency, we have a SPD matrix $\widetilde{\mathbf{A}}(\mathbf{p})$, which depends on the parameter vector $\mathbf{p}$ and arises from the finite difference discretization of the PDE (4.1). In our application, the changes in $\widetilde{\mathbf{A}}(\mathbf{p})$ are quite small relative to the magnitude of the matrix coefficients as $\mathbf{p}$ changes, see (4.37). However, the changes become significiant over multiple optimization steps. In addition, we also need to solve for multiple right hand sides for each matrix. Therefore, Krylov subspace recycling is a useful tool to speed up the convergence of each linear system [2, 27, 47]. In particular, we use the recycling version of MINRES (RMINRES) for symmetric matrices with no complex shifts. The details of RMINRES can be found in [42, 60]. To speed up convergence further, we use preconditioners. For SPD matrices, a standard preconditioning is the incomplete Cholesky factorization of $\widetilde{\mathbf{A}}(\mathbf{p})$ [41]. We use the Matlab function $\texttt{ichol}(\widetilde{\mathbf{A}})$ with no fill to compute the preconditioner. We also need to solve the systems

$$\left(\frac{\imath\omega_j}{\nu}\,\mathbf{E} + \widetilde{\mathbf{A}}(\mathbf{p})\right)\hat{\mathbf{Y}} = \mathbf{B} \text{ and } \left(\frac{\imath\omega_j}{\nu}\,\mathbf{E} + \widetilde{\mathbf{A}}(\mathbf{p})\right)^T\widehat{\mathbf{Z}} = \mathbf{C}^T, \tag{5.2}$$

for each nonzero frequency ω_j with n_s and n_d number of right hand sides.

For nonzero frequencies, we separate the real and imaginary parts of the matrix to avoid complex arithmetic as in [45]

$$\mathbf{K} = \begin{bmatrix} \widetilde{\mathbf{A}}(\mathbf{p}) & -\frac{\omega}{\nu}\mathbf{I} \\ -\frac{\omega}{\nu}\mathbf{I} & -\widetilde{\mathbf{A}}(\mathbf{p}) \end{bmatrix}, \tag{5.3}$$

where $\mathbf{K} \in \mathbb{R}^{2n\times 2n}$ and $\left(\frac{\omega}{\nu}\mathbf{I}\right) \in \mathbb{R}^{n\times n}$ corresponds to the complex shifts in (5.1). Since $\left(\frac{\omega}{\nu}\right)$ is relatively small compared with $\|\widetilde{\mathbf{A}}\|$, we base the preconditioning on the incomplete Cholesky factorization of the $(1,1)$ block, $\widetilde{\mathbf{A}}(\mathbf{p}) = \mathbf{M}_1\mathbf{M}_1^T$. Preconditioning from the left and right, we obtain

$$\begin{bmatrix} \mathbf{M}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_1^{-1} \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{A}}(\mathbf{p}) & -\frac{\omega}{\nu}\mathbf{I} \\ -\frac{\omega}{\nu}\mathbf{I} & -\widetilde{\mathbf{A}}(\mathbf{p}) \end{bmatrix} \begin{bmatrix} \mathbf{M}_1^{-T} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_1^{-T} \end{bmatrix}, \tag{5.4}$$

where the preconditioned matrix, $\mathbf{M}_1 = \texttt{ichol}(\widetilde{\mathbf{A}})$. However, the resulting matrix is indefinite with n positive and n negative eigenvalues, which typically leads to slow convergence. To move the negative eigenvalues of $\mathbf{K}$ to the right half plane, we combine this preconditioning with the following diagonal preconditioning [15, 55]

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{M}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_1^{-1} \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{A}}(\mathbf{p}) & -\frac{\omega}{\nu}\mathbf{I} \\ -\frac{\omega}{\nu}\mathbf{I} & -\widetilde{\mathbf{A}}(\mathbf{p}) \end{bmatrix} \begin{bmatrix} \mathbf{M}_1^{-T} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_1^{-T} \end{bmatrix}. \tag{5.5}$$

An alternative approach for multiple shifts is discussed in [37].

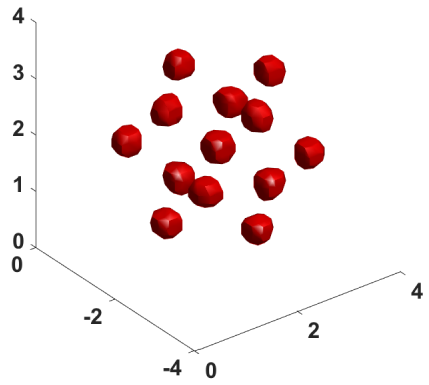


Figure 5.1: Initial configuration with 27 basis functions arranged in a $3 \times 3 \times 3$ grid where 13 basis functions have a positive expansion factors (visible) and 14 basis functions have negative expansion factors (invisible).

5.2 Numerical Experiments

In this section, we present two proof-of-concept experiments for 3D DOT inversion to analyze how sensitive the quality of the reduced order model is to the chosen tolerance. The experimental set up we use is that described in Section 4.5. For comparison, we also include reconstruction results for the FOM and the ROM using all sources and detectors using Algorithm 1. In both experiments, we use 60 stochastic sources and detectors as the right and left tangential interpolation directions to compute global basis using Algorithm 2. Using the SVD of the candidate bases, we discard the directions corresponding to the zero singular values so that interpolation is exact at the parameter and frequency sample points. For ROM using stochastic sources and detectors, interpolation is exact along the tangential directions. We report the relative interpolation errors due to the iterative solvers in Table 5.4. We stop the optimization when the residual norm falls below 1.2 times the noise level.

We use 27 CSRBFs to reconstruct the anomaly, leading to 135 parameters. The initial absorption image used in both experiments is given in Figure 5.1, where 13 basis functions have a positive expansion coefficient (visible as high absorption regions) and 14 basis functions have a negative expansion coefficient (invisible).

Experiment 1. The mesh is $32 \times 32 \times 32$, which gives $32,768$ degrees of freedom in the forward model (2.2). We use 5 parameter sample points and only the zero frequency to construct the ROM basis. To solve (5.1) and (5.2), we use RMINRES with the preconditioner described in Section 5.1.

We use five different values for the relative residual tolerance of 10^{-10}, 10^{-8}, 10^{-6}, 10^{-4}

tolerance	ROM stoch. srcs/dets	ROM all srcs/dets
10^{-10}	40	40
10^{-8}	32	31
10^{-6}	24	23
10^{-4}	15	15

Table 5.1: Number of preconditioned RMINRES iterations on average per right-hand side for each tolerance for Experiment 1 using the ROM computed with 60 stochastic sources and detectors and the standard ROM using all sources and detectors.

tolerance	$\frac{\|\mathbf{\Psi} - \mathbf{\Psi}_r\|}{\|\mathbf{\Psi}\|}$	$\frac{\|\mathbf{L}^T\mathbf{\Psi} - \mathbf{L}^T\widetilde{\mathbf{\Psi}}_r\|}{\|\mathbf{L}^T\mathbf{\Psi}\|}$	$\frac{\|\mathbf{\Psi}\mathbf{R} - \widetilde{\mathbf{\Psi}}_r\mathbf{R}\|}{\|\mathbf{\Psi}\mathbf{R}\|}$
10^{-10}	2.2×10^{-13}	2.8×10^{-11}	2.8×10^{-12}
10^{-8}	3.3×10^{-13}	3.5×10^{-9}	1×10^{-10}
10^{-6}	1.7×10^{-10}	3.1×10^{-7}	3.8×10^{-8}
10^{-4}	1.2×10^{-7}	3×10^{-5}	2.5×10^{-5}

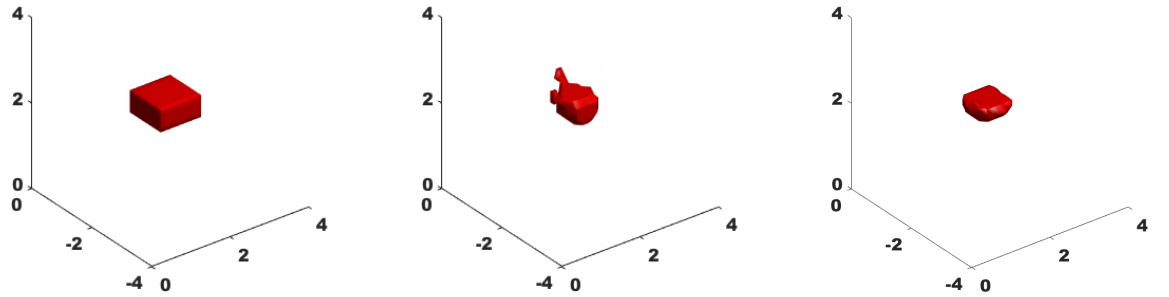
Table 5.2: Relative interpolation errors for Experiment 1 using the ROM computed with 60 stochastic sources and detectors and the standard ROM computed with all sources and detectors .

and $\texttt{10}^{-3}$ to test the quality of the reduced order models. For $\texttt{10}^{-3}$, the ROM computed with stochastic sources and detectors does not converge to the noise level while the ROM computed with all sources and detectors still converges.

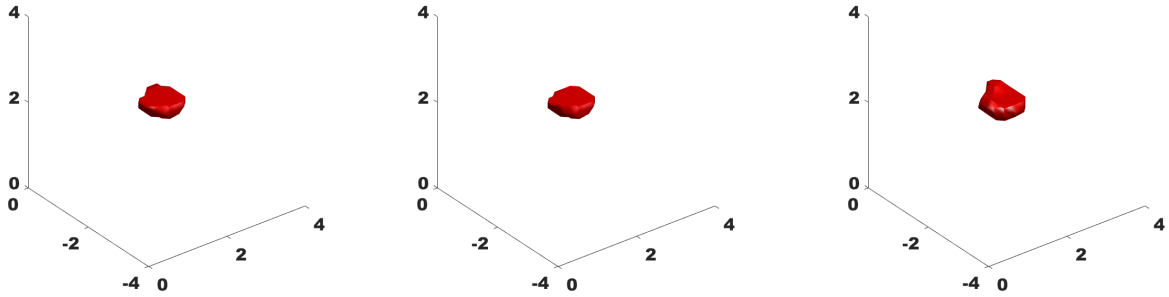
Figure 5.2 demonstrates that using iterative methods, we can obtain similar quality reconstruction results using 60 stochastic sources and detectors by reducing the solver tolerance. Similarly, substantial computational savings are possible using ROMs computed with all sources and detectors, see Figure 5.3.

In Table 5.1, we give the number of iterations for preconditioned RMINRES. On average, solver converges in 40 steps per right-hand side for both systems in (5.1) with $\texttt{tol} = \texttt{10}^{-10}$ and in 15 steps with $\texttt{tol} = \texttt{10}^{-4}$ per right-hand side without degrading the accuracy of the solution to the inverse problem. If we use different numbers of parameter sample points, the results are similar. We also give the relative interpolation errors for each tolerance in Table 5.2. The results justify that high accuracy solves are not necessary to compute accurate ROMs.

Experiment 2. We use the same model as in Experiment 1. To construct ROM basis using Algorithm 1 and Algorithm 2, we use 4 parameter sample points and two frequencies, $\omega_1 = 0$ and $\omega_2 = 10^8$. We use the same resulting relative residual tolerances, 10^{-10}, 10^{-8}, 10^{-6}, and $\texttt{10}^{-4}$ to test the quality of the reduced order models.

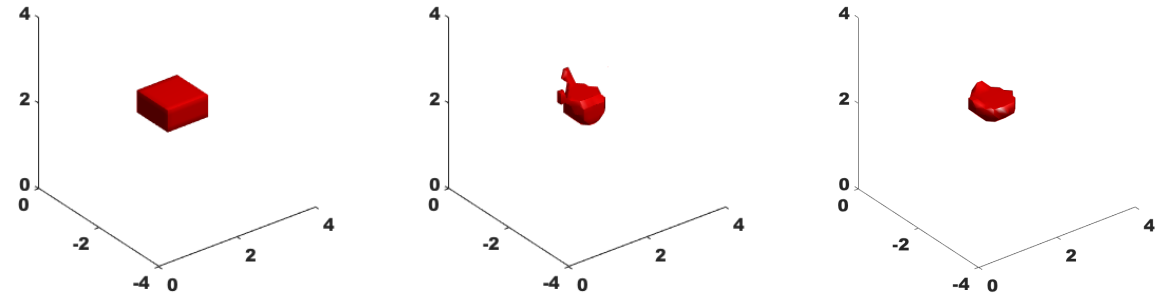


(a) (b) (c) $tol = 10^{-10}$.

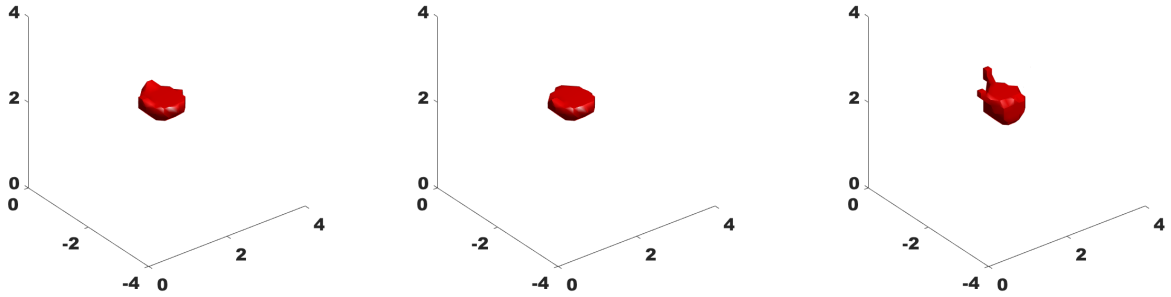


(d) $tol = 10^{-8}$. (e) $tol = 10^{-6}$. (f) $tol = 10^{-4}$.

Figure 5.2: Results for Example 1. Reconstruction of a test anomaly on the $32 \times 32 \times 32$ mesh with 225 sources and detectors, 27 basis functions, and using only the zero frequency. (a) True shape of the anomaly. (b) Reconstruction using the FOM. (c)-(f) Reconstructions using the ROM computed with 60 stochastic sources and detectors for the chosen tolerance.



(a) (b) (c) $tol = 10^{-10}$.



(d) $tol = 10^{-8}$. (e) $tol = 10^{-6}$. (f) $tol = 10^{-4}$.

Figure 5.3: Results for Example 1. Reconstruction of a test anomaly on the $32 \times 32 \times 32$ mesh with 225 sources and detectors, 27 basis functions, and using only the zero frequency. (a) True shape of the anomaly. (b) Reconstruction using the FOM. (c)-(f) Reconstructions using the ROM computed with all sources and detectors for the chosen tolerance.

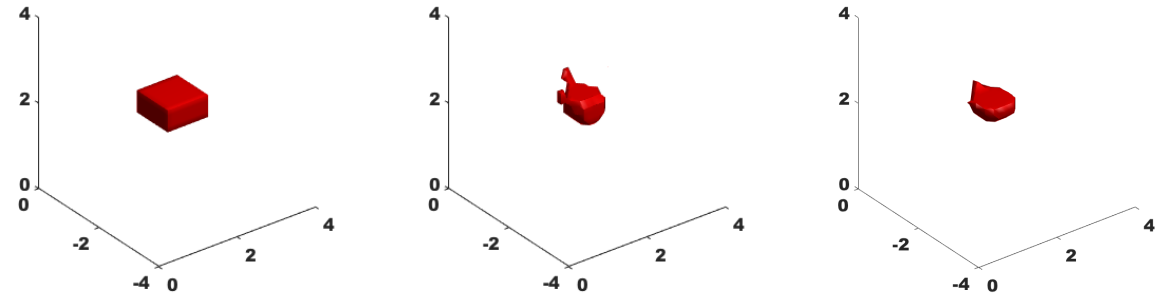
tolerance	ROM stoch. srcs/dets	ROM all srcs/dets
10^{-10}	56	56
10^{-8}	46	46
10^{-6}	35	35
10^{-4}	21	22

Table 5.3: Number of preconditioned RMINRES iterations on average per right-hand side for each tolerance for Experiment 2 using the ROM computed with 60 stochastic sources and detectors and the standard ROM using all sources and detectors.

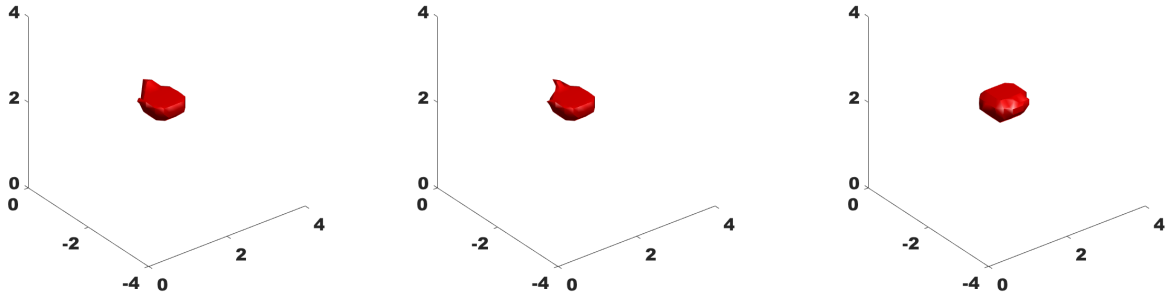
tolerance	$\frac{\|\mathbf{\Psi} - \mathbf{\Psi}_r\|}{\|\mathbf{\Psi}\|}$	$\frac{\|\mathbf{L}^T\mathbf{\Psi} - \mathbf{L}^T\widetilde{\mathbf{\Psi}}_r\|}{\|\mathbf{L}^T\mathbf{\Psi}\|}$	$\frac{\|\mathbf{\Psi}\mathbf{R} - \widetilde{\mathbf{\Psi}}_r\mathbf{R}\|}{\|\mathbf{\Psi}\mathbf{R}\|}$
10^{-10}	4.1×10^{-13}	2.3×10^{-13}	1.2×10^{-13}
10^{-8}	4.9×10^{-13}	2.2×10^{-11}	1.4×10^{-12}
10^{-6}	1.9×10^{-12}	1.2×10^{-9}	7.7×10^{-10}
10^{-4}	3.1×10^{-9}	7×10^{-6}	7.9×10^{-7}

Table 5.4: Relative interpolation errors for Experiment 2 using the ROM computed with 60 stochastic sources and detectors and the standard ROM computed with all sources and detectors .

Figure 5.4 and Figure 5.5 show that the reconstruction results are almost indistinguishable for these tolerances. In Table 5.3, we give the number of iterations for preconditioned RMINRES. For both systems in (5.1), for the nonzero frequency, the solver converges, on average, in 56 steps per right-hand side with `tol` $= 10^{-10}$ and in 21 steps with `tol` $= 10^{-4}$, without degrading the accuracy of the solution to the inverse problem. The numbers of iterations are shown in Table 5.3 for the zero frequency. We also give the relative interpolation errors for each tolerance in Table 5.4. The order of the errors remains relatively constant as $\mathbf{p}_k$ and ω_j varies. Hence, we only report the results for one parameter sample point and the nonzero frequency. The results suggest that high accuracy solves are not necessary to compute accurate ROMs.

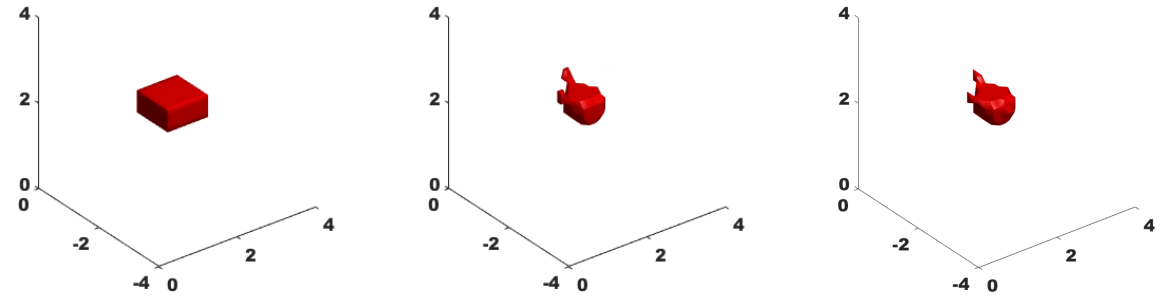


(a) (b) (c) $tol = 10^{-10}$.

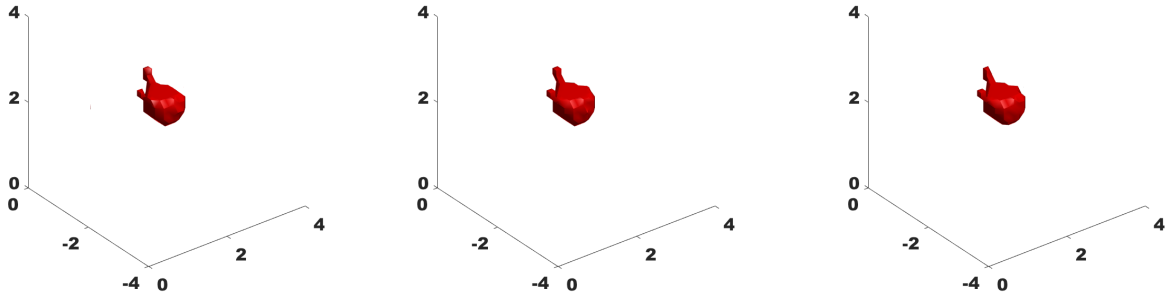


(d) $tol = 10^{-8}$. (e) $tol = 10^{-6}$. (f) $tol = 10^{-4}$.

Figure 5.4: Results for Example 2. Reconstruction of a test anomaly on $32 \times 32 \times 32$ mesh with 225 sources and detectors, 27 basis functions, and using two frequencies. (a) True shape of the anomaly. (b) Reconstruction using the FOM. (c)-(f) Reconstructions using the ROM computed with 60 stochastic sources and detectors for chosen tolerance.



(a) (b) (c) $tol = 10^{-10}$.



(d) $tol = 10^{-8}$. (e) $tol = 10^{-6}$. (f) $tol = 10^{-4}$.

Figure 5.5: Results for Example 2. Reconstruction of a test anomaly on $32 \times 32 \times 32$ mesh with 225 sources and detectors, 27 basis functions, and using two frequencies. (a) True shape of the anomaly. (b) Reconstruction using the FOM. (c)-(f) Reconstructions using the ROM computed with all sources and detectors for chosen tolerance.

Chapter 6

Conclusions

In this thesis, we introduce two new effective and computationally highly efficient methods to reduce the cost of the computing forward models and its derivatives.

In the first method, we use the SAA approach to estimate the objective function and the Jacobian using only a few random simultaneous sources and detectors in DOT problems. After solving to a modest tolerance, we use a few simultaneous sources and detectors that are optimized to maximize the Frobenius norm of the sampled Jacobian to improve the rate of convergence and obtain more accurate solutions. We complement these optimized sources and detectors by random linear combinations of the sources and detectors constrained to a complementary subspace. With combining simultaneous and optimized sources and detectors, we observed faster convergence and good quality reconstructions and robustness.

Moreover, we provide an alternative approach to compute optimized simultaneous sources and detectors. In this approach, when a chosen intermediate tolerance is reached, similar to the previous approach, we compute the full Jacobian once. Then, we replace the least effective randomized sources and detectors by simultaneous sources and detectors that are optimized to maximize the Frobenius norm of the sampled Jacobian. Numerical experiments show that this approach improves the rate of convergence of the optimization and the quality of the inverse solution while reducing the number of large-scale linear systems solves.

To compute optimized sources and detectors, we need to compute the full Jacobian once. In the future, we aim to reduce this additional cost by estimating the Jacobian. Although these approaches have proved successful experimentally, we aim to understand the underlying theory better. We plan to analyze what are the most effective simultaneous sources and detectors for fast convergence of the inverse problem: randomized, optimized (and in what sense), and their combination.

In the second method, we propose to use randomization to drastically reduce the number of large linear solves needed for constructing the global ROM bases without degrading the accuracy of the solution to the inversion problem. Moreover, we provide a theoretical justi-

fication for exploiting low rank structure in the candidate basis, and connect our approach to randomization in computing the interpolatory model reduction bases to tangential interpolation. Numerical experiments show that even for a small 3D problem, our approach drastically reduces the large solver cost, by about a factor 40. We expect much larger gains in larger problems with many sources and detectors and using multiple frequencies.

Bibliography

[1] A. Aghasi, E. Miller, and M. E. Kilmer. Parametric level set methods for inverse problems. *SIAM Journal on Imaging Science*, 4(2):618–650, 2011. (Cited on pp. 5, 9, 10, 15, 22, 39, 50, 51, 71)

[2] K. Ahuja, P. Benner, E. de Sturler, and L. Feng. Recycling BiCGStab with an application to parametric model order reduction. *SIAM J. Sci. Comput.*, 37(5):S429–446, 2015. (Cited on p. 79)

[3] A. Antoulas. *Approximation of Large-Scale Dynamical Systems*. SIAM, 2005. (Cited on p. 64)

[4] A. Antoulas, C. Beattie, and S. Gugercin. Interpolatory model reduction of large-scale dynamical systems. In J. Mohammadpour and K. Grigoriadis, editors, *Efficient Modeling and Control of Large-Scale Systems*, pages 2–58. Springer-Verlag, 2010. (Cited on p. 52)

[5] S. R. Arridge. Optical tomography in medical imaging. *Inverse Problems*, Vol. 16:R41–R93, 1999. (Cited on pp. 1, 8, 50)

[6] S. S. Aslan, E. de Sturler, and M. E. Kilmer. Randomized approach to nonlinear inversion combining simultaneous random and optimized sources and detectors. *submitted for publication, see: arXiv preprint:1706.05586*, 2017. (Cited on pp. 9, 49, 50, 54)

[7] R. Barrett, M. W. Berry, T. F. Chan, J. Demmel, a. D. J. Donato, V. Eijkhout, R. Pozo, C. Romine, and H. van der Vorst. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*. SIAM, 1993. (Cited on p. 78)

[8] U. Baur, P. Benner, C. Beattie, and S. Gugercin. Interpolatory projection methods for parameterized model reduction. *SIAM J. Sci. Comput*, 33:2489–2518, 2011. (Cited on pp. 52, 54, 55)

[9] C. J. Beasley. A new look at marine simultaneous sources. *Leading Edge*, 27:914–917, 2008. (Cited on pp. 2, 6, 10)

[10] C. Beattie and S. Gugercin. *Model Reduction and Approximation*, chapter 7: Model Reduction by Rational Interpolation, pages 297–334. SIAM, 2017. (Cited on p. 52)

[11] C. Beattie, S. Gugercin, and S. Wyatt. Inexact solves in interpolatory model reduction. *Linear Algebra and its Applications*, 436:2916–2943, 201. (Cited on pp. 3, 78)

[12] P. Benner, A. Cohen, M. Ohlberger, and K. Willcox. *Model Reduction and Approximation: Theory and Algorithms*. SIAM, 2017. (Cited on p. 52)

[13] P. Benner, S. Gugercin, and K. Willcox. A survey of model reduction methods for parametric systems. *SIAM Review*, 57(4):483–531, 2015. (Cited on pp. 52, 53)

[14] P. Benner, M. Ohlberger, A. Patera, G. Rozza, and K. Urban. *Model Reduction of Parametrized Systems*. Springer, 2017. (Cited on p. 52)

[15] M. Benzi and V. Simoncini. On the eigenvalues of a class of saddle point matrices. *Numer. Math.*, 103:173–196, 2006. (Cited on p. 79)

[16] D. Boas, D. Brooks, E. Miller, C. DiMarzio, M. Kilmer, R. Gaudette, and Q. Zhang. Imaging the body with diffuse optical tomography. *IEEE Signal Processing Magazine*, 18(6):57–75, 2001. (Cited on pp. 1, 8)

[17] R. Bollapragada, R. Byrd, and J. Nocedal. Exact and inexact subsampled Newton methods for optimization. *arXiv preprint arXiv:1609.08502*, 2016. (Cited on p. 15)

[18] B. N. Bond and L. Daniel. Parameterized model order reduction of nonlinear dynamical systems. In *IEEE/ACM Internat.Conf. on Computer-Aided Design, 2005. ICCAD-2005*, pages 487–494, 2005. (Cited on p. 52)

[19] T. Bui-Thanh, K. Willcox, and O. Ghattas. Model reduction for large-scale systems with high-dimensional parametric input space. *SIAM J. Sci. Comput.*, 30(6):3270–3288, 2008. (Cited on p. 52)

[20] T. Bui-Thanh, K. Willcox, and O. Ghattas. Parametric reduced-order models for probabilistic analysis of unsteady aerodynamic applications. *AIAA Journal*, 46(10):2520–2529, 2008. (Cited on p. 52)

[21] J. Bushberg, J. Seibert, E. Leidholdt Jr, J. Boone, and E. Goldschmidt Jr. *The essential physics of medical imaging*, volume 30. Lippincott Williams & Wilkens, 2003. (Cited on p. 1)

[22] R. H. Byrd, G. M. Chin, J. Nocedal, and Y. Wu. Sample size selection in optimization methods for machine learning. *Mathematical programming*, 134(1):127–155, 2012. (Cited on p. 15)

[23] E. de Sturler, S. Gugercin, M. E. Kilmer, S. Chaturantabut, C. Beattie, and M. O'Connell. Nonlinear parametric inversion using interpolatory model reduction. *SIAM J. Sci. Comput*, 37(3):B495–B517, 2015. (Cited on pp. 3, 6, 9, 10, 21, 22, 39, 49, 50, 53, 54, 56, 60, 71)

[24] E. de Sturler and M. E. Kilmer. A regularized Gauss-Newton trust region approach to imaging in diffuse optical tomography. *SIAM J. Sci. Comput.*, 33:3057 – 3086, 2011. (Cited on pp. 6, 9, 15, 22, 32, 39, 51, 71)

[25] J. Dennis and R. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations*. SIAM, 1996. (Cited on p. 9)

[26] L. Feng and P. Benner. A robust algorithm for parametric model order reduction based on implicit moment matching. *Proc. Appl. Math. Mech.*, 7:10215.01–10215.02, 2008. (Cited on p. 52)

[27] L. Feng, P. Benner, and J. G. Korvink. Subspace recycling accelerates the parametric macro-modeling of MEMS. *International Journal for Numerical Methods in Engineering*, 94(1):84–110, 2013. (Cited on p. 79)

[28] M. S. Gockenbach. *Partial Differential Equations, Analytical and Numerical Results*. SIAM, 2002. (Cited on pp. 63, 70)

[29] S. Gugercin, A. Antoulas, and C. Beattie. $\mathcal{H}_2$ model reduction for large-scale linear dynamical systems. *SIAM J. Matrix Anal. Appl.*, 30(2):609–638, 2008. (Cited on p. 55)

[30] S. Gugercin, T. Stykel, and S. Wyatt. Model reduction of descriptor systems by interpolatory projection methods. *SIAM J. Sci. Comput.*, 35(5):B1010–B1033, 2013. (Cited on p. 59)

[31] P. Gunupudi, R. Khazaka, M. Nakhla, T. Smy, and D. Celo. Passive parameterized time-domain macromodels for high-speed transmission-line networks. *IEEE Trans. Microwave Theory and Techniques*, 51(12):2347–2354, 2003. (Cited on p. 52)

[32] E. Haber, M. Chung, and F. Herrmann. An effective method for parameter estimation with PDE constraints with multiple right-hand sides. *SIAM J. Optim.*, 22(3):739–757, 2012. (Cited on pp. 2, 6, 10, 11, 12, 30, 54)

[33] N. Halko, P. G. Martinsson, and J. A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53(2):217–288, 2011. (Cited on p. 54)

[34] J. S. Hesthaven, G. Rozza, and B. Stamm. *Certified reduced basis methods for parametrized partial differential equations*. Springer, 2016. (Cited on p. 52)

[35] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, 1991. (Cited on pp. 13, 17)

[36] M. F. Hutchinson. A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines. *Commun. Statist. Simulation Comput.*, 18(3):1059–1076, 1990. (Cited on p. 11)

[37] M. E. Kilmer and E. de Sturler. Recycling subspace information for diffuse optical tomography. *SIAM J. Sci. Comput.*, 27(6):2140–2166, 2006. (Cited on pp. 21, 54, 57, 79)

[38] J. R. Krebs, J. E. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M.-D. Lacasse. Fast full-wavefield seismic inversion using encoded sources. *Geophysics*, 74(6):WCC177–WCC188, 2009. (Cited on p. 6)

[39] Z. Li, Z. Qiao, and T. Tang. *Numerical Solution of Differential Equations: Introduction to Finite Difference and Finite Element Methods*. Cambridge University Press, 2018. (Cited on p. 62)

[40] A. Louis. *Medical imaging: state of the art and future development*, volume 8. Institute of Physics Publishing, 1992. (Cited on p. 1)

[41] J. A. Meijerink and H. A. van der Vorst. An Iterative Solution Method for Linear Systems of Which the Coefficient Matrix is a Symmetric M-Matrix. *Mathematics of Computation*, 31(137):148–162, 1977. (Cited on p. 79)

[42] L. A. M. Mello, E. de Sturler, G. H. Paulino, and E. C. N. Silva. Recycling Krylov subspaces for efficient large-scale electrical impedance tomography. *Comput. Methods Appl. Mech. Engrg.*, 199:3101–3110, 2010. (Cited on p. 79)

[43] S. A. Morton and C. C. Ober. Faster shot-record depth migrations using phase encoding. *68th Annual International Meeting, Society of Exploration Geophysicists, Expanded Abstracts*, 37:1131–1134, 1998. (Cited on pp. 2, 10)

[44] R. Neelamani, C. E. Krohn, J. R. Krebs, J. K. Romberg, M. Deffenbaugh, and J. E. Anderson. Efficient seismic forward modeling using simultaneous random sources and sparsity. *Geophysics*, 75(6):WB15–WB27, 2010. (Cited on pp. 2, 6, 10)

[45] M. O'Connell, M. E. Kilmer, E. de Sturler, and S. Gugercin. Computing reduced order models via inner-outer Krylov recycling in diffuse optical tomography. *SIAM J. Sci. Comput.*, 39(2):B272–B297, 2017. (Cited on pp. 49, 54, 55, 57, 79)

[46] C. C. Paige and M. A. Saunders. Solution of sparse indefinite systems of linear equations. *SIAM J. Numer. Anal.*, 12(4):617–629, 1975. (Cited on p. 78)

[47] M. Parks, E. de Sturler, G. Mackey, D. Johnson, and S. Maiti. Recycling Krylov subspaces for sequences of linear systems. *SIAM J. Sci. Comput.*, 28:1651–1674, 2006. (Cited on pp. 54, 79)

[48] F. Roosta-Khorasani and M. W. Mahoney. Sub-sampled Newton methods I: Globally convergent algorithms. *arXiv preprint arXiv:1601.04737*, 2016. (Cited on p. 15)

[49] F. Roosta-Khorasani and M. W. Mahoney. Sub-sampled Newton methods II: Local convergence rates. *arXiv preprint arXiv:1601.04738*, 2016. (Cited on p. 15)

[50] F. Roosta-Khorasani, K. van den Doel, and U. Ascher. Stochastic algorithms for inverse problems involving PDEs and many measurements. *SIAM J. Sci. Comput.*, 36(5):S3–S22, 2014. (Cited on pp. 6, 15)

[51] Y. Saad. *Iterative Methods for Sparse Linear Systems, second ed.* SIAM, second edition, 2003. (Cited on p. 78)

[52] Y. Saad and M. H. Schultz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. and Stat. Comput.,*, 7(3):856–869, 1986. (Cited on p. 78)

[53] O. Semerci and E. L. Miller. A parametric level-set approach to simultaneous object identification and background reconstruction for dual-energy computed tomography. *IEEE Transactions on Image Processing*, 21(5):2719–2734, 2012. (Cited on p. 1)

[54] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory.* SIAM, 2009. (Cited on pp. 2, 6, 14, 23, 40)

[55] A. Sidi. A zero-cost preconditioning for a class of indefinite linear systems. *WSEAS Transactions on Mathematics*, 2:142–150, 2003. (Cited on p. 79)

[56] R. Snieder and J. Trampert. *Inverse problems in geophysics.* Springer, 1999. (Cited on p. 1)

[57] H. van der Vorst. *Iterative Krylov Methods for Large Linear Systems.* Cambridge University Press, 2003. (Cited on p. 78)

[58] T. van Leeuwen, A. Y. Aravkin, and F. J. Herrmann. Seismic waveform inversion by stochastic optimization. *International Journal of Geophysics*, 2011. (Cited on p. 6)

[59] C. R. Vogel. *Computational Methods for Inverse Problems.* SIAM, Philadelphia, 2002. (Cited on pp. 5, 9, 51)

[60] S. Wang, E. de Sturler, and G. Paulino. Large-scale topology optimization using preconditioned Krylov subspace methods with recycling. *International Journal for Numerical Methods in Engineering*, 69:2441–2461, 2007. (Cited on p. 79)

[61] A. Webb and G. Kagadis. *Introduction to biomedical imaging*, volume 30. Wiley-IEEE Press, 2003. (Cited on p. 1)

[62] Z. Ying, R. Naidu, and C. R. Crawford. Dual energy computed tomography for explosive detection. *Journal of X-Ray Science and Technology*, 14:235–256, 2006. (Cited on p. 1)

[63] M. Zhdanov. *Geophysical inverse theory and regularization problems.* Elsevier Science Ltd, 2002. (Cited on p. 1)

Appendices

Appendix A

Randomization for Efficient Reduced Order Models

A.1 Perturbations in Gramians: Derivations

In this section, we include the detailed steps we skipped in section 4.4.3. We start with derivations for the proof of Lemma 4.2. We use the following properties of the trace operator:

i. Let $\mathbf{M}$ and $\mathbf{N}$ be square matrices, then

$$\operatorname{trace}(\mathbf{M}+\mathbf{N}) = \operatorname{trace}(\mathbf{M}) + \operatorname{trace}(\mathbf{N}) \tag{A.1}$$

and

$$\operatorname{trace}(c\mathbf{M}) = c\operatorname{trace}(\mathbf{M}), \tag{A.2}$$

where c is a scalar.

ii. Let $\mathbf{K}$ be an $m \times n$ matrix and $\mathbf{L}$ be an $n \times m$ matrix,

$$\operatorname{trace}(\mathbf{KL}) = \operatorname{trace}(\mathbf{LK}), \tag{A.3}$$

which is known as the cyclic property of trace.

iii. Let $\mathbf{K}$ be an $m \times n$ matrix and $\mathbf{L}$ be an $n \times m$ matrix. If $\mathbf{K}$ is a positive semi-definite matrix, then

$$\operatorname{trace}(\mathbf{KL}) \leq \|\mathbf{L}\| \operatorname{trace}(\mathbf{K}). \tag{A.4}$$

We aim to show that

$$\|\Delta\mathbf{P}\|^2 \leq 4\|\mathbf{Q}\|^2\|\mathbf{P}_i\|^2, \tag{A.5}$$

where $\mathbf{Q}$ is the unique positive semi-definite solution of the Lyapunov equation (4.69), and the integral representation of $\mathbf{Q}$ is defined as

$$\mathbf{Q} = \int_0^\infty e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t}\, \widetilde{\Delta\mathbf{A}}\, e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} dt. \tag{A.6}$$

Then,

$$\begin{aligned}
\|\Delta\mathbf{P}\|^2 \;=\; & \underbrace{\int_0^\infty\int_0^\infty \operatorname{trace}\left(e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t}\widetilde{\Delta\mathbf{A}}\mathbf{P}_i e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s}\widetilde{\Delta\mathbf{A}}\mathbf{P}_i e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s}\right) dsdt}_{\mathcal{I}_1} \\
& + \underbrace{\int_0^\infty\int_0^\infty \operatorname{trace}\left(e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t}\widetilde{\Delta\mathbf{A}}\mathbf{P}_i e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s}\mathbf{P}_i\widetilde{\Delta\mathbf{A}} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s}\right) dsdt}_{\mathcal{I}_2} \\
& + \underbrace{\int_0^\infty\int_0^\infty \operatorname{trace}\left(e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t}\mathbf{P}_i\widetilde{\Delta\mathbf{A}} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s}\widetilde{\Delta\mathbf{A}}\mathbf{P}_i e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s}\right) dsdt}_{\mathcal{I}_3} \\
& + \underbrace{\int_0^\infty\int_0^\infty \operatorname{trace}\left(e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t}\mathbf{P}_i\widetilde{\Delta\mathbf{A}} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s}\mathbf{P}_i\widetilde{\Delta\mathbf{A}} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s}\right) dsdt.}_{\mathcal{I}_4}
\end{aligned} \tag{A.7}$$

In the following derivations, we use the cylic property of trace multiple times to factor out $\|\mathbf{P}_i\|^2$, then we interchange the trace and integral operators to obtain (A.6). Consider the

first term in (A.7)

$$
\begin{aligned}
\mathcal{I}_1 &= \int_0^\infty \int_0^\infty \operatorname{trace}\left(e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})t} \widetilde{\Delta \mathbf{A}} \mathbf{P}_i \underbrace{e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})t} e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})s} \widetilde{\Delta \mathbf{A}} \mathbf{P}_i e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})s}}_{\text{cyclic prop.}} \right) ds\, dt \\
&= \int_0^\infty \int_0^\infty \operatorname{trace}\left(e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})t} e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})s} \widetilde{\Delta \mathbf{A}} \mathbf{P}_i e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})s} e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})t} \widetilde{\Delta \mathbf{A}} \underbrace{\mathbf{P}_i}_{\text{prop iii}} \right) ds\, dt \\
&\leq \|\mathbf{P}_i\| \int_0^\infty \int_0^\infty \operatorname{trace}\left(e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})t} e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})s} \widetilde{\Delta \mathbf{A}} \mathbf{P}_i \underbrace{e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})s} e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})t} \widetilde{\Delta \mathbf{A}}}_{\text{cyclic prop.}} \right) ds\, dt \\
&= \|\mathbf{P}_i\| \int_0^\infty \int_0^\infty \operatorname{trace}\left(e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})s} e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})t} \widetilde{\Delta \mathbf{A}} e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})t} e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})s} \widetilde{\Delta \mathbf{A}} \underbrace{\mathbf{P}_i}_{\text{prop iii}} \right) ds\, dt \\
&\leq \|\mathbf{P}_i\|^2 \int_0^\infty \int_0^\infty \operatorname{trace}\left(e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})s} e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})t} \widetilde{\Delta \mathbf{A}} e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})t} e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})s} \widetilde{\Delta \mathbf{A}} \right) ds\, dt \\
&= \|\mathbf{P}_i\|^2 \int_0^\infty \int_0^\infty \operatorname{trace}\left(\underbrace{e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})s}}_{\text{cyclic prop.}} e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})t} \widetilde{\Delta \mathbf{A}} e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})t} e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})s} \widetilde{\Delta \mathbf{A}} \right) ds\, dt \\
&= \|\mathbf{P}_i\|^2 \int_0^\infty \int_0^\infty \operatorname{trace}\left(e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})t} \widetilde{\Delta \mathbf{A}} e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})t} e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})s} \widetilde{\Delta \mathbf{A}} e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})s} \right) ds\, dt \\
&= \|\mathbf{P}_i\|^2 \operatorname{trace}\left(\underbrace{\int_0^\infty e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})t} \widetilde{\Delta \mathbf{A}} e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})t} dt}_{\mathbf{Q}} \underbrace{\int_0^\infty e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})s} \widetilde{\Delta \mathbf{A}} e^{\tilde{\mathbf{A}}(\mathbf{p}_{i+1})s} ds}_{\mathbf{Q}} \right),
\end{aligned}
\tag{A.8}
$$

which gives the following bound,

$$
\mathcal{I}_1 \leq \|\mathbf{P}_i\|^2 \|\mathbf{Q}\|^2. \tag{A.9}
$$

Next, we consider $\mathcal{I}_2$

$$
\begin{aligned}
\mathcal{I}_2 &= \int_0^\infty \int_0^\infty \operatorname{trace}\left(e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} \widetilde{\Delta\mathbf{A}} \mathbf{P}_i \underbrace{e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s} \mathbf{P}_i \widetilde{\Delta\mathbf{A}} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s}}_{\text{cyclic prop.}} \right) ds\, dt \\
&= \int_0^\infty \int_0^\infty \operatorname{trace}\left(e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s} \mathbf{P}_i \widetilde{\Delta\mathbf{A}} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} \widetilde{\Delta\mathbf{A}} \underbrace{\mathbf{P}_i}_{\text{prop iii}} \right) ds dt \\
&\leq \|\mathbf{P}_i\| \int_0^\infty \int_0^\infty \operatorname{trace}\left(e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s} \mathbf{P}_i \underbrace{\widetilde{\Delta\mathbf{A}} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} \widetilde{\Delta\mathbf{A}}}_{\text{cyclic prop.}} \right) ds\, dt \\
&= \|\mathbf{P}_i\| \int_0^\infty \int_0^\infty \operatorname{trace}\left(\widetilde{\Delta\mathbf{A}} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} \widetilde{\Delta\mathbf{A}} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s} \underbrace{\mathbf{P}_i}_{\text{prop iii}} \right) ds\, dt \\
&\leq \|\mathbf{P}_i\|^2 \int_0^\infty \int_0^\infty \operatorname{trace}\left(\underbrace{\widetilde{\Delta\mathbf{A}} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s}}_{\text{cyclic prop.}} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} \widetilde{\Delta\mathbf{A}} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s} \right) ds\, dt \\
&= \|\mathbf{P}_i\|^2 \int_0^\infty \int_0^\infty \operatorname{trace}\left(e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} \widetilde{\Delta\mathbf{A}} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s} \widetilde{\Delta\mathbf{A}} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s} \right) ds\, dt \\
&= \|\mathbf{P}_i\|^2 \operatorname{trace}\left(\underbrace{\int_0^\infty e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} \widetilde{\Delta\mathbf{A}} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})t} dt}_{\mathbf{Q}} \underbrace{\int_0^\infty e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s} \widetilde{\Delta\mathbf{A}} e^{\widetilde{\mathbf{A}}(\mathbf{p}_{i+1})s} ds}_{\mathbf{Q}} \right),
\end{aligned}
\tag{A.10}
$$

which gives the following bound

$$\mathcal{I}_2 \leq \|\mathbf{P}_i\|^2 \|\mathbf{Q}\|^2. \tag{A.11}$$

Similarly, we can bound $\mathcal{I}_3$ and $\mathcal{I}_4$ using the properties of the trace operator such that

$$\mathcal{I}_3 \leq \|\mathbf{P}_i\|^2 \|\mathbf{Q}\|^2, \tag{A.12}$$

and

$$\mathcal{I}_4 \leq \|\mathbf{P}_i\|^2 \|\mathbf{Q}\|^2. \tag{A.13}$$

Combining (A.9), (A.11), (A.12), and (A.13), we obtain

$$\|\Delta\mathbf{P}\|^2 \leq 4\|\mathbf{Q}\|^2 \|\mathbf{P}_i\|^2, \tag{A.14}$$

leading to the desired upper bound in Lemma 4.5

$$\frac{\|\Delta\mathbf{P}\|}{\|\mathbf{P}_i\|} \leq 2\|\mathbf{Q}\|. \tag{A.15}$$