Investigations in Diffuse Optical Tomography

Chetan Mehta

Supervisor: Alexander Barnett

Dartmouth College

`chetan.mehta@dartmouth.edu`

March 25, 2008

1   Introduction

Diffuse Optical Tomography (DOT) is a non-invasive, non-ionizing method for medical imaging whose applications include neuro-imaging, breast tumor detection, tracking muscular oxygenation and arthritic joint imaging. DOT includes fast temporal resolution of 10 ms and can determine spatial information to a depth of several centimeters, making it of great use in brian imaging applications. The apparatus for DOT is inexpensive, portable and capable of continuous monitoring. The DOT process involves illuminating the tissue with multiple light sources and measuring the light leaving the tissue at different detectors (Figure 1). The available signals comprise light intensities at each detector due to each source. For each of these source-detector pairs, it is possible to measure a dc intensity (continuous-wave systems), or obtain the intensity distribution as a function of time of flight (time-domain systems). In this paper, we delve into the mathematical model used to resolve the propagation of light within the tissue, parametrized in terms of scattering and absorption as a function of the position in the tissue [1].

The key challenge of DOT is that of extracting spatial maps of the optical properties (absorption coefficient μ_a and reduced scattering coefficient μ_s') within a highly scattering tissue volume by coupling multiple light sources and multiple detectors to the surface of the skin. The path taken by photons from source to detector is diffusive rather than straight since the tissue

dimensions are much larger than the photon mean free path. Recovery of $\mu_a(\mathbf{r})$ and $\mu_s(\mathbf{r})'$ from measured signals requires the solving of an inverse problem, non-linear in the optical parameters and known to be ill-posed [1]. In this paper, we concern ourselves solely with measuring absorption changes.

Due to the complexities of real-world physical situations, we approach the problem through analyzing simple geometries in which the analytical forward model solutions can be obtained. We use Bayesian inference to handle the uncertainty in the inverse problem. Our goal is to find the relative improvement in the quality of our reconstructed image when considering various physical configurations of sources and detectors.

1.1 History of DOT

Optical imaging of biological tissue dates to the late 1920s, when use of continuous wave light to detect breast lesions was originally proposed by Cutler. This procedure was not very successful since it overheated the patient's skin. Diaphanography, in which the breast is positioned between a visible (or infrared) light- source and the physician, was introduced in the 1970s, with several advancements made during the 1980s. Due to a Swedish study in 1992 that found the method inferior to traditional imaging techniques, optical imaging of the breast was abandoned in the early 90s. Later developments in the field, made possible through development of pulse oximetry (which provides accurate information on blood oxygen saturation), laser Doppler blood-flowmetry, and near-infrared spectroscopy, renewed interest in optical imaging in the late 1990s. Research on the possibility of extending photon migration spectroscopy to imaging by solving the inverse problem was done in the late 1980s. Modern DOT imaging evolved from these various, disparate developments [7].

1.2 Outline

Resolving the DOT problem is a multi-stage procedure. As such, the next section will discuss the Forward Problem – simulation of flux measurements at the detectors. We ground the analysis in solid theoretical terms, providing analytical solutions to the forward problem in 2D by solving the steady-state diffusion approximation using Green's funtion for the Laplacian. The forward problem in 3D and in cases with non-zero baseline tissue absorption

are dealt with in the Appendix. Probing the forward problem allows us to gain a greater understanding of the eventual optimization process involving positions of sources and detectors.

In Section 3, we analyze the Inverse Problem, providing the theoretical basis for image reconstruction based on the measurements obtained through the forward model. We use Bayesian inference to model the inverse problem. The goal is to update our knowledge of absorption at voxels using our forward model measurements; we use the posterior distribution thus obtained in order to differentiate between different configurations of sources and detectors. Such differentiation allows us to identify the configuration that will reconstruct the image in the best possible manner. In this section we go into detail about the Singular Value Decomposition and how our work connects to Tikhonov regularization, a method for smoothing the data. In particular, we show that the reconstructed image obtained through Tikhonov regularization is the same as the posterior mean of a product of multivariate Gaussians under certain conditions. Lastly, we provide support for using the Volume Ratio (equivalent to the Occam factor in the literature [3]) of the posterior distribution in relation to the prior to evaluate the boost in information we receive from our measurements for a particular source-detector configuration.

Sections 4 and 5 provide a description of the experimental setup in MATLAB and a discussion of the results, respectively. In Section 4, we describe our setup for simulating different arrays of voxels, finding the measurement matrix, and, consequently, the singular values. In addition, we describe minimization of the objective function using $fmincon$, the MATLAB function used to find global minima of a function of several variables. In Section 5, we discuss the results of our simulations, showing images of different source-detector configurations over the voxel geometries and analyzing the relative improvement in image reconstruction of adding more sources and detectors. We conclude with a summary of our findings, applicability to real-world problems, and directions for future research.

2   The Forward Problem

At its heart, the forward problem involves the simulation of flux measurements at certain distances from the sources, given an optical model of the un-
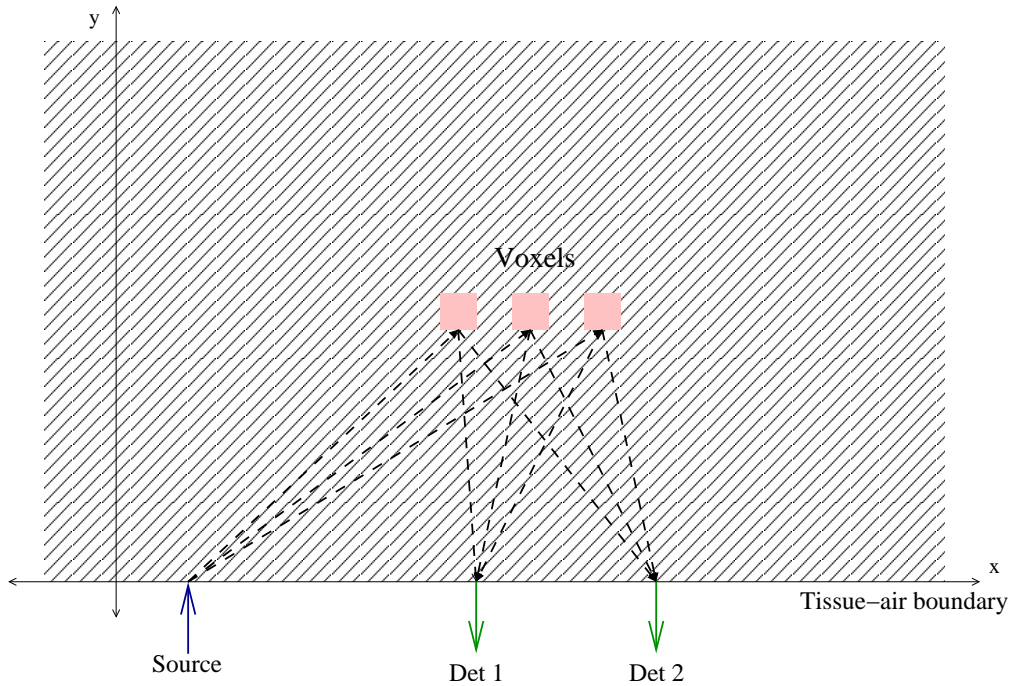
Figure 1: Source-Detector Pairs Setup

derlying tissue. In our case, the geometry of the forward problem is modeled using a semi-infinite slab with sources and detectors placed at the tissue-air boundary. In this section, we use a 2D representation of the model, with sources and detectors placed along the x-axis and the tissue region located in the upper-half plane (see Figure 1).

Real-world applications of DOT involve various geometries and a baseline absorption level ≥ 0 for the tissue. While our approach may seem limiting, we show in the Appendix that solving the 2D problem without baseline absorption provides a template for obtaining an analytic solution, allowing us to easily generalize to three dimensions and to cases with a non-zero baseline absorption.

In our representation, voxels, representing tissues with properties possibly differing from the homogeneous medium, are placed at a certain depth along the y-axis. Photon transport in a homogeneous, absorbing medium is well described by the time-dependent diffusion approximation (DA):

4

$$\nabla \cdot (\kappa \nabla u) - \mu_a(\mathbf{x})u = u_t \tag{1}$$

where $u(\mathbf{x})$ ($\mathbf{x} := (x, y)$) is the photon density, $\mu_a(\mathbf{x})$ is the local absorption function with units L^{-1}, and $\kappa(\mathbf{x})$ is the diffusion coefficient[1]. Since the physical system we are modeling reaches steady-state rather quickly, taking $u_t = 0$ in (1) we obtain the steady-state diffusion approximation:

$$\nabla \cdot (\kappa \nabla u) - \mu_a(\mathbf{x})u = 0 \tag{2}$$

The semi-infinite slab can be represented by $\Omega = (-\infty, \infty) \times (0, \infty)$. The measurement model for the detectors at some $\mathbf{r} = (x_d, 0)$ involves the normal derivative of the flux at the boundary [2].

$$m_d = \frac{du}{dy}(x_d, 0) \tag{3}$$

Using the [model approximation] homogeneous scattering medium $\kappa = 1$ (in suitable dimensionless units) and the Dirchlet approximation for tissue-air boundary conditions, (2) becomes:

$$\begin{cases} \Delta u - \mu_a(\mathbf{x})u = 0 \\ u \to 0 \text{ when } y \to \infty \\ u(x, 0) = 0 \ \forall x \end{cases} \tag{4}$$

In this problem, we will consider small changes from uniformity. Imagine a simple region has a μ_a change of functional form $F(\mathbf{x})$. Then, $\mu_a(\mathbf{x}) = \epsilon F(\mathbf{x})$, where $F(\mathbf{x})$ is the voxel shape function.

Doing perturbation theory for a general, small absorption change, we can postulate that the solution is of the form:

$$u = u_0 + \epsilon u_1 + O(\epsilon^2) \tag{5}$$

Substituting (5) into (4), we have the PDE:

$$\Delta(u_0 + \epsilon u_1 + ...) - \epsilon F(\mathbf{x})(u_0 + \epsilon u_1 + ...) = 0 \tag{6}$$

In (6) the zero-th order solution is just the unperturbed problem ($\mu_a(\mathbf{x}) = 0$), which amounts to solving Laplace's equation in the upper-half plane:

---

[1]Note: $\nabla \cdot (\kappa \nabla u) = \kappa \Delta u$ when κ is a constant.

$$\begin{cases} \Delta u = 0 \\ x \in \mathbb{R}, y > 0 \\ u(x,0) = f(x) \\ f(x) = \delta(x - x_s) \end{cases} \tag{7}$$

where x_s is the position of the source along the x-axis and δ is the Dirac delta function. We will now find the Green's function, which solves (7) for a general $f(\mathbf{x})$. Taking the Fourier transform of (7) on x with y as a parameter:

$$\begin{aligned} \widehat{u}_{yy} - \xi^2 \widehat{u} &= 0 & (8) \\ \widehat{u}(\xi, y) &= a(\xi)e^{-\xi y} + b(\xi)e^{\xi y} & (9) \\ &= c(\xi)e^{-|\xi|y} & (10) \end{aligned}$$

Using the boundary conditions from (7) we can see that

$$\widehat{u}(\xi, y) = \widehat{f}(\xi)e^{-|\xi|y} \tag{11}$$

Applying the Convolution Theorem:

$$\begin{aligned} u(x,y) &= \frac{y}{\pi} \frac{1}{x^2 + y^2} * f & (12) \\ &= \frac{y}{\pi} \int_{-\infty}^{\infty} \frac{f(\eta)d\eta}{(x-\eta)^2 + y^2} & (13) \end{aligned}$$

When $f(\eta) = \delta(\eta - x_s)$, we get the Poisson Kernel for the half-plane:

$$u_0(x,y) = \frac{y}{\pi} \frac{1}{(x - x_s)^2 + y^2} \tag{14}$$

this is u_0, the zeroth-order solution in (5).

It should be noted that (14) is actually twice the normal derivative at the boundary (due to using the Method of Images, explained later in this section) of the 2D free space Green's function. We now demonstrate this claim. As we know, the free space Green's function for the 2-dimensional Laplacian is given by

$$G(r) = \frac{1}{2\pi} \ln r \qquad (15)$$

where $r = \sqrt{x^2 + y^2}$ is the distance (not to be confused with the x and y coordinates used above). Taking the derivative of (15) with respect to y, multiplying by 2, and substituting in the source and voxel coordinates:

$$2\frac{d(\frac{1}{2\pi} \ln r)}{dy} = 2(\frac{1}{2\pi}\frac{y}{r^2}) \qquad (16)$$

$$= \frac{y}{\pi}\frac{1}{(x - x_s)^2 + y^2} \qquad (17)$$

See Appendix for a derivation of (15).

Next, at $O(\epsilon)$, we have Poisson's equation with homogeneous boundary conditions:

$$\begin{cases} \Delta u_1 = F(\mathbf{x})u_0 \\ \text{BCs: } u_1(x, 0) = 0 \end{cases} \qquad (18)$$

Note that by matching the $O(\epsilon)$ terms:

$$u_0(x, 0) + \epsilon u_1(x, 0) = 0 \; \forall x$$
$$\Rightarrow u_1(x, 0) = 0$$

Let $g(\mathbf{x}) = F(\mathbf{x})u_0$ be the right hand side of (18). The problem can then be solved by using the Green's function in two dimensions. Assume the Green's function for our BCs and $g(\mathbf{x})$ has the form $G_0(\mathbf{r}, \mathbf{s})$ where $\mathbf{r} = (x_1, y_1)$ and $\mathbf{s} = (x_2, y_2)$, in cartesian coordinates, and $G_0(\cdot, \mathbf{s})$ satisfies the boundary conditions for all $\mathbf{s} \in \Omega$. The condition for the Green's function is given by

$$\Delta_{\mathbf{r}} G_0(\mathbf{r}, \mathbf{s}) = \delta(\mathbf{r} - \mathbf{s}) \qquad (19)$$

If such a Green's function exists, we can solve (18). Using the above form of Green's function, the solution is

$$u_1(\mathbf{r}) = \int_\Omega G_0(\mathbf{r}, \mathbf{s})g(\mathbf{s})d\mathbf{s} \qquad (20)$$
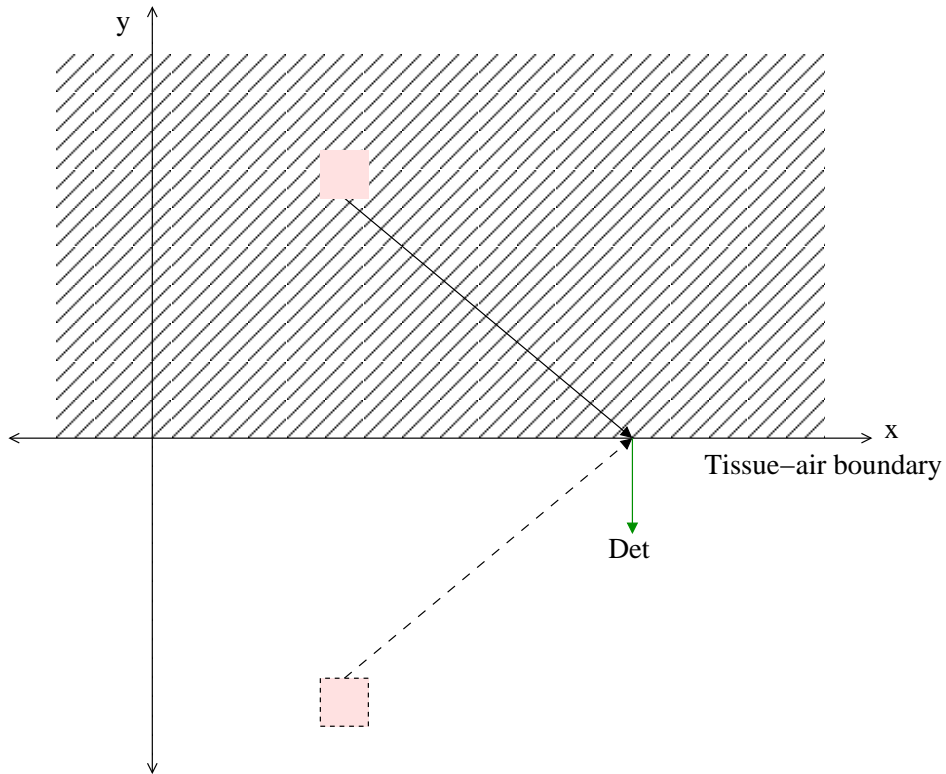
7

Figure 2: Method of Images

Since we know that $\Delta u_1 = g$, this implies

$$\Rightarrow \Delta u_1(\mathbf{r}) = \int \delta(\mathbf{r} - \mathbf{s})g(\mathbf{s})d\mathbf{s} = g(\mathbf{r}) \tag{21}$$

showing that (20) solves (18).

In free space, that is, if there were no boundary conditions, $G_0(\mathbf{r}, \mathbf{s}) = G_0(\mathbf{r} - \mathbf{s})$. Let $\rho = |\mathbf{r} - \mathbf{s}|$, then, as we saw before, the Green's function for the 2-dimensional Laplacian is

$$G_0(\rho) = \frac{1}{2\pi} \ln \rho \tag{22}$$

Referring back to (18), the BCs require that the Green's function is 0 when $y = 0$. As such, we use the Method of Images to find the solution.

Placing an imaginary voxel that emits photons to the detector in the lower half the plane, we make $G_0 = 0$ along the [x-axis] tissue-air boundary (Figure 2). We know that if $\mathbf{s} = (x_2, y_2)$, then $\mathbf{s}' = (x_2, -y_2)$. Which implies

$$G_0(\mathbf{r}, \mathbf{s}) = \frac{1}{2\pi} \ln(|\mathbf{r} - \mathbf{s}|) - \frac{1}{2\pi} \ln(|\mathbf{r} - \mathbf{s}'|) \tag{23}$$

Substituting (23) in (20), we get

$$
\begin{aligned}
u_1(\mathbf{r}) &= \frac{1}{2\pi} \int_\Omega (\ln(|\mathbf{r} - \mathbf{s}|) - \ln(|\mathbf{r} - \mathbf{s}'|)) g(\mathbf{s}) d\mathbf{s} \tag{24} \\
&= \frac{1}{2\pi} \int_\Omega (\ln(|\mathbf{r} - \mathbf{s}|) - \ln(|\mathbf{r} - \mathbf{s}'|)) F(\mathbf{s}) u_0(\mathbf{s}) d\mathbf{s} \tag{25}
\end{aligned}
$$

Going back to (19), we now make a simplifying assumption about the form of the voxel shape function, namely $F(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{r}_v)$. This corresponds to assuming each voxel is much smaller than other distances in the problem. This enables us to remove the integral using the Delta function's sifting property and substituting $\mathbf{r} = (x_1, y_1)$, $\mathbf{r}_v = (x_v, y_v)$, and $\mathbf{r}'_v = (x_v, -y_v)$, and since contributions for many different voxels can be added together in the 1st order case, we see that the solution for (18) is

$$u_1(\mathbf{r}) = \frac{1}{2\pi} (\ln(|\mathbf{r} - \mathbf{r}_v|) - \ln(|\mathbf{r} - \mathbf{r}'_v|)) u_0(\mathbf{r}_v) \tag{26}$$

The actual measurement at the detector positioned at $\mathbf{r} = (x_d, 0)$ can be obtained from the measurement model (3).

$$m_d = \frac{1}{2\pi} \frac{d}{dy_1} [(\ln[(x_1 - x_v)^2 + (y_1 - y_v)^2] - \ln[(x_1 - x_v)^2 + (y_1 + y_v)^2]) \frac{1}{\pi} u_0(\mathbf{r}_v)] \tag{27}$$

Defining

$$h_1(\mathbf{r}, \mathbf{r}_v) := \ln[(x_1 - x_v)^2 + (y_1 - y_v)^2] \tag{28}$$

the derivative is

$$\frac{dh_1}{dy_1} = \frac{2(y_1 - y_v)}{(x_1 - x_v)^2 + (y_1 - y_v)^2} \tag{29}$$

Evaluating this at $x_1 = x_d$ and $y_1 = 0$

$$\frac{dh_1}{dy_1} = \frac{-y_v}{(x_d - x_v)^2 + y_v^2} \tag{30}$$

Similarly, defining

$$h_2(\mathbf{r}, \mathbf{r}'_v) := -\ln[(x_1 - x_v)^2 + (y_1 + y_v)^2] \tag{31}$$

the derivative is

$$\frac{dh_2}{dy_1} = -\frac{2(y_1 + y_v)}{(x_1 - x_v)^2 + (y_1 + y_v)^2} \tag{32}$$

Evaluating this at $x_1 = x_d$ and $y_1 = 0$

$$\frac{dh_2}{dy_1} = \frac{-y_v}{(x_d - x_v)^2 + y_v^2} \tag{33}$$

Finally, substituting this back into (27) gives us the simplified expression for the linearized sensitivity of a measurement to changes in ϵ, the absorption perturbation:

$$m_d = -\frac{1}{\pi^2} \frac{y_v}{(x_d - x_v)^2 + y_v^2} \frac{y_v}{(x_v - x_s)^2 + y_v^2} \tag{34}$$

where m_d is the signal detected at a detector placed at x_d from a voxel at $\mathbf{r}$ that is hit by light from a source at x_s. Note that (34) has the form

$$m_d = -u_0{}^{(s)}(\mathbf{r}_v) u_0{}^{(d)}(\mathbf{r}_v) \tag{35}$$

where $u_0{}^{(s)}(x, y)$ is the same as u_0 from before and evaluating $u_0{}^{(d)}$ simply involves replacing x_s with x_d. This is a useful relation that can be generalized for other geometries and for cases with a positive baseline absorption. See Appendix for details.

Using the formula given in (34), we can construct our measurement matrix, which we will call A. This matrix simply measures how changes in μ_a at the voxels cause changes in the detected signals.

Say there are N voxels with locations $\mathbf{r_j}$, $j = 1 \ldots N$. If $\mathbf{w}$ is a vector of absorption changes at the voxels, that is, $F(\mathbf{r}) = \sum_{\mathbf{j}} \mathbf{w_j} \delta(\mathbf{r} - \mathbf{r_j})$, then the first-order perturbation result derived above implies that $\mathbf{z}$, the change in measurement vector, is $\mathbf{z} = \mathbf{Aw}$. The dimensions of A correspond to

the number of source-detector pairs (rows) and the total number of voxels (columns). Each entry of the matrix – $A(i, j)$ – corresponds to the ith source-detector pair for an individual voxel (j).

3   The Inverse Problem

The inverse problem involves using the forward problem iteratively to get information about $\mu_a(\mathbf{x})$ – the absorption at the actual voxel positions (or image) – using the measurements (34) from and knowledge of the forward problem. In our case, it is a series of source-detector measurements of the absorption at the voxels. Since we restricted to first order perturbation (linear changes), the inverse problem is a linear one.

The inverse problem can be described as follows:

$$\mathbf{z} = A\mathbf{w} + \mathbf{n} \tag{36}$$

$$\mathbf{n} = \mathbf{z} - A\mathbf{w} \tag{37}$$

Here, $\mathbf{z}$ is a column vector of the measurements, of which there are as many as there are source-detector pairs; $\mathbf{w}$ is a column vector of μ_a values which multiply delta functions at the voxels; and $\mathbf{n}$ is a column vector of the measurement noise.

We will use a Bayesian framework in order to model the inverse problem. At its heart, the Bayesian process quantifies our knowledge about the validity of a hypothesis before evidence to support it has been collected, and then updates this estimate using the collected evidence. Bayesian inference adjusts probabilities with the following relation

$$p(hypothesis|data, \text{ I}) \quad \propto \quad p(data|hypothesis, \text{ I}) \times p(hypothesis| \text{ I}) \tag{38}$$

$$p(H|D, \text{ I}) \quad = \quad \frac{p(D|H, \text{ I}) \times p(H| \text{ I})}{p(D| \text{ I})} \tag{39}$$

where I is our assumed model. Our unknowns or hypothesis (in this case, the absorption measurements at the voxels) have a posterior probability distribution function (PDF), which is obtained by multiplying the prior PDF –

our state of knowledge about the problem – with a likelihood function (the conditional probability function) which uses our data measurements.

The likelihood function for additive noise is given by a Gaussian in $\mathbf{n}$, that is:

$$p(\mathbf{z}|\mathbf{w}, \sigma^2) = \frac{1}{(2\pi)^{N/2}\sqrt{\det \Lambda}} e^{\frac{-1}{2}(\mathbf{z}-A\mathbf{w})^T \Lambda^{-1}(\mathbf{z}-A\mathbf{w})} \tag{40}$$

where the noise has zero mean and a covariance matrix Λ. A common model is that (37) is normally distributed with variance β^2, that is, $\mathbf{n} \sim N(0, \beta^2)$. When the covariance matrix is of the simpler form $\Lambda = \beta^2 I$, we have the following form

$$\frac{1}{(2\pi\beta^2)^{N/2}} e^{\frac{-1}{2\beta^2}|\mathbf{z}-A\mathbf{w}|^2} \tag{41}$$

In order to evaluate the boost in information we receive, in updating from the prior to the posterior PDF, we evaluate the *volume* of the posterior in relation to that of the prior. Although a Gaussian PDF does not have compact support, it only has significant size in a certain region of parameter space. The volume of this region we may define as follows.

Definition. *The volume of a N-dimensional Gaussian Probability Density function ($p_g(\mathbf{x}) = \frac{1}{(2\pi)^{N/2}\sqrt{det\Gamma}} e^{\frac{-1}{2}\mathbf{x}^T\Gamma^{-1}\mathbf{x}}$) is the square-root of the determinant of its covariance matrix (Γ)*

$$\text{volume}(p_g(\mathbf{x})) = \sqrt{\det \Gamma} \tag{42}$$

This concept of volume has some intuitive appeal, as it is not only 1/max value of the PDF, it is proportional to the volume of a multi-dimensional ellipsoid.

An ellipsoid is obtained when a matrix acts upon a multi-dimensional sphere. For any ellipse, the area is proportional to the lengths of its semi-axes. Similarly, for an m-dimensional ellipsoid the volume is proportional to the lengths of its m semi-axes. The length of the semi-axes for an ellipsoid obtain by acting upon the unit sphere N with an $m \times m$ matrix Γ (covariance matrices are by definition square matrices) is given by the square root of the eigenvalues (λ_i) of Γ; this is a standard result in the literature. The volume
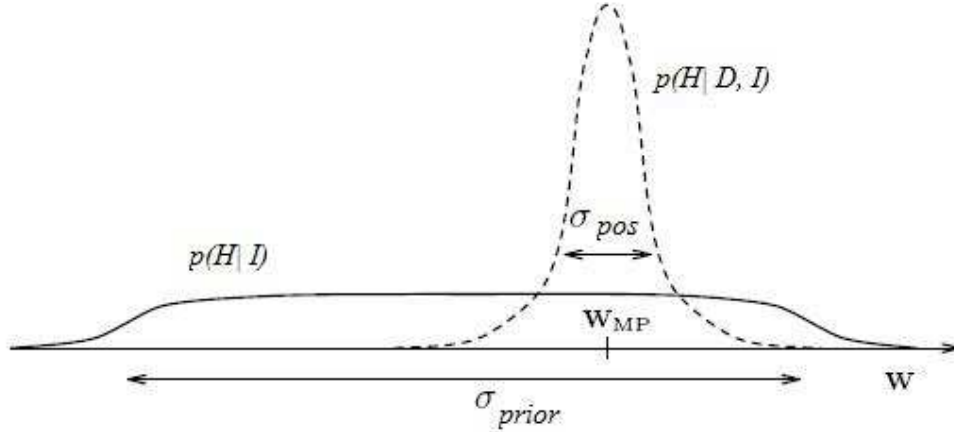
Figure 3: Occam Factor

of the corresponding ellipsoid ΓN is proportional to the product of those lengths:

$$\sqrt{\det \Gamma} = \prod_i \sqrt{\lambda_i} \tag{43}$$

The ratio of the volume of the prior PDF to the posterior PDF is a natural measure for evaluating the boost in information upon measurement of the data. If we take σ_{pos} (proportional to covariance matrix of the posterior) to be the posterior uncertainty in the reconstructed image and σ_{prior} to be the prior uncertainty in the image, we can define the ratio as the Occam factor [3]

$$\text{Occam Factor} = \frac{\sigma_{pos}}{\sigma_{prior}} \tag{44}$$

David MacKay [3] describes the Occam factor as follows

The Occam factor is equal to the ratio of the posterior accessible volume of I's parameter space to the prior accessible volume, or the factor by which I's hypothesis space collapses when the data arrive.

13

Note that we do not believe this definition of imaging success has been applied in DOT before. Using the above definition (42), we can calculate the volume for our posterior PDF. The image can be described as a set of absorption measurements μ_a at the voxels ($\mathbf{w}$ in $\mathbb{R}^N$) which are unknown. What we do know is the measurement matrix (A): how changes in μ_a at the voxels cause changes in the detected signals. Given S sources and D detectors, we get $S \times D$ measurements at each voxel from our array of source-detector pairs that form the components of the matrix A.

Assuming a Gaussian prior with a mean centered at $\mathbf{w}_0$ and covariance matrix Γ, we obtain the posterior distribution by multiplying it with the maximum likelihood function (41).

$$p(\mathbf{w}|\mathbf{z}, \beta^2) \propto e^{-\frac{1}{2}(\mathbf{w}-\mathbf{w}_0)^T\Gamma^{-1}(\mathbf{w}-\mathbf{w}_0)}e^{-\frac{1}{2}(\mathbf{z}-A\mathbf{w})^T\Lambda^{-1}(\mathbf{z}-A\mathbf{w})} \tag{45}$$

$$\Rightarrow p(\mathbf{w}|\mathbf{z}, \beta^2) \propto \exp{-\frac{1}{2}\left[\begin{array}{l} \mathbf{w}^T\Gamma^{-1}\mathbf{w} - \mathbf{w}_0{}^T\Gamma^{-1}\mathbf{w} - \mathbf{w}^T\Gamma^{-1}\mathbf{w}_0 \\ +\mathbf{w}_0{}^T\Gamma^{-1}\mathbf{w}_0 + \mathbf{z}^T\Lambda^{-1}\mathbf{z} - \mathbf{w}^TA^T\Lambda^{-1}\mathbf{z} \\ -\mathbf{z}^T\Lambda^{-1}A\mathbf{w} + \mathbf{w}^TA^T\Lambda^{-1}A\mathbf{w} \end{array}\right]}$$

We can ignore the prefactors since they can always be worked out later given the covariance matrix for a normalized Gaussian PDF. As such, we focus on the quadratic form in the exponent (quadratic function of $\mathbf{w}$).

The term in the exponent can be simplified as follows:

$$\mathbf{w}^T(\Gamma^{-1} + A^T\Lambda^{-1}A)\mathbf{w} - 2(\mathbf{w}_0\Gamma^{-1} + \mathbf{z}^T\Lambda^{-1}A)\mathbf{w} + (\mathbf{w}_0{}^T\Gamma^{-1}\mathbf{w}_0 + \mathbf{z}^T\Lambda^{-1}\mathbf{z}) \tag{46}$$

The last term is a constant. Since we are multiplying two Gaussians, the product – the posterior distribution – will also take the form of a Gaussian. The analogy, in this case, is completing the square. Since the exponent in (45) is of a quadratic form, the posterior will have be of quadratic form, implying the existence of a Gaussian distribution for the posterior.

Let us assume the posterior [having unknown mean value $\mathbf{w}_p$ and unknown variance Γ_p] is of the following form:

$$e^{-\frac{1}{2}(\mathbf{w}-\mathbf{w}_p)^T\Gamma_p^{-1}(\mathbf{w}-\mathbf{w}_p)} \tag{47}$$

Rewriting the exponent in the posterior as follows,

$$\mathbf{w}^T\Gamma_p^{-1}\mathbf{w} - 2\mathbf{w}_p\Gamma_p^{-1}\mathbf{w} + \mathbf{w}_p{}^T\Gamma_p^{-1}\mathbf{w}_p \tag{48}$$

we can see that the last term is a constant. Matching terms between the (46) and (48) we get,

$$\Gamma_p^{-1} = \Gamma^{-1} + A^T\Lambda^{-1}A \tag{49}$$

$$\mathbf{w_p} = (\Gamma^{-1} + A^T\Lambda^{-1}A)^{-1}(\mathbf{w}_0\Gamma^{-1} + \mathbf{z}^T\Lambda^{-1}A) \tag{50}$$

These are the rules for Bayesian updating within a Gaussian Approximation [4].

3.1 Calculating the Volume Ratio

Given (50), we are basically updating the prior to the posterior covariance matrix. If we choose $\Lambda = \beta^2 I$, implying constant variance (noise) at each measurement, we update $\Gamma_p^{-1} = \Gamma^{-1} + \beta^{-2}A^TA$. We will assume this case from now on. Our aim is to find the volume of this posterior distribution using the volume definition (42) from above.

In order to find the volume, we must first simplify the term A^TA in (50). This can be done using the Singular Value Decomposition of the matrix A, since every matrix has a SVD [5].

3.1.1 Singular Value Decomposition

The motivation for the Singular Value Decomposition comes from considering the unit sphere in n-dimensional space. Any $m \times n$ matrix M will map this sphere onto an ellipsoid in m-dimensional space. We define σ_i as the length of a semi-axis of this ellipsoid and u_i as the unit vector in the direction that semi-axis. There are $r = rank(M)$ different lengths of the semi-axes – known as singular values $(\sigma_1, ..., \sigma_r)$ – of the ellpisoid. There are just as many directions $(u_1, ..., u_r)$, known as the left singular vectors of M. The original matrix also has a set of preimages of the axes $(v_1, ..., v_r)$ defined as the right singular vectors of M [5]. Given these values, the relation

$$Mv_i = \sigma_i u_i \tag{51}$$

is satisfied. The SVD factorizes any $m \times n$ matrix M into three parts:

$$M = U\Sigma V^* \tag{52}$$

- A matrix U in $\mathbb{C}^{m \times m}$ consisting of m orthonormal columns that comprise the left singular vectors.

- A matrix V in $\mathbb{C}^{n \times n}$ consisting of n orthonormal columns that comprise the right singular vectors.

- A diagonal matrix Σ in $\mathbb{R}^{m \times n}$ consisting of the singular values in decreasing order.

Note that U and V are simply the vectors $u_1, ..., u_r$ and $v_1, ..., v_r$ with, respectively, $m - r$ and $n - r$ additional columns. The matrix Σ only has r positive diagonal entries, with the rest being 0. Also, both U and V are unitary matrices, which implies

$$U^T = U^{-1}$$
$$V^T = V^{-1}$$

Using this result, we can decompose our measurement matrix A and simplify $A^T A$. The following Lemma helps us do that [5].

Lemma 3.1. *For an $m \times n$ matrix A, the determinant of $A^T A$ is the product of the square of the singular values of A.*

Proof. Using the SVD of A, we can decompose $A^T A$.

$$A^T A = (U\Sigma V^T)^T (U\Sigma V^T) = V\Sigma^T U^T U \Sigma V^T = V\Sigma^T \Sigma V^T$$

Since V is a unitary matrix, its determinant is 1. Taking the determinant of the last expression, we obtain

$$\det A^T A = \det V\Sigma^T \Sigma V^T = \det V \det \Sigma^T \Sigma \det V^T = \det \Sigma^T \Sigma = \prod_i \sigma_i^2$$

$\square$

Using Lemma 3.1, we return to (42) and calculate the volume of the posterior. If our prior distribution is characterized as $\Gamma = \upsilon^2 I$ and our noise model by $\Lambda = \beta^2 I$, then our update rule is as follows:

$$
\begin{aligned}
\Gamma_p{}^{-1} &= \upsilon^{-2}I + \beta^{-2}A^T A & (53) \\
&= \beta^{-2}(\frac{\beta^2}{\upsilon^2}I + A^T A) & (54)
\end{aligned}
$$

Defining $\epsilon^2 := \frac{\beta^2}{\upsilon^2}$ in our case, we calculate volume of the posterior distribution. The following lemma helps us find the singular values when a constant is added to the diagonal of a matrix .

Lemma 3.2. *For a $m \times m$ matrix M, adding cI, where I is the identity matrix and c is a constant, shifts each eigenvalue by c.*

Proof. Taking the SVD of M:

$$
M = V D V^T
$$

D consists of the ordered eigenvalues of M along the diagonal. We also know that VV^T is the just identity matrix, so

$$
\begin{aligned}
M + cI &= V D V^T + cI \\
&= V(D + cI)V^T
\end{aligned}
$$

Thus, each eigenvalue is shifted by c. $\qquad\square$

As we know, $\det A$ is the product of the eigenvalues of A and that the eigenvalues of $A^T A$ are the squares of the singular values of A. Ignoring prefactors as before

$$\begin{align}
volume(\Gamma_p) &= volume(\Gamma_p^{-1})^{-1} \tag{55} \\
&= \frac{1}{\sqrt{\det \beta^{-2}(\epsilon^2 I + A^T A)}} \tag{56} \\
&= \frac{\beta^N}{\sqrt{\prod_i (\lambda_i(A^T A) + \epsilon^2)}} \tag{57} \\
&= \frac{\beta^N}{\sqrt{\prod_i (\sigma_i^2(A) + \epsilon^2)}} \tag{58} \\
&= \prod_{i=1}^{N} \frac{\beta}{\sqrt{\sigma_i^2(A) + \epsilon^2}} \tag{59}
\end{align}$$

The term $\sigma_i^2(A)$ represents the square of the singular values of A and (59) is our expression for the volume of the posterior PDF.

As we mentioned before, the Occam factor quantifies the boost in information upon measurement of the data. As such, in order to find the volume ratio, we need to find the volume of the prior PDF. Since our prior PDF is characterized as $\Gamma = \upsilon^2 I$, we can use (42) to find the volume of our prior:

$$\begin{align}
volume(\Gamma) &= \sqrt{\det \Gamma} \tag{60} \\
&= \sqrt{\prod_{i=1}^{N} \upsilon^2} \tag{61} \\
&= \prod_{i=1}^{N} \upsilon \tag{62}
\end{align}$$

Since we have analytical expressions for both the prior and posterior volumes, we can calculate the volume ratio as follows:

18

$$\text{Volume Ratio} \quad = \quad \left(\frac{\text{volume of posterior}}{\text{volume of prior}} \right) \tag{63}$$

$$= \quad \left(\prod_{i=1}^{N} \frac{\frac{\beta}{\upsilon}}{\sqrt{\sigma_i{}^2(A) + \epsilon^2}} \right) \tag{64}$$

$$= \quad \prod_{i=1}^{N} \frac{\epsilon}{\sqrt{\sigma_i{}^2(A) + \epsilon^2}} \tag{65}$$

We can use this number as a guide to the improvement obtained through different source-detector configurations. If we rewrite (65) as

$$\prod_{i=1}^{N} \frac{\epsilon}{\sigma_i} \sqrt{\frac{\sigma_i{}^2}{\sigma_i{}^2 + \epsilon^2}} \tag{66}$$

then the expression underneath the square-root is the Tikhonov filter function (which we explain in the next section) with parameter ϵ^2.

3.2 Connection to Tikhonov Regularization

Under certain conditions the posterior mean obtained using Bayesian inference is the same as the regularized approximation to the true image obtained by doing Tikhonov regularization on the singular values. If we return to our data model from (37), we know that $\mathbf{w}$ represents the *true image* whereas $\mathbf{n}$ is the additive Gaussian noise or error in the data (recall, the noise had a mean of zero and a variance of β^2). As we have previously established, all matrices have a singular value decomposition .

The SVD for our measurement matrix A can be represented as follows,

$$A = U\Sigma V^T \tag{67}$$

where Σ is a diagonal matrix with positive, decreasing singular values, which we can represent as σ_i. Using the properties of the SVD discussed earlier, we know that the pseudo-inverse of A [5] can be represented as,

$$A^{-1} = V\Sigma^{-1}U^T \tag{68}$$

Returning to (37), we multiply both sides by A^{-1} [9]

19

$$V\Sigma^{-1}U^T\mathbf{w} = \mathbf{z} + \sum_{i=1}^{n} \sigma_i^{-1}(\mathbf{u}_i^T\mathbf{n})\mathbf{v}_i \tag{69}$$

As we can see in (69), we are dividing by very small singular values, often very close to zero. The consequent multiplication by large factors distorts the answers and can lead to meaningless results in practice. Thus, it is desirable to somehow remedy this situation by modifying the small singular values. One way to do it is by regularizing the singular values using a filter function known as the Tikhonov Filter Function, defined as

$$\omega_\alpha(\sigma^2) = \frac{\sigma^2}{\sigma^2 + \alpha} \tag{70}$$

The α in (70) is called the *regularization parameter*. One can think of it as determining a sort of *cut-off* for the filter. The size of α directly correlates with the degree of filtering, with a small α filtering out little noise and large values of α filtering out most of it. Of course, as with any filter, Tikhonov regularization filters out a good part of the solution and smooths the function too much at high values of α. The correct value of α to use is generally unknown. Armed with this filter, we multiply $\omega_\alpha(\sigma^2)$ with the singular values in (69) and obtain an approximation $\mathbf{w}_p$ for $\mathbf{w}$ [9],

$$\begin{aligned}
\mathbf{w}_p &= V\omega_\alpha(\sigma^2)\Sigma^{-1}U^T\mathbf{z} & (71) \\
&= \sum_{i=1}^{n} \frac{\sigma_i(\mathbf{u}_i^T\mathbf{z})\mathbf{v}_i}{\sigma_i^2 + \alpha} & (72) \\
&= (A^TA + \alpha I)^{-1}A^T\mathbf{z} & (73)
\end{aligned}$$

Note that the second term in (69) has been dropped since $\mathbf{n}$ has a mean of 0. The equation in (73) seems quite familiar, and it should, since under certain assumptions, it is equivalent to (50), the posterior mean obtained through Bayesian inference.

Theorem 3.1. *The mean of the posterior distribution obtained by doing Bayesian inference with constant variance in the noise level ($\Lambda = \beta^2 I$), constant variance in the prior ($\Gamma = v^2 I$), and a zero mean in the prior ($\mathbf{w}_0 = \mathbf{0}$) is equivalent to the regularized approximation to the true image obtained by Tikhonov regularization using a Tikhonov filter function with regularization parameter $\alpha = \frac{\beta^2}{v^2}$.*

Proof. The regularized approximation to $\mathbf{w}_p$ using the Tikhonov filter function is shown in (73). Going back to (50), we substitute in $\Gamma^{-1} = v^{-2}I$, $\Lambda^{-1} = \beta^{-2}I$, and $\mathbf{w}_0 = \mathbf{0}$. Solving for $\mathbf{w}_p$, we obtain.

$$
\begin{aligned}
\mathbf{w}_p &= (\frac{1}{v^2}I + \frac{1}{\beta^2}IA^TA)^{-1}(\mathbf{0} + \frac{1}{\beta^2}\mathbf{z}^TA) \\
&= \beta^2(A^TA + \alpha I)^{-1}\frac{1}{\beta^2}(A^T\mathbf{z}) \\
&= (A^TA + \alpha I)^{-1}A^T\mathbf{z}
\end{aligned}
$$

Thus, with the given noise and prior models, the posterior mean obtained by Bayesian inference is equivalent to approximation obtained through Tikhonov regularization.

$\square$

4 Optimization of the Source and Detector Locations

As we have mentioned, *volume* of the posterior gives a measure of the quality of the reconstructed image or the information context of the measurement process. We are concerned with the ratio of the volume of the posterior to the prior in order to optimize across sources and detector locations. Better configurations will yield a lower ratio, as the posterior volume contracts and the prior volume remains the same.

If $N = \text{rank}(A)$, then we can take the log of (65) in order to specify our objective function as follows:

$$
\text{Objective function} = J(A) = \log\left(\prod_{i=1}^{N} \frac{\epsilon}{\sqrt{\sigma_i^2(A) + \epsilon^2}}\right) \tag{74}
$$

$$
= \sum_{i=1}^{N} \log\left(\frac{\epsilon}{\sqrt{\sigma_i^2(A) + \epsilon^2}}\right) \tag{75}
$$

We choose a prior PDF of order 1, implying a prior volume of 1:

$$\nu \sim 1 \Rightarrow \det \Gamma = 1$$

This simplifies (75), giving us our final objective function

$$\sum_{i=1}^{N} \log \left(\frac{\beta}{\sqrt{\sigma_i{}^2(A) + \beta^2}} \right) \tag{76}$$

We want to make this expression more negative in magnitude, as that would indicate a smaller number (hence a tighter posterior volume) in the log expression. The Gaussian noise β is a key parameter. For our model and apparatus, this measurement error is roughly $\beta = 10^{-3}$. This value comes from real-world uses of the DOT apparatus.

4.1 Generating the Singular Values

In order to simulate real-world measurements, we wrote a program in MAT-LAB to simulate the geometry of the problem and model different source and detector configurations. The problem was simulated in both 2D and 3D geometries, with and without a baseline absorption. As we will see later, the presence of baseline absorption did not significantly affect the optimum configurations. Since our objective function takes singular values as inputs, our first step is to generate those values.

4.1.1 Voxel Configurations

While there are several possible configurations of voxels that can be encountered in the real world, we wanted configurations that could be generalized or would allow us to test particular properties. As such, we restrict the analysis to four particular configurations that we believe are representative of both real-world applications and the overall problem.

1. In the 2D case, we optimize over a 5×5 grid consisting of 25 voxels equally spaced in the square with corners (-0.5, 1), (-0.5, 2), (0.5, 2), and (0.5, 1) (see Figure 4a).

2. In 3D, we first optimize over a $5 \times 5 \times 5$ cubic lattice consisting of 125 voxels contained within a box of height 1 unit at depth 1 unit below the surface, with length and breadth given by the square used above (see Figure 4b).

3. In 3D, we remove depth from the above lattice and use a voxel sheet at depth 0.5 units below the surface, with length and breadth same as above (see Figure 4c). This is simply the first sheet of the above lattice at a depth of 1 unit.

4. Lastly, in 3D, we use a column of 20 voxels centered at the origin, equally placed at depths between 1 and 3 along the z-axis (see Figure 4d).

The voxel grid and lattice are directly applicable to brain imaging, where researchers are often interested in the properties of a small region below the surface. The voxel column allows us to explore pure depth sensitivity.

These voxel configurations are centered at the origin in order to highlight the symmetries we see in our result. As such, these configurations can be translated to anywhere else in the plane (or space) to yield identical results.

4.1.2 Filling the Measurement Matrix

Any set of singular values correspond to a certain measurement matrix A. The first step was to fill the A measurement matrix, which corresponded to a certain source-detector configuration and a particular voxel grid. As such, we wrote a program that takes cartesian coordinates for the source and detector locations and the voxel grid as inputs and returns the singular values of the measurement matrix as outputs. Using the measurement model (35), the program computes the measurement at each voxel for a particular source-detector pair and fills in the corresponding $A(i, j)$ entry in the measurement matrix where i is the source-detector pair and j is the voxel.

4.1.3 Normalization of Measurements Based on Distance

Having computed the measurement matrix, we proceed to *normalize* each row of the matrix by multiplying each entry of the row by a factor that scales the values based on the distance from the source to detector. This factor is an attempt to model the intrinsic gain of the system, which is defined for each source-detector pair. This scaling models the correct form of light intensity, preventing a nearby source from flooding a detector with light signals. Defining the factor as $\eta_{s_i d_i}$ for a source-detector pair i

$$\text{Factor}_{S-D} = \eta_{s_i d_i} = \frac{1}{\frac{1}{(x_s - x_d)^2 + (y_s - y_d)^2} + 1} \tag{77}$$

(a) 5×5 Voxel Grid

(b) $5 \times 5 \times 5$ Voxel Lattice

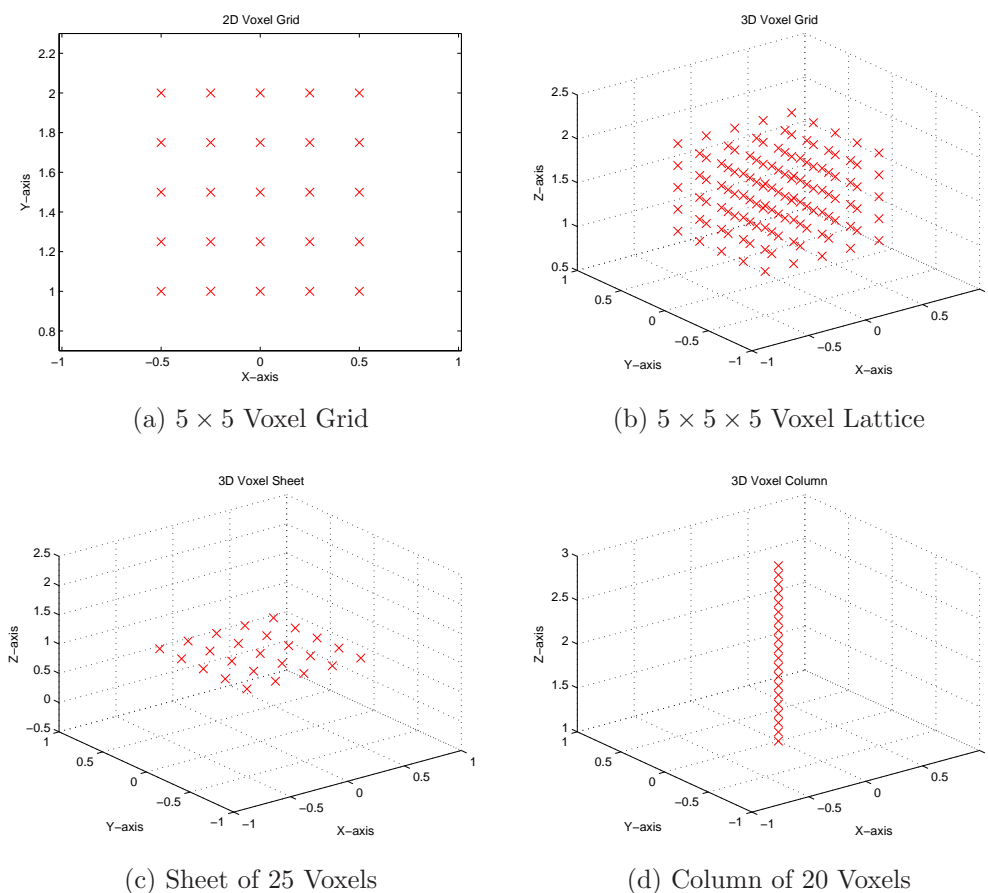(c) Sheet of 25 Voxels

(d) Column of 20 Voxels

Figure 4: Voxel configurations

where (x_s, y_s) and (x_d, y_d) are the coordinates of the particular source-detector pair that corresponds to that row[2]. We then proceed to normalize the measurement matrix:

$$\widetilde{A_{ij}} = \eta_{s_i d_i} A_{ij} \ \forall i, j \tag{78}$$

We can see the evolution of the gain factor in 2D and 3D as the distance between the particular source and detector increases (see Fig. 5). The behaviour is similar to what we see in the real world, as we cannot set the gain

---

[2]Note: y_s and y_d are both zero in the 2D case.

24

(a) 2D Normalization   (b) 3D Normalization
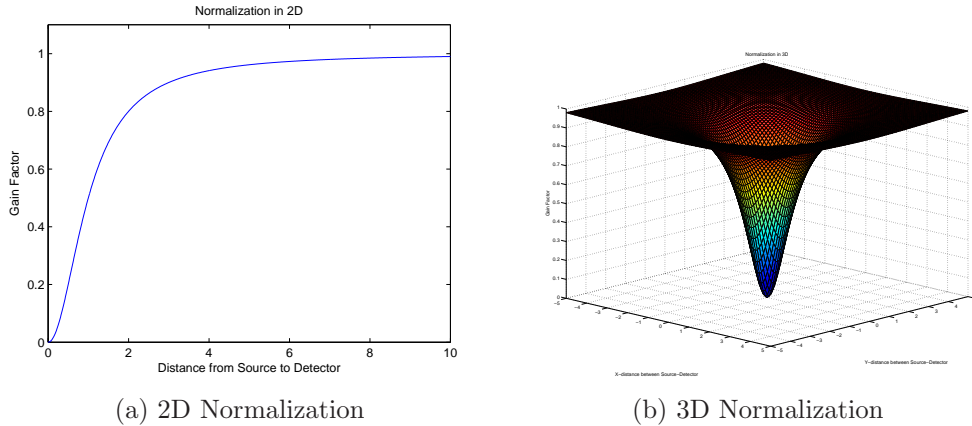
Figure 5: Normalization as a function of distance between source and detector

arbitrarily high. From here on, A will refer to the normalized measurement matrix.

4.2   Minimizing the Objective Function

Having found the singular values for a particular source-detector configuration, we proceed to find the volume ratio corresponding to that configuration using (76). In order to find the unique minima, we use the function optimization algorithm $fmincon$ built into MATLAB, which finds the constrained minimum of a function of several variables. If $\mathbf{x}$ is a vector of source-detector coordinates corresponding to a particular configuration and $F_{obj}(\mathbf{x})$ is our objective function, then $fmincon$ uses a given start point vector $\mathbf{x}_0$ in order to find the minimum subject to specified parameters and constraints.

For our optimization, we constrain our solutions to realistic bounds of $x \in$ [-10, 10] in the 2D case and $x \in$ [-10, 10] and $y \in$ [-10, 10] in the 3D case. The baseline absorption is set to 0.1 dimensionless units, which is realistic for a homogeneous tissue medium. In addition, we specify a tolerance of 10^{-6} for both $\mathbf{x}$ and the evaluated function $F_{obj}(\mathbf{x})$ at that minima. Basically, we can identify two different configurations and their respective values of calculated objective functions to an accuracy of 10^{-6}. Since the function $fmincon$ finds minima through repeated iteration, results can differ from one iteration to the next. As such, we find the minima for $n = 3000$ iterations in the 2D cases

and $n = 1000$ in the 3D cases, designating the configuration that minimizes the objective function to the greatest degree as the *Best Configuration.* The above values of n have been chosen based on repeated experimentation aimed at finding a reasonable number of iterations required to find the unique global minima using $fmincon$. As such, although we can identify the unique global minima to a high degree of certainty, we are not as certain about the 2nd and 3rd best solutions. Repeated iterations, perhaps of several magnitudes greater, are required to pin down solutions close to the first-best.

In addition, the program found many duplicate solutions due to the inherent symmetry of our setup. For example, in 2D, source 1 at x_1 and source 2 at x_2 is exactly the same as source 2 at x_1 and source 1 at x_2; the is true for detectors. We report only one of the many solutions, making it clear that a switch between source and detector positions along the x-axis will yield a similar result. Similarly, the optimum solution in 3D often had rotational symmetry; its many variants were excluded from the results and treated as a single solution. Thus, the 2nd and 3rd best solutions are both spatially distinct from the first-best.

5   Results

Using the setup described above, we simulate optimization for our different voxel configurations. This section is divided into subsections based on the configuration being used. Note that in the images, a red 'X' denotes a voxel, while a blue '*' denotes a source, and an 'O' denotes a detector.

Detector   ○

Source   ✳

Voxel   ✕

Figure 6: Legend

26
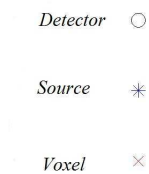
5.1 Voxel Configuration: 5×5 Grid

We first find the optimum source-detector placement in the case of 2 sources and 2 detectors in 2D. The three best placement schemes are shown in Figure 7.

In Table 1, we can see the locations of the sources and detectors (along the x-axis) for each configuration. We can see that the best solution has a remarkable symmetry; indeed, several variants of the first-best with switched coordinates for sources and detectors turned up as optimal solutions. While the 2nd and 3rd best solutions aren't entirely symmetric, they nevertheless aren't too far from the positioning of the best configuration. Indeed, all configurations have both a source and detector on either side of the y-axis.

Since the objective function is measuring the collapse of the prior distribution over the posterior, a natural way to compare different configurations is to compare by how much they collapse the parameter space as compared to the best configuration. If we denote the value of the objective function for the best configuration as F_{best} and for any other configuration i as F_i, then the value *Percent of Best* column is calculated using the following formula:

$$\text{Percent of Best}_i = 100 \times \exp(F_{best} - F_i)$$

As such, the second column in Table 1 shows how well the 2nd and 3rd best configurations compare to the first-best.

Rank	Percent of Best	Obj. Function	S (1)	S (2)	D (1)	D (2)
1	100	-12.9817	-1.3908	0.4257	-0.4257	1.3908
2	95.4493	-12.9351	-1.395	0.3942	-0.3942	1.2231
3	94.7019	-12.9272	-0.3346	1.5158	-1.298	0.4158

Table 1: Optimum configurations for a 5×5 grid with 2 sources and 2 detectors, no baseline absorption

The simulation is repeated with a baseline absorption and results are shown in Table 2. As we can see in Table 2, adding a baseline absorption of $\mu_a = 0.1$ does not change the first-best results markedly, as the locations for the optimum configuration are less than 1 percent away from the optimum configuration without baseline absorption. As we expect, we get better
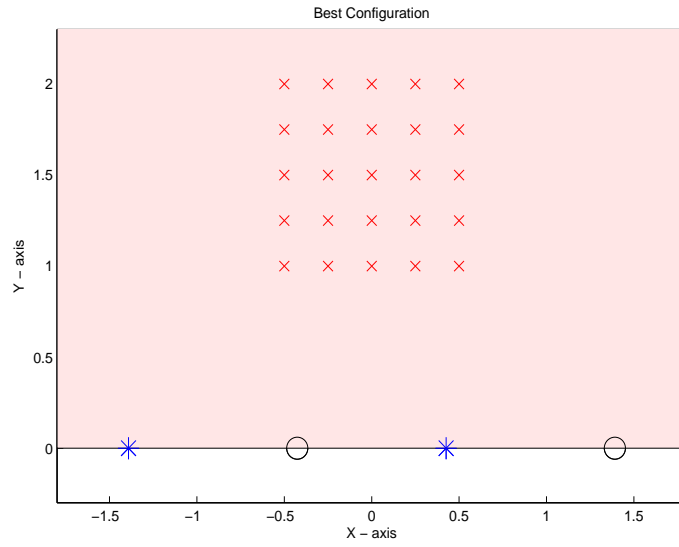
(a) Best configuration
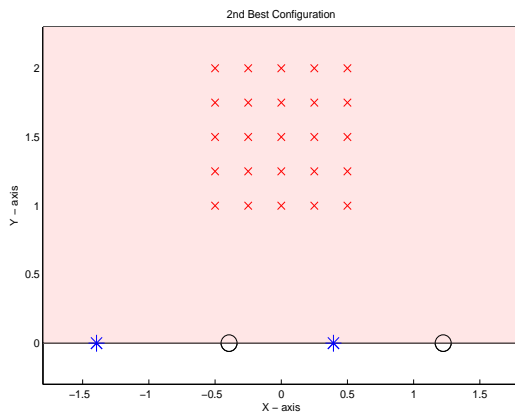
(b) 2nd Best configuration
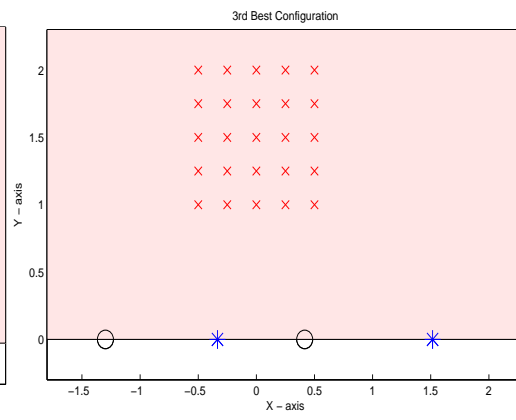
(c) 3rd Best configuration

Figure 7: Optimum configurations for a 5×5 grid with 2 sources and 2 detectors, no baseline absorption

results without baseline absorption, since the objective function at each configuration is lower, indicating a smaller ratio of posterior volume to prior.

Rank	Percent of Best	Obj. Function	S (1)	S (2)	D (1)	D (2)
1	100	-12.7903	-0.4217	1.3753	-1.3753	0.4217
2	99.1474	-12.7818	-0.4219	1.3076	-1.4083	0.4219
3	96.0299	-12.7498	-1.3724	0.3588	-0.3932	1.2326

Table 2: Optimum configurations for a 5×5 grid with 2 sources and 2 detectors, with baseline absorption

One thing to note is that the collapse in the parameter space for the best configuration in Table 2 is only 82.58 percent of the best configuration in Table 1. As such, adding baseline absorpotion markedly affects the quality of our reconstructed image. From here on, we restrict our analysis in 2D to cases without baseline absorption.

This particular configuration consists of an equal number of sources and detectors, so a natural question to ask is what happens when we add or remove a source or a detector? We can see the optimum configurations for the case with 3 sources and 1 detector and 1 source and 3 detectors in Figure 8.

We can see clearly that the two cases are entirely symmetric, implying that sources and detectors are spatially interchangeable. In Table 3 we compare the best configuration for these 2 cases with the optimum configuration in the 2 source and 2 detector case. Note that here *Percent of Best* refers to the best configuration in the symmetric case with 2 sources and 2 detectors. As we can see, our results are indifferent to an extra source or detector. Having an equal number of sources and detectors is more beneficial than having an extra one of either when the total number of sources and detectors in the system remains constant. In addition, the improvement from having a symmetric configuration is very drastic: unequal numbers produce roughly 10 percent of the resolution of a symmetric system.

We will see later that these results hold in cases with non-zero baseline absorption and in 3D geometry. As such, we keep the focus on cases where there are equal numbers of sources and detectors (or one less source in odd-
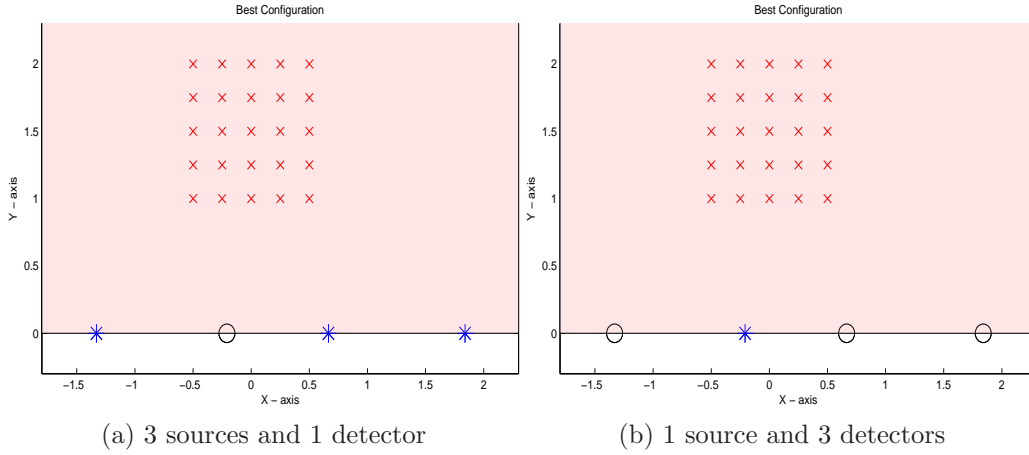
(a) 3 sources and 1 detector   (b) 1 source and 3 detectors

Figure 8: Comparison of optimum configurations for a 5 × 5 grid with unequal numbers of sources and detectors, no baseline absorption

Number of Detectors	Percent of Best	Obj. Function	S (1)	S (2)	D (1)	D (2)
2	100	-12.9817	-1.3908	0.4257	-0.4257	1.3908
1	10.5315	-10.7309	-1.3315	0.6645	1.8393	-0.2087
3	10.5315	-10.7309	-0.2087	-1.3315	0.6645	1.8393

Table 3: Comparison of optimum configurations for a 5 × 5 grid, no baseline absorption

number cases).

Another question to ask is the degree of improvement in our reconstructed image upon changing the total number of sources and detectors. Naturally, we expect that the marginal improvement gained by an extra source or detector (we have already established that they are substitutable) should decrease as the total number increases. This is exactly what we see in Table 4

Despite the improvements in the objective function at higher numbers of sources and detectors, we notice something peculiar about the positioning of the sources and detectors in the optimum configurations at when we have 7 or more in our system. As we can see in Figure 9, at S-D pairs ≥ 7, we hit a wall in terms of optimal placement. We need to place certain sources

30

Number of Sources	Number of Detectors	Percent of Best Case	Obj. Function
5	5	100.00	-20.3504
4	4	45.075	-19.5536
4	4	17.766	-18.6226
3	4	6.7653	-17.6571
3	3	2.8065	-16.7722
2	3	0.5469	-15.1418
2	2	0.0631	-12.9816

Table 4: Comparison of improvement in reconstructed image for higher numbers of sources and detectors for a 5 × 5 grid, no baseline absorption

and detectors on top of existing sources and detectors. It appears that given this geometry, there is an optimum spatial configuration for 7 sources and detectors, with any greater numbers simply being piled on to existing locations.

As we see in Figure 10, this is not the case when the number of source and detector pairs is less than 7. Perhaps this is why the boost in information for higher numbers of sources and detectors appears to level off more rapidly after we have 3 sources and 3 detectors (see Table 4).

(a) 2 sources and 3 detectors

(b) 3 sources and 3 detectors
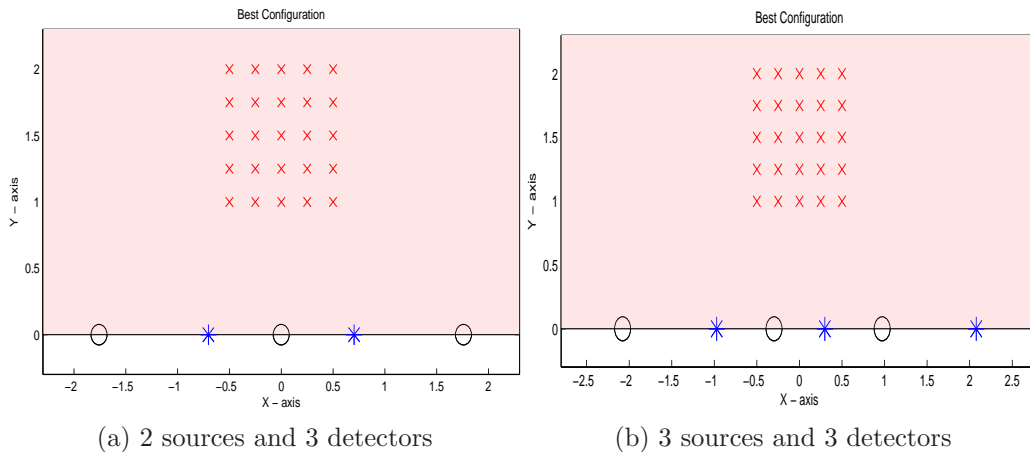
Figure 10: Comparison of optimum configurations for a 5 × 5 grid, no baseline absorption

(a) 3 sources and 4 detectors

(b) 4 sources and 4 detectors

(c) 4 sources and 5 detectors

(d) 5 sources and 5 detectors
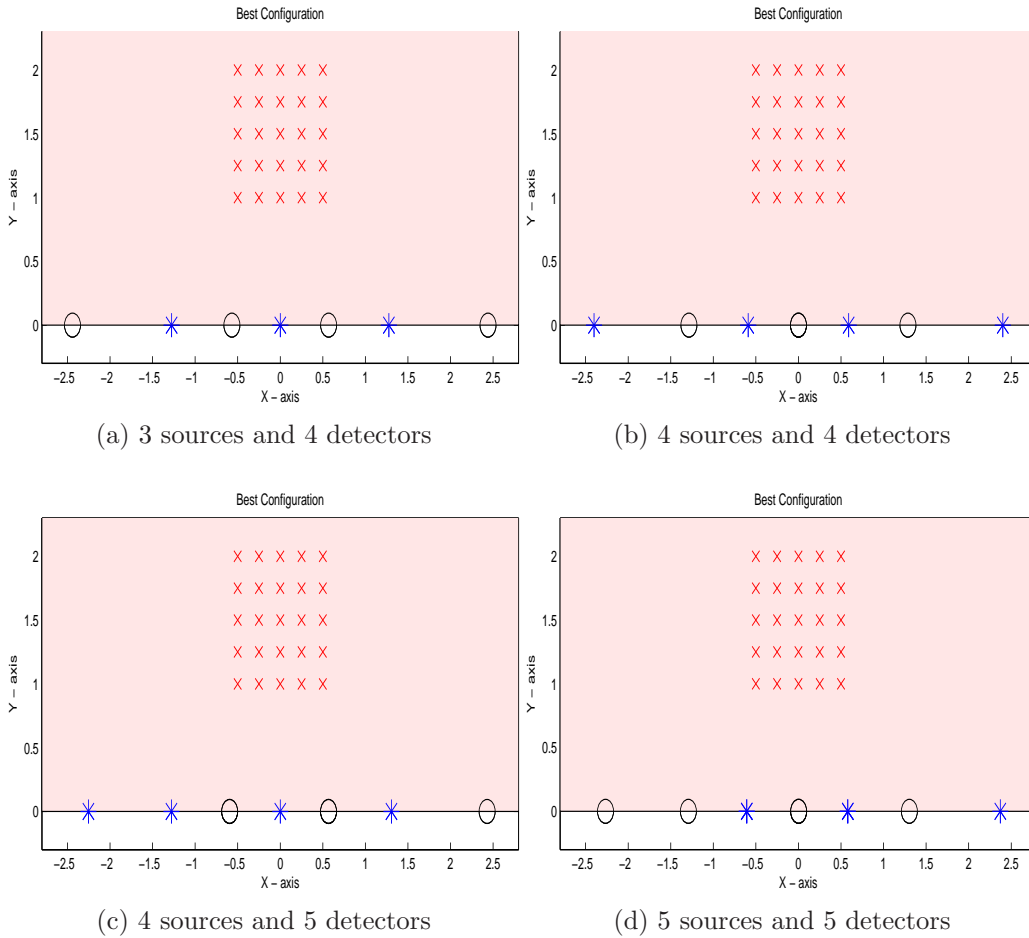
Figure 9: Comparison of optimum configurations for a 5×5 grid for ≥ 7 sources and detectors, no baseline absorption

5.2 Voxel Configurations: $5 \times 5 \times 5$ Lattice and Voxel Sheet

Once again, in 3D geometry we first find the optimum source-detector placement in the case of 2 sources and 2 detectors. The three best placement schemes are shown in Figure 11. As we can see, the symmetry from the 2D case appears to hold quite well in the best case. The second and third best cases aren't quite symmetric, like the 2D case, but appear to be quite close,

with the sources and detectors switched.

One thing of note that can be seen in Table 5 is that the 2nd and 3rd best configurations in the 3D case are not as good in comparison to the best as in the 2D case. Due to the increase in the number of dimensions over which we are optimizing, it becomes increasingly complex to find global and local minima. As such, it is possible that we have not found the true 2nd and 3rd best configurations in this case, which, if found, would yield a substantially better reconstructed image than our current results and would compare more favourably to the best case.

Rank	Percent of Best	Obj. Function
1	100.000	-10.4613
2	89.7171	-10.3528
3	87.3787	-10.3264

Table 5: Optimum configurations for a $5 \times 5 \times$ cubic lattice with 2 sources and 2 detectors, no baseline absorption

Once again, we test to see if baseline absorption affects the results significantly and find little change in the optimum configuration. Images showing the three best configurations in three dimensions with absorption can be found in the Appendix (see Figure 15). From here on, we will restrict our discussion to cases without baseline absorption, unless otherwise noted.

Remarkably, the symmetry between sources and detectors remains in 3D: their numbers and positions on the surface are entirely interchangeable, as long as the total number of sources and detectors remains constant, as we can see in Figure 12 and Table 6. Indeed, the contribution of a source or detector to the reconstructed image is equivalent, making us once more indifferent between extra sources and detectors. As in the 2D case, unequal numbers of sources and detectors provide a reconstructed image that is almost an order of magnitude worse than the best case. It appears symmetry in numbers of sources and detectors is at a premium in the 3D case as well.
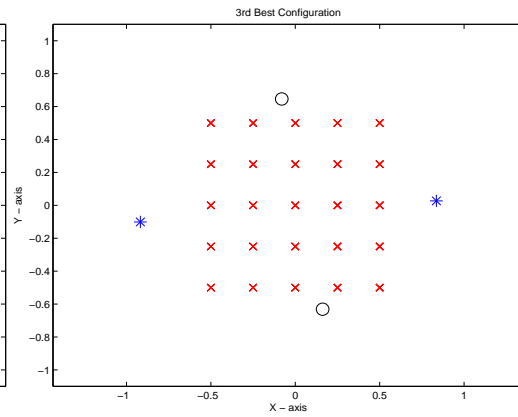
The computation for the 3D lattice is quite intensive, since the program has to continually find the SVD of a matrix with 125 columns. As such, we decided to replace the 3D lattice with a Voxel Sheet of 25 voxels at a depth
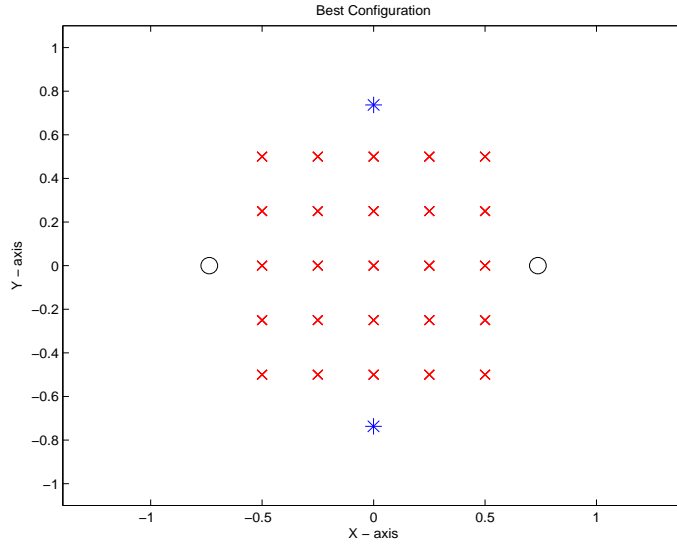
(a) Best Configuration
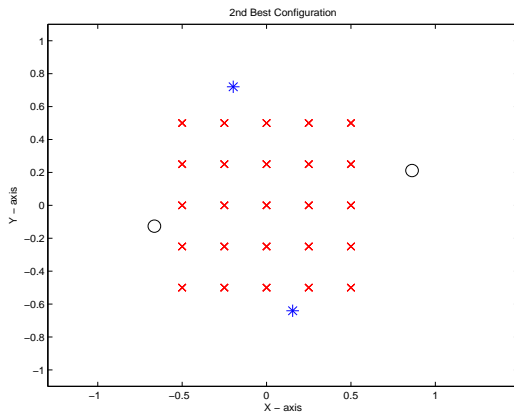
(b) 2nd Best Configuration

(c) 3rd Best Configuration

Figure 11: Optimum configurations for a $5 \times 5 \times 5$ cubic lattice, no baseline absorption

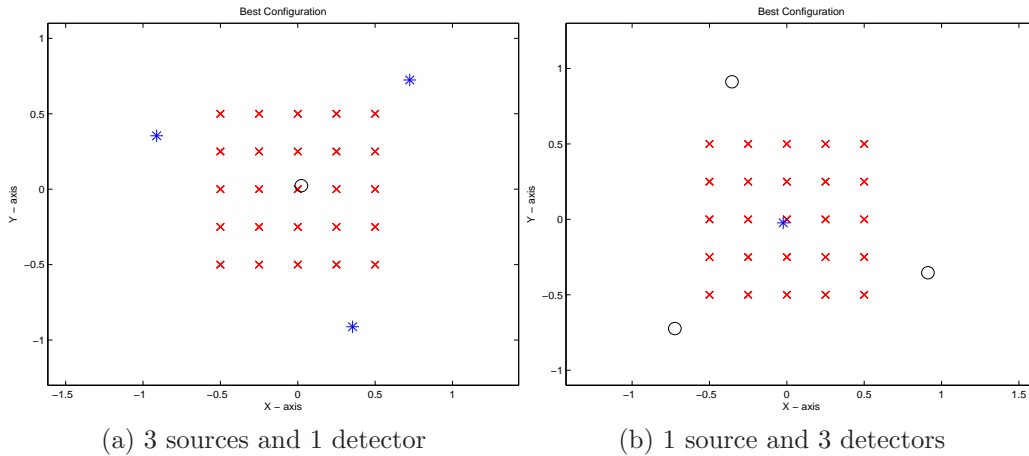(a) 3 sources and 1 detector   (b) 1 source and 3 detectors

Figure 12: Comparison of optimum configurations for a $5 \times 5 \times 5$ cubic lattice with unequal numbers of sources and detectors, no baseline absorption

Number of Detectors	Percent of Symmetric Case	Obj. Function	S (1)	S (2)	D (1)	D (2)
2	100.000	-10.461	(0, 0.73)	(0, -0.73)	(-0.73, 0)	(0.73, 0)
1	12.2456	-8.3614	(-0.023, -0.023)	(0.91, -0.35)	(-0.72, -0.72)	(-0.35, 0.91)
3	12.2456	-8.3614	(-0.35, 0.91)	(-0.72, -0.72)	(0.91, -0.35)	(-0.023, -0.023)

Table 6: Comparison of optimum configurations for a $5 \times 5 \times 5$ cubic lattice with unequal numbers of sources and detectors, no baseline absorption

of 1 unit from the surface. Indeed, we find that optimizing on the sheet is less computationally intensive and yields the roughly the same optimal configurations as the 3D lattice. See Figure 16, for 2 sources and 2 detectors, and Figure 17, for 2 sources and 5 detectors, in the Appendix[3]. We pick the latter case in order to check if symmetry holds with unequal numbers of sources and detectors in 3D.

5.3 Voxel Configuration: Voxel Column

While our results from the Voxel Sheet indicate that removing the depth parameter does not affect the results significantly, we now explore pure depth

---

[3]Note that while we do not show figures of the close approximation in optimum configurations for the cubic lattice and the voxel sheet for all higher numbers of sources and detectors, we have experimentally confirmed that to be true upto 5 sources and 5 detectors

sensitivity using the voxel column setup.

Interestingly, we find, as in the latter cases for 2D setup, the optimum configuration often involves multiple sources and detectors superimposed on one another. As we can see in Figure 13, the 2 source 2 detector optimum configuration has the two sources in the same region. All four sources and detectors are roughly along a straight line through the voxel column, with sources on one side and detectors on the other.
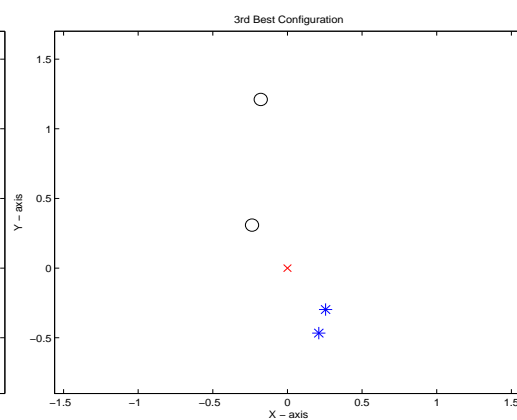
In Figure 14 we can see the comparison for higher numbers of sources and detectors. It appears the symmetry holds and we can't seem to get beyond four distinct spatial coordinates (there is rotational symmetry across the configurations).
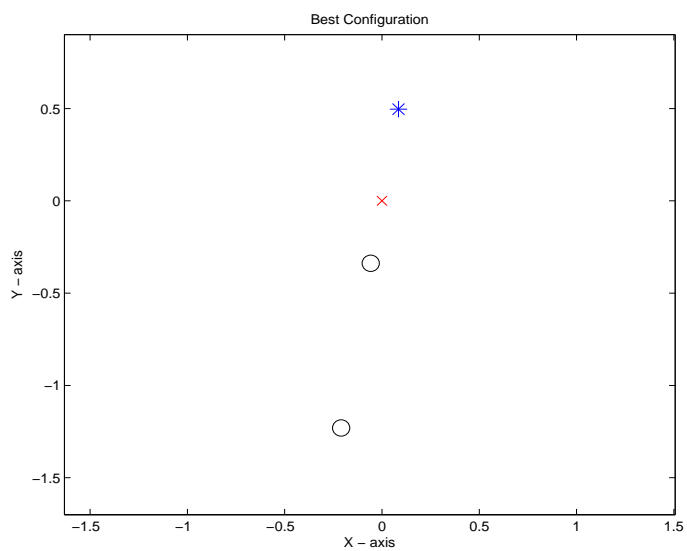
(a) Best Configuration
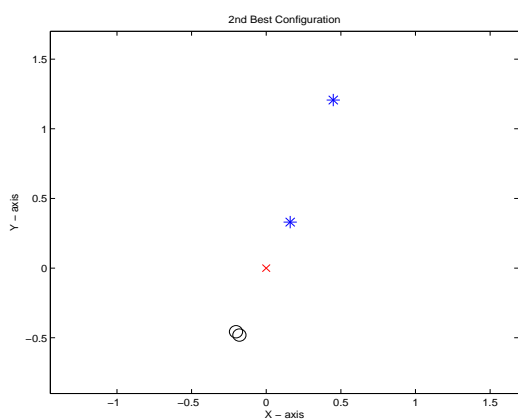
(b) 2nd Best Configuration

(c) 3rd Best Configuration

Figure 13: Optimum configurations for a Voxel Column, no baseline absorption
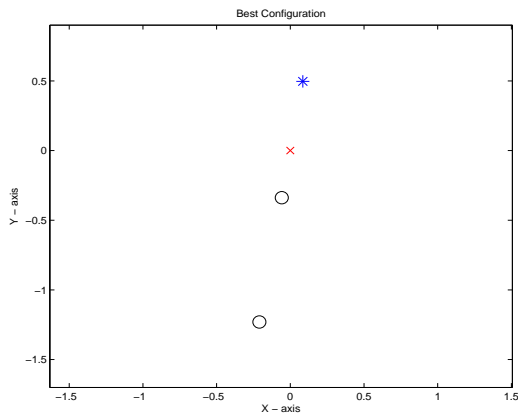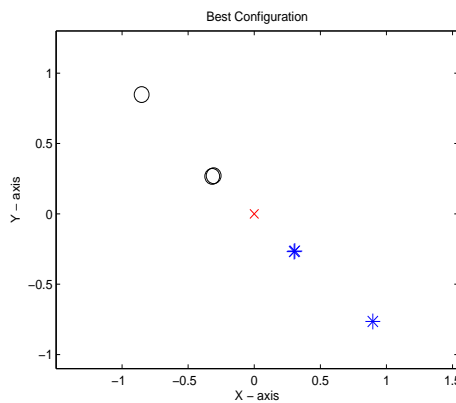
(a) 2 sources and 2 detectors
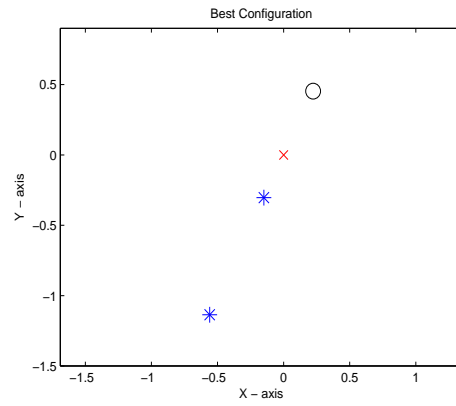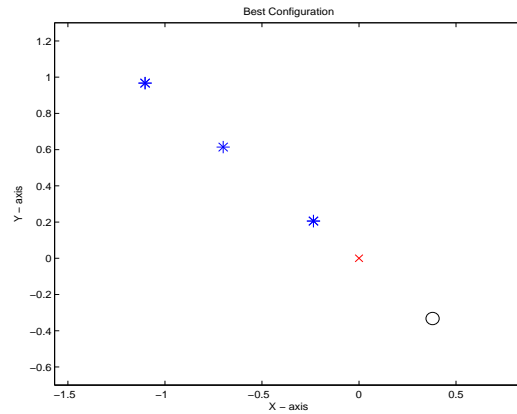
(b) 2 sources and 3 detectors
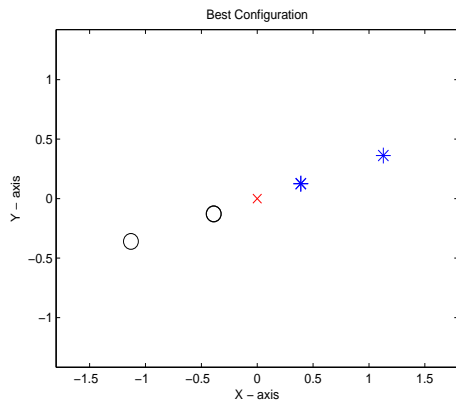
(c) 3 sources and 3 detectors
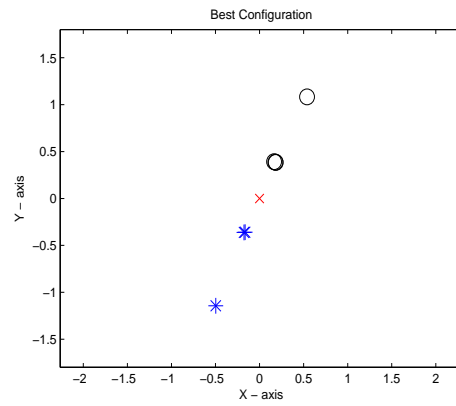
(d) 3 sources and 4 detectors

(e) 4 sources and 4 detectors

(f) 5 sources and 5 detectors

Figure 14: Comparison of optimum configurations for a Voxel Column, no baseline absorption

38

6 Conclusion

In this paper we have found optimal configurations for sources and detectors given certain voxel geometries. We have shown the existence of a unique optimal configuration in a multitude of settings. In addition, our analysis has shown that this configuration provides a substantial improvement over 2nd and 3rd best configurations, with a higher difference in three-dimensions. We find that the optimal configuration is almost always symmetric, with sources and detectors interchangeable and able to be rotated on some axis. There is also a premium on equal numbers of sources and detectors, with a symmetric configuration producing results that are almost an order of magnitude better than configurations with unequal numbers of sources and detectors (given the total number is constant). This is likely due to the fact that equal numbers of sources and detectors will form the most S-D pairs, resulting in a larger measurement matrix that will likely have more and larger singular values.

The baseline absorption at the level of $\mu_a = 0.1$ does not affect the results significantly, with the optimum configuration hardly moving. That said, background absorption significantly degrades the quality of our reconstructed image, which is not surprising. In the 2D case, we notice that we have a diminishing marginal benefit from adding more sources and detectors. This is particularly important for researchers who need to make a tradeoff between the higher cost of conducting the experiment with more sources and detectors and the better image quality obtained from the marginal source or detector.

We noticed some peculiar behaviour in two cases: when the total number of sources and detectors is equal to or greater than 7 in the 2D Voxel Grid and the optimal configurations for a 3D voxel column. In both cases, we see that the optimal configuration involves superimposition of sources and detectors. In the former case, we also see that there is a natural configuration at 7 sources and detectors. In the latter case, the optimal configuration always lies on a straight line through the Voxel Column for sources and in cases with more than 5 sources and detectors, there are only 4 unique positions to place them.

Since we have restricted ourselves to a few voxel geometries, future work can explore different voxel configurations and larger numbers of sources and

detectors. It would be interesting to find out if there is a natural optimal spatial configuration for any given voxel configuration no matter how many sources and detectors we have. The practical applications of our work are substantial, as we can see that there is a sizable difference in reconstructed image quality between even the first-best and third-best configurations for any geometry. Our method for optimization can be of use in determining both optimum configurations and marginal benefits for additional sources and detectors given any voxel configuration.
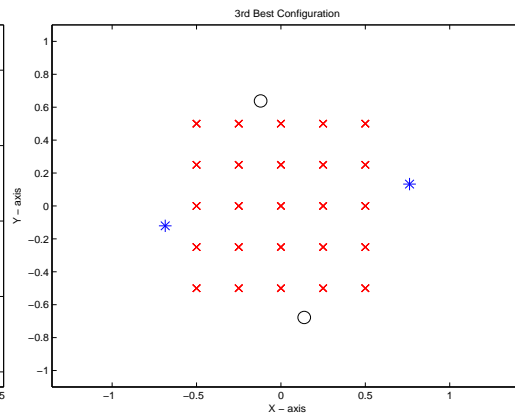
7 Appendix

7.1 Figures for Optimum Configurations
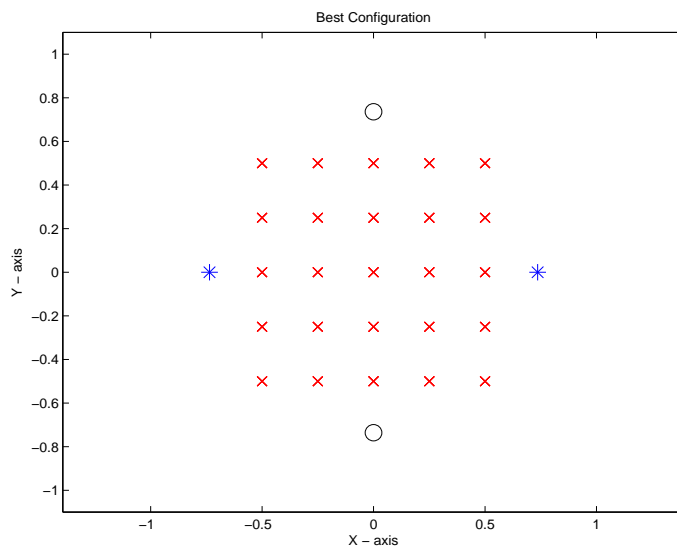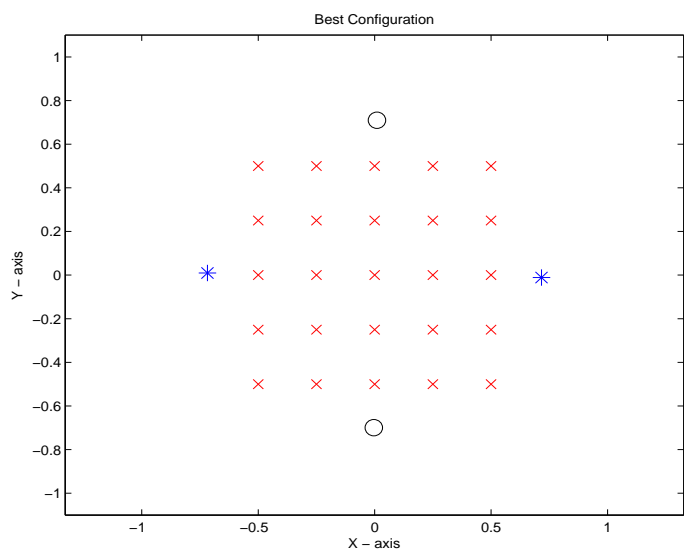

(a) Best Configuration
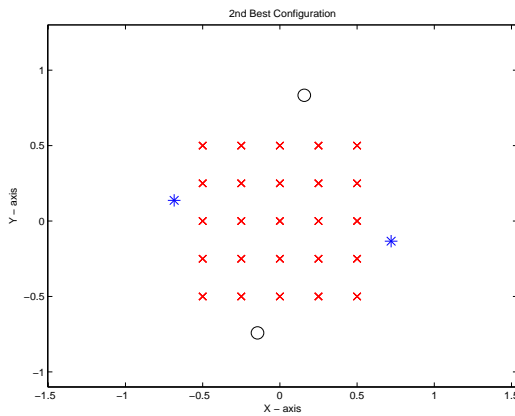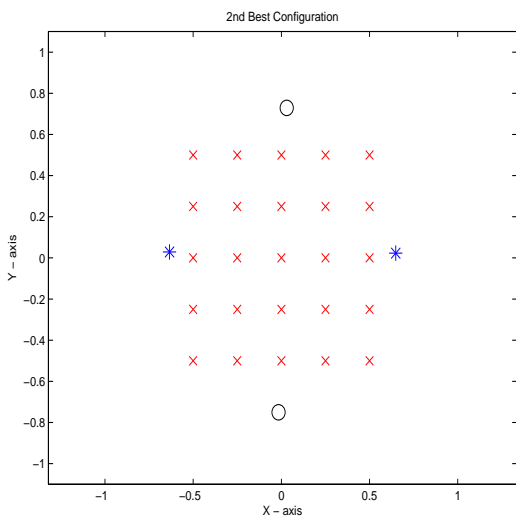

(b) 2nd Best Configuration
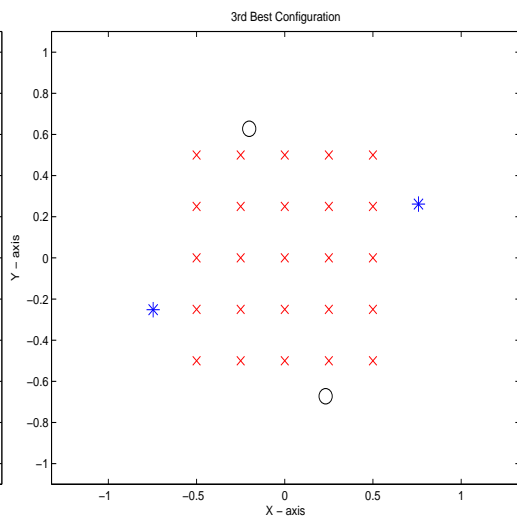

(c) 3rd Best Configuration

Figure 15: Optimum configurations for a $5 \times 5 \times 5$ cubic lattice, with baseline absorption

41

(a) Best Configuration

(b) 2nd Best Configuration

(c) 3rd Best Configuration

Figure 16: Optimum configurations for a Voxel Sheet, no baseline absorption

(a) Best Configuration

(b) 2nd Best Configuration

(c) 3rd Best Configuration

Figure 17: Optimum configurations for a Voxel Sheet, no baseline absorption
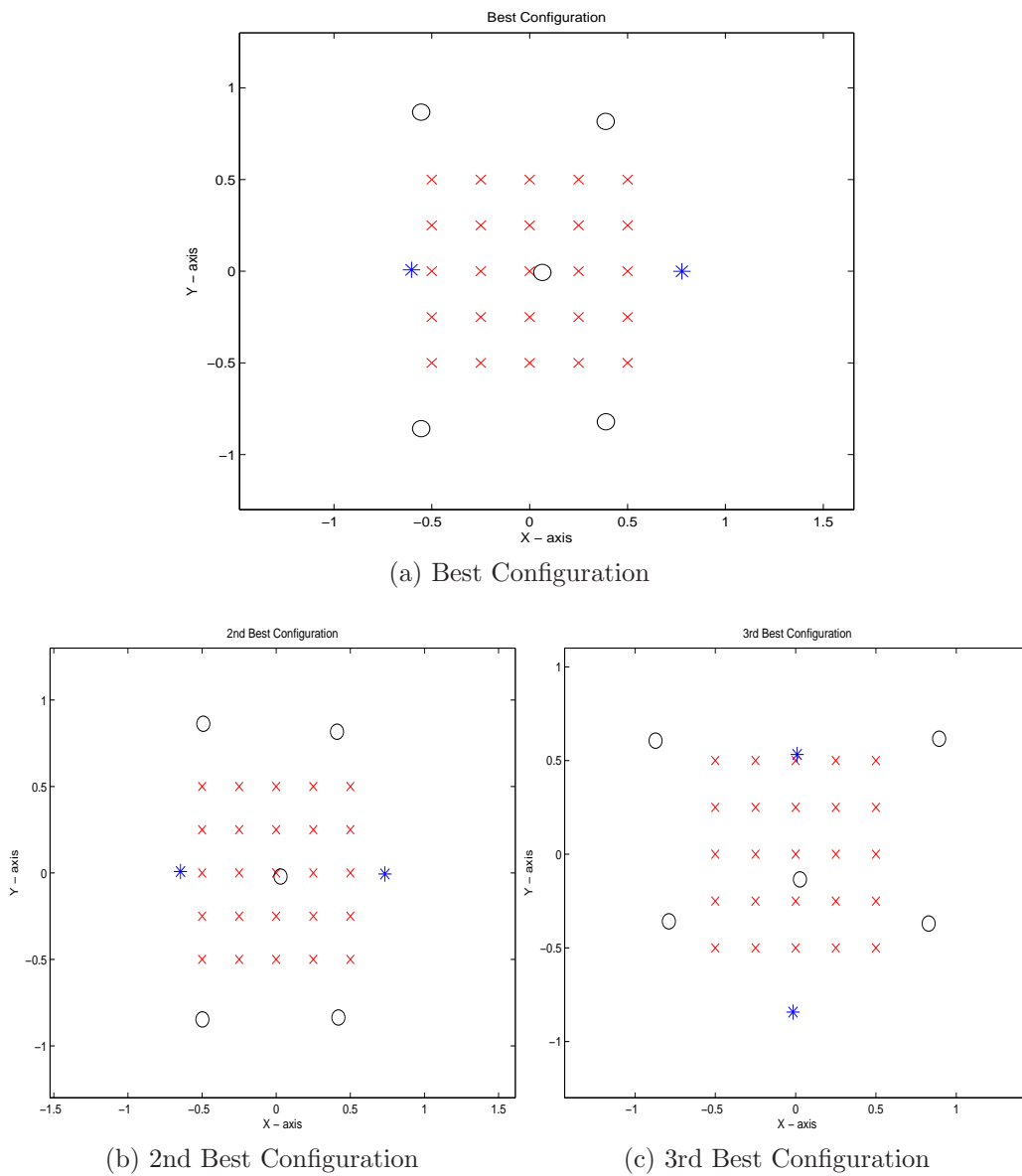
7.2 Deriving the Green's function for the 2D Laplacian

We know that the Green's function $G(\mathbf{r}, \mathbf{s})$ for the 2D Laplacian must satisfy the following condition:

$$\Delta G(\mathbf{r}, \mathbf{s}) = \delta(\mathbf{r} - \mathbf{s}) \tag{79}$$

Since we are in free space, we can take $\rho = |r - s|$, substituting that in, we get

$$G(\mathbf{r}, \mathbf{s}) = G_0(\mathbf{r} - \mathbf{s}) \tag{80}$$

Integrating both sides of (79), we see that integrating the delta function over a disk with radius greater than 0 is equal to 1

$$1 = \int\int_A \Delta G_0 dA \tag{81}$$

We then use the Divergence theorem to obtain

$$1 = \int_S G_{0\rho} dS \tag{82}$$

$$= 2\pi\rho G_{0\rho} \tag{83}$$

$$\Rightarrow G_{0\rho} = \frac{1}{2\pi\rho} \tag{84}$$

$$\Rightarrow G_0 = \frac{\ln\rho}{2\pi} \tag{85}$$

Thus, (85) is our expression for the free space Green's function in 2D without baseline absorption.

7.3 Deriving the Green's Function for the 3D Laplacian

Our results can be easily extended to the three-dimensional case without absorption. Returning to (35), we see that the measurement model requires us to know only u_0, the zero-th order solution from (5), in order to compute the source-detector measurement at an voxel. When considering no baseline absorption ($\mu_a = 0$), we are simply finding the Green's function for the 3-dimensional Laplacian. As such, (79) remains valid in 3D. We once again integrate both sides of (79)

44

$$1 = \int\int\int_V \Delta G_0 dV \tag{86}$$

$$= \int\int_S G_{0\rho} dS \tag{87}$$

$$= 4\pi\rho^2 G_{0\rho} \tag{88}$$

$$\Rightarrow G_{0\rho} = \frac{1}{4\pi\rho^2} \tag{89}$$

$$\Rightarrow G_0 = -\frac{1}{4\pi\rho} \tag{90}$$

Thus, (90) is our expression for the free space Green's function in 3D without baseline absorption.

7.4 The 3D Forward Problem

At its heart, the forward problem in three dimensions is very similar to that in two dimensions. In addition, we noted in (15)–(17) that twice the derivative at the surface of the free space Green's function is equivalent to u_0. Using these two handy facts, we need only to know the free-space Green's function for the Laplacian in 3D in order to complete the measurement model and fill our measurement matrix.

As we have shown, the free-space Green's function for the Laplacian in three-dimensions is

$$G(\rho) = -\frac{1}{4\pi\rho} \tag{91}$$

where $\rho = \sqrt{x^2 + y^2 + z^2}$. As such, taking twice the normal derivative on the surface, we get:

$$2\frac{dG}{dz} = -\frac{1}{2\pi}\frac{z_v}{\left((x_s - x_v)^2 + (y_s - y_v)^2 + (z_v)^2\right)^{3/2}} \tag{92}$$

Our solution for the 3D function is quite similar to the 2D case, and using (35), we compute the measurement function

45

$$m_d = -u_0^{(s)}(\mathbf{r}_v)u_0^{(d)}(\mathbf{r}_v) \tag{93}$$

$$= \frac{1}{4\pi^2} \frac{z_v}{\left((x_s - x_v)^2 + (y_s - y_v)^2 + (z_v)^2\right)^{3/2}} \frac{z_v}{\left((x_d - x_v)^2 + (y_d - y_v)^2 + (z_v)^2\right)^{3/2}} \tag{94}$$

Something to note is the continuation of the symmetry from the 2D case: switching the positions of the sources and detectors does not alter our result.

7.5 The Forward Problem With Non-Zero Background Absorption

So far we've only considered results with a background absorption of 0. But, in the real-world most mediums have a baseline background absorption that needs to be factored into the results. As such, (1) and (2) are no longer entirely valid. We must modify the diffusion equation in the following way and obtained a new steady-state PDE

$$(\Delta - \kappa^2)u = 0 \tag{95}$$

where we've made the substitution $\mu_a = \kappa^2$ since we know $\mu_a > 0$. This is the familiar modified Helmholtz equation.

7.5.1 The Forward Problem in 2D

In the 2D case, we can work in polar coordinates:

$$(\Delta - \kappa^2)u = 0 \tag{96}$$

$$u_{rr} + \frac{1}{r}u_r - \kappa^2 u = 0 \tag{97}$$

Using dimensional analysis, we can simplify the problem. In this case, we know that κ^2 has units L^{-2} so κ has units $1/L$. We define a natural length scale in the given parameter

$$r_c = \frac{1}{\kappa} \tag{98}$$

$$x = \frac{r}{r_c} = \kappa r \tag{99}$$

Substituting the scaled values into (97)

$$u_{xx} + \frac{1}{x}u_x - u \quad = \quad 0 \tag{100}$$

$$x^2 u_{xx} + x u_x - x^2 u \quad = \quad 0 \tag{101}$$

The expression in (101) is a version of the Modified Bessel's differential equation, which is given by [6]:

$$x^2 \frac{d^2 y}{dx^2} + x \frac{dy}{dx} - (x^2 + \alpha^2)y = 0 \tag{102}$$

where α is a real or complex number. When α is an integer, it is known as the *order* of the equation. There are two known linearly independent solutions to the Modified Bessel Equation[6]:

$$I_\alpha(x) \quad = \quad i^{-\alpha} J_\alpha(ix) \tag{103}$$

$$K_\alpha(x) \quad = \quad \frac{\pi}{2} \frac{I_{-\alpha}(x) - I_\alpha(x)}{\sin \alpha \pi} \tag{104}$$

where J_α is the Bessel function of the first kind and a solution to the unmodified Bessel Equation. It can be expressed as a Taylor expansion around 0

$$J_\alpha = \sum_{m=0}^{\infty} \frac{(-1)^m}{m! \Gamma(m + \alpha + 1)} \left(\frac{x}{2}\right)^{2m+\alpha}$$

where $\Gamma(x)$ is the Gamma function. In (101), we get the modified Bessel equation with $\alpha = 0$. Since the behaviour of the δ shows a peak at zero, we will need to use (104), the modified Bessel function of the second kind, which also diverges at zero.

The free space Green's function for the modified Helmholtz in 2D must satisfy the following condition:

$$(\Delta - \kappa^2)G(\mathbf{r}, \mathbf{s}) = \delta(\mathbf{r} - \mathbf{s}) \tag{105}$$

We will not go through the entire derivation of the Green's function for modified Helmholtz in 2D, as it is quite similar to the 2D Laplacian. As we saw earlier, the Green's function for the 2D Laplacian was simply the

fundamental solution multiplied by a factor of $\frac{1}{2\pi}$. Since we have the same geometry for the modified Helmholtz equation, we use the same factor. Once again, taking $\rho = |\mathbf{r} - \mathbf{s}|$ and rescaling $x = \kappa\rho$, we obtain the free space Green's function for the modified Helmholtz equation

$$G(\rho) = \frac{1}{2\pi}K_0(\kappa\rho) \tag{106}$$

Thus, (106) is our expression for the free-space Green's function in 2D for the Modified Helmholtz equation. Note that we can rewrite (106) as follows:

$$\frac{1}{2\pi}K_0(\kappa\rho) = \frac{1}{2\pi}K_0(\kappa\sqrt{x^2 + y^2}) \tag{107}$$

Taking twice the derivative in the y-direction, we get an expression of u_0, and plugging that into (35), we compute the measurement function

$$m_d = -u_0{}^{(s)}(\mathbf{r}_v)u_0{}^{(d)}(\mathbf{r}_v) \tag{108}$$

$$u_0{}^{(s)}(\mathbf{r}_v) = \frac{\kappa}{\pi}\frac{y_v}{\sqrt{(x_s - x_v)^2 + y_v^2}}K_1(\kappa\sqrt{(x_s - x_v)^2 + y_v^2}) \tag{109}$$

$$u_0{}^{(d)}(\mathbf{r}_v) = \frac{\kappa}{\pi}\frac{y_v}{\sqrt{(x_d - x_v)^2 + y_v^2}}K_1(\kappa\sqrt{(x_d - x_v)^2 + y_v^2}) \tag{110}$$

7.5.2 The Forward Problem in 3D

The procedure of deriving the Green's function for the 3D Modified Helmholtz equation is quite similar to the one above. Since a derivation does not enhance our understanding of the problem, we will forgo it in this paper. The Green's function in this case is known to be[6]:

$$G(\rho) = \frac{\exp(-\kappa\rho)}{4\pi\rho} \tag{111}$$

Once again, using (35), the measurement function can be computed by taking twice the normal derivative on the surface to find an expression for u_0.

References

[1] A.H. Barnett, J.P. Culver, A.G. Sorensen, A.Dale, D.Boas, *Robust inference of baseline optical properties of the human head with three-*

dimensional segmentation from magnetic resonance imaging, Applied Optics, vol. 42, no. 16, pp.3095-108, Jun. 2003

[2] R.C. Haskell, L.O. Svaasand, T.-T. Tsay, T.-C. Feng, M.S. McAdams, B.J. Tromberg, *Boundary Conditions for the diffusion equation in radiative transfer*, J. Opt. Soc. Am. A 11 (1994) 2727-2741

[3] MacKay, D. J. C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press.

[4] Kaipio, J., Somersalo, E. (2005). *Statistical and computational inverse problems*. New York: Springer.

[5] Trefethen, L. N., Bau, D. (1997). *Numerical Linear Algebra*. Philadelphia: Society of Industrial and Applied Mathematics.

[6] Arfken, G. (1985). *Mathematical Methods for Physicists*. Academic Press, Inc.

[7] D.A. Boas, D.H. Brooks, E.L. Miller, C.A. DiMarzio, M. Kilmer, R.J. Gaudette, Q. Zhang, *Imaging the Body with Diffuse Optical Tomography*, IEEE Signal Processing Magazine, vol. 18, issue. 6, pp. 57-75, Nov. 2001

[8] A.H. Barnett, *A fast numerical method for time-resolved photon diffusion in general stratified turbid media*, Journal of Computational Physics 201, pp.771-792, 2004

[9] Vogel, C. (2002). *Computational Methods for Inverse Problems*. 1st ed. Soc for Industrial and Applied Math