Model-Based Image Reconstruction for THz Imaging Systems

by

Malakeh Musheinesh

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering)
in The University of Michigan
2011

Doctoral Committee:

Professor Theodore B. Norris, Chair
Professor Jeffrey A. Fessler
Professor Herbert Graves Winful
Assistant Professor Alexander George Roy Thomas
Research Scientist John F. Whitaker

2011

To my parents…

Acknowledgements

Although a PhD is largely a solitary undertaking, the work presented in this thesis would not have been possible without the numerous contributions from my colleagues and my supervisors. I would like to thank my mentor Professor Theodore Norris for his guidance, inspiration, support, and patience during my time as a graduate student. Ted introduced me to the rich world of terahertz and gave me the opportunity to work on a project that has given me tremendous exposure to optics and image reconstruction. Ted encouraged successful collaborations with others when my own projects reached an impasse. I am immensely grateful for his continuous flow of ideas, his advice, and his perpetual optimism during my graduate time.

I would also like to thank Professor Jeffrey Fessler because without his collaboration and his time, this Thesis would not have been possible. His outstanding course on Image Reconstruction method gave me the opportunity to grasp the theory behind model-based image reconstruction and therefore the ability to apply it to the THz domain. I am truly grateful for the existence of his image reconstruction toolbox which facilitated the development of the many algorithms used in this thesis. Lastly, Professor Fessler was extremely gracious in providing time from his busy schedule to discuss my results and to offer ideas and advices. I cannot emphasize how important Professor Fessler's role has been in this thesis but I hope I have been successful in expressing my sincere gratitude to him.

I would also like to thank my committee members, Dr. Herb Winful, Dr. John Whitaker and Dr. Alec Thomas, for all your contributions and suggestions in the development of my research. I am extremely thankful to my fellow research group members for the often helpful and always entertaining discussions we've had over the years: Chuck Divin, Dong Sun, Momchil Mihnev, Jessica Ames, and Pacha Mongkolwongrojn. In particular, I'm deeply indebted to Chuck for his help during my

graduate years. His mentorship on THz generation made it possible to carry out the experiments in this thesis. He was always willing to help me troubleshoot hardware issues and discuss my results. He has also been an inspiration and a role model.

I would like to thank my parents, Abdul and Aicheh Musheinesh, for their love, encouragement, and support over the years. I am also grateful for the support, encouragement, love, and friendship of my eldest brother, Muhammed. I am also grateful to the rest of my siblings, Malek, Wasfeh, Ahmad, and Dina. My late uncle, Dr. Ahmad Musheinesh, gave me the inspiration to pursue a doctoral degree in engineering and I never really had a chance to express my sincere thanks to him for his encouragement and his advice. He is truly missed.

Lastly, I am also thankful to my fiancé David for his love, encouragement, and patience through all the long nights in the lab. He has been there for me from start to finish, so thank you.

Table of Contents

List of Figures

List of Appendices

Abstract

Broadband THz imaging is an emerging technology with a plethora of promising applications in biological imaging, nondestructive testing, security screening, and process control. Despite the inherent potential of THz technology, there are many factors that severely constrain THz systems from large-scale market introduction. Most notable among these factors is the low power of THz emitters and the measurement speeds of THz imaging systems. To live up to its enormous inherent potential, THz imaging has to dramatically increase its acquisition speed without compromising the signal to noise ratio (SNR) of the imaging system. This thesis presents research in the area of inverse THz imaging architectures via model-based image reconstruction. In inverse or indirect imaging, one can use a single point measurement system with a very high SNR to collect a fraction of the measurements needed in direct imaging modalities in order to reconstruct the object.

This thesis begins by building on the initial work of time-reversal THz imaging, a simple inverse algorithm, by adapting a waveguide approach first pioneered in ultrasound to effectively increase the numerical aperture of the THz imaging system without compromising the acquisition speed of the system. The waveguide approach was a success because of the 2.6 x improvement in intensity and the approximate 30% improvement in resolution while maintaining the same acquisition time. The second part of the thesis presents the theoretical framework of model-based image reconstruction in the context of THz imaging. Sophisticated algorithms based on this paradigm are developed for image reconstruction in both transmission-mode and reflection-mode THz systems and for both dielectric and metallic objects. The reconstructed images via the model-based algorithms are shown to be significantly better both quantitatively and qualitatively than those obtained via the time-reversal technique.

Chapter 1
Introduction

On February 1, 2003 the world watched as the Columbia space shuttle disintegrated upon re-entry into the earth's atmosphere, resulting in the death of all seven of its crew members [1]. The loss of Columbia is attributed to a piece of foam insulation the size of a small briefcase that broke off during launch and subsequently struck the leading edge of the left wing, damaging the shuttle's thermal protection system [2]. The foam was dislodged as a result of disbonds or voids underneath it and one can only speculate on what might have transpired that day had NASA had the capability to image the insulation of the shuttle. It is very plausible to think that a time-domain terahertz system in a backscatter configuration with the potential to penetrate and image tiny defects underneath the shuttle's foam could have averted the shuttle's tragic end.

Several attractive properties of THz wave radiation have spurred decades of research to develop this frequency band for imaging. One such property is terahertz's ability to penetrate and image inside most dielectric materials, which may be opaque to visible light and or may be low contrast to X-rays, making it applicable to airport screening, manufacturing quality control and product inspection [3-5]. Terahertz radiation has low photon energies, a million times lower than X-ray photons, and therefore does not cause ionization of the biological tissue. This attractive property makes THz systems particularly relevant in biomedical settings [6-7]. Lastly, the coherent time-domain mapping of the amplitude and phase of THz electric field provides THz systems with the capability of providing spectroscopic images unlike microwave and X-ray imaging modalities.

Despite its unique properties and the potential it holds to shed new light on relatively unexplored areas, sensing and imaging with terahertz remains in its infancy with many challenges. Various factors severely constrain the impact that terahertz technology can

have on our lives. Several of these challenges, including signal to noise (SNR), acquisition rate, and resolution, reflect common problems in many THz systems [8]. Current efforts to resolve these issues through higher power THz sources, higher sensitivity detectors, and improved understanding of the interaction between THz and matter are critical to advancing THz technology. Equally important but largely overlooked is the development of high speed imaging architectures particularly through algorithms and signal processing tools that accurately processing THz data into quality images. Hence the focus of this thesis is on advancing THz inverse imaging by developing the theoretical foundation for and the software needed to reconstruct quality THz images from scattered THz fields. We begin by first providing a brief history of THz and THz imaging in order to specifically point out the challenges and hurdles faced by THz.

History of THz and THz Imaging

THz radiation, residing between microwave and infrared light, is defined by the frequency range of .1 to 10 THz. Until recently, it has proven to be quite elusive because of it resistance to well established techniques from neighboring bands. That is, it was difficult to efficiently generate and detect THz radiation. On the generation side, most THz sources were either low-brightness emitters or single frequency molecular vapor lasers and detection usually relied on bolometeric methods which required cryogenic operation and generally provided low sensitivity [9, 10].

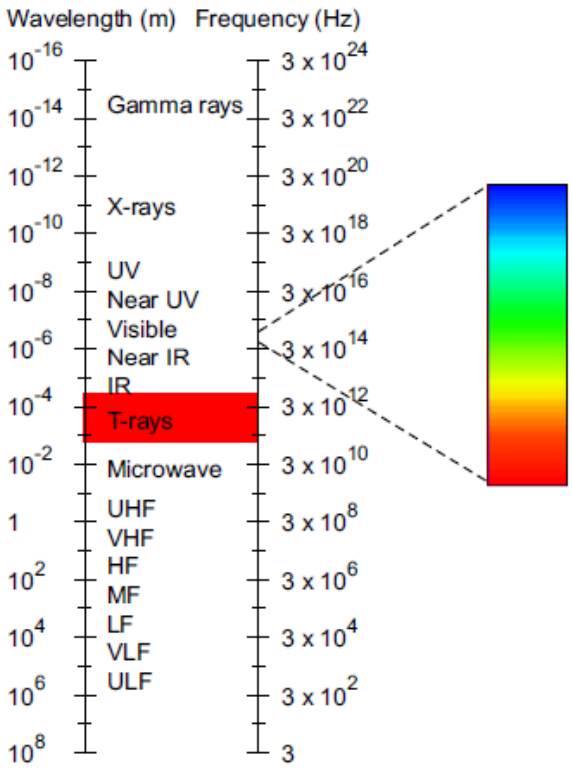


Figure 1.1. The THz (T-ray) frequency range is situated between microwave frequencies and the infrared portion of the optical spectrum.

Recently, the field of THz has experienced revitalization because of newly discovered generation and detection schemes. Equally important in the resurgence of THz science has been the demonstrated potential for advanced research and commercial applications of THz systems to numerous diverse fields including medical, semiconductors, manufacturing and defense. One of the first important uses of terahertz radiation is in an optical technique called terahertz time-domain spectroscopy (THz-TDS) [11]. The central components of a THz-TDS system are a femtosecond laser and a specially designed emitter and detector. By gating the emitter, on can generate THz transients on the order of sub-picoseconds and because the transients are nearly single cycle of the electromagnetic field, they span a very broad range of frequencies ≈ 100 GHz to 2 or 3 THz are routine and more than 5 THz have been demonstrated [12]. The coherent detection of THz transients is orders of magnitude more sensitive than typical bolometric detection and requires no cooling of any kind [13, 14].

Equally important as spectroscopic studies using THz-TDS systems is the application of THz radiation to imaging, which has lately garnered considerable attention. The development of many imaging modalities based on THz radiation has been spurred in large part because of the unique properties of THz which makes it an attractive candidate for imaging in a wide range of applications. Research in terahertz imaging, or T-ray imaging for short, have increased significantly since its first demonstration by Hu and Nuss in 1995 and it is expected that T-ray imaging will continue to grow and find new niche applications in a multitude of diverse fields [15, 16]. Most of the T-ray imaging modalities developed are extensions of THz-TDS and therefore do not require any cryogenics or shielding for the detector. Thus T-ray imaging systems have the potential to be portable, compact and robust enough for practical application in real world settings.

T-raying imaging remains an immature technology with principle challenges. Several of these key issues, including SNR, acquisition rate, and resolution reflect common issues confronting a number of direct T-ray imaging modalities. Inverse THz imaging is a relatively new imaging modality which also has the potential to address the hurdles facing the field of THz. Inverse imaging is largely an inverse problem and therefore requires sophisticated algorithms for image reconstruction. Inverse problems in physics often pose questions such as, "Can one hear the shape of a drum?" [17]. The inverse method seeks to analyze scattered light to infer an object's shape. Typically in inverse problems, one considers the diffraction or scattering of narrow-band wave in the frequency domain. However, in our inverse method we consider the scattering of broadband THz transients in the time domain for object reconstruction. Prior to discussing both direct and inverse imaging modalities, we briefly discuss the critical issues facing the field of THz in more detail.

Principle Challenges of THz Systems

The lack of a high power, low cost, and portable THz source is the most significant limitation in THz systems. The majority of broadband THz sources are based on the excitation of different materials with ultrashort pulses. Conversion efficiencies for THz sources are very low (10^{-6}), consequently average THz power tends to be very low as compared to the average power of the femtosecond laser. Photoconductive and optical

rectification are two of the most common approaches for generating broadband THz pulses. The physical mechanism for THz pulse generation in photoconductive antennas begins with an ultrafast laser pulse which creates electron-hold pairs in the photoconductor [18]. The free carriers then accelerate in the static bias field to form a transient photocurrent and this fast-varying current radiates electromagnetic waves. A conventional photoconductive emitter which is pumped by a Ti:sapphire oscillator typically achieves an average power on the order of 1 μW [19, 20]. Optical rectification is an alternative mechanism for broadband THz generation [21, 22]. Again, femtosecond laser pulses are required but in contrast to photoconducting elements where the optical beam functions as a trigger, the energy of the THz radiation in the optical rectification comes directly from the laser pulse excitation. The conversion efficiency in optical rectification depends on the material's nonlinear coefficient and the phase matching condition. Optical rectification is a second-order nonlinear effect which relies on an electro-optic crystal with non-zero order $\chi^{(2)}$ coefficient. The pump pulse induces an ultrafast transient polarization, $P(t)$, which radiates THz. The temporal THz pulse profile is given by the second time derivative of the polarization transient [23].

A high power THz source will improve both the signal-to-noise ratio (SNR) and the dynamic range of imaging and sensing systems by providing the capability to penetrate deeper into strongly scattering or absorbing materials. Furthermore, high power THz sources promise to drastically reduce the data acquisition time at current SNR, which opens up the possibility for real-time imaging of objects. The typical power of broadband THz source, although is on the order of a μW, can provide an SNR of 10^5 or higher for a single pixel detector [24]. However, for a detector array system for real-time 2D imaging, the available THz power is spread over multiple detectors and the dynamic range and SNR are considerably reduced.

Over the past 10 years, the THz power scaling has closely followed improvements to the pump lasers. Further average power scaling for a conventional photoconductive emitter that is pumped by a Ti:sapphire oscillator is significantly impeded by saturation effects at low repetition rates, and thermal dissipation of the pump at higher repetition rates [25]. In contrast, optical rectification of a femtosecond pulse train in a nonlinear crystal offers much better THz power scalability [26-29]. One approach for THz power

scaling is to scale the pump laser to very low repetition rates and very high pulse energies. In this scheme, the average pump power is held constant, $I_{pump} = 1/N_{rr}$, which leads to an overall inverse relationship with the system repetition rate. This method has proven very successful at generating high average power THz pulses, however the low repetition rate prevents the use of many common noise reduction techniques [30].

Charles Divin, a fellow research group member, has demonstrated a novel technique that scales by using a high-power, high-repetition-rate, pump laser in conjunction with long crystals fashioned into a waveguide. By focusing on a system with a high repetition rate, he maintained the ability to use mature noise reduction techniques so that the system SNR scales with increasing power. Furthermore, increasing the pump average power from a few watts to ten or one hundred watts requires transitioning from Ti:sapphire lasers to Yb-doped fiber systems. For a 1 μm laser system, GaP has been theoretically demonstrated to be the optimal optical rectification crystal. GaP offers many advantages over ZnTe, such as a broad velocity-matching bandwidth, absence of two-photon absorption at the pump wavelength, and negligible nonlinear refractive index effect. At an optical power of 10 W, he was able to generate broadband THz pulses with an average power of 150 μW from a large waveguide and 120 μW from a narrower waveguide. This represented a 12-15x improvement over a 1mm crystal [31].

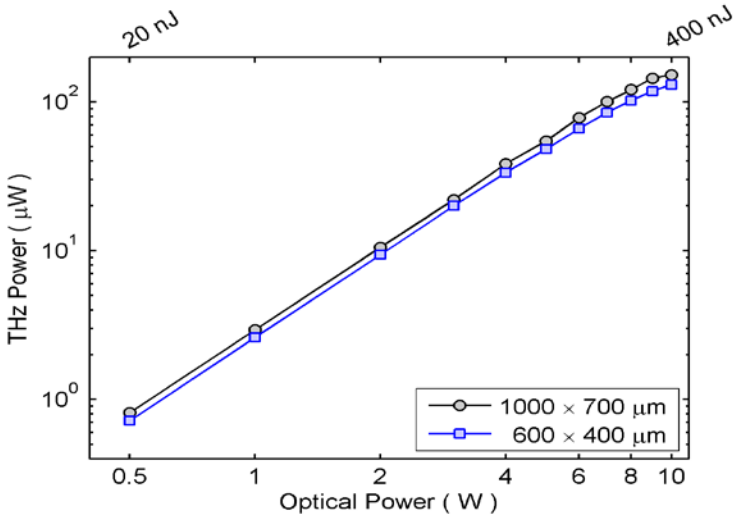


Figure 1.2. Shows the THz power dependence on pump pulse energy.

Another challenge faced by THz imaging systems is the signal to noise ratio. This is inherently tied to the average power of the THz emitter. In THz time domain spectroscopy systems, a very high SNR on the order of 10^5 can be achieved. However, in imaging applications, a number of factors combine to dramatically reduce the SNR to the point where it becomes a limiting concern. Some of these factors include the need to accelerate the imaging acquisition speed and the high absorption of many materials. Significant advances are required in the acquisition speed of THz systems in order to achieve real-time imaging. Conventional THz imaging systems rely on scanning the sample in the x and z dimensions to obtain an image. This places a severe limit on the available acquisition speed. Recently, two-dimensional (2D) electro-optic sampling has been used together with a CCD camera to provide a dramatic increase in imaging speed. However, a lock-in-amplifier cannot be synchronized to multiple pixels resulting in a significant reduction in SNR as compared to scanned approach [32].

Another challenge constraining THz imaging systems is the limited frequency bandwidth available to the system. Ideally a THz imaging system would allow spectroscopic responses to be measured up into the infrared region. This would not only allow broader signature to be observed but more importantly, for medical imaging, would reduce the potential for water attenuation, which falls dramatically as the frequency increases. Typically a conventional photoconductive antenna sources are limited to frequencies below 5 THz. Optical rectification does proved a wider bandwidth with demonstrated generation and detection bandwidths in excess of 30 THz. However this is at the expense of THz power and thus SNR. The scaling of THz systems via the method developed by Dr. Divin present the potential to achieve larger bandwidths without reducing the average THz power of the GaP emitter.

Lastly, scattering is also a challenge for THz systems. Scattering is a common problem encountered by many imaging modalities. Scattering of X-ray photons causes artifacts in the reconstructed images and in optical tomography, it is the main transport phenomena [33, 34]. The reconstruction algorithms in optical tomography are based on using a diffusive process to model the photon propagation. THz photons experience less Rayleigh scattering than optical and X-ray photons due to the longer wavelength.

However, scattering remains an issue for T-ray imaging and accurately modeling the scattering process may aid future imaging algorithms.

THz Imaging Architectures

THz imaging systems can be classified as either having a direct or inverse imaging architecture. The majority of imaging systems belong to the former group. The most basic of direct THz imaging modality is the traditional scanning THz imaging system first demonstrated by Hu and Nuss in 1995 and is a very simple extension of a standard THz-TDS system. In its simplest form, the sample mount is replaced with a 2D translation stage and the rest of the system is unaltered. As the object is raster-scanned, the THz spectrum is measured at every position. This method does provide extremely high SNR but has prohibitively long acquisition times. Attempts to improve the acquisition time of the system via using a longer time constant for the lock-in-amplifier or speeding up the motorized translation stage significantly reduce the SNR. For example, if we were to consider a small image of 60 by 60 pixels, 3600 single measurements are required and depending on the delay line concept, the desired lock-in time constants, and other factors, to record a high-quality THz pulse can take longer than 30 s and can result in a total measurement time of 30 h for a full image [3].

A number of variations and alternatives to the traditional raster-scanning THz imaging systems have been developed to overcome the prohibitively long acquisition times. However, for such direct imaging modalities, the SNR has been compromised because of the relegated use of a lock-in-amplifier. To improve acquisition speed, two-dimensional electro-optic detection of THz pulses was developed to provide a parallel detection capability and to remove the need to scan the object [32]. The experimental setup is similar to THz-TDS with electro-optic sampling, except that rather than focusing the THz pulse on the sample, plane wave illumination is used. Furthermore, the probe beam is expanded to a diameter greater than that of the THz beam and the two pulses are incident on the EO detector crystal. The polarization of the probe beam is modulated by the THz electric field via the EO effect. This polarization is converted into an intensity modulation and a CCD camera records the 2D spatial distribution of the THz modulated probe beam. Although in principle the acquisition speed of this system is limited by the

frame rate of the CCD camera, the SNR of the system is extremely poor because a lock-in amplifier cannot be used with the CCD.

Dynamic subtraction with 2D electro-optic detection of THz pulses was developed by Jiang as a means to improve the SNR of such a system [35, 36]. In this method, the CCD is set to trigger at a fixed sample rate, the trigger out signal from the CCD is then taken as the input to a frequency divider circuit, which halves the frequency and this signal is used to trigger the chopper. For example, with a CCD frame rate of 30 frames per second the THz signal would be amplitude modulated at a frequency of 15 Hz. This chopper provides a 50% duty cycle and therefore every second frame measures the THz signal amplitude, while every other frame simply measures the probe laser power without the THz field which corresponds to background noise. Every second frame is subtracted from the previous one and thereby the laser background noise is subtracted from each frame to compensate for the long term background drift of the laser. It is unknown quantitatively how well this method compares in performance to that of single point measurement systems with a lock-in amplifier. Jiang's paper does not cite any figure of merit for their method.

Another method for potentially decreasing the acquisition time of a direct THz imaging system is based on the EO detection of THz pulses using a chirped pulse probe [37, 38]. In conventional THz-TDS, the full temporal profile of the THz pulse is measured by having the probe pulse probe the instantaneous THz field at a certain time delay, adjusting the delay between the probe pulse and THz pulse, and then repeating the measurement. Using a chirped pulse allows the full waveform to be measured simultaneously rather than requiring a rapid scanner to scan the temporal profile. A linear chirp can be applied to the probe pulse using a diffraction grating. The different wavelength components of the probe pulse traverse different path lengths due to the variation in the first order diffraction angle with wavelength, λ. The output from the grating pair is a pulse with a longer duration and a wavelength that varies linearly with time. Conceptually the chirped probe pulse can be seen as a succession of short pulses each with a different wavelength. Thus when the chirped probe pulse is modulated by a THz pulse in EO detection, each of its wavelength components encodes a different

portion of the THz pulse. A spectrometer spatially separates the different wavelength components and retrieves the temporal profile of THz pulse. The spatial signal output from the spectrometer is measured using a CCD. There are a number of inherent drawbacks to this method particularly the distortion of THz spectrum. The high frequency components of the recovered spectrum are strongly attenuated and this causes significant broadening of the THz pulse. In comparison with time scanned THz detection, the chirped measurement technique suffers from a reduced temporal resolution and reduced frequency bandwidth.

The focus of this thesis is on the advancement of inverse THz imaging via model-based image reconstruction. In inverse or indirect imaging, one can use a single point measurement system with a very high SNR to a collect a fraction of the measurements needed in direct imaging modalities for image reconstruction. However, inverse imaging architectures require more sophisticated algorithms, as compared to direct methods, in order to reconstruct target images. Time reversal imaging is an early innovative inverse imaging method demonstrated with broadband THz by Ruffin et al. at the University of Michigan. By exploiting the time-reversal symmetry of Maxwell's wave equations they derived an image reconstruction algorithm based on the time-domain Huygens-Fresnel diffraction integral. This method allowed them to achieve fast reconstruction of 1D, 2D, and 3D amplitude and phase contrast objects based on measurements of the diffracted THz field at multiple angles. However, a principle drawback to this method is the significant temporal ringing after the main pulse. This ringing greatly degrades the quality of the reconstructed image. This thesis attempts to resolve this issue by developing algorithms that account for the non-ideal response of the THz imaging system to mitigate the effect of the ringing in the reconstruction of the target images. To do so, we introduce the concept of model-based imaging to field of THz which allows us to account for the impulse response of our THz imaging system via a system model. To reconstruct images via this technique requires that we "invert" our system model. However, the process of "inversion" is not as straightforward as it may sound and thus we develop the mathematical formulism that makes such an inversion possible. We assess the performance of the model-based method by comparing its performance to that of the time-reversal method. Both methods have the same acquisition times, which are

significantly shorter than the acquisition time of traditional scanning THz systems, but the model-based reconstruction algorithm yields significantly better images than the time-reversal algorithm.

Dissertation Outline

In chapter 2, we present the theory behind time-reversal THz imaging and demonstrate two methods that improve upon this technique. We first adapt a waveguide approach previously pioneered in ultrasound to increase the effective numerical aperature of the system without decreasing the data acquisition speed of the THz system. Secondly, we implement a simple 1D algorithm based on the principle of model-based reconstruction that uses the actual impulse response of the THz system. Deferring the mathematical formulism needed for the development of this algorithm till the next chapter, we demonstrate it performance as compared to the time-reversal algorithm for Young's double slit for the cases with and without the waveguide.

Chapter 3 focuses exclusively on the mathematical formulism of model-based image reconstruction. The model-based method through regularization provides a method for "inverting" a forward model that incorporates the actual measured impulse response of our THz system. The model-based method is a 5 step approach in which we:

A. Parameterize the object using a parametric model
B. Develop a system model that relates the unknown image to the expectation of each detector measurement
C. Develop a statistical model for how the detector measurements vary about their expectations
D. Choose a cost function with an appropriate regularization term based on the tools of estimation theory
E. Choose an iterative algorithm for minimizing the cost function.

We provide a thorough discussion of each of the bulleted points in the context of THz imaging in preparation for the experimental studies presented in the following chapter.

In chapter 4, we the principles of model-based image reconstruction, developed in the previous chapter, to a 2D THz system operating in reflection mode in order to show fast reconstruction of metallic and dielectric objects. We then compare the performance of

this new inverse imaging technique to the time-reversal method in order to show that we can obtain better quality images without compromising the acquisition speed of the THz imaging system.

Final conclusions are presented in Chapter 5. Following the chapter, Appendices A–C provide derivations for some of the ideas presented in chapter 3 and chapter 4.

Dissertation Contributions

This thesis makes a number of significant contributions to the body of THz imaging. Chapter 2 presents 1D THz imaging in transmission mode. Two methods were presented that improved upon the time-reversal algorithm. The first method adapted a waveguide technique first pioneered in ultrasound to increase the effective numerical aperture of the system without decreasing the data acquisition speed of the THz system. The waveguide approach was a success because of the 2.6 x improvement in intensity and the approximate 30% improvement in resolution while maintaining the same acquisition time. The second approach demonstrated for the first time the application of model-based image reconstruction to THz systems. An algorithm based on the principles of model-based image reconstruction that made use of the actual impulse response of the THz system used to reconstruct an image of a double slit. The use of impulse response of the system greatly led to a quantifiable improvement in the peak SNR of the reconstructed images. Model-based image reconstruction is quite ubiquous in the biomedical world. However, the application of model-base image reconstruction to the field of THz science has been relatively unexplored. Thus, chapter 3 developed the theoretical framework of model-based image reconstruction in the context of THz imaging. In chapter 4, we applied the principles of model-based image reconstruction, developed in the previous chapter, to a 2D THz system operating in reflection mode in order to show fast reconstruction of metallic and dielectric objects. The experiments

conducted in chapter 4, clearly showed the qualitative and quantitative improvement of the model-based algorithms as compared to the time-reversal technique.

Chapter 2
Time Reversal and Model-Based THz Imaging in a Waveguide

Sensing and imaging using broadband THz radiation has been widely and rapidly recognized as a promising non-destructive and non-contact detection technique. THz radiation can penetrate through most of the commonly used dielectric materials, such as paper, plastic, leather and wood. Many chemical materials such as explosives and bio-chemicals have resonant peaks due to intermolecular or phonon vibrations within the THz region [39, 40]. Research in terahertz imaging, or T-ray imaging for short, has increased significantly since it first debuted and it is expected that T-ray imaging will continue to grow, due to technological advances, and find new niche applications in a multitude of diverse fields. The Coherent detection of THz pulses either in transmission or reflection mode does not require any cryogenics or shielding for the detector and thus T-ray imaging systems have the potential to be portable, compact and robust enough for practical application in real world settings.

Direct THz imaging modalities are by far the most prevalent imaging architecture available. The most basic imaging setup belonging to this group is the traditional scanning THz imaging system first demonstrated by Hu and Nuss in 1995. This modality is a very simple extension of a standard THz time-domain spectroscopy system (THz-TDS). In its simplest form, the sample mount is replaced with a 2D translation stage and the rest of the system is unaltered. As the object is raster-scanned, the THz spectrum is measured at every position. This method provides extremely high SNR but has prohibitively long acquisition times. Attempts to improve the acquisition time of raster scanning systems via parallel detection methods such as 2D free space electro-optic sampling or via chirped pulse imaging do so, but at the expense of degrading the SNR of the images [35-38]. Thus to use the high SNR benefit of single point detection schemes

while reducing the number of measurements needed to generate images of targets, indirect imaging methods are required.

Time reversal imaging is an early innovative inverse imaging method demonstrated with broadband THz by Ruffin et al. at the University of Michigan [41]. By exploiting the time-reversal symmetry of Maxwell's equations they derived an image reconstruction algorithm based on the time-domain Huygens-Fresnel diffraction integral [42]. This method has allowed them to achieve fast reconstruction of 1D, 2D, and 3D amplitude and phase contrast objects based on measurements of the diffracted THz field at multiple angles [43]. However, a principle drawback to this method is the significant temporal ringing after the main pulse. This ringing greatly degrades the quality of the reconstructed image.

In this chapter, we demonstrate two methods that improve upon the time-reversal technique. We first adapt a waveguide approach previously pioneered in ultrasound to increase the effective numerical aperture of the system without decreasing the data acquisition speed of the THz system. Secondly, we implement a simple 1D algorithm based on the principle of model-based reconstruction that uses the actual impulse response of the THz system. However, prior to discussing the both methods, we briefly review coherent electro-optic THz detection and time-reversal THz imaging.

Electro-optic Detection

One of the primary methods for measuring the temporal shape of propagating THz fields is through free space electro-optic sampling (FS-EOS). This method uses the linear electro-optic effect to rotate the polarization of an ultrafast probe beam in proportion to the instantaneous THz field. Using a differential intensity technique, the induced ellipticity in the probe beam and hence the instantaneous THz field can be measured with high precision. By scanning the relative delay between the THz and optical pulses, the complete THz waveform can be sequentially sampled.

The determination of the refractive indices for propagation in anisotropic materials is accomplished by using a construct known as the index ellipsoid. If x_1, x_2, and x_3 are axes aligned along the principal directions of the anisotropic material, then the index ellipsoid is defined by

$$\frac{x_1^2}{n_1^2}+\frac{x_2^2}{n_2^2}+\frac{x_3^2}{n_3^2}=1 \tag{2.1}$$

The relationship between the THz field and the differential probe intensity can be derived by beginning with the index ellipsoid equation. In crystals belonging to the $\bar{4}3m$ point group, such as ZnTe and GaP, the equation for the index ellipsoid is given by:

$$\frac{x^2+y^2+z^2}{n^2}+2r_{41}\left(yzE_z+xzE_y+xyE_z\right)=1 \tag{2.2}$$

It can be shown that a pair of coordinate transforms can recast the index ellipsoid in a rotated coordinate system [44]. For optical beam propagating along the (110) crystal direction and polarized at an angle ϕ with respect to the (001) axis, as shown in Fig. 2.1, THz induced birefringence is:

$$\begin{aligned} n_{y''}(\alpha) &\approx n+\frac{n^3}{2}E_{THz}r_{41}[\cos\alpha\sin^2\theta+\cos(\alpha+2\theta)], \\ n_{z''}(\alpha) &\approx n+\frac{n^3}{2}E_{THz}r_{41}[\cos\alpha\cos^2\theta-\cos(\alpha+2\theta)] \\ 2\theta &= -\arctan(2\tan\alpha) \end{aligned} \tag{2.3}$$

where $n_{y''}$ and $n_{z''}$ are the indices of refraction in the new coordinate system, and α is the angle between the THz electric field and the crystal (001) axis. For ZnTe, the index of refraction (n) and the linear electro-optic coefficient (r_{41}) are approximately 2.79 and 4 pm/V.

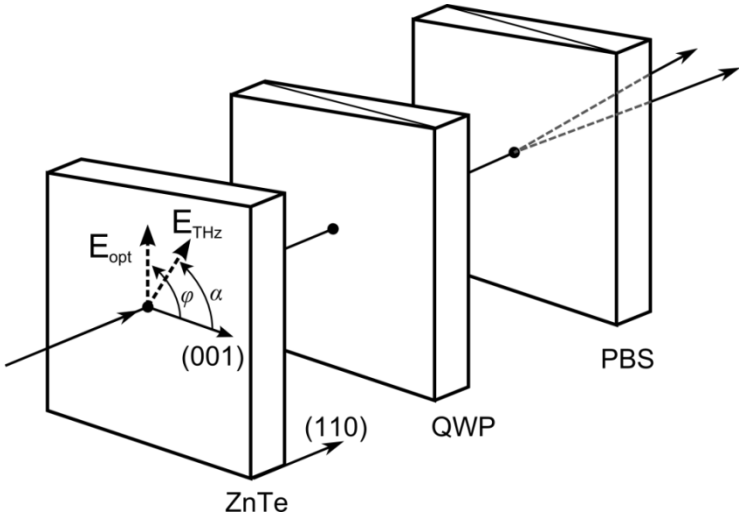


Figure 2.1. Free space electrooptic sampling setup.

As the optical probe beam copropagates with the THz beam inside the detection crystal, the probe beam acquires an elliptical polarization due to the THz induced

birefringence. In a standard balanced electro-optic detection setup, the optical beam is polarized at 45° to the THz induced birefringent axes. As the beam propagates, the acquired ellipticity is measured by passing the probe beam through a quarter wave plate and a polarizing beam splitter. The difference in the intensity between the horizontal and vertical polarizations is detected using a pair of photodiodes and the differential intensity is given by,

$$\frac{\Delta I(\alpha)}{I_p} \approx \left(\frac{2\pi}{\lambda}\right) n^3 E_{THz} r_{41} L(\cos\alpha \sin 2\varphi + 2\sin\alpha \cos 2\varphi) \qquad (2.4)$$

The electro-optic sensitivity is maximized when the THz is polarized along the $(1\bar{1}1)$ axis, $\alpha = 90°$, and the optical beam is either co-polarized or orthogonal to the THz polarization, $\varphi = 0°, 90°$.

Time-Reversal THz Imaging

Time reversal imaging is a novel inverse imaging method demonstrated with broadband THz radiation by Ruffin et al. at the University of Michigan [41, 42]. By exploiting the time-reversal symmetry of Maxwell's equations, they derived an image reconstruction algorithm that is based on the time-domain Huygens-Fresnel diffraction. It was then shown via this method that fast reconstruction of 1D, 2D, and 3D amplitude and phase contrast objects was possible.

A time domain view of diffraction is important for developing the idea of time-reversal of diffracted waveforms. A pictorial view of diffraction in the time domain is shown in Fig. 2.2 for the classical double slit problem.

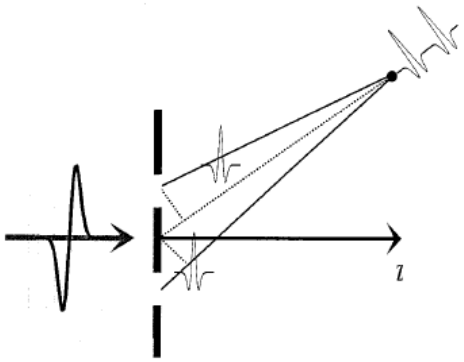


Figure 2.2. Time-domain view of diffraction from a double slit. Impulses originating from adjacent points in the aperature plane will add coherently and give rise to the multicycle waveforms at off-axis positions.

The frequency domain analysis of this example maintains that the phase differences between the apertures lead to constructive or destructive interference at off-axis, far-field positions. However, these same phase differences correspond to temporal delays; thus intuitively, each temporal feature of the scattered transient corresponds to the arrival of an impulse from a different region of the aperatures as shown in Fig.2.2.

The time-domain Huygens-Fresnel diffraction integral describes the diffraction of broadband electromagnetic pulses in free space

$$u(P_0, t) = \iint_{\Sigma} \frac{\cos(\theta)}{2\pi r_{01}} \frac{\partial}{\partial t} u\left(P_1, t - \frac{r_{01}}{c}\right) d\sigma \tag{2.5}$$

where $u(\boldsymbol{r}, t)$ is the electric field as a function of position and time, P_1 is a point on the object, P_0 is a far-field point on the measurement plane, r_{01} is the distance between these two points, c is the speed of light and θ is the zenith angle made by line joining P_0 and P_1 with the wave vector of the incident radiation [45]. This integral can calculate the diffracted field, $u(P_0, t)$, on a distant plane by integrating the time derivative of the field at the object plane, $u\left(P_1, t - \frac{r_{01}}{c}\right)$, over the aperature Σ.

The time-reversal symmetry of Eq. (2.4) can be exploited to allow the field at the object plane $P_1$to be calculated based on the measurements of the diffracted field P_0 at several off-axis far-field positions. The scattered transients correspond to the electric field amplitude and so they contain all the time delay (phase) information needed to reconstruct the spatial distribution of the electric field scattered from the object.

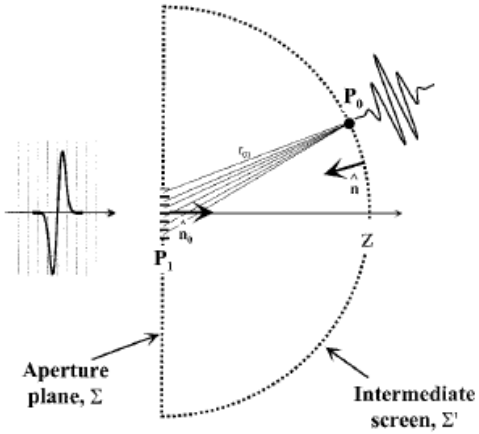


Figure 2.3. Pictorial view of Diffraction.

For the geometry shown in Fig. 2.3, the reconstruction algorithm obtained by Ruffin et al. for the field at the object plane was shown to be:

$$u(P_1,t) = -\frac{1}{4\pi c}\iint_{\Sigma'}(1+\cos(\theta))\frac{\partial}{\partial t}u\left(P_0, t+\frac{r_{01}}{c}\right)d\sigma' \tag{2.6}$$

where $u\left(P_0, t+\frac{r_{01}}{c}\right)$ is the time reversed measured scattered field and the integral is performed over the measurement semicircle [42].

This method was extended to allow 2D objects to be imaged by fixing the detector at a given zenith angle, ($\theta = 12^0$), rotating the object about the optical axis and measuring the diffracted field for 72 azimuthal positions (ϕ) between 0 and 2π. The 2D reconstruction algorithm was applied to a 10 mm star and the resultant reconstructed image is shown in Fig. 2.4.

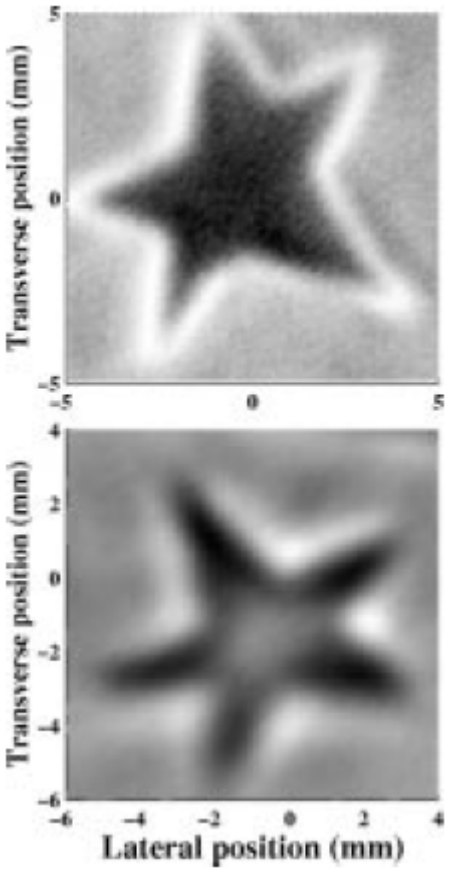


Figure 2.4. Images of a star pattern via (top) conventional scanned THz imaging and via (bottom) time-reversal technique.

It is clear from Fig. 2.4 that the time-reversed ϕ-scanning technique produces images of quality comparable to the direct imaging technique [4]. The greatest difference between these two techniques is the number of measurements required to reproduce the object. For direct scanning, the 1.1 mm pixel size required 8100 data points to produce a 90 x 90 image where as the time-reversal method required only 72 measurements. This represents a considerable saving in acquisition time and demonstrates the power of this technique.

The resolution of the method was derived using the Sparrow criterion which states that two peaks are resolved if there is a clear local minimum between the principal peaks of the two waveforms [46]. Consider two temporal waveform measurements that are separated in time by Δt, as shown in Fig. 2.5., then according to the sparrow criterion, the two waveforms are resolved if there is a clear local minimum between their principal peaks.

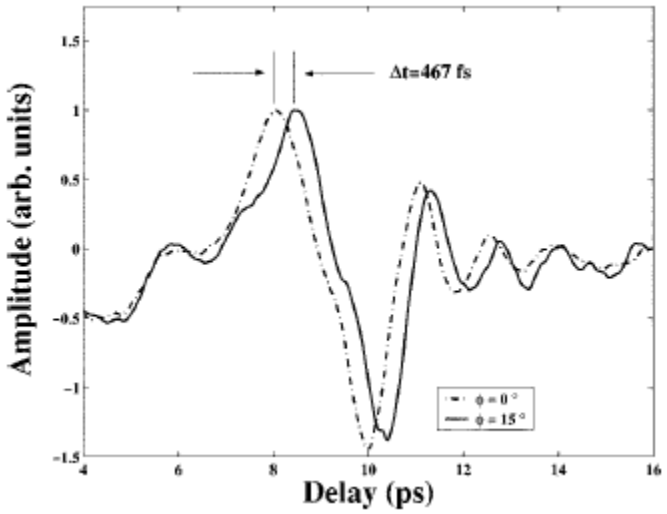


Figure 2.5. Time domain adaption of Sparrow criterion.

This temporal resolution Δt can be translated into a later spatial resolution Δx in the object plane from simple geometric considerations

$$\Delta x = \frac{c\Delta t}{\sin(\theta)} \tag{2.7}$$

Using this method, a resolution of 674 μm was demonstrated, which was significantly smaller than the mean wavelength of the THz source used.

As stated earlier, one of the main objectives of this chapter is to improve upon the time-reversal technique. The first method seeks to increase the numerical aperture of the

THz imaging system without decreasing its acquisition speed. This is done by adapting a waveguide approach first in pioneered in ultrasound. It may not be clear how the introduction of multipath can effectively increase the numerical aperture of the system and thus we present the theory behind using a waveguide in our system. The waveguide technique in the ultrasound domain has shown super-resolution focusing and this has largely motivated us to use this approach.

Ultrasound Waveguide

Time-reversal invariance of the acoustic wave equation has spurred the development of time reversal mirrors to focus pulsed waves through heterogeneous media. Time reversal mirrors (TRM) are made of reversible transducer arrays, allowing the incident acoustic field to be sampled, time reversed and re-emitted [47]. A time-reversal experiment conducted in an ultrasonic waveguide by Roux et al showed the substantial improvement in the spatio-temporal focusing of the time-reversed field [48]. The principle of the experiment is presented in Fig. 2.6. The water channel is bounded by two plane and parallel interfaces (steel). The total length of the channel along *y* axis is *L*=740 mm and a water depth *H*=40 mm along the *x* axis. A point-like ultrasonic source (a single transducer) is located on one side of the waveguide and can be used either as a source or as a hydrophone. A time reversal mirror, made of a linear array of 96 transducers identical to the transducer source, is located a distance *L* from the source and is located along the *x* axis.

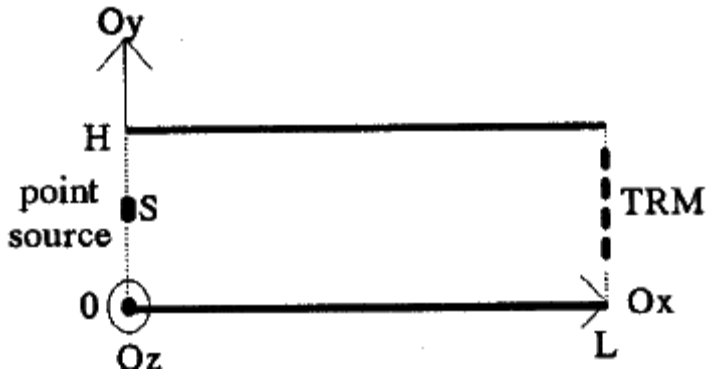


Figure 2.6. Schematic of acoustic waveguide

The transducer source located at *S* transmits a pulsed wave. The data recorded by the array is time reversed and retransmitted along the same channel. The time-reversed field is measured in the source plane by the transducer source which is used as a hydrophone.

The hydrophone can be translated around the initial position *S* along the *x* axis to measure step by step the time-reversed pressure field $p_{tr}(x,t)$ in order to show the substantial improvement in the temporal compression and in spatial focusing of the time-reversed field around the initial source [48].

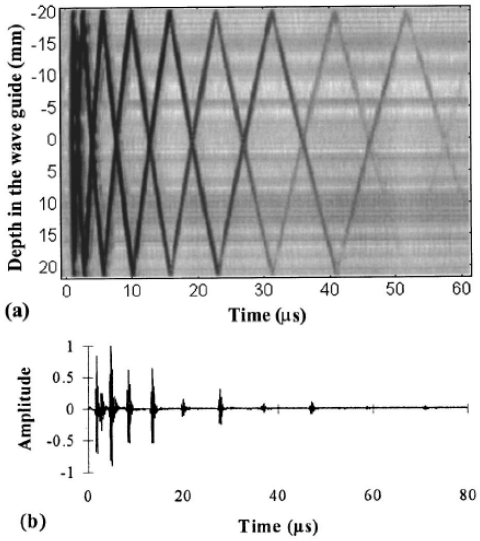


Figure 2.7. (a) B scan of incident acoustic field. (b) Temporal signal measured on one transducer.

The transmitted field recorded by the array after propagation through the channel is shown in Fig.2.7(a). A set of multipath signals corresponding to multiple reflections of the incident wave on the interfaces are shown to be arriving after the arrival of the first cylindrical wavefront which corresponds to the direct path [48]. The signal recorded on one transducer of the TRM is shown in Fig.2.7(b).

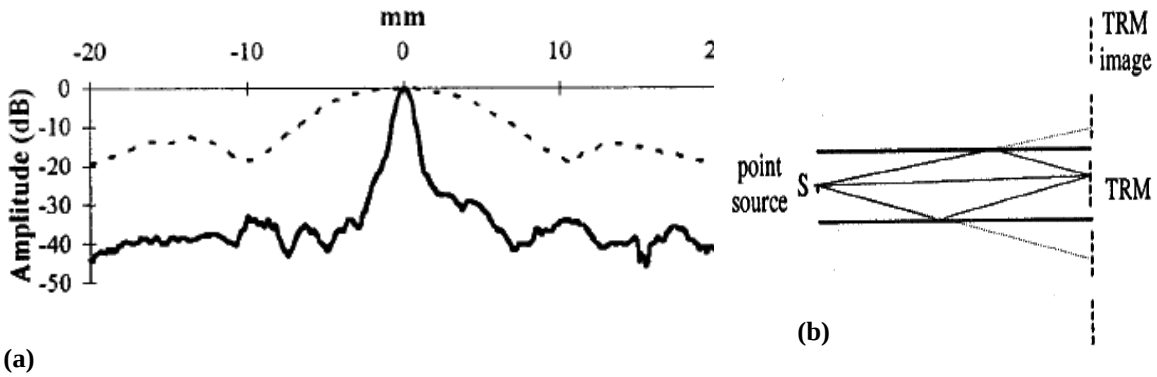


Figure 2.8. Directivity pattern of time-reversed field (a) in the plane of the point source. Principle of mirror images (b) applied to the waveguide.

The data recorded by the array is time reversed and retransmitted along the same channel. The directivity pattern of the time-reversed beam observed in the source plane is shown in Fig. 2.8(a). It shows that the time-reversed beam is focused on a spot which is much smaller than the one observed in the unbounded case and this can be easily explained by the principle of mirror images as shown in Fig. 2.8(b). At the source point, the TRM can be considered as a periodic vertical array of TRM images. Each reflected pulse received by the real TRM corresponds to one image of the TRM. Thus when taking into account the first ten echoes receive, the theoretical aperture of the TRM is ten time larger than the real aperture since the first echo corresponds to the real TRM and the following nine can be interpreted as nine TRM images [48].

Time-Reversal THz Imaging in a Waveguide

We adapt the waveguide approach from the ultrasound experiment to a THz imaging system to demonstrate the potential this technique has at increasing the effective numerical aperture of the imaging system. The experimental setup, as given in Fig. 2.9, is a typical electro-optic THz sampling system with an additional stage in the pump arm to compensate for the horizontal translation of one of the imaging parabolas.

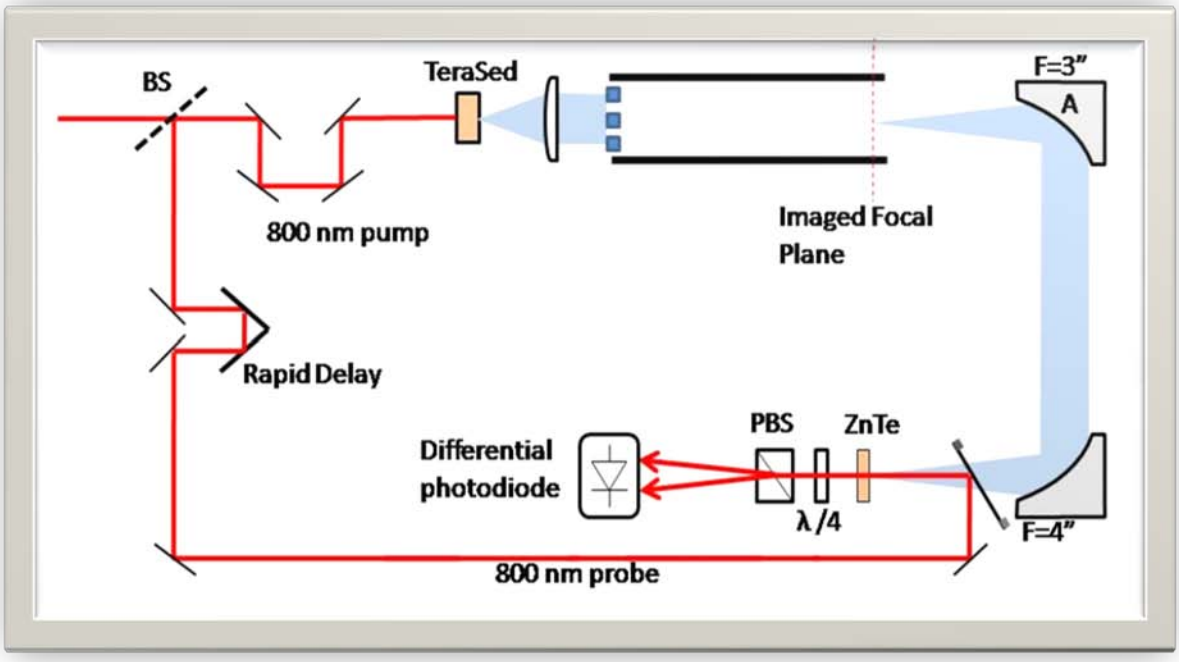


Figure 2.9. Experimental setup

A femtosecond laser pulse is split into a pump and a probe pulse by a beam splitter. The delayed pump pulse illuminates a large-area photoconductive emitter (TeraSED, GigaOptics GmbH) to generate a nearly single cycle THz pulse. The THz beam is then collimated by a polyethylene lens with a focal length of 7.6 cm. The collimated THz beam is used to illuminate an object, which in this experiment consists of two metal slits with dimensions of 1 mm x 8 mm and a spacing of 2.0 mm. The slits are bounded by two 12 inch flat mirrors which act as a planar waveguide redirecting the THz scattered at large angles and therefore delivering higher spatial frequencies into the detection region of the imaging parabolas. The focal plane of parabola A is imaged onto the electro-optic (EO) crystal by scanning parabola A across the exit face of the waveguide. A pellicle reflects the probe pulse to propagate collinearly with the THz pulse in the EO crystal at the focus of parabola B. The EO crystal is a (1 1 0) cut ZnTe crystal which velocity matches the THz pulse and near IR probe pulse to enable coherent detection of the THz pulse. The THz pulse induces a birefringence in the crystal through the linear electro-optic effect (Pockels effect), which is probed by the linearly polarized sampling pulse. The induced phase modulation of the probe pulse is converted into an intensity modulation and detected by a differential photodiode [49].

The resolution of an imaging system is limited by its numerical aperture (NA). A higher NA can be obtained by collecting data at larger angles which requires scanning the detector over more spatial positions and hence leads to a longer acquisition time. However the waveguide technique can effectively increase the numerical aperture of the imaging setup without increasing the number of spatial scan positions [48]. We can intentionally introduce multipath into our THz imaging setup by bounding our object with two planar mirrors and time reversing both the direct signal and the multipath signals. We invoke the principle of mirror images to explain why bounding our object with the planar mirrors can effectively increase the numerical aperture of our imaging system. Each reflected pulse that is detected corresponds to a virtual detector position. Hence, we can effectively double our numerical aperture with virtual detector positions by simply capturing the first set of reflections off the mirrors and accounting for their proper time delays in the time-reversal algorithm [48]. Furthermore, the reflected pulses which diffracted at larger angles than the direct path signals have a higher spatial

frequency content and thus by using a waveguide to redirect them we can improve the resolution of our system. Thus, sampling more spatial points translates to simply scanning longer in time.

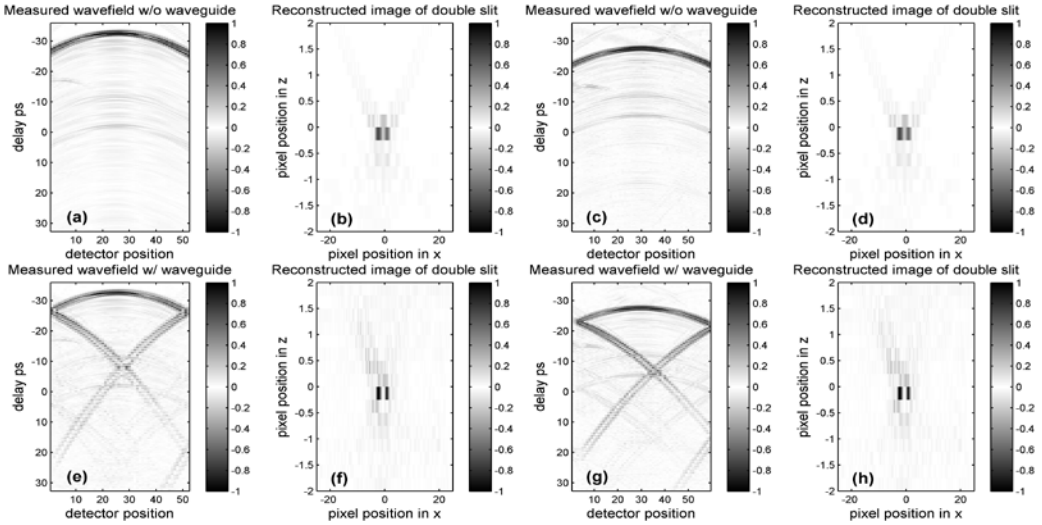


Figure 2.10. Measured THz wavefield plots without a waveguide and with a symmetric and asymmetric placed waveguide and the corresponding reconstructed images based on time-reversal.

In our experimental demonstration, we illuminated a double slit and measured the scattered THz radiation in the far field. The THz wavefield plots shown in Fig. 2.10(a) & 2.10(c) were obtained by scanning parabola A horizontally in increments of one millimeter over a range of 52 mm and a range of 59 mm respectively. At each position, the time domain THz waveform was measured by scanning the delay between pump and probe pulses over a 40 s acquisition time. Hence, the y-axis represents time delay and the x-axis represents effective detector position at the exit of the waveguide. We then carried out the same experiment with the waveguide symmetrically and asymmetrically placed about the object as shown in Fig 2.10(e) and Fig 2.10(g). The THz wavefield plots in the waveguide cases show, in addition to the direct path signals, pulses arriving at a later time corresponding to a single reflection from the waveguide mirrors. By the principle of mirror images, each reflected pulse that is detected corresponds to a virtual detector position. Hence, the first set of reflected signals should effectively double the numerical

aperture of the THz system. In practice, as the angle of an image point increases corresponding to the arrival of reflected pulses later in time, the signal strength decreases due to the limited acceptance angle of the imaging parabola. Thus only the first set of reflections is captured in our system and it is not possible to achieve a full doubling of the numerical aperture of the system. In ultrasound, the acoustic detectors are isotropic, enabling more reflections to be captured; the use of a shorter focal length parabola and a waveguide with a smaller aspect ratio should yield more reflections and thus enable a larger numerical aperture for the THz imaging system.

The reconstructed images in Fig. 2.10(b), 2.10(d), 2.10(f) and 2.10(h) were obtained by numerically back-propagating the corresponding experimental wavefield plots using the delay-sum algorithm. By accounting for the proper time delays of the reflected signals, we were able to achieve better spatio-temporal compression of the time-reversed fields with the waveguide in place regardless of its symmetry about the object. We can quantify the temporal and spatial improvement as a result of using a waveguide by taking a horizontal slice through the reconstructed images via the time-reversal technique. Fig. 2.11(a) & 2.11(c) show an intensity enhancement for the symmetric and asymmetric waveguides of 2.6 and 1.9 respectively. These values were computed by taking the ratio of the peaks of the waveguide curve (red) to the non-waveguide curve (blue). The increase in intensity for the waveguide cases is attributed to an enhancement of the fields. This enhancement is the result of the coherent addition upon back-propagation of the pulses reflected from the waveguide walls to the direct pulses, resulting in a larger constructive interference of their maxima as well as a larger destructive interference of side lobes. This is in contrast to the case without the waveguide, in which only the direct pulses are available for back-propagation and coherent addition. Fig. 2.11(b) & 2.11(d) also shows that the waveguide in both the symmetric and asymmetric cases has led to better spatial focusing of the time-reversed fields. From the resolution plot for the symmetric waveguide case, Fig. 2.11(b), we computed a full width half maximum (FWHM) value of 1.08 mm for the first peak of the red curve and a FWHM of 1.64 mm for the respective peak of the blue curve. The slits that we imaged, as mentioned earlier, have widths of 1 mm.

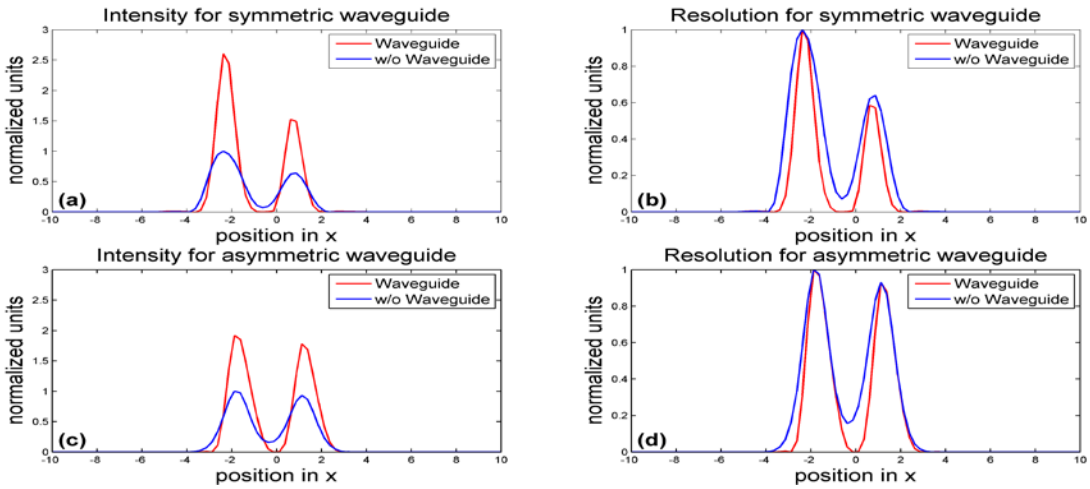


Figure 2.11. Intensity plots for symmetric (a) and asymmetric (c) waveguide configurations. Resolution plots for symmetric (b) and asymmetric (d) waveguide configurations.

Likewise, for the asymmetric waveguide case, Fig. 2.11(d), we computed a FWHM of 1.16 mm for the first peak of the red curve and a FWHM of 1.36 mm for the respective blue curve. In both waveguide cases, the time-reversal reconstruction of the two slits yielded a reconstructed object with dimensions much closer to the true dimensions of the two slits. However, in the asymmetric case the blue curve of Fig. 2.11(d) has a smaller FWHM than the blue curve in the symmetric case of Fig. 2.11(b) and this is consistent with the fact that in the asymmetric case we scanned more positions and hence we should have achieved better reconstruction. However, the FWHM of the red curve in the asymmetric case, Fig. 2.11(d), did not match the FWHM of the red curve of the symmetric case, Fig. 2.11(b), even though we sampled more detector positions. The reason for this discrepancy is evident in the wave field plot for the asymmetric case, Fig. 2.10(g). One side of the waveguide was much closer to the object than the other side. Hence we have more reflected pulses coming from the farther side than the closer side which will significantly contribute to a sharper rise on that side of the reconstructed object. Hence the right part of the red curve in the asymmetric case is overlapped with the corresponding part of the blue curve. As mentioned previously, achieving a larger angular acceptance in the THz detection will enable a higher effective NA and would alleviate this problem. In our experimental demonstration, we were still far from the ultimate diffraction limit of our system which has a λ_{peak}= 429 μm and a λ_{mean} = 119 μm.

Model-based THz Imaging in a Waveguide

Ideally the time-reversal algorithm enables the realization of an optimal spatio-temporal filter as a result of the reciprocity theorem, which states that the position of a source and receiver can be interchanged without altering the resulting field [48]. We have shown that the introduction of a waveguide has effectively increased the NA of our setup and thus we have achieved a better spatio-temporal compression of our time-reversed fields than without the waveguide. However, the performance of the time-reversal algorithm is nonetheless degraded by the presence of temporal ringing on the THz pulse. Ringing arises in the system due to reflections in the THz emitter and detector, atmospheric absorption lines, and the non-ideal response of the ZnTe electro-optic crystal. Although our THz pulse is far from being a clean single cycle pulse, we can mitigate the effect of the ringing in our reconstruction algorithm by taking into account the measured impulse response of our THz imaging system. That is, we can approach image reconstruction from scattered fields as a model-based inverse problem in which we try to recover some underlying function that describes the object from the collected data in a "best fit" manner without overly fitting the noise [50]. Hence we have investigated replacing the time-reversal reconstruction algorithm with a more general statistical algorithm that estimates what the object is from the data collected [50].

At every detector position, the received signal is just a superposition of THz pulses from every point in the object plane with an appropriate delay parameter. That is the observed signal at the *m*th detector position can be expressed as:

$$S_m(t) = \sum_{n=1}^{N} h(t - \tau_{mn})\theta_n \tag{2.8}$$

where θ_n denotes the unknown value of the object's transmissivity at the *n*th sample position in the object plane, and *h(t)* is a THz pulse that is delayed by a known parameter τ_{mn}. By concatenating our observed signals into one vector, we can recast the above equation as:

$$\begin{bmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_M(t) \end{bmatrix} = \begin{bmatrix} h(t-\tau_{11}) & h(t-\tau_{12}) & \cdots & h(t-\tau_{1N}) \\ h(t-\tau_{21}) & h(t-\tau_{22}) & \cdots & h(t-\tau_{2N}) \\ \vdots & \vdots & \ddots & \vdots \\ h(t-\tau_{M1}) & h(t-\tau_{2M}) & \cdots & h(t-\tau_{MN}) \end{bmatrix} \tag{2.9}$$

or more succinctly as:

$$\boldsymbol{y} = \boldsymbol{A\theta} \tag{2.10}$$

where **y** is a vector consisting of observed signals, **A** is a known system matrix and **θ** is a vector of unknown parameters. We could find an estimate for **θ** from **y** by minimizing the following least-squares criterion:

$$\widehat{\boldsymbol{\theta}} = \min_{\theta} \|\boldsymbol{y} - \boldsymbol{A\theta}\|_2^2 \tag{2.11}$$

However, since our goal is not only to obtain an estimate of **θ** but also to reduce the presence of artifacts in our reconstructed images, we minimize instead the following regularized least-squares (RLS) cost function:

$$\widehat{\boldsymbol{\theta}} = \min_{\theta} \|\boldsymbol{y} - \boldsymbol{A\theta}\|_2^2 - \beta \sum_{n=1}^{N} (\theta_n - \theta_{n-1})^2 \tag{2.12}$$

The additional term is a regularizing penalty term and its effect is to discourage disparities in neighboring pixel values while the effect of the first term is to encourage a best fit of the measured data [50]. Since these two effects are conflicting the adjustable parameter β controls the tradeoff between the two and controls the balance between spatial resolution and noise in the final estimate [50].

We implemented a one-dimensional reconstruction algorithm based on the RLS criterion given in Eq. (2.12) and compared its performance to the performance of the time-reversal algorithm in the case without a waveguide and in the case with an asymmetrically placed waveguide. The algorithm for both experiments took on average 5 iterations to converge. In the case without the waveguide, the RLS algorithm achieves a better reconstruction of the object than the time-reversal algorithm as evident by the 4.5 x improvement in intensity as shown in Fig. 2.12(a) and we have also calculated a peak SNR ratio improvement of 2.2. The RLS algorithm has also improved the resolution of the system. The FWHM for the red curve is 1.03 mm and the FWHM for blue curve is 1.36 mm. Furthermore, the accuracy of our system model can be determined by how

well the simulated wavefield data, which can be obtained by multiplying the reconstructed object from the RLS algorithm by the system matrix A, matches the measured wavefield data. The measured wavefield data of Fig. 2.12(c) and the simulated wavefield data in Fig. 2.12(d) are well matched, and the presence of spurious signals due to imperfections in the imaging optics have been removed from the measured data. Furthermore, we can better show the accuracy of our system model by taking any arbitrary vertical time slice from wavefield plots of Fig. 2.12(c) & 2.12(d) and determining how well the measured THz and the simulated THz pulses match at a particular detector position. Fig. 2.12(e) shows a comparison between the measured and the forward-propagated RLS-reconstructed THz pulse at detector position 10. The simulated pulse shows strong agreement with the measured direct path THz pulse at -25 ps, while showing strong suppression of spurious signals, most notably the one present at approximately -15 ps.

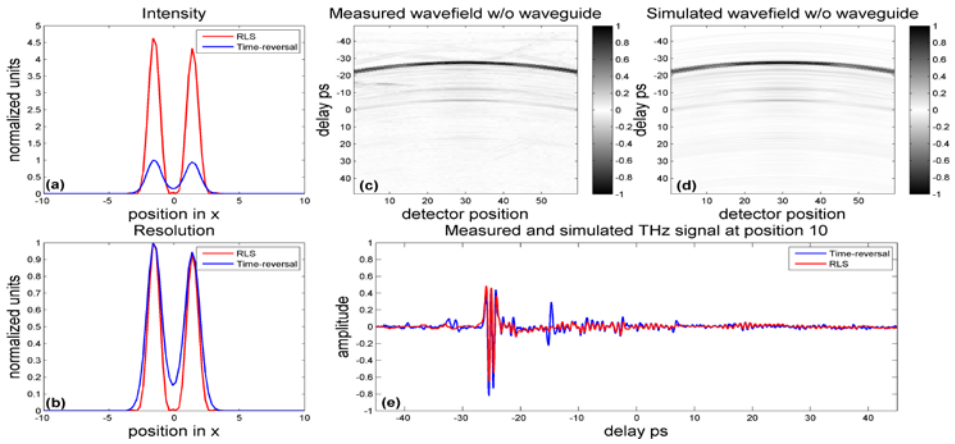


Figure 2.12. Absolute (a) and normalized intensity (b) obtained by the RLS algorithm (red curve) versus the time-reversal algorithm (blue curve). Fig. 2.12(c) & (d) show the similarity of the measured and simulated fields. Fig. 2.12(e) shows the measured and forward-propagated RLS-reconstructed THz waveform.

The model-based reconstruction algorithm works well in reconstructing the scattering object from the measured data. However in the waveguide case, the algorithm did not yield a substantial improvement over the time-reversal algorithm as seen in the intensity and the resolution plots of Fig. 2.13(a) & 2.13(b) respectively. We computed a peak SNR improvement of 1.47 for model-based algorithm over the time-reversal algorithm

for the case with the waveguide. The lack of a substantial improvement in the waveguide case can be attributed to a combination of two factors. The first factor is the sensitivity of the RLS approach to modeling errors in the system matrix A. If there is a phase shift that we are not accounting for in the waveguide case, the impulse function used to create the A matrix will not be able to completely model the waveguide system, particularly the reflections off the mirrors. We can see in Fig 2.13(c) & 2.13(d) that there is a discrepancy between the measured and simulated wavefield data indicating the presence of an unknown phase shift. Fig. 2.13(e) further shows that although we modeled the direct part of the signal accurately as evident by the strong agreement between the direct parts of the measured and simulated signals, there is a mismatch between the reflected parts. The other factor preventing a substantial improvement in reconstruction for the waveguide case is the diffraction limit of the system. The presence of the waveguide has improved the NA of the system beyond the acceptance angle of the imaging parabolas and hence the RLS algorithm has very little to improve upon as it asymptotically approaches the diffraction limit of the system. Nonetheless, the model-based algorithm in conjugation with the waveguide performs better than the time-reversal algorithm with the waveguide and we anticipate better performance provided we can generate the correct system matrix.

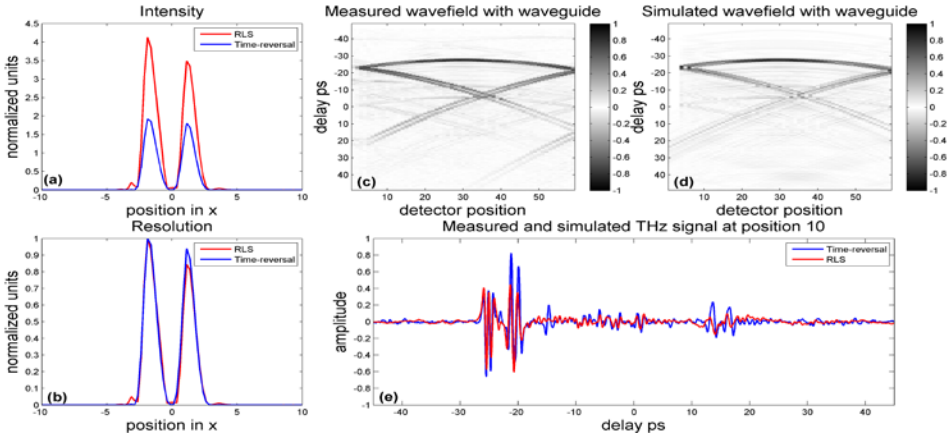


Figure 2.13. Absolute (a) and normalized intensity (b) obtained by the RLS algorithm (red curve) versus time-reversal with waveguide (blue curve). Both plots show the peak intensity relative to the time-reversal algorithm without the waveguide. Fig. 2.13(c) & (d) show the slight discrepancy between the measured and simulated wavefield data. Fig. 2.13(e) shows the measured and simulated THz signals at detector position 10.

Conclusion

In conclusion we have presented two methods for improving the time-reversal imaging technique. We first used a waveguide to increase the effective numerical aperture of the system. The waveguide technique not only yields an improvement in the numerical aperture of the system, but more generally illustrates how techniques used in ultrasound may be fruitfully adapted to THz imaging technology. Secondly, we implemented a model-based reconstruction technique that uses the actual impulse response of the experimental THz system and is therefore better suited for reconstructing the object and eliminating spurious signals than the simple time-reversal algorithm. We have demonstrated the model-based algorithm for a THz system operating in transmission mode; however, we can easily extend this algorithm for systems operating in reflection mode provided that the impulse response for the system is known in advance in order to construct the system matrix.

Chapter 3
Principles of Model-Based Imaging

Inverse problems in physics often pose such questions as, “Can one hear the shape of a drum?” Inverse problems are very important in several domains of applied science such as medical diagnostics, seismic exploration, atmospheric remote sensing, microscopy, and so on. The relevance of these problems has stimulated, since the beginning of the 1960s, the development of theoretical and practical methods for determining approximate and stable solutions [51]. In Optics, the inverse method seeks to analyze scattered light to infer or reconstruct an object’s shape.

Previously, we have shown a novel form of object reconstruction via numerically time-reversing broad-band THz fields. The extremely broad-band nature of the scattered fields is manifested in the coherent detection of nearly single-cycle pulses in the far-field. Hence, the time-reversal technique is a time-domain approach in which we simply time-reverse and back-propagate scattered transients to the object plane. Although the time-reversal method does lead to fast reconstruction of objects, the reconstructed images are often cluttered. The reason for such background clutter lies in the fact that the scattered pulses are nearly single cycle but not exactly single cycle. The presence of temporal ringing after the main pulse leads to significant clutter in the reconstructed images. Simply stated, the time-reversal method does not take into account the nature of the detected signal. However, by incorporating the actual measured THz impulse response of our system into a system matrix that captures the imaging process, we can obtain a more accurate estimate of the target’s profile and hence reduce artifacts due to the ringing on the THz pulse. In this chapter, we provide a theoretical framework for model-based image reconstruction for 2D reflection mode THz system. In the next chapter, we present experimental results showing the substantial improvement of this technique over time-reversal method for a 2-D reflection mode THz imaging system..

The scattering of THz fields can be loosely regarded as the convolution of a kernel and of the scattering object's spatial function. That is, the forward problem is the well-known Fredholm integral equation of the first kind and according to Hadamard is a well-posed problem [52, 53]. A mathematical problem is well posed if its solution a) exits, b) is unique, and c) is stable [53]. Inverse problems are usually not well-posed. In most cases an "inverse" problem can be obtained from the "direct" or "forward" one by exchanging the role of the solution and the data. In our case, we are interested in the object's spatial function given a known kernel and the observed function (data) [51]. Hence, our problem is an inverse problem and thus ill-posed. It is ill-posed for two significant reasons. The first reason is that the solution is the derivative of the data and hence the solution does not depend continuously on the data. To provide an intuitive reason for why such a reason leads to the problem being ill-posed, we consider a simple case. Suppose for example, we have a function $f(x)$ that is perturbed by a very small noise term $\varepsilon \sin \omega x$. The functions $f(x)$ and $f(x) + \varepsilon \sin \omega x$ can be arbitarly close for small ε but their derivatives may be very different if ω is large enough. This simply means that differentiation "amplifies" high frequency noise and thus our solution will have large oscillations.

Although, the first reason does present a challenge for recovering an object's profile as a function of the recorded data, a more pressing reason for the ill-posedness of our problem stems from its underdetermined nature. The unknown object is a continuous function of space but the recorded measurements are finite [54]. Therefore, there are uncountably many solutions that agree exactly with the measured data even if the data is noiseless and thus nonuniquness is the principle challenge. The question now is how can we recover a physically correct solution that is robust against noise given the insufficient data at our disposal? The answer is regularization theory [54]. Regularization provides techniques that exploit a priori information to counter noise in the data and to fill-in wherever data is missing. Model-based image reconstruction falls under the broader framework of regularization theory. Reconstructing an image of a scattering object's spatial function from scattered THz fields is an ill-posed problem for the two reasons cited earlier. Although, we had success using the time-reversal method to reconstruct an image of the scattering object's spatial function from scattered field, despite not having a

model to "invert", the reconstructed images as noted earlier were cluttered. The model-based method through regularization provides a method of "inverting" a forward model that incorporates the actual measured impulse response of our THz system. The model-based method is a 5 step approach [50] in which we:

A. Parameterize the object using a parametric model
B. Develop a system model that relates the unknown image to the expectation of each detector measurement
C. Develop a statistical model for how the detector measurements vary about their expectations
D. Choose a cost function with an appropriate regularization term based on the tools of estimation theory
E. Choose an iterative algorithm for minimizing the cost function.

Theory

This section provides a theoretical treatment of model-based image reconstruction by providing a mathematical discussion of its five components.

A. Object Parameterization and B. System Model

The initial steps of model-based image reconstruction are to develop a system model for the imaging process and then to parameterize the object space. The goal of image reconstruction is to find the unknown object, $f(x,z)$, from a set of measurements $\{y_\theta\}_{\theta=1}^{\theta=N}$. The unknown object is a continuous function of space but the recorded measurements are finite and hence there are an infinite collection of solutions for $f(x,z)$. Such a scenario necessitates that we impose some type of prior knowledge about f in order to proceed. One of the simplest and most straightforward assumptions to make is that the unknown object f is bandlimited spatially [54]. By parameterizing our unknown object using a set of basis functions, we are making the assumption that f lies in subspace of $\mathcal{L}_2$. If our object is bandlimited, then according to the 2D sampling theorem, there exist sample distances Δ_x and Δ_z such that

$$f(x,z) = \sum_{n=1}^{n=N} \sum_{m=1}^{m=M} f[n,m]\beta(x - n\Delta_{\mathrm{x}}, \mathrm{z} - \mathrm{m}\Delta_{\mathrm{z}}) \tag{3.1}$$

where $f[n,m] = f(n\Delta_x, m\Delta_z)$ and β is a basis function. The sampling theory for 2D band-limited functions requires an infinite double sum. However for practical purposes,

we truncate the double sum. For THz image reconstruction our basis functions will be the comb function, a function consisting of spatially shifted delta functions.

$$f(x,z) = \sum_{n=1}^{n=N} \sum_{m=1}^{m=M} f_{nm}\delta(x - n\Delta_{\mathrm{x}}, \mathrm{z} - \mathrm{m}\Delta_{\mathrm{z}}) \tag{3.2}$$

Another natural choice for a basis function is the rectangular function (pixels). For simplicity and proof of concept, it easier to work with the comb function. Once a suitable expansion for our unknown object is chosen, it is inserted into the system model. Assuming that our time domain Terahertz system is linear and shift invariant, the output of our system can be expressed as a convolution integral. That is the detected signal as a function of angle can be expressed as a convolution between the object's spatial profile and the measured impulse response of our THz system.

$$y_\theta(t) = \iint f(x,z)h(t - \tau_{\theta xz})\, dx\, dz \tag{3.3}$$

where $f(x,z)$ is the object's spatial profile and $h(t)$ is the impulse response of the THz system that is delayed by a known parameter $\tau_{\theta xz} = \frac{x\sin(\theta+\alpha)+z\cos(\theta+\alpha)+x\sin(\theta-\alpha)+z\cos(\theta-\alpha)}{c}$ with respect to the pivot point of the system. The rotating geometry and the plane wave detection our THz setup is captured by the delay parameter $\tau_{\theta xz}$. In the following chapter, we discuss the geometry of our setup in detail. However, for now it suffices to note that $\tau_{\theta xz}$ is a known quantity. By inserting Eq. (3.1) into Eq. (3.3), the continuous forward model can be expressed discretely as:

$$y_\theta(t) \approx \iint \sum_{m=1}^{m=M} \sum_{n=1}^{n=N} f_{nm}\beta(x - n\Delta_x, z - m\Delta_z)\, h(t - \tau_{\theta xz})\, dx\, dz \tag{3.4}$$

$$y_{\theta(t)} \approx \sum_{m=1}^{m=M} \sum_{n=1}^{n=N} \left[\iint \beta(x - i\Delta_x, z - j\Delta_z)h(t - \tau_{\theta xz})dx\, dz\right] f_{nm} \tag{3.5}$$

$$y_\theta(t) \approx \sum_{m=1}^{m=M} \sum_{n=1}^{n=N} A_{\theta mn}(t) f_{nm} \tag{3.6}$$

where

$$A_{\theta mn}(t) = \iint \beta(x,z)h(t-\tau_{\theta xz})dx\,dz$$
$$A_{\theta mn}(t) = \iint \delta(x-n\Delta_x, z-m\Delta_z\, h(t-\tau_{\theta xz}) \tag{3.7}$$
$$A_{\theta mn}(t) = h(t-\tau_{\theta mn})$$

Hence according to the above equations by parameterizing our object, we have essentially discretized our object plane into a rotating grid of possible point scatterers with reflectivity given byf_{nm}. Furthermore, the convolution integrals have now become finite summations. Since regularization is a concept borrowed from numerical linear algebra, it is more convenient to work in matrix-vector form. Lexicographic ordering is an operation in which a finite sized image, $g[n,m]$, with domain $N\ x\ M$ is represented by a vector y of length NM and with ith element given by

$$y_i = g[n(i), m(i)], \quad i = 1, \ldots, NM \tag{3.8}$$

where vector index i maps to pixel coordinates $[n(i), m(i)]$ as follows:

$$n(i) \triangleq (i-1) mod N \tag{3.9}$$

$$m(i) = \left\lfloor \frac{i-1}{N} \right\rfloor \tag{3.10}$$

The following table summarizes the relationship between the vector y and the 2D image $g[n,m]$.

Table 1: illustrates the mapping between a 2D image and a 1D vector

$y_1 = g[0,0]$	$y_2 = g[1,0]$	$\cdots$	$y_N = g[N-1,0]$
$y_{N+1} = g[0,1]$	$y_{N+2} = g[1,1]$	$\cdots$	$y_{2N} = g[N-1,1]$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$y_{N(M-1)+1} = g[0,M-1]$	$y_{N(M-1)+2} = g[1,M-1]$	$\cdots$	$y_{NM} = g[N-1,M-1]$

By lexicographically ordering our grid points and concatenating our observed signals, we can cast Eq. (3.6) into matrix form as:

$$y = Af \tag{3.11}$$

The vector y consists of observed signals and has dimension n_d, the number of detected measurements. The known system matrix A has dimensions $n_d\ x\ n_p$, where $n_p = NM$, and f is a vector of unknown parameters of the object's profile and has dimension n_p. Furthermore, we often assume that the size of y is greater than or equal to

f, that is the parmeterization scheme yields either an exact-determined or over-determined linear set of equations. Previously, our goal had been to recover a continuous space function, $f(x,z)$, from a finite set of measurements, $\{y_\theta\}_{\theta=1}^{\theta=N}$ and given the inherent under-determined nature of the problem, there had been an infinite number of possible solutions from which to choose. However, by choosing a basis and linearly parameterizing our object with respect to the chosen basis, we have reduced our problem to one of estimation. That is our new goal is to determine the object's parameter vector, f, from a finite set of measurements, y [54].

C. Statistical Model

Although the "deterministic" model given in Eq. (3.11) can be solved algebraically for f, it is incomplete because it does not specify a statistical model for the noisy measurements. Hence despite the possibility of incurring additional computation time, we must incorporate into Eq. (3.11) a statistical model that describes the distribution of each measurement about it mean and then find a measure of similarity between the actual measurements and the calculated projections from the forward model [54]. There are several statistical models but the most prevalent and relevant one is the additive noise model, $i.e.$,

$$y = Af + \varepsilon \tag{3.12}$$

where $y, \varepsilon \in \mathbb{R}^{n_d}, f \in \mathbb{R}^{n_p}$, and $A \in \mathbb{R}^{n_d \times n_p}$. The noise statistics can be described by the $n_d \times n_d$ covariance matrix of ε given by:

$$\boldsymbol{K}_\varepsilon = Cov\{\varepsilon\} = \boldsymbol{E}\left[(\varepsilon - \boldsymbol{E}[\varepsilon])(\varepsilon - \boldsymbol{E}[\varepsilon])^T\right] \tag{3.13}$$

where $\boldsymbol{E}[\cdot]$ denotes statistical expectation. The elements of $\boldsymbol{K}_\varepsilon$ are given by

$$[\boldsymbol{K}_\varepsilon]_{ij} = e_i' \boldsymbol{K}_\varepsilon e_j = Cov\{\varepsilon_i, \varepsilon_j\} = \boldsymbol{E}\left[(\varepsilon_i - \boldsymbol{E}[\varepsilon_i])(\varepsilon_j - \boldsymbol{E}[\varepsilon_j])\right] \tag{3.14}$$

where e_i denotes the i^{th} unit vector of length n_d. Assuming that the additive noise has zero mean and a known covariance matrix, $\boldsymbol{K}_\varepsilon$, that is both symmetric and positive definite, then our measurement vector, y, is a Gaussian random vector with probability density function given by:

$$p(y|f) = \frac{1}{\sqrt{(2\pi^{n_d}|\boldsymbol{K}_\varepsilon|)}} \exp\left(-\frac{1}{2}(y - Af)^{'T} \boldsymbol{K}_\varepsilon^{-1}(y - Af)\right) \tag{3.15}$$

Although we have had made the assumption that the randomness of our measurements is due to electronic noise only, which has a Gaussian distribution, it is possible that in other optical imaging systems, the random quantum effects of photon may also dominate [54]. The statistics of photons are modeled by a Poisson distribution for such optical imaging systems, and thus a more realistic model would account for both distributions.

D. Estimation theory, Regularization, and Cost Functions

By incorporating a statistical model for the noise into our system model, our objective becomes one of estimating the unknown object f from the noisy measurements in$y$. Hence we can apply the tools of estimation theory to find an estimator $\hat{f} = \hat{f}(y)$ of f. The most frequently studied methods of estimation theory are maximum-likelihood (ML), Bayesian, and Penalized likelihood [54]. By illustrating the drawbacks to the maximum likelihood and the Bayesian estimation methods, we provide suitable grounds for using the penalized-likelihood method to find an estimator $\hat{f}$ for f.

Maximum Likelihood Estimation

We can quantify agreement between the measurement vector y and a candidate object vector f by considering the likelihood term $p(y|f)$. For maximum likelihood estimation, we can find a suitable $\hat{f}$ that best fits the data by maximizing a log-likelihood measure:

$$L(f) \triangleq \log p(y|f) \tag{3.16}$$

We typically drop the dependence of L on the data y notationally because we have only one vector y in a given experiment but numerous candidate vectors for f. The ML estimator is then defined by

$$\hat{f}_{ML} = \underset{f}{\operatorname{argmax}}\, L(f) \tag{3.17}$$

To best illustrate the drawbacks to ML, we consider the Gaussian case for which the log-likelihood term is expressed as:

$$L(x) = -\frac{1}{2}(y - Af)^T \boldsymbol{K}_{\varepsilon}^{-1}(y - Af) - \frac{1}{2}\log((2\pi)^{n_d}|\boldsymbol{K}_{\boldsymbol{\varepsilon}}|) \tag{3.18}$$

Neglecting the constant term and assuming Gaussian white noise, $\boldsymbol{K}_{\boldsymbol{\varepsilon}} = \sigma^2 \boldsymbol{I}$, we can further simplify the log-likelihood expression as:

$$L(f) = -\frac{1}{2}(y - Af)^T (y - Af) \tag{3.19}$$

where ^{T} denotes transpose, f is a vector of the object's parameters (pixels), A is the system matrix, and y is a measurement vector. From a geometric standpoint, we can see that the log-likelihood measure can be expressed more compactly as:

$$L(f) = -\|y - Af\|_2^2 \tag{3.20}$$

Hence the log-likelihood measure can be seen as the negative of the L_2 norm of the difference between the measurement vector and the forward projection vector and assesses the degree of fit between the data and the model [54]. Thus in order to maximize the log-likelihood measure, we must minimize the L_2 norm of the *data mismatch* term [54]. We can redefine our new likelihood term as simply:

$$L(f) = \|y - Af\|_2^2 \tag{3.21}$$

To carry out the minimization procedure, we differentiate the norm with respect to the vector f and set the result to zero. This yields the so-called normal equations:

$$A^T Af = A^T y \tag{3.22}$$

The solution is then

$$\hat{f} = arg\min_f L(f) \tag{3.23}$$

$$\hat{f} = \left(\boldsymbol{A^T A}\right)^{-1} \boldsymbol{A}^T y \tag{3.24}$$

Assuming that $(\boldsymbol{A^T A})$ is invertible, this estimator is just the conventional least squares estimator (LS). If y is corrupted by noise and our system matrix is ill-conditioned, then our estimate $\hat{f}$ is unstable. To illustrate the instability of this estimation technique, we consider a simple image restoration problem. Suppose we wish to recover an image that has been corrupted by a shift-invariant blur. Furthermore, suppose our system matrix is square and invertible. Thus Eq. (3.24) collapses to:

$$\hat{f} = A^{-1} y \tag{3.25}$$

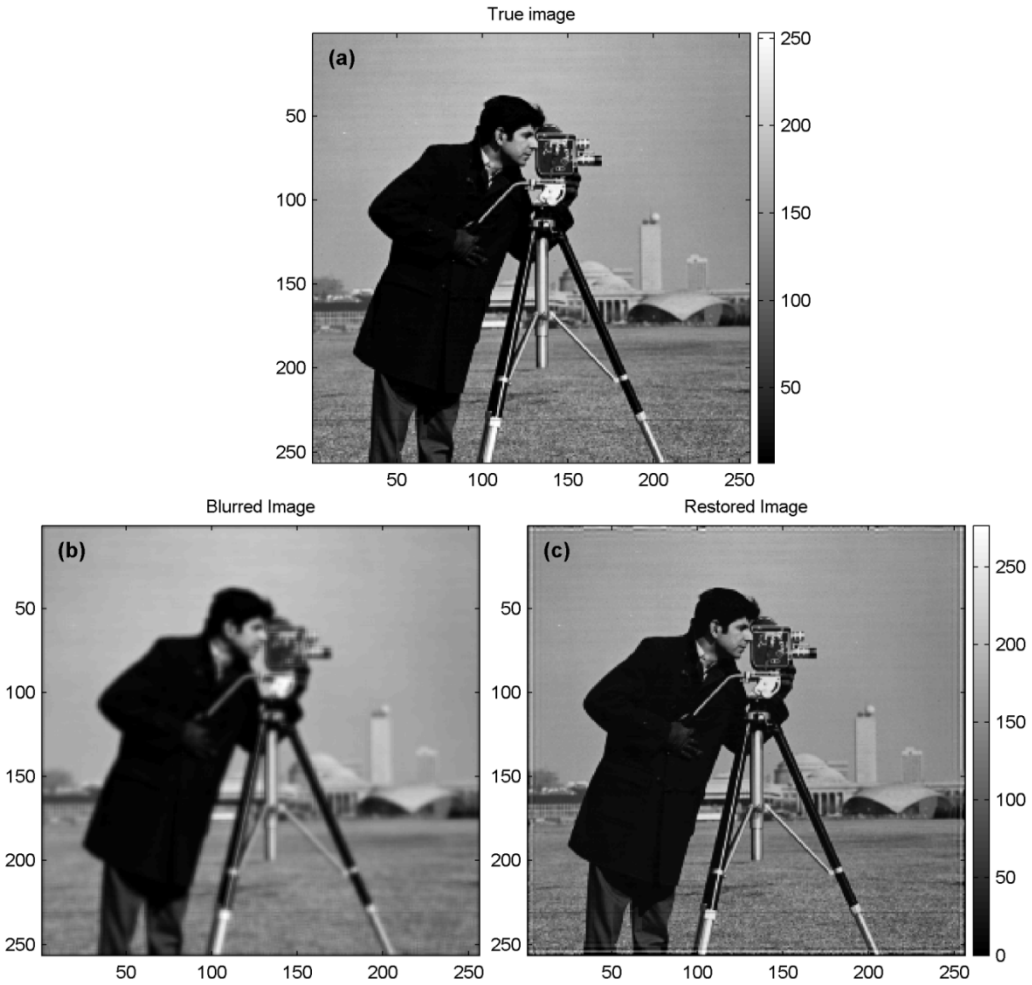


Figure 3.1. Shows a) true image, b) blurred image, c) restored image.

We can restore the true image Fig. 3.1a) from the corrupted image, Fig. 3.1b), by obtaining the inverse (pseudo inverse if A is rectangular) of our system matrix and applying it to the corrupted image. The restore image is shown in Fig. 3.1c). In the noiseless case, applying the inverse matrix technique restored the blurred image to the original image. However, the inverse method or the pseudo inverse method is not robust against noise. If in addition to blurring, our image is corrupted by zero mean Gaussian noise, then the inverse method will not restore the image as was previously shown in the noiseless case.

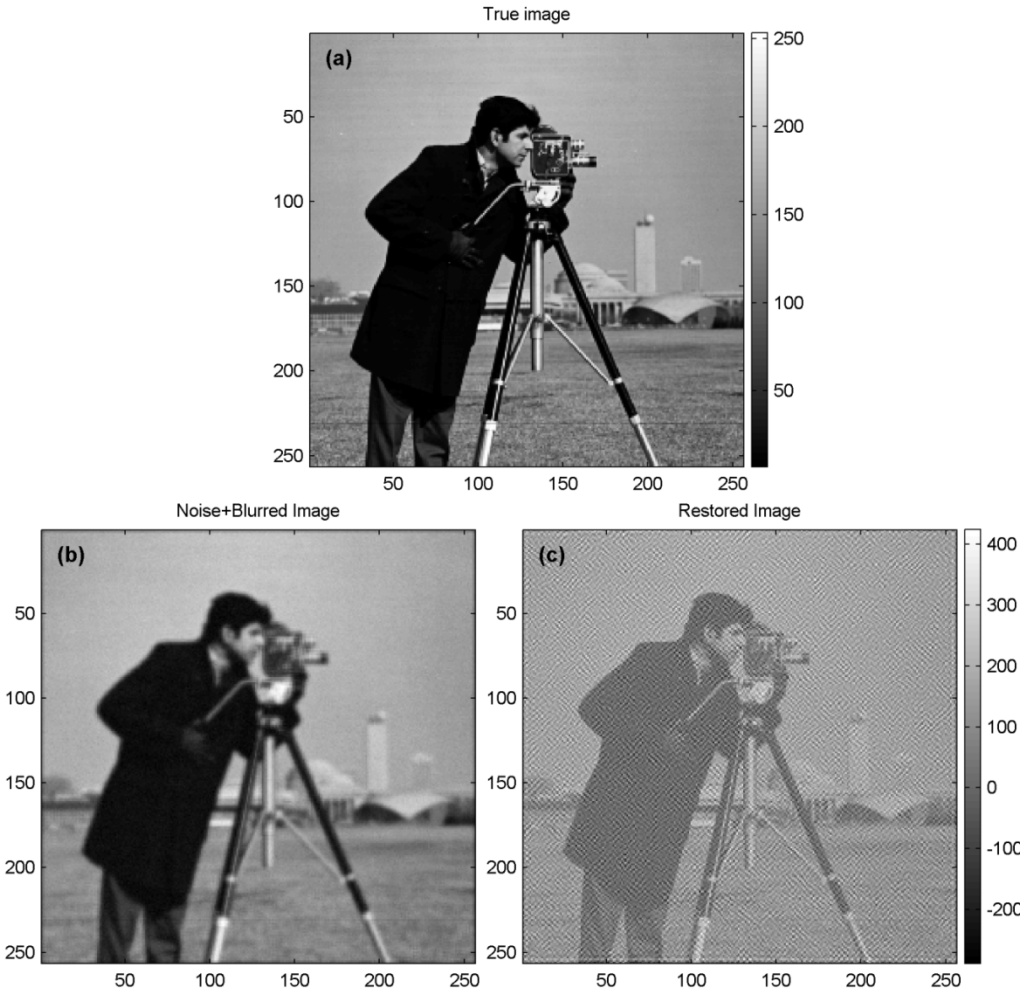


Figure 3.2. Shows a) true image, b) blurred and noisy image, c) restored image.

As shown in Fig. 3.2c) the simple inverse filter approach greatly amplifies high-spatial frequency noise and thus the restored image is noisier than the corrupted image, Fig. 3.2b). To offset the amplification of noise, we must impose constraints on our estimate. For example, a reasonable assumption on our estimate is that it lacks large fluctuations between neighboring pixels. Enforcing such an assumption forces neighboring pixels to have similar values and thus prevents amplification of noise. Bayesian estimation and Penalized-likelihood estimation are two very similar methods of incorporating a priori information about our estimate.

Bayesian Estimation

For Bayesian estimation, in addition to having a statistical model for the measurements, the likelihood term $p(y|f)$, we must also have a probability distribution $p(f)$ for the unknown object vector f called a prior distribution. In Bayesian estimation,

it is assumed the object properties are known prior to acquiring the measurement vector y. Given a statistical model in the form of a likelihood and a prior, we can devise an estimator that minimizes an expected cost function averaged over the family of possible objects f [54]. The simplest cost function is the mean-squared error (MSE) of an estimator $\hat{f}$ given by

$$MSE(\hat{f}) = E\left[\|\hat{f} - f\|^2\right] = \iint \|\hat{f}(y) - f\|^2 p(y|f)p(f)dydf \tag{3.26}$$

A classical result in Bayesian estimation is that the minimum means-squared error (MMSE) estimator for f given y is the conditional mean:

$$\hat{f}_{MMSE} = \underset{f}{\operatorname{argmin}}\, MSE(\hat{f}) = E[f|y] = \int f p(f|y)df \tag{3.27}$$

In general the MMSE estimator is nonlinear and hence the Bayesian estimator that is often used is the maximum a posteriori (MAP) estimator, defined as the maximizer of the posterior distribution $p(f|y)$:

$$\hat{\hat{f}}_{map} = arg\max_f p(f|y) \tag{3.28}$$

This estimator finds the image that has the highest posterior probability given the measurement vector y. Given a prior distribution $p(f)$ for an unkown object vector f, by Bayes rule the posterior probability can be expressed more conveniently as:

$$posterior \qquad p(f|y) = \frac{p(y|f)p(f)}{p(y)} \tag{3.29}$$

The logarithm function is monotone and thus we can simplify the above expression to the following:

$$\log p(f|y) = \log \mathrm{p}(\mathrm{y}|\mathrm{f}) + \log \mathrm{p}(\mathrm{f}) - \log \mathrm{p}(\mathrm{y}) \tag{3.30}$$

The negative log-likelihood, $-\log \mathrm{p}(\mathrm{y}|\mathrm{f})$, corresponds to a data mismatch term and the term $-\log \mathrm{p}(\mathrm{f})$ corresponds to a penalty term that penalizes our estimate depending on how far it departs from our assumption, the prior. The term $\log \mathrm{p}(\mathrm{y})$ is a constant and can be dropped. The MAP estimator in terms of log likelihood terms can be given by:

$$\hat{\hat{f}}_{MAP} = \underset{f}{\operatorname{argmax}}[\log p(y|f) + \log p(f)] \tag{3.31}$$

In a few special cases there are analytical expressions for the MAP estimator. We consider one of those special cases in hopes of illustrating that one of the drawbacks to

Bayesian estimation is that the Bayesian image priors rarely represent the object of interest. For example, suppose we assume additive Gaussian noise and further assume a Gaussian prior distribution for f. In such a case, we can obtain a closed form solution for the MAP estimator. More specifically, suppose

- $y = Ax + \varepsilon$ (Linear Model)
- $\varepsilon \sim N(O, \boldsymbol{K_\varepsilon})$ (Gaussian Noise)
- $f \sim N(\boldsymbol{\mu_f}, \boldsymbol{K_f})$ (Gaussian Prior)
- ε and f are independent
- μ_f, $\boldsymbol{K_\varepsilon}$, and $\boldsymbol{K_f}$ are all known

The likelihood term can again be expressed as in Eq. (3.15). The prior distribution is Gaussian and can be expressed as:

$$p(f) = \frac{1}{\sqrt{(2\pi)^{n_p}\left|\boldsymbol{K_f}\right|}} e^{-\frac{1}{2}(\boldsymbol{f}-\boldsymbol{\mu_f})^{\boldsymbol{T}}\boldsymbol{K_f^{-1}}(\boldsymbol{f}-\boldsymbol{\mu_f})} \tag{3.32}$$

Ignoring irrelevant constants, the negative log likelihood can be combined with the negative log prior to obtain the following form for the MAP estimator:

$$\hat{f}_{MAP} = arg\min_{\boldsymbol{f}} \left(\frac{1}{2}(y - \boldsymbol{Af})^{\boldsymbol{T}}\boldsymbol{K_\varepsilon^{-1}}(y - \boldsymbol{Af}) + \frac{1}{2}(\boldsymbol{f} - \boldsymbol{\mu_f})^{\boldsymbol{T}}\boldsymbol{K_f^{-1}}(\boldsymbol{f} - \boldsymbol{\mu_f}) \right) \tag{3.33}$$

$$\hat{f}_{MAP} = arg\min_{\boldsymbol{f}} \boldsymbol{\Psi}(f) \tag{3.34}$$

This cost function unlike the one given in Eq. (3.19) consists of a data fit term and a prior. The quadratic cost function given in Eq. (3.13) has a closed form solution that can be obtained by equating the gradient of $\Psi(f)$ with zero. The gradient is a vector field that, for a given point z points in the direction of greatest increase of a given function. The column gradient is defined to be:

$$\nabla_f = \begin{bmatrix} \frac{\partial}{\partial f_1} g(\boldsymbol{f}) \\ \frac{\partial}{\partial f_2} g(\boldsymbol{f}) \\ \frac{\partial}{\partial f_3} g(\boldsymbol{f}) \\ \vdots \\ \frac{\partial}{\partial f_n} g(\boldsymbol{f}) \end{bmatrix} \tag{3.35}$$

It follows that

$$\begin{aligned} &\nabla_f \left[\frac{1}{2} (y - \boldsymbol{A}\boldsymbol{f})^{\boldsymbol{T}} \boldsymbol{K}_{\boldsymbol{\varepsilon}}^{-\boldsymbol{1}} (y - \boldsymbol{A}\boldsymbol{f}) \right] \\ &= \nabla_f \left[\frac{1}{2} y^T \boldsymbol{K}_{\boldsymbol{\varepsilon}}^{-\boldsymbol{1}} y - y^T \boldsymbol{K}_{\boldsymbol{\varepsilon}}^{-\boldsymbol{1}} \boldsymbol{A} f + \frac{1}{2} f^T \boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{K}_{\boldsymbol{\varepsilon}}^{-\boldsymbol{1}} \boldsymbol{A} f \right] \\ &= -\boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{K}_{\boldsymbol{\varepsilon}}^{-\boldsymbol{1}} f + \boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{K}_{\boldsymbol{\varepsilon}}^{-\boldsymbol{1}} \boldsymbol{A} f \\ &= -\boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{K}_{\boldsymbol{\varepsilon}}^{-\boldsymbol{1}} (y - \boldsymbol{A} f) \end{aligned} \tag{3.36}$$

It further follows that

$$\nabla_f \Psi(\boldsymbol{f}) = -\boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{K}_{\boldsymbol{\varepsilon}}^{-\boldsymbol{1}} (y - \boldsymbol{A} f) + \boldsymbol{K}_{\boldsymbol{f}}^{-\boldsymbol{1}} (\boldsymbol{f} - \boldsymbol{\mu}_{\boldsymbol{f}}) \tag{3.37}$$

Equating the gradient term to zero yields the following MAP estimator

$$\hat{f}_{MAP} = \mu_f + \left[\boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{K}_{\boldsymbol{\varepsilon}}^{-\boldsymbol{1}} \boldsymbol{A} + \boldsymbol{K}_{\boldsymbol{f}}^{-\boldsymbol{1}} \right]^{-1} \boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{K}_{\boldsymbol{\varepsilon}}^{-\boldsymbol{1}} (y - \boldsymbol{A}\mu_f) \tag{3.38}$$

If our confidence in the prior information decreases; that is $\boldsymbol{K}_{\boldsymbol{f}} \longrightarrow \infty$ $\left(\boldsymbol{K}_{\boldsymbol{f}}^{-\boldsymbol{1}} \longrightarrow 0\right)$, then

$$\hat{f}_{MAP} = [\boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{K}_{\boldsymbol{\varepsilon}}^{-\boldsymbol{1}} \boldsymbol{A}]^{-1} \boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{K}_{\boldsymbol{\varepsilon}}^{-\boldsymbol{1}} y$$

which is the conventional weighted least-squares (WLS) estimator which also happens to be the ML estimator in this case. If in addition, we have additive white Gaussian noise, $\boldsymbol{K}_{\boldsymbol{\varepsilon}} = \sigma^2 \boldsymbol{I}$, we obtain the ML or least squares solution given in Eq. (3.24) which was not very robust against noise. Thus, the prior information is essential to avoid the noise problems associated with the ML solution.

We can study the stabilizing effect of incorporating a prior term in our cost function by first supposing that the object vector is zero mean, $\mu_x = 0$, and has independent and identically $\boldsymbol{K}_{\boldsymbol{f}} = \frac{1}{\beta} \boldsymbol{I}$ distributed components (iid). We can further suppose that the noise is again white and Gaussian, $\boldsymbol{K}_{\boldsymbol{\varepsilon}} = \sigma^2 \boldsymbol{I}$ and thus can rewrite our cost function compactly as:

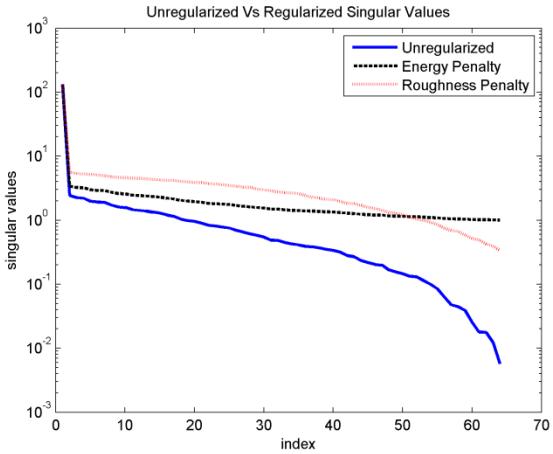
$$\Psi(f) = \frac{1}{2}(y - \boldsymbol{A}f)^{\boldsymbol{T}} \frac{1}{\sigma^2}\boldsymbol{I}(y - \boldsymbol{A}f) + \frac{1}{2}(\mathbf{f} - \mu_f)^{\boldsymbol{T}}\beta\boldsymbol{I}(f - \mu_f)$$
$$= \frac{1}{2\sigma^2}\|y - \boldsymbol{A}f\|^2 + \frac{1}{2}\beta\|f\|^2 \tag{3.39}$$

We see that under these unrealistic assumptions (neighboring pixels are not necessarily iid), the cost function consists of a least squares data term (negative log-likelihood) and an energy penalty term. The MAP estimator for this case is

$$\hat{f}_{MAP} = [\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A} + \beta\sigma^2\boldsymbol{I}]^{-1}\boldsymbol{A}^{\boldsymbol{T}}y \tag{3.40}$$

Previously we had addressed the non-uniqueness of our inverse problem by parameterizing our object which resulted in a set of linear equations. However, as we saw in the ML scenario, the discrete system is ill-condition and the presence of noise can result in large errors in our estimate of the object parameters. As Eq. (3.38) shows, the presence of a penalty term perturbs the singular values of the original ill-conditioned system by pulling them away from zero. Since β is usually much greater than the smallest singular value of the original system, the new system has all its singular values greater than or equal to β and thus the new system is better conditioned.

We further reinforce this point through a simple simulation. We generated a random system matrix, $\boldsymbol{A}$, that is 120x64 and so $\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}$ is 64x64. Plotted in Fig. 3.3a) are the singular value spectra of the original system, and the new systems with an energy penalty term and a roughness penalty.



	σ_1	σ_{64}	K
Original System	133.184	0.006	23268.792
Energy Penalty	134.184	1.006	133.420
Roughness Penalty	133.193	0.335	397.404

Figure 3.3. Shows a) plot of singular values of original system and new systems with energy and roughness penalty and b) table listing the largest and smallest singular values and condition numbers for the original system and the new systems.

The roughness penalty term, similar to the energy term, embodies prior information about our object vector. However, in the context of Bayesian estimation under which the energy penalty term was derived, a stochastic model was chosen for both the signal and the noise process. However, in the roughness penalty case, our unknown object vector is represented as a deterministic spatial function while the noise is still represented by a stochastic process. The roughness penalty term is derived under the framework of penalized likelihood estimation which is very similar to Bayesian estimation but differs in its interpretation of the penalty term and will be discussed in the next section.

According to the table in Fig. 3.3b) the energy term improves the condition number of the regularized system and thus there does not appear to be any issues with assuming a Gaussian prior with zero mean. However for large β values, the estimator assumes our confidence in the prior mean is high and thus will drive our solution to the mean, which happens to be zero. Shrinking the estimate to zero is not consistent with most imaging problems. Hence, it would seem that a Gaussian prior with zero mean is an unrealistic model form most imaging problems. In general, it is very difficult to formulate realistic prior but it is easier to penalize estimates that deviate from prior assumptions.

Penalized Likelihood Estimation

Often ill-posed problems are ill-conditioned when discretized and thus any perturbations due to noise in the system matrix or in the measurement vector will lead to large errors in our estimate of $\hat{f}$. Diffraction inherently acts as a low-pass filter in the forward direction and since image reconstruction involves inverting diffracted fields, we would expect the pseudo inverse of our system matrix to be a high pass filter. The amplification of noise will lead to large image oscillations contrary to our prior expectations of what are images should look like. Noise-reduction methods can be employed to reduce the effect of noise on our estimate. One can either modify the data by prefiltering or "denoising" or one can modify the algorithm derived for the ill-conditioned problem. That is, we can incorporate a filtering step at every iteration or modify the update to dampen high-spatial frequencies. However, the two approaches to noise-reduction cannot guarantee a unique and stable solution. Conversely, true regularization can result in a unique and stable estimate.

By redefining the problem to eliminate ill-conditioning we avoid the need to modify the data or the algorithm. There are several options for regularization and they all have their apparent advantages and disadvantages. However, for image reconstruction, a straightforward method of regularization is adding a penalty term to the cost function very similar to what was done in MAP estimation [54]. This approach termed penalized likelihood estimation is generally more attractive than MAP estimation because it is often very difficult to derive realistic priors than penalty functions for images [54]. More specifically, we seek an estimate, $\hat{\boldsymbol{f}}$, by minimizing a cost function of the following form:

$$\hat{f} = arg\min_{f} \Psi(f)$$
$$\Psi(f) = -\log p(y|f) + \beta R(f) \tag{3.41}$$

The negative of the log-likelihood term quantifies disagreement between f and the measurement vector y and thus this term should be small. The second term is a regularizing penalty function $R: \mathbb{R}^{np} \rightarrow \mathbb{R}$ that penalizes an object vector f by how much it departs from our assumptions about image properties [54]. The regularization parameter β controls the tradeoff between the fit to the data and our assumptions of how $\hat{f}$ should behave. For very small β, $\hat{f}$ will closely fit the data, which usually means very good spatial resolution in the absence of noise, but very noisy estimates in the presence of noise [54]. On the other hand, for large β, the emphasis will be on minimizing $R(f)$, which usually results in a smoother (filtered) estimate with low noise. In short the amount of regularization should correlate with the amount of noise present. Selecting a penalty function is often subjective and problem dependent. For example if we expect the object vector to have small values then the energy penalty, corresponding to a Gaussian prior as discussed earlier, is an option

$$R(f) = \frac{1}{2}\|f\|^2 = \sum_{j=1}^{n_p} \frac{1}{2}|f_j|^2 \tag{3.42}$$

However for most problems we expect the object to have nonzero energy so this may not be a good penalty function for image reconstruction. It is essential to have a penalty term because the lack of such a term in the cost function leads to the inverse solution given in by maximum likelihood estimation which greatly amplifies noise and leads to large image oscillations. These oscillations are contrary to our prior expectations about

what images should look like. Therefore penalty functions that discourage highly oscillatory images are more natural than an energy penalty. An apparent disadvantage is the subjective choice of the penalty function and in the apparent difficulty in choosing the penalty parameter, β. For image reconstruction, we generally seek a penalty or a potential function that discourages high spatial frequency oscillation and in our search for such a function, we must consider whether it will be:

1. Separable vs. nonseperable
2. Quadratic vs nonquadratic

Separable penalty functions are simpler to minimize and can be expressed as follows:

$$R(f) = \sum_{j=1}^{n_p} \psi(f_j) \tag{3.43}$$

where $\psi(t)$ is a potential function. For example for the energy penalty function, the potential function is given by $\psi(t) = \frac{t^2}{2}$ and thus:

$$R(f) = \sum_{j=1}^{n_p} \frac{f_j^2}{2} = \frac{1}{2} f^t \boldsymbol{I} f \tag{3.44}$$

As we have seen earlier, the penalty energy function penalizes large values of f but does not explicitly enforce smoothness. One way to enforce smoothness constraint is to penalize anenergy functional that measure the variation with in a function. In order to find smooth solutions, we can define a norm on the solution space. For a one-dimensional function $f(x)$, we can integrate the squared first or second derivative of the function,

$$\begin{aligned} R_1 &= \int f_x^2(x)\, dx \\ R_2 &= \int f_{xx}^2(x)\, dx \end{aligned} \tag{3.45}$$

In two dimensions, the corresponding smoothness functionals are given by:

$$\begin{aligned} R_1 &= \int f_x^2(x,y) + f_y^2(x,y)\, dx\, dy = \int \|\nabla f(x,y)\|^2 dx\, dy \\ R_2 &= \int f_{xx}^2(x,y) + 2f_{xy}^2(x,y) + f_{yy}^2(x,y)\, dx\, dy \end{aligned} \tag{3.46}$$

where the mixed term $2f_{xy}^2$ is needed to make the measure rotationally invariant.

The above first order and second order roughness penalty functions for one and two-dimensions are for a continuous space image $f(x,y)$. There are different methods for obtaining discrete forms of the above continuous roughness measures; finite differences, splines, and finite element methods. The finite difference method is the simplest among the three. Using finite difference methods, the first and second order derivative of a one-dimensional function can be expressed as

$$f_x = f_j - f_{j-1} \quad and \quad f_{xx} = f_{j+1} - 2f_j + f_{j-1} \tag{3.47}$$

The discrete forms of the first and second order derivatives in the two-dimensional case are analogous to the one-dimensional case. The finite difference values at boundary points must be specified by boundary conditions. The discrete forms of roughness penalty function measure the variation among neighboring pixels and penalize large variations between neighboring pixels, thereby discouraging roughness and favoring a spatially smoother estimate. To study the effect of such penalty functions, we can consider the one-dimensional case and generalize our results to the two-dimensional case of images.

A 1st-order roughness penalty measures the variation among neighboring pixels thereby improving on the energy penalty and can be expressed as:

$$R(f) = \sum_{j=2}^{N} \frac{1}{2} \left| f_j - f_{j-1} \right|^2 \tag{3.48}$$

In matrix-vector form, the above equation can be re-written using *N-1xN differencing matrix, C*. To begin with, *C* can be expressed as:

$$\boldsymbol{C} = \boldsymbol{D}_N = \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ & & \ddots & \ddots & & \\ 0 & \cdots & 0 & -1 & 1 & 0 \\ 0 & \cdots & 0 & 0 & -1 & 1 \end{bmatrix}, \quad so\ that\ \boldsymbol{C}f = \begin{bmatrix} f_2 - f_1 \\ \vdots \\ f_N - f_{N-1} \end{bmatrix} \tag{3.49}$$

With $\boldsymbol{C}$ being defined so, we see that $[\boldsymbol{C}f]_k = f_{k+1} - f_k$, so that

$$R(f) = \sum_{k=1}^{N-1} \frac{1}{2} |[\boldsymbol{C}f]_k|^2 = \frac{1}{2} \|\boldsymbol{C}f\|^2 = \frac{1}{2} f' \boldsymbol{C}' \boldsymbol{C} f = \frac{1}{2} f' \boldsymbol{R} f \tag{3.50}$$

The matrix $\boldsymbol{R}$ is the Hessian of the roughness penalty $R(f)$; the matrix of its second partial derivatives. The matrix for $\boldsymbol{R}$ can be expressed as:

$$\boldsymbol{R} = \boldsymbol{C}'\boldsymbol{C} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 \\ & & \ddots & \ddots & \ddots & \\ 0 & \cdots & 0 & -1 & 2 & -1 \\ 0 & 0 & \cdots & 0 & -1 & 1 \end{bmatrix} \tag{3.51}$$

This type of penalty function is a quadratic penalty because $R(f)$ is a quadratic form in f and but not separable since it couples neighboring pixels. Quadratic penalties have nice attractive properties such as simpler optimization through conjugate gradient. They provide a global smoothing of the estimate and hence it is easier to quantify the noise-resolution tradeoff. They are also particularly convenient when the measurements have a Gaussian distribution. The roughness penalty R in the forward domain is a highpass filter but acts as a lowpass filter in the inverse domain. Hence when applied to an image, it has the drawback of smoothing it. As a result, we can reduce noise only by compromising the spatial resolution and the only way we can only overcome this noise-resolution tradeoff is by considering nonquadratic penalty functions.

Combining the log-likelihood corresponding to the linear Gaussian case given by Eq. (3.15) with the quadratic penalty yields the following cost function for penalized likelihood estimation:

$$\Psi(f) = \frac{1}{2}(y - \boldsymbol{A}\boldsymbol{f})^T \boldsymbol{K}_\varepsilon^{-1}(y - \boldsymbol{A}f) + \frac{\beta}{2} f^T \boldsymbol{R} f \tag{3.52}$$

By taking the gradient of Eq. (3.52) and setting it to zero, we obtain the following expression for the minimizer:

$$\hat{f} = [\boldsymbol{A}^T \boldsymbol{K}_\varepsilon^{-1} \boldsymbol{A} + \beta \boldsymbol{R}]^{-1} \boldsymbol{A}^T \boldsymbol{K}_\varepsilon^{-1} y \tag{3.53}$$

In comparing Eq. (3.53) to Eq. (3.38), we see that the quadratically penalized weighted least-squares (QPWLS) estimator is a special case of the MAP estimator with $\mu_f = 0$ and $\beta \boldsymbol{R} = \boldsymbol{K}_f^{-1}$. In Bayesian terminology the roughness penalty function corresponds to the following improper Gaussian prior:

$$p(f) = ce^{-\frac{\beta}{2} f^T \boldsymbol{R} f} \tag{3.54}$$

This prior is improper because the hessian $\boldsymbol{R}$ is singular and in general most useful quadratic penalty functions have singular hessians which is acceptable in the framework of penalized-likelihood estimation but not in Bayesian [54].

To gain insight into the filtering properties of the QPWLS estimator, we apply circulant analysis to Eq. (3.53) [55]. We must first alter the form of $R(f)$ given in Eq. (3.38) to make it circularly shift-invariant and convenient for analysis:

$$R(f) = \frac{1}{2}|f_1 - f_N|^2 + \sum_{j=2}^{N} \frac{1}{2}\left|f_j - f_{j-1}\right|^2 \tag{3.55}$$

The differencing matrix $\boldsymbol{C}$ is now $N \times N$ and has the following circular form

$$\begin{aligned} \boldsymbol{C} = \boldsymbol{D}_N &= \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ & & \ddots & \ddots & \\ 0 & \cdots & 0 & -1 & 1 \\ 1 & 0 & \cdots & 0 & -1 \end{bmatrix}, \quad so\ that\ \boldsymbol{C}f \\ &= \begin{bmatrix} f_2 - f_1 \\ \vdots \\ f_N - f_{N-1} \\ f_1 - f_N \end{bmatrix} \end{aligned} \tag{3.56}$$

Likewise, $\boldsymbol{R}$ is a circulant matrix with the following form:

$$\boldsymbol{R} = \boldsymbol{C}'\boldsymbol{C} = \begin{bmatrix} 2 & -1 & 0 & \cdots & 0 & -1 \\ -1 & 2 & -1 & 0 & \cdots & 0 \\ & & \ddots & \ddots & \ddots & \\ 0 & \cdots & 0 & -1 & 2 & -1 \\ -1 & 0 & \cdots & 0 & -1 & 2 \end{bmatrix} \tag{3.57}$$

Since $\boldsymbol{R}$ is circulant, it can be diagonalized by the discrete Fourier transform (DFT); that is, it's eigenvalue decomposition is given by $\boldsymbol{R} = \boldsymbol{Q\Omega Q^T}$ where $\boldsymbol{Q}$ is the $N \times N$ orthonormal DFT matrix with elements

$$Q_{nk} = \frac{1}{\sqrt{N}} e^{i\frac{2\pi}{N}kn}, \quad k, n = 0, \ldots\ldots, N-1 \tag{3.58}$$

and $\boldsymbol{\Omega}$ is a diagonal matrix with eigenvalues that are the DFT coefficients of the first column of $\boldsymbol{R}$. $\boldsymbol{Q}'$ performs an *fft* on an input vector, $\boldsymbol{\Omega}$ scales vector, and $\boldsymbol{Q}$ transforms the vector back to the space domain. The first column of $\boldsymbol{R}$ can be expressed as:

$$R = 2\delta[m] - \delta[m-1] - \delta[m+1] \tag{3.59}$$

The DFT of Eq. (3.59) shows that the eigenvalues of $\boldsymbol{\Omega}$ are

$$R_k = 2 - e^{-i\frac{2\pi k}{N}} - e^{i\frac{2\pi k}{N}} = 2 - 2\cos(2\pi k/N) \tag{3.60}$$

Suppose we further assume that the noise is white, $\boldsymbol{K}_\varepsilon = \sigma^2 \boldsymbol{I}$ and that the system matrix is also circulant and hence diagonalizable by the DFT, $\boldsymbol{A} = \boldsymbol{Q\Gamma Q^T}$. We then substitute our circulant approximation of $\boldsymbol{A}$ and $\boldsymbol{R}$ into the expression for the QPWLS estimator to yield the following:

$$\begin{aligned}\hat{f} &= [\boldsymbol{A^T A} + \beta\sigma^2 \boldsymbol{R}]^{-1}\boldsymbol{A^T} y \\ &= \boldsymbol{Q}[\boldsymbol{\Gamma^T \Gamma} + \beta\sigma^2 \boldsymbol{\Omega}]^{-1}\boldsymbol{\Gamma^T Q^T} y = \boldsymbol{Q} diag\{L_k\}\boldsymbol{Q^T} y\end{aligned} \tag{3.61}$$

The expression above shows that the QPWLS estimator is a linear, circularly shift-invariant, filter with frequency response given by:

$$L_k = \frac{B_k^*}{|B_k|^2 + \beta\sigma^2 R_k} \tag{3.62}$$

We illustrate the overall low pass nature of the QPWLS estimator with a simple simulation. Suppose our system matrix is circulant with impulse response given by:

$$b[n] = \frac{1}{4}\delta[n-1] + \frac{1}{2}\delta[n] + \frac{1}{4}\delta[n+1] \tag{3.63}$$

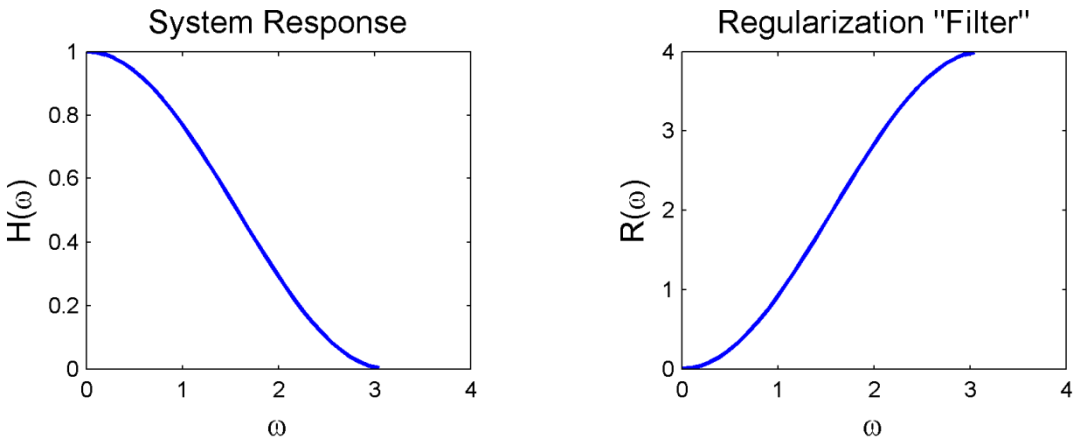


Figure. 3.4. Shows a) System response and b) regularization filter.

The system response, illustrated in Fig. 3.4a), is a low pass filter, $fft(\boldsymbol{A}e_1)$, and so the inverse of such a filter would be amplify high frequency noise. Hence it would make intuitive sense that the regularization filter, Fig. 3.4b), has a high-pass nature and thus acts low-pass filter in the inverse case thereby mitigating the effect of the system response. The competing filters are balanced by the β term. We illustrate in the following figure the overall response of both filters for both a small β, Fig. 3.5a), and a large β value, Fig. 3.5b).

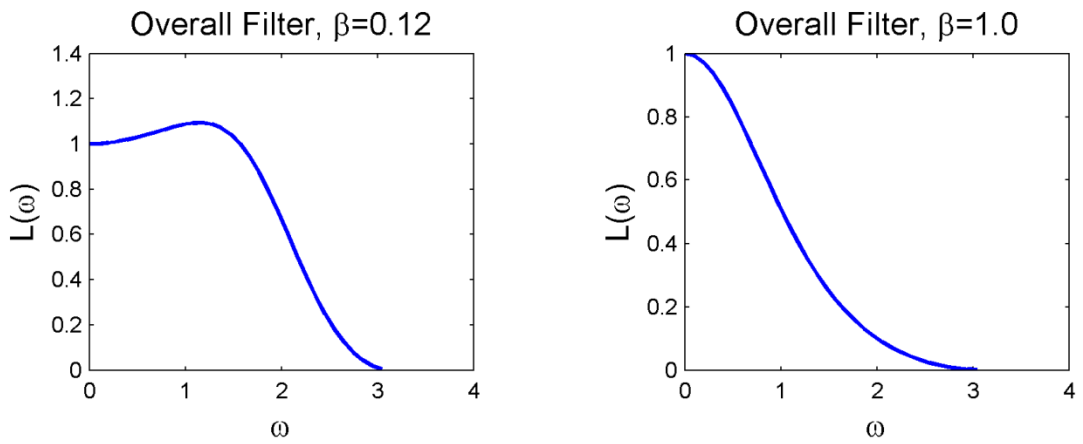


Figure 3.5. Overall system response for a) β=0.12 and b) $\beta = 1.0$.

We see that for both large and small β values, the overall system response is a low pass filter and so we can reduce noise only by compromising spatial resolution. We further see that for large β the cutoff frequency of the overall filter is larger than the cutoff frequency for the same system but with a smaller β value leading to less noise but more smoothing of the image. There is a fundamental tradeoff between noise and resolution that is controlled by β. We can illustrate the dependence of this tradeoff on β by analyzing the mean and variance of the QWPLS estimator. Furthermore, by analyzing the mean or bias and the variance of the QWPLS estimator, we can quantify the mean-square-error (MSE) of our estimate. To obtain an expression for the bias, we note that the estimator is linear and can be written in terms of an operator as $\hat{f} = \boldsymbol{L}y$ where

$$\boldsymbol{L} = [\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A} + \beta\sigma^2\boldsymbol{R}]^{-1}\boldsymbol{A}^{\boldsymbol{T}} \tag{3.64}$$

We have assumed in Eq. (3.64) that the noise is white Gaussian, $\varepsilon \sim N(0, \sigma^2\boldsymbol{I})$. By linearity of the expectation operator, we have that:

$$E[\hat{f}|f] = E[\boldsymbol{L}y|f] = \boldsymbol{L}E[y|f] = \boldsymbol{L}\boldsymbol{A}f = [\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A} + \beta\sigma^2\boldsymbol{R}]^{-1}\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}f \tag{3.65}$$

Assuming that the system response is linear and shift invariant, ($\boldsymbol{A}$ is circulant), and since by construction $\boldsymbol{R}$ is circulant, we can use circulant analysis to simplify Eq. (3.65) to:

$$E[\hat{f}|f] = \boldsymbol{Q}[\boldsymbol{\Gamma}^{\boldsymbol{T}}\boldsymbol{\Gamma} + \beta\sigma^2\boldsymbol{\Omega}]^{-1}\boldsymbol{\Gamma}^{\boldsymbol{T}}\boldsymbol{\Gamma}\mathbf{Q}^{\mathbf{T}}f = \boldsymbol{Q}\text{diag}\{M_k\}\boldsymbol{Q}'f \tag{3.66}$$

Hence, the expectation of $\hat{f}$ is a filtered version of the true object f and the frequency response of the filter is given by

$$M_k = \frac{|B_k|^2}{|B_k|^2 + \beta\sigma^2 R_k} \tag{3.67}$$

where $\boldsymbol{\Gamma} = \text{diag}\{B_k\} = fft(\boldsymbol{A}e_1)$ and $\boldsymbol{\Omega} = \text{diag}\{R_k\} = fft(\boldsymbol{R}e_1)$. Ideally, for good spatial resolution, we would want this filter to pass all spatial frequencies with a gain of unity, which means we would like $\beta\sigma^2 R_k$ to be small.

The covariance of $\hat{f}$ for a given true object f is given by:

$$\begin{aligned} \text{Cov}\{\hat{f}|f\} &= \text{Cov}\{\boldsymbol{L}y|f\} = \boldsymbol{L}\text{Cov}\{\text{y}|\text{x}\}\boldsymbol{L}^{\boldsymbol{T}} = \boldsymbol{L}(\sigma^2\boldsymbol{I})\boldsymbol{L}^{\boldsymbol{T}} \\ &= \sigma^{\boldsymbol{2}}[\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A} + \beta\sigma^2\boldsymbol{R}]^{-1}\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}[\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A} + \beta\sigma^2\boldsymbol{R}]^{-1} \end{aligned} \tag{3.68}$$

Using circulant analysis, Eq. (3.68) simplifies to:

$$\begin{aligned} \text{Cov}\{\hat{f}|f\} &= \sigma^{\boldsymbol{2}}\boldsymbol{Q}[\boldsymbol{\Gamma}^{\boldsymbol{T}}\boldsymbol{\Gamma} + \beta\sigma^2\boldsymbol{\Omega}]^{-1}\boldsymbol{\Gamma}^{\boldsymbol{T}}\boldsymbol{\Gamma}[\boldsymbol{\Gamma}^{\boldsymbol{T}}\boldsymbol{\Gamma} + \beta\sigma^2\boldsymbol{\Omega}]^{-1}\boldsymbol{Q}^T \\ &= \sigma^{\boldsymbol{2}}\boldsymbol{Q}\text{diag}\left\{\frac{|B_k|^2}{(|B_k|^2 + \beta\sigma^2 R_k)^2}\right\}\boldsymbol{Q}^{\boldsymbol{T}} \end{aligned} \tag{3.69}$$

For low-noise image reconstruction, we would like for the variance of the individual components, $\hat{f}_j$, of $\hat{f}$ to be small. The variance of $\hat{f}_j$ is given by:

$$\begin{aligned} \text{Var}\{\hat{f}_j|f\} &= e_j'\text{Cov}\{\hat{f}|f\}e_j \\ &= \frac{\sigma^2}{N}\sum_{k=0}^{N-1}\left|\frac{B_k}{|B_k|^2 + \beta\sigma^2 R_k}\right|^2 \end{aligned} \tag{3.70}$$

According to Eq. (3.70), in order to have small variances for $\hat{f}_j$, βR_k must be large. Comparing Eq. (3.67) and Eq. (3.70), we see an apparent tradeoff between bias (resolution) and variance (noise). For good spatial resolution, we want Eq. (3.67) to be approximately unity, which means $\beta R_k \approx 0$, but for low noise, Eq. (3.70) requires that βR_k be large. Hence, we need to choose β such that the variance and bias of our estimator are balanced.

Although there are several methods for computing the optimal regularization parameters that will result in a balance between bias and variance, these methods are computational intensive and hence it is easier to extract a range of reasonable βvalues graphically [56]. Having obtained expressions for the bias and variance, we are in a position to quantify the distance between an estimate, $\hat{f}$, and the true object,f, through the mean squared error (MSE). The MSE consists of both a bias and a variance term.

The MSE is initially due entirely to the variance of the noise in an estimate. To reduce the MSE, we would need to use a larger β value to obtain a smoother image. However, the smoothing introduces distortion and hence the bias term increases. The larger the regularization parameter,β, the more distortion. Since MSE is the sum of the bias and

variance, changes in both terms will affect the overall MSE. When we use a small β the reduction of variance outweighs the increase in the bias. As a result, the MSE is reduced. As β increases, the variance decreases but the bias will increase. Eventually, the increase in bias will exceed the reduction of variance and the MSE will get larger as β increases. We illustrate the tradeoff via a simple simulation of a 1D deconvolution problem in which the blur is given by:

$$b[n] = \frac{1}{4}\delta[n+1] + \frac{1}{2}\delta[n] + \frac{1}{4}\delta[n-1] \tag{3.71}$$

We can show the tradeoff between spatial resolution by measuring the full width half maximum (FWHM) of an impulse and noise variance over a range of values for β for both first order and second order finite differences. The arrow in the figure shows the direction of increasing β.

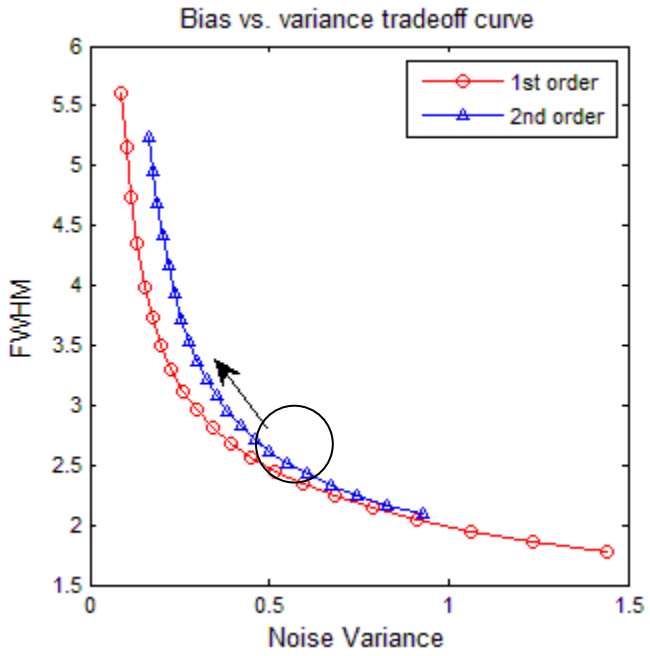


Figure 3.6. Bias vs. Variance tradeoff curve for 1st and 2nd order differences.

The distance between a point on one of the curves in Fig. 3.6 and the origin is the $\sqrt{MSE}$ for a particular value of the regularization parameter, β. The minimum MSE is the point on a bias-variance curve that has the minimum distance from the origin.

Although we have provided sufficient reasons for using a roughness penalty by considering a 1D example, we require a 2D roughness penalty function for image reconstruction. We can easily extrapolate Eq. (3.48) to 2D. The simplest 2D quadratic

roughness penalty discourages large variations between horizontal and vertical neighboring pixels, referred to as a first-order neighborhood roughness penalty. In concise notation, we can write the penalty function as:

$$R(f) = \sum_{j=1}^{n_p} \sum_{l \in \mathcal{N}_j} \frac{1}{2} |f_j - f_l|^2 \tag{3.72}$$

where $\mathcal{N}_j$ denotes half of the set of neighbors of the jth pixel. For a first-order neighborhood, $\mathcal{N}_j$ is the following set:

$$\mathcal{N}_j = \{j-1, j-N\} \tag{3.73}$$

If the quadratic roughness penalty is to include diagonal pixels as well, the 2nd order neighborhood set would be defined as follows:

$$\mathcal{N}_j = \{j-1, j-N, j-N-1, j-N+1\} \tag{3.74}$$

In matrix-vector representation and using the first-order neighborhood, we can write the 2D roughness penalty function as:

$$R(f) = \frac{1}{2} \|\boldsymbol{C}_1 f\|^2 + \frac{1}{2} \|\boldsymbol{C}_2 f\|^2 \tag{3.75}$$

Because of the lexicographic ordering of 2D image into a 1D object vector, the elements of $\boldsymbol{C}_1$ are as follows:

$$[\boldsymbol{C}_1]_{kj} = \begin{cases} 1, & k = j = 1 + n + mN \\ -1 & k = 1 + n + mN, \quad j = k - 1 \\ 0 & otherwise \end{cases} \tag{3.76}$$

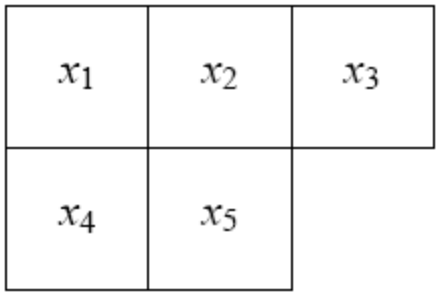
for $n = 1, \dots, N-1, m = 0, \dots, M-1$, for $k = 2, \dots, n_p = NM$. We can define $\boldsymbol{C_2}$ similarly. We can write the quadratic penalty given in Eq. (3.75) even more concisely as:

$$R(f) = \frac{1}{2} \|\boldsymbol{C} f\|^2 \tag{3.77}$$

with $\boldsymbol{C}$ has dimensions $2NM \times NM$ and the following form

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{C_1} \\ \boldsymbol{C}_2 \end{bmatrix} \tag{3.78}$$

It's worth seeing an example of a 2D roughness penalty in order to understand the matrix-vector expressions. Suppose our images has the following pixels, then for the roughness penalty given by Eq. (3.77) $\boldsymbol{C}$ must have the following form:



$$\boldsymbol{C}f = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ -1 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \end{bmatrix} = \begin{bmatrix} f_2 - f_1 \\ f_3 - f_2 \\ f_5 - f_4 \\ f_4 - f_1 \\ f_5 - f_2 \end{bmatrix}$$

and so,

$$R(f) = \sum_{k=1}^{5} |[\boldsymbol{C}f]_k|^2 = |f_2 - f_1|^2 + |f_3 - f_2|^2 + |f_5 - f_4|^2 + |f_4 - f_1|^2 + |f_5 - f_2|^2$$

As we have seen the quadratic roughness penalty, albeit nonseparable, yields behaviors that are better than the separable energy penalty. Furthermore, it's quadratic and hence its convexity makes it easy to minimize. However, as discussed earlier, a quadratic roughness penalty causes blurring of edges. In order to preserve edges, we would need to replace the quadratic function with a nonquadratic function and even possibly a non-convex function. Nonquadratic or edge-preserving potential functions are quadratic near zero, but are approximately linear far from zero. This property will encourage most neighboring pixels to be similar, but will also allow them to be different in image locations where there are sufficient discrepancies between neighbors, such as image edges. In general, it is possible to use different potential functions for different spatial locations and thus we have the following general form for the penalty function:

$$R(f) = \sum_{k=1}^{K} \psi_k([\boldsymbol{C}f]_k) \tag{3.79}$$

where $[\boldsymbol{C}f]_k = \sum_{j=1}^{n_p} c_{kj} f_j$. The matrix $\boldsymbol{C}$ is $K \times n_p$ where $n_p = NM$. For example, for the 2D case with vertical and horizontal neighbors, we have $K = 2NM$.

Although the analysis of the properties of penalized-likelihood estimator $\hat{f}$ is more difficult when the regularizer is nonquadratic, we can still get some insight into the edge-

preserving characteristics by deriving a recursive expression for $\hat{f}$. Despite not having a quadratic penalty function, the penalized least-squares cost function will still have the following form:

$$\Psi(f) = \frac{1}{2}\|y - \boldsymbol{A}f\|^2 + \beta \boldsymbol{R}(f) \tag{3.80}$$

where the penalty function has the general form of Eq. (3.79) but with $\psi_k = \psi$; that is we only use one potential function for all locations of the image and that the noise variance has been absorbed by β.

The following plot illustrates two examples of non-quadratic but convex edge-preserving potential functions that can be used as penalty functions. The quadratic penalty function was also plotted to show that these two edge-preserving functions behave much like the quadratic penalty functions for neighboring pixel values whose variations are relatively small. However, far from zero, the behavior of the edge-preserving penalty functions deviate from the shape of a parabola. Thus the edge-preserving penalty functions assign a lower cost for large differences and hence this leads to edge preservations. We plotted the Huber function which has the following form:

$$\psi(t) = \begin{cases} \frac{1}{2}|t|^2, & |t| < \delta \\ \delta|t| - \frac{1}{2}|\delta|^2, & |t| > \delta \end{cases} = \min\left(\frac{1}{2}|t|^2, \delta|t| - \frac{1}{2}|\delta|^2\right) \tag{3.81}$$

and the hyperbola function which has the following form

$$\psi(t) = \delta^2\left(\sqrt{1 + \left(\frac{t}{\delta}\right)^2} - 1\right) \tag{3.82}$$

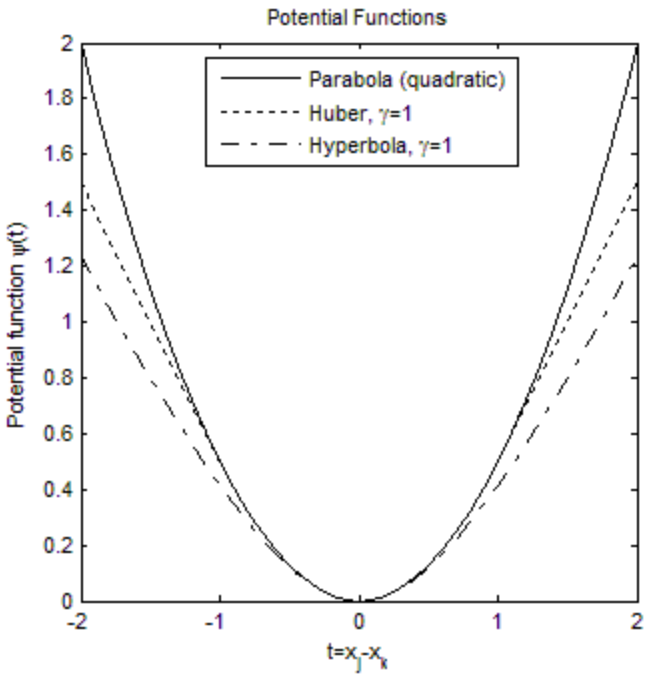


Figure 3.7. Illustration of potential functions ψ used for regularization: the quadratic function, and two nonquadratic edge-preserving functions, the Huber function and the "broken parabola" function.

Typical edge-preserving potential functions have the property that they are nearly quadratic near zero and roughly linear when the argument exceeds δ. The parameter δ controls the transistion between smoothing and edge-preserving. Hence for nonquadratic edge-preserving functions, we now have an additional parameter δ that we must set.

For edge preserving functions, circulant approximations cannot be used because such penalty functions introduce local shift-variant effects. However, we make the following assumptions with regards to the potential function ψ:

- ψ is differentiable with derivative $\dot{\psi}$.
- The following weighting function is defined (finite) and nonnegative for all t:

$$\omega_\psi(t) \triangleq \frac{\dot{\psi}(t)}{t}$$

Fig. 3.8 illustrates the weighting function ω_ψ for the potential functions shown in Fig. 3.7. As stated earlier, the parameter δ controls the transition between smoothing and edge-preserving and so one must have in advance a rough idea of the anticipated differences between neighboring pixels that straddle edge boundaries, or use trial and error to find δ.

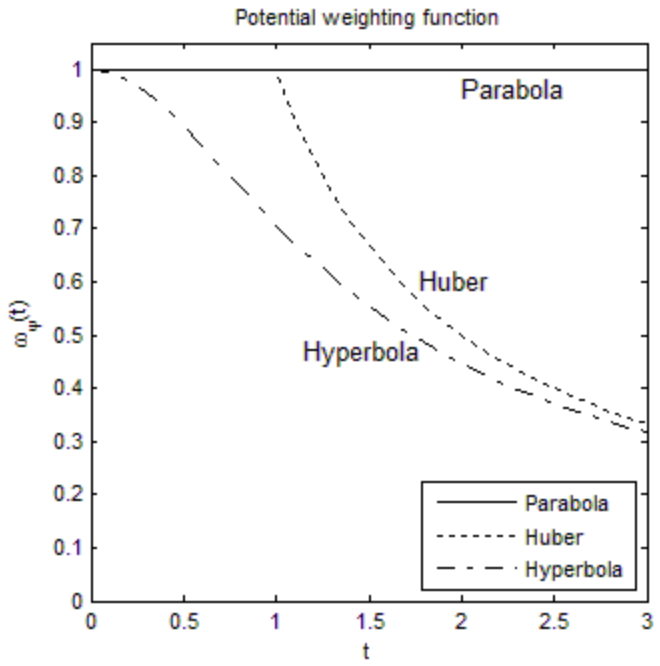


Figure 3.8. The potential weighting function $\omega_\psi(t)$ for a quadratic potential, Huber potential, and a hyperbola potential for $\delta = 1$.

To obtain a recursive expression for the nonquadratic estimator, we set the gradient of the cost function to zero. We begin by first taking the gradient of the penalty function is:

$$\begin{aligned}\nabla R(f) &= \sum_{k=1}^{K} \nabla\psi\left(c_k^T f\right) \\ &= \sum_{k=1}^{K} c_k \dot{\psi}([\boldsymbol{C}f]_k) = \sum_{k=1}^{K} c_k \omega_\psi([\boldsymbol{C}f]_k)[\boldsymbol{C}f]_k = \boldsymbol{C^T D}(f)\boldsymbol{C}f\end{aligned} \tag{3.83}$$

where $c_k^T = e_k^T \boldsymbol{C}$ denotes the kth row of **C** and we define the following $K \times K$ diagonal weighting matrix:

$$\boldsymbol{D}(f) = \text{diag}\{\omega_\psi([\boldsymbol{C}f]_k)\} \tag{3.84}$$

Thus the expression of the gradient for the cost function can be written as:

$$\nabla\Psi(f) = -\boldsymbol{A^T}(y - \boldsymbol{A}f) + \beta \boldsymbol{C^T D}(f)\boldsymbol{C}f \tag{3.85}$$

Equating the gradient of the cost function to zero and rearranging terms yields the following expression for the estimator $\hat{f}$:

$$\hat{f} = \left[\boldsymbol{A^T A} + \beta \boldsymbol{C^T D}\left(\hat{f}\right)\boldsymbol{C}\right]^{-1} \boldsymbol{A^T} y \tag{3.86}$$

Although, the expression for the estimator is a recursive expression because $\hat{f}$ is present on the right-hand side, it does provide insight into the properties of the nonquadratic

penalty functions. First we note that in the quadratic case where $(t) = {t^2}/{2}$, we have that $\omega_\psi(t) = 1$ so that $\boldsymbol{D}(\hat{f}) = \boldsymbol{I}$ and we have the usual quadratic penalized least-squares estimator given in Eq. (3.53).

For more insight into the nonquadratic estimator, we consider a 1D problem. If we knew where the edges were, we would use a spatially weighted penalty function [74]:

$$R(f) = \sum_{j=2}^{N} \omega_j \psi(f_j - f_{j-1}) \tag{3.87}$$

where we would set $\omega_j = 0$ between pixels straddling an edge, and $\omega_j = 1$ for pixels in uniform regions and this would preserve the edges. In the quadratic case where $\psi(t) = {t^2}/{2}$, this regularizer would have the form:

$$\hat{f} = \left[\boldsymbol{A^T A} + \beta \boldsymbol{C^T} \text{diag}\{\omega_j\} \boldsymbol{C}\right]^{-1} \boldsymbol{A^T} y \tag{3.88}$$

However, in practice we usually do not know where the edges are in advance, so we must have the algorithm locate them. Comparing Eq. (3.88) and Eq. (3.86), we see that the only real difference is that in Eq. (3.86) the weights depend on the estimate $\hat{f}$. If we consider the Huber potential function and its corresponding weighting function, we see that when the difference $[\boldsymbol{C}f]_k$ between neighboring pixels exceeds δ, the corresponding weight is reduced from 1 by ω_ψ similar to the effect of Eq. (3.88). Thus instead of needing to know the edge locations in advance, a nonquadratic penalty function can provide estimate-based weighting.

E. General Purpose Minimization Algorithms

In general, closed-form solutions are unavailable for the minimizer $\hat{f}$ of the cost function Eq. (3.41), so iterative algorithms are required. An iterative algorithm is a procedure that starts with an initial guess $f^{(0)}$ for $\hat{f}$ and then recursively generates a sequence $f^{(1)}, f^{(2)}, \ldots,$ also denoted $\{f^{(n)}\}$ [57]. In order to minimize a cost function, we make the following assumptions about $\Psi(f)$:

- $\Psi(f)$ is differentiable
- $\Psi(f)$ has a finite global minimizer

$$\exists\, \hat{f} \in \mathbb{R}^{n_p} : -\infty < \Psi(\hat{f}) \leq \Psi(f), \qquad \forall f \in \mathbb{R}^{n_p} \tag{3.89}$$

A minimizer of a differentiable cost function $\Psi(f)$ is necessarily the solution of the following n_p equations in n_p unknowns:

$$\nabla\Psi(f)|_{f=\hat{f}} = 0 \tag{3.90}$$

In general there are no closed-form solutions to the system of equations, unless we are working with a least squares or a weighted least squares problem, where closed form solutions do exist. However, these direct solutions are computation intensive and thus an iterative algorithm is generally required even in those cases. When developing algorithms for image reconstruction, there are many design considerations such as:

- Convergence rate
- Computation time per iteration
- Sensitivity to numerical errors
- Storage requirements

However, even more critical than the properties listed above, is the important property of monotonicity which plays a significant role in determining whether an algorithm will even converge [57]. An algorithm is monotone if it generates a sequence $\{f^{(n)}\}$ that decreases $\Psi(f)$ at every iteration. That is, if

$$\Psi\left(f^{(n+1)}\right) \leq \Psi\left(f^{(n)}\right), \qquad n = 0,1,2, \ldots. \tag{3.91}$$

Fixed Point Iteration

A necessary condition that leads to fixed point iteration algorithms can be obtained by multiplying both sides of Eq. (3.90) by some $n_p \times n_p$ matrix $M(f)$ of choice, that may or may not depend on f, and then subtracting both sides from $\hat{f}$ yielding:

$$\hat{f} = \hat{f} - M(f)\nabla\Psi(f) \tag{3.92}$$

If we then replace the equality in the above equation with a recursion, we obtain fixed point iterations or the method of successive approximations.

$$f^{(n+1)} = f^{(n)} - M\left(f^{(n)}\right)\nabla\Psi\left(f^{(n)}\right) \tag{3.93}$$

Fix point iterations are rarely globally convergent, although they are sometime locally convergent and if our choice for $M(f)$ is ad hoc, then the fixed point algorithm will most likely diverge [57]. An algorithm is globally convergent if $f^{(n)} \to \hat{f}$ for any starting point f^0. An algorithm is locally convergent if $f^{(n)} \to \hat{f}$ for some nonempty set of initial guesses x^0 that are sufficiently close to $\hat{f}$.

Preconditioned Gradient Descent (PGD)

If we choose $M(f)$ to be a positive scalar, α, multiplying a fixed preconditioning matrix $\boldsymbol{P}$, we obtain the following preconditioned gradient descent (PGD) algorithm [57]:

$$f^{(n+1)} = f^{(n)} - \alpha \boldsymbol{P} \nabla \Psi\left(f^{(n)}\right) \tag{3.94}$$

If we further assume certain properties of $\Psi(f)$ beyond those given in Eq. (3.89), we can specify conditions on α and $\boldsymbol{P}$ to ensure that the PGD algorithm will monotonically decrease the cost function.

Theorem 5.1 *If the gradient of $\Psi(f)$ satisfies a Lipschitz condition of the following form*

$$\|\nabla \Psi(f) - \Psi(z)\|_{P^{1/2}} \leq \|\mathbf{G}(\mathrm{f} - \mathrm{z})\|_{P^{1/2}}, \qquad \forall f, z \in \mathbb{R}^{n_p} \tag{3.95}$$

where $\boldsymbol{P}$ is a symmetric positive definite matrix, $\boldsymbol{G}$ *is some $n_p \times n_p$ matrix and where*

$$0 < \alpha < 2 / \left\| \boldsymbol{P}^{1/2} \boldsymbol{G} \boldsymbol{P}^{1/2} \right\| \tag{3.96}$$

then the PGD algorithm will monotonically decrease $\Psi(f)$.]. The matrix $\boldsymbol{G}$ should serve as an upper bound on the difference between the gradients at two points but should also be small so that the step size is large in order to achieve faster convergence. For example, suppose we want to find the smallest $\boldsymbol{G}$ for the regularized least-squares problem.

$$\Psi(f) = \frac{1}{2}\|y - \boldsymbol{A}f\|^2 + \frac{\beta}{2}\|\boldsymbol{C}f\|^2$$

$$\Psi(f) = \frac{1}{2}(y - \boldsymbol{A}f)'(y - \boldsymbol{A}f) + \frac{\beta}{2} f' \boldsymbol{C}' \boldsymbol{C} f$$

$$\nabla \Psi(f) = -\boldsymbol{A}'(y - \boldsymbol{A}f) + \beta \boldsymbol{C}' \boldsymbol{C} f$$

Subtracting the gradients at f and z yields the following expression:

$$\nabla \Psi(f) = (\boldsymbol{A}'\boldsymbol{A} + \beta \boldsymbol{C}'\boldsymbol{C})(f - z)$$

It suffices for regularized least squares to choose the $\boldsymbol{G} = (\boldsymbol{A}'\boldsymbol{A} + \beta \boldsymbol{C}'\boldsymbol{C})$ and to choose the maximum step size to be $\alpha < 2 / \left\| \boldsymbol{P}^{1/2} \boldsymbol{G} \boldsymbol{P}^{1/2} \right\|$. Thus for the regularized least-squares problem, in which the preconditioning matrix, $\boldsymbol{P} = \boldsymbol{I}$, if the step size is chosen to be less than the maximum eigenvalue of $\boldsymbol{G}$, the PGD algorithm will monotonically decrease the cost function. In this example we chose the preconditioning matrix $\boldsymbol{P}$ to be the identity matrix. This may not be the best choice. The best preconditioning matrix is the one which leads to the fastest asymptotic convergence rate of the PGD algorithm, i.e,

the one that leads to the least number of iterations. The convergence rate is defined by how fast $\|f^{(n)} - \hat{f}\| \to 0$.

For quadratic cost functions, we can find the exact asymptotic convergence rate but for non-quadratic cost functions, we must assume that $\Psi(f)$ is locally quadratic near the minimizer, i.e., for $f = \hat{f}$. The following graph illustrates a 1D non-quadratic cost function that is approximately parabolic near the minimizer point.

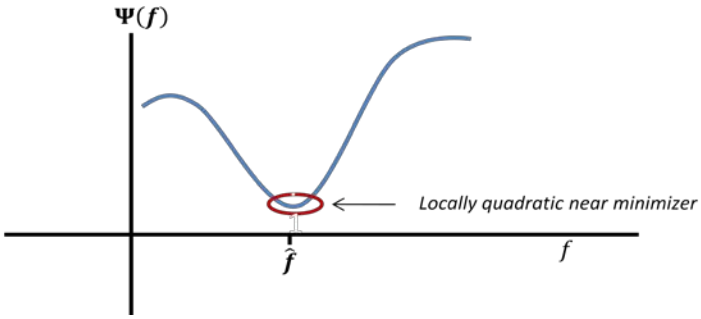


Figure 3.9. Local quadratic approximation for non-quadratic 1D cost functions.

Near the minimizer $\hat{f}$, the cost function can be expanded using a 2nd order Taylor series as follows:

$$\Psi(f) \cong \Psi(\hat{f}) + \nabla\Psi(\hat{f})(f - \hat{f}) + \frac{1}{2}(f - \hat{f})^{\boldsymbol{T}}\boldsymbol{H}(\hat{f})(f - \hat{f}), \qquad for\ f \approx \hat{f} \tag{3.97}$$

where $\boldsymbol{H} = \nabla^2\Psi(f)$ is the Hessian, the matrix of 2nd order derivatives, and has elements given by:

$$h_{ij}(f) = \frac{\partial^2}{\partial f_j \partial f_k}\Psi(f), \qquad k, j = 1, \dots, n_p \tag{3.98}$$

Furthermore, the necessary condition given in Eq. (3.90) eliminates the second term in Eq. (3.97). Therefore near the minimizer, we have the following approximate expression:

$$\Psi(f) \cong \Psi(\hat{f}) + \frac{1}{2}(f - \hat{f})^{\boldsymbol{T}}\boldsymbol{H}(\hat{f})(f - \hat{f}), \qquad for\ f \approx \hat{f}$$

and

$$\nabla\Psi(\hat{f}) = \boldsymbol{H}(\hat{f})(f - \hat{f}) \tag{3.99}$$

Inserting the expression for the gradient in Eq. (3.99) into the expression given by Eq. (3.94) and assuming that $f^{(n)}$ is sufficiently close to $\hat{f}$, we obtain the following:

$$f^{(n+1)} \approx f^{(n)} - \alpha\boldsymbol{P}\boldsymbol{H}(\hat{f})(f^{(n)} - \hat{f}) \tag{3.100}$$

Subtracting $\hat{f}$ from both sides of Eq. (3.100)

$$f^{(n+1)} - \hat{f} \approx \left(\boldsymbol{I} - \alpha \boldsymbol{PH}\left(\hat{f}\right)\right)\left(f^{(n)} - \hat{f}\right) \tag{3.101}$$

Using Cauchy-Schwarz inequality, we obtain an approximate expression for the asymptotic convergence rate:

$$\left\|f^{(n+1)} - \hat{f}\right\| \leq \left\|\boldsymbol{I} - \alpha \boldsymbol{PH}\left(\hat{f}\right)\right\| \left\|\left(f^{(n)} - \hat{f}\right)\right\| \tag{3.102}$$

and hence to within approximation:

$$\left\|f^{(n+k)} - \hat{f}\right\| \leq \left\|\left(\boldsymbol{I} - \alpha \boldsymbol{PH}\left(\hat{f}\right)\right)^{k}\right\| \left\|\left(f^{(n)} - \hat{f}\right)\right\| \tag{3.103}$$

Thus, we see that the asymptotic convergence rate of the PGD method is governed by the spectral radius of $\left(\boldsymbol{I} - \alpha \boldsymbol{PH}\left(\hat{f}\right)\right)$, provided that the norm is the $\mathcal{L}_2$ norm. Eq. (3.103) provides an upper bound on how fast the distance to the minimzer is decreasing at each iteration and that rate is governed by the norm of $\left(\boldsymbol{I} - \alpha \boldsymbol{PH}\left(\hat{f}\right)\right)$.

Ideally we would like the matrix, $\left(\boldsymbol{I} - \alpha \boldsymbol{PH}\left(\hat{f}\right)\right)$, to be as close to 0 as possible. For a given preconditioner $\boldsymbol{P}$, the best step size α satisfies $\alpha_* = \frac{2}{\lambda_{min}\left(\boldsymbol{PH}\left(\hat{f}\right)\right) + \lambda_{max}\left(\boldsymbol{PH}\left(\hat{f}\right)\right)}$, in which case the spectral radius of $\left(\boldsymbol{I} - \alpha_* \boldsymbol{PH}\left(\hat{f}\right)\right)$ is given by:

$$\rho\left(\boldsymbol{I} - \alpha_* \boldsymbol{PH}\left(\hat{f}\right)\right) = \frac{\lambda_{min}\left(\boldsymbol{PH}\left(\hat{\boldsymbol{f}}\right)\right) - \lambda_{max}\left(\boldsymbol{PH}\left(\hat{\boldsymbol{f}}\right)\right)}{\lambda_{min}\left(\boldsymbol{PH}\left(\hat{\boldsymbol{f}}\right)\right) + \lambda_{max}\left(\boldsymbol{PH}\left(\hat{\boldsymbol{f}}\right)\right)} = \frac{\kappa - 1}{\kappa + 1} \tag{3.104}$$

where $\kappa \triangleq \lambda_{max}\left(\boldsymbol{PH}\left(\hat{f}\right)\right) / \lambda_{min}\left(\boldsymbol{PH}\left(\hat{f}\right)\right)$ denotes the condition number of $\boldsymbol{PH}\left(\hat{f}\right)$**.** Hence fast convergence happens when $\boldsymbol{P}$ is chosen to minimize the condition number of $\boldsymbol{PH}\left(\hat{f}\right)$**.** For quadratic cost functions, the ideal precondition would be $\alpha \boldsymbol{P} = \boldsymbol{H}^{-1}\left(\hat{f}\right)$ and convergence is achieved in one step because $\left(\boldsymbol{I} - \alpha \boldsymbol{PH}\left(\hat{f}\right)\right) = 0$. For non-quadratic $\Psi(f)$, the inverse-Hessian preconditioner $\alpha \boldsymbol{P} = \boldsymbol{H}^{-1}\left(\hat{f}\right)$ would yield superlinear convergence rate.

For large scale applications, general purpose optimization methods should always be used with a preconditioner. We briefly digress a bit to give a brief intuitive understanding of preconditioning. As its name suggest, precondition is meant to improve the condition number of a matrix [58]. Suppose that $\boldsymbol{M}$ is a symmetric, positive definite matrix that approximates A but it is easier to invert. We can solve the unregularized system of equations, $\boldsymbol{A}f = y$ indirectly by solving $\boldsymbol{M}^{-1}\boldsymbol{A}f = \boldsymbol{M}^{-1}y$. If $\kappa(\boldsymbol{M}^{-1}\boldsymbol{A}) \ll$

$\kappa(\boldsymbol{A})$, or if the eigenvalues of $\boldsymbol{M}^{-1}\boldsymbol{A}$ are better clustered than those of $\boldsymbol{A}$, we can achieve faster convergence by using the preconditioner $\boldsymbol{M}^{-1}$.

Geometrically speaking, precondition is an attempt to stretch the quadratic cost function (or the locally quadratic approximation of a cost function) to make it appear more spherical so that the gradient vector points toward the minimizer and that the eigenvalues are close to each other. A perfect precondition is $\boldsymbol{M} = \boldsymbol{A}$, which scales the quadratic form along its eigenvector axes causing it to be perfectly spherical. Thus the solution is reached in one iteration. However, such a preconditioner is not useful at all since we would still have to solve $\boldsymbol{M}f = y$.

The simplest preconditioner is a diagonal matrix whose diagonal entries are identical to those of $\boldsymbol{A}$ The process of applying this preconditioner is equivalent to scaling the quadratic form along the coordinate axes which could prove to beneficial for large systems. Furthermore, a diagonal matrix is trivial to invert. There are many more sophisticated preconditioners that have been developed and in general one should use a preconditioner for large scale systems.

Preconditioned Steepest Descent (PSD)

An apparent improvement in the PGD algorithm, is to vary the step size in each iteration. If we define a search direction using the preconditioned gradient vector, we can then seek the minimizer of $\Psi(f)$ along that direction [57]. The search starts at an arbitrary point $\boldsymbol{f}_0$and then we slide down the gradient until we are close enough to the solution. Mathematically, the iterative procedure is given by

$$f^{(n+1)} = f^{(n)} + \alpha_n d^{(n)} \tag{3.105}$$

where $d^{(n)} = -P\nabla\Psi\left(f^{(n)}\right)$ is the preconditioned gradient at a given iteration. Now, the question, is how big should the step taken in that direction be? This line search should obviously move us to the point where the cost function takes on a minimum value. Hence a necessary condition for finding the minimizing step size α_n is that

$$\frac{\partial}{\partial\alpha}\,\Psi\left(f^{(n)} + \alpha d^{(n)}\right)|_{\alpha=\alpha_n} = 0 \tag{3.106}$$

Applying the chain rule, we get the following expression:

$$\nabla\Psi\left(f^{(n)} + \alpha d^{(n)}\right)^T d^{(n)}|_{\alpha=\alpha_n} = \nabla\Psi\left(f^{(n+1)}\right)^T d^{(n)}|_{\alpha=\alpha_n} = 0 \tag{3.107}$$

The above expression indicates that the next gradient $\nabla\Psi\left(f^{(n+1)}\right)$ and the current search direction $d^{(n)}$ are to be orthogonal if we are to find an exact minimizer. What we really have here is actually a minimization problem along a line, where the line is given by Eq. (3.105) for different values of α. This is usually solved as stated earlier by doing a line search; that is searching for a minimum point along a line. Hence, the search for a minimum of $\Psi(f)$ is reduced to a sequence of linear searches. We can obtain an expression for minimizing step as follows:

$$\alpha_n = arg\min_{\alpha}\Psi\left(f^{(n)} + \alpha d^{(n)}\right) = \nabla\Psi\left(f^{(n+1)}\right)^T d^{(n)} \tag{3.108}$$

Near the minimizer, the cost function is approximately parabolic and so $\nabla\Psi(f) = \boldsymbol{H}\left(\hat{\boldsymbol{f}}\right)\left(f - \hat{f}\right)$. Thus we have that:

$$\nabla\Psi\left(f^{(n+1)}\right)^T d^{(n)} = \left(f^{(n)} - \hat{f} + \alpha d^{(n)}\right)^T \boldsymbol{H}\left(\hat{\boldsymbol{f}}\right) d^{(n)} = 0 \tag{3.109}$$

Solving for α, we get the following expression for the minimizing step size:

$$\begin{aligned}\alpha_n = \nabla\Psi\left(f^{(n+1)}\right)^T d^{(n)} &= \frac{\left(f^{(n)} - \hat{f}\right)^T \boldsymbol{H}\left(\hat{f}\right) d^{(n)}}{d^{(n)\boldsymbol{T}} \boldsymbol{H}\left(\hat{f}\right) d^{(n)}} \\ &= \frac{\left(g^{(n)}\right)^T d^{(n)}}{d^{(n)\boldsymbol{T}} \boldsymbol{H}\left(\hat{f}\right) d^{(n)}}\end{aligned} \tag{3.110}$$

where $g^{(n)} = \nabla\Psi\left(f^{(n)}\right) = \boldsymbol{H}\left(\hat{f}\right)\left(f^{(n)} - \hat{f}\right)$. The preconditioned steepest descent can now be expressed as follows:

PSD

$$d^{(n)} = -P\nabla\Psi\left(f^{(n)}\right)$$

$$\alpha_n = arg\min_{\alpha}\Psi\left(f^{(n)} + \alpha d^{(n)}\right)$$

$$f^{(n+1)} = f^{(n)} + \alpha_n d^{(n)}$$

Choosing the preconditioner $\boldsymbol{P}$ follows similar analysis as given in the earlier section. It can be shown that for the fastest asysmptotic convergence rate, one should choose $\boldsymbol{P}$ to minimize the condition number of the product of $\boldsymbol{PH}$. If we define the error vector $\delta^{(\boldsymbol{n})} = \boldsymbol{P}^{-1/2}\left(f^{(n)} - \hat{f}\right)$ and the preconditioned Hessian matrix $\widetilde{\boldsymbol{H}} = \boldsymbol{P}^{1/2}\boldsymbol{H}\boldsymbol{P}^{1/2}$, one can show that error norm decreases each iteration by at least as much as the following:

$$\left\|\delta^{(n+1)}\right\|_{\tilde{\boldsymbol{H}}^{1/2}} \leq \frac{\kappa-1}{\kappa+1}\left\|\delta^{(n)}\right\|_{\tilde{\boldsymbol{H}}^{1/2}} \tag{3.111}$$

where $\kappa = \frac{\lambda_{max}(\tilde{\boldsymbol{H}})}{\lambda_{min}(\tilde{\boldsymbol{H}})}$ is the condition number of $\tilde{\boldsymbol{H}}$. Hence the closer κ is to unity the faster the convergence should be and hence we should choose $\boldsymbol{P}$ to minimize the condition number of the product of $\boldsymbol{PH}$. For a more thorough analysis of the Steepest Descent algorithm for quadratic cost functions and without preconditioning, refer to Appendix A.

The method of steepest descent is simple, easy to apply, and each iteration is fast. It is also very stable; if the minimum points exit, the method is guaranteed to locate them after at least an infinite number of iterations.

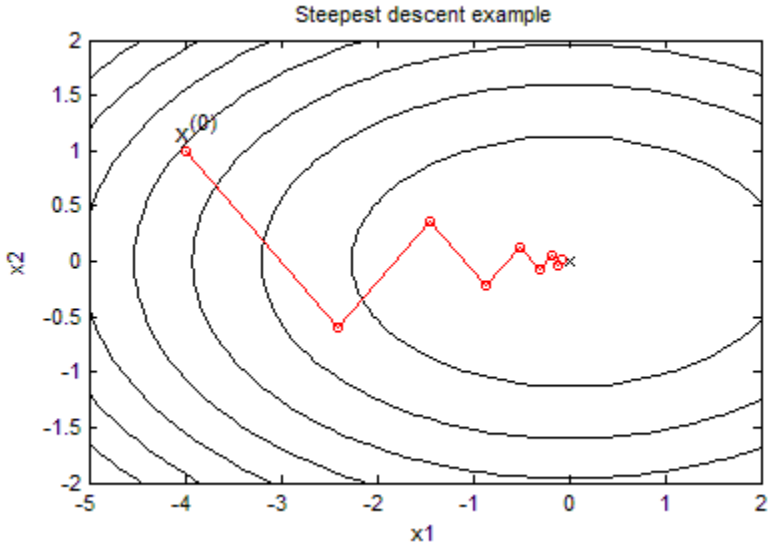


Fig. 3.10. Illustration of slow convergence of steepest descent. The PSD begins at $f^{(0)}$ and the iterates $\{f^{(n)}\}$ converge slowly to the minimizer $\hat{f} = 0$.

But even with all these positive characteristics, the method has one very important drawback; it generally has slow convergence. Fig. 3.10 illustrates the slow convergence of the PSD algorithm even for simple quadratic cost functions. For badly scaled systems; i.e. if the eigenvalues of the Hessian matrix at the solution point are different by several orders of magnitude, the method could end up spending an infinite number of iterations before locating a minimum point. It may start out with a reasonable convergence rate, but the progress gets slower and slower as the step size gets smaller and smaller near the

stationary point. The method may converge fast for such badly scaled systems, but is then very much dependent on a good choice of starting point. In short, the steepest descent method can be used where one has an indication of where the minimum is, but is generally considered to be a poor choice for many optimization problems. It is mostly only used in conjunction with other optimizing methods.

Preconditioned Conjugate Gradients

As seen in the previous subsection, the reason why the method of steepest descent converges slowly is that it has to take a right angle turn after each step and consequently search in the same direction as earlier steps. The method of conjugate gradients (CG) is an attempt to mend this problem by "learning from experience" [57, 58]. To further accelerate convergence, CG methods modify the search directions to ensure that they are mutually conjugate or approximately so for nonquadratic problems. Conjugacy means that two unequal vectors, d^i and d^j, are orthogonal with respect to symmetric positive definite matrix, $\boldsymbol{Q}$.

$$d_i^T \boldsymbol{H} d_j = 0 \tag{3.112}$$

This can be looked upon as a generalization of orthogonality, for which **H** is the unitary matrix. The idea is to let each search direction d_i be dependent on all the other directions searched to locate the minimum of $\Psi(f)$. A set of such search directions is referred to as a $\boldsymbol{H}$-orthogonal set and it will take a positive definite n_p-dimensional quadratic function such as our penalized least squares cost function to its minimum point in at most n_p exact linear searches for a quadratic cost function.

We would like to choose a search direction $d^{(n)}$ that is $\boldsymbol{H}_{(n-1)}$-orthogonal to the previous search direction $d^{(n-1)}$, where $\boldsymbol{H}_n$ is the Hessian of $\Psi(f)$ at $f^{(n)}$. For quadratic cost function, $\boldsymbol{H}_n = \boldsymbol{H}$ is just constant at every iteration. Specifically, we would like to choose the search direction $d^{(n)}$ such that the following condition is true

$$\left(d^{(\boldsymbol{n})}\right)^T \boldsymbol{H}_{n-1} d^{(n-1)} = 0 \tag{3.113}$$

We can design the search direction $d^{(n)}$ to achieve conjugacy using the recursion

$$d^{(n)} = -\boldsymbol{P} g^{(n)} + \gamma_n d^{(n-1)} \tag{3.114}$$

where $g^{(n)} = \nabla\Psi\left(f^{(n)}\right)$ and we choose γ_n to satisfy Eq. (3.113). Substituting Eq. (3.114) into Eq. (3.113) and simplifying yields

$$\gamma_n = \frac{\left(d^{(n-1)}\right)^T \boldsymbol{H}_{(n-1)} \boldsymbol{P} g^{(n)}}{(d^{(n-1)})^T \boldsymbol{H}_{(n-1)} d^{(n-1)}} \tag{3.115}$$

This choice for γ_n will satisfy the conjugacy relation given in Eq. (3.113) but is inconvenient particularly for non-quadratic cost functions because it depend on the Hessian $\boldsymbol{H}_{(n-1)}$**.** We would like to have an expression for γ_n that does not depend on the Hessian $\boldsymbol{H}_{(n-1)}$ even if $\boldsymbol{H}_{(n-1)} = \boldsymbol{H}$ is a constant**.**

It can be shown that given a set of linearly independent vectors such as the gradient vectors, $g^{(n)}$ at every iteration, can be used to construct a set of mutually **H**-conjugate directions $d^{(0)}, d^{(1)}, \ldots d^{(n-1)}$. Furthermore, we can obtain the following recursive expression for the update γ_n that is independent of the Hessian matrix.

$$\gamma_n^{FR} = \frac{\left(g^{(n)}\right)^T \boldsymbol{P} g^{(n)}}{(g^{(n-1)})^T \boldsymbol{P} g^{(n-1)}} \tag{3.116}$$

For a detailed discussion on the generation of such a set and on the derivation of the update expression, refer to Appendix B. This is referred to as the Fletcher-Reeves update and it shows that the search direction $d^{(n)}$ can be generated with knowledge of only the current and previous gradients $g^{(n)}$ and $g^{(n-1)}$ and the previous search direction $d^{(n-1)}$. This method uses only vector multiplication and is of vital importance when n_p is large and computer storage is not an option. This method when applied to a positive definite quadratic function of n_p variables will find the minimum in at most n_p iterations.

The conjugate gradient method is used to solve a set of positive definite symmetric linear equations. The linear conjugate-gradient algorithm takes the following form, for $n = 0,1,\cdots$, given $f^{(0)}$ and

$$g^{(0)} = \nabla\Psi\left(f^{(0)}\right) \tag{3.117}$$

then for $n = 0,1,\cdots$,

$$\gamma_n^{FR} = \frac{\left(g^{(n)}\right)^T \boldsymbol{P} g^{(n)}}{(g^{(n-1)})^T \boldsymbol{P} g^{(n-1)}} \tag{3.118}$$

$$d^{(n)} = -P g^{(n)} + \gamma_n d^{(n-1)} \tag{3.119}$$

It can be shown (appendix B) that the current gradient vector $g^{(n)}$ is orthogonal to the previous search direction, $d^{(n-1)}$. That is ${g^{(n)}}^T d^{(n-1)} = 0$. Thus

$$\alpha_n = \alpha = \frac{\left(g^{(n)}\right)^T d^{(n)}}{d^{(n)^T} \boldsymbol{H} d^{(n)}} = \frac{\left(g^{(n)}\right)^T \boldsymbol{P} g^{(n)}}{d^{(n)^T} \boldsymbol{H} d^{(n)}} = \frac{\left\|g^{(n)}\right\|^2_{\boldsymbol{P}^{1/2}}}{d^{(n)^T} \boldsymbol{H} d^{(n)}} \tag{3.120}$$

$$f^{(n)} + \alpha d^{(n)} \tag{3.121}$$

The impressive performance of PCG is demonstrated in the Fig. 3.11. Using the same quadratic function as in the PSD example (Fig. 3.10), we see that we converge to the solution in two iterations which is consistent with the size of the parameter space.

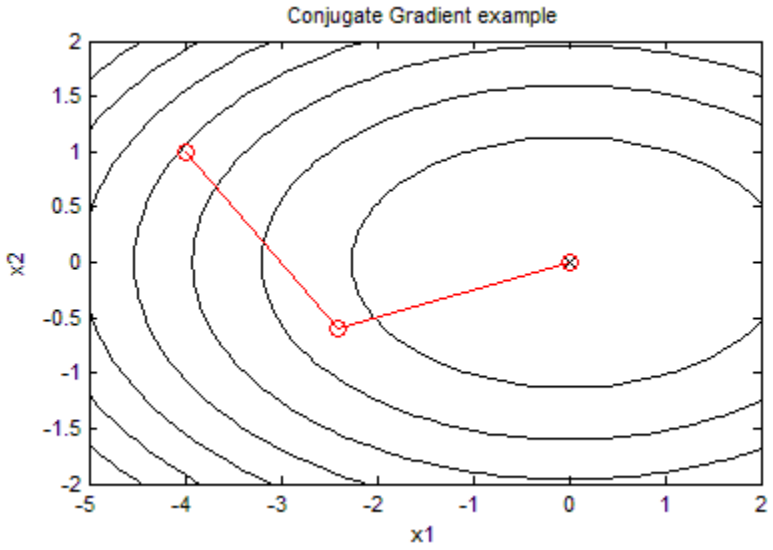


Fig. 3.11. Illustrates the fast convergence of the PCG method in which $P = I$. Unlike PSD, the algorithm converges in two iterations.

The PCG method can also be used to solve systems where A is not symmetric, not positive-definite, and not even square. For example, the unregularized least squares problem

$$\hat{f} = \underset{f}{\operatorname{argmin}} \|y - \boldsymbol{A}f\|^2 \tag{3.122}$$

can be solved by setting the derivate of Eq. (3.122) to zero which yields the normal equations

$$\boldsymbol{A^T A} f = \boldsymbol{A^T} y \tag{3.123}$$

If $\boldsymbol{A}$ is not square and $\boldsymbol{A}f = y$ is overconstrained, more linearly independent equations than variables, then there may or may not be a solution to $\boldsymbol{A}f = y$, but it is always possible to find a value of f that minimizes Eq. (3.122). $\boldsymbol{A^T A}$ is symmetric and positive

because for any f, $f^T \boldsymbol{A^T A} f = \|Af\|^2 \geq 0$. If $\boldsymbol{A}f = y$ is not underconstrained, then $\boldsymbol{A^T A}$ is nonsingular and PCG can be used to solve Eq. (3.123). The only issue in doing so is that condition number of $\boldsymbol{A^T A}$ is the square of that of A and so convergence is significantly slower. Furthermore, an important technical point is that the matrix $\boldsymbol{A^T A}$ is even denser than $\boldsymbol{A}$ and should not be formed explicitly.

The rate of convergence of PCG is determined by the location of the spectrum of the system matrix $\boldsymbol{A}$. Roughly speaking a good spectrum leads to faster convergence. A good spectrum may happen for two reasons: the eigenvalues maybe grouped in small clusters or they may lie well separated in a relative sense from the origin. First, we suppose that the eigenvalues are perfectly clustered but assuming nothing about these locations. If $\boldsymbol{A}$ has only n distinct eigenvalues, then the CG converges in n steps. At the other extreme, suppose we know nothing about any clustering of the eigenvalues but only that their distances from the origin vary by at most a factor of $\kappa \geq 1$. In other words, suppose we know only the 2-norm condition number, $\kappa = \frac{\lambda_{max}}{\lambda_{min}}$, then the following is true:

$$\frac{\left\|f^{(n)} - \hat{f}\right\|_A}{\left\|f^{(0)} - \hat{f}\right\|_A} \leq 2\left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)^n \approx 2\left(1 - \frac{2}{\sqrt{\kappa}}\right)^n \tag{3.124}$$

The PCG method when applied to general nonlinear functions loses the property of finding the minimum in at most n_p iterations and becomes just an iterative method. The Fletcher-Reeves algorithm suggests reverting periodically after a cycle of n linear searches to the direction of steepest descent and discarding all previous information on directions. Later, Polak and Ribiere introduced another expression for γ_n given by:

$$\gamma_n^{PR} = \frac{\left(g^{(n)} - g^{(n-1)}\right)^t \boldsymbol{P} g^{(n)}}{(g^{(n-1)})^t \boldsymbol{P} g^{(n-1)}}$$

The Polak-Ribier algorithm coincides with the Fletcher-Reeves algorithm whenever it is applied to a quadratic function but differs from it when applied to nonlinear functions. Due to nonquadratic terms in the objective function and possible inaccurate line searches, conjugacy of the directions generated by the CG will be lost and so the Polak-Ribier

method is meant to resolve this issue. The PCG method using Fletcher-Reeves update is summarized as follows:

PCG Algorithm

$$g^{(n)} = \nabla\Psi\left(f^{(n)}\right) \qquad (gradient)$$

$$p^{(n)} = \boldsymbol{P}g^{(n)} \qquad (precondition)$$

$$\gamma_n^{FR} = \begin{cases} 0 & n = 0 \\ \dfrac{\left(g^{(n)}\right)^T \boldsymbol{P} g^{(n)}}{(g^{(n-1)})^T \boldsymbol{P} g^{(n-1)}} & n > 1 \end{cases}$$

$$d^{(n)} = -p^{(n)} + \gamma_n d^{(n-1)} \qquad (search\ direction)$$

$$\alpha_n = \min_{\alpha \in \mathbb{R}} \Psi\left(f^{(n)} + \alpha\right) d^{(n)} \qquad (step)$$

$$f^{(n+1)} = f^{(n)} + \alpha_n d^{(n)} \qquad (update)$$

Conclusion

In this chapter, we have provided a detailed discussion of Model-based image reconstruction in the context of penalized-likelihood estimation. First, we showed the necessary steps to take for developing suitable system and statistical models. We then presented explicit reasons for using penalized-likelihood estimation as opposed to other estimation techniques in deriving a suitable cost function with an appropriate regularizing penalty term. We also gave a general discussion of the properties of the QPWLS estimator, particularly its low-pass filtering nature. We also provided a brief discussion of edge-preserving penalty functions. Lastly, we discussed general purpose optimization algorithms and showed the impressive performance of conjugate gradients method for quadratic cost functions.

Chapter 4
Model-Based THz Imaging for 2D Reflection Mode Geometry

Imaging with terahertz is still in its infancy with many complex challenges. Various factors severely constrain plausible scenarios of applications. Several of these challenges, including signal to noise ratio, SNR, acquisition speed, and resolution, reflect common problems that a number of imaging modalities attempt to address. In direct imaging modalities, there is a direct tradeoff between SNR and acquisition speed. For example, the most basic coherent imaging can be achieved by raster scanning a sample through the THz focus and recording the full spectroscopic information at each pixel. If we were to consider a small image of 60 by 60 pixels, 3600 single measurements are required. Depending on the delay line concept, the desired lock-in time constants, and other factors, to record a high-quality THz pulse can take longer than 30 s and can result in a total measurement time of 30 h for this small image. To increase the acquisition speed, a two-dimensional (2D) electro-optic sampling has been used together with a CCD camera to provide a dramatic increase in imaging speed and rates as high as 5000 pixels/second are feasible. However, a lock-in amplifier cannot be synchronized to multiple pixels. The relegation of the lock-in amplifier results in a significant reduction in SNR compared to the scanned approach.

Inverse imaging architectures have the potential to increase the acquisition speed without compromising SNR. In chapter 2, we discussed time reversal imaging which is an inverse imaging scheme and showed that fast reconstruction of objects can be obtained. However, we also addressed that one of the limitation of time-reversal imaging is the lack of a system model to "invert" and hence the impulse response of the system with its temporal ringing seriously degrades the quality of the reconstructed images. In the previous chapter, we developed the theoretical framework of model-based imaging and showed the method by which one can develop a system model and also the steps one

must take to “invert such model”. In this chapter, we apply the principles of model-based imaging to a 2D THz system operating in reflection mode to show fast reconstruction of metallic and dielectric objects. At the heart of the model-based image reconstruction algorithm, is the minimization of a quadratically penalized least-squares (QPLS) cost function via conjugate gradients. QPLS is a special form of the QPWLS, discussed in the previous chapter, but has equal weighting. We compare the reconstructions by this algorithm to that of the time-reversal algorithm for metallic and dielectric objects and show that having a system model to “invert” leads to better quality images without compromising acquisition speed.

Experimental Setup

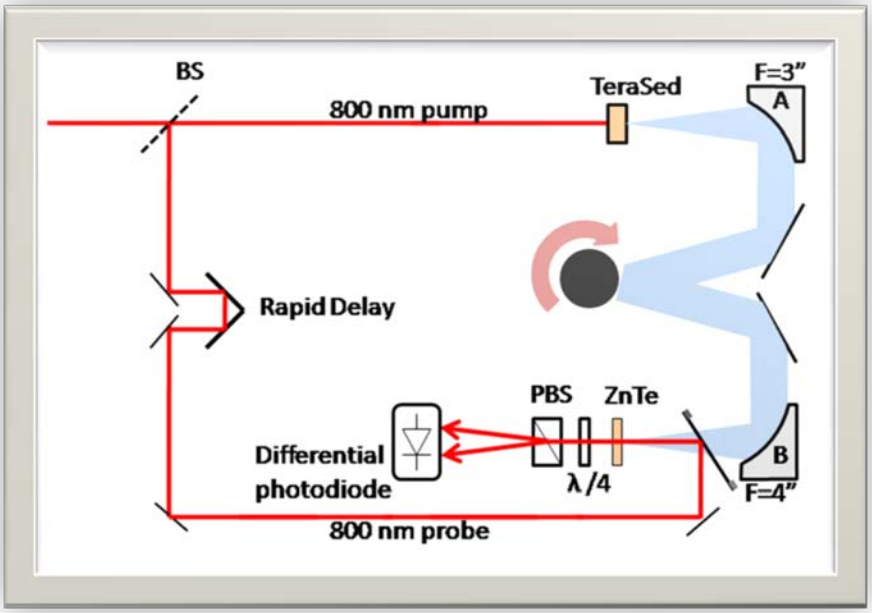


Figure 4.1. Illustrate the 2D THz imaging setup in reflection mode.

The experimental setup for 2D reflection mode THz imaging is depicted in Fig. 4.1. A Coherent regeneratively amplified laser is used to generate near-infrared (NIR) pulses with a pulse duration of approximately 100fs. The pulse energy is 4 μJ at a repetition rate of 250 kHz, providing an average power of 1W. The laser pulses are split into pump and probe beams using a beamsplitter. The pump beam illuminates a large-area photoconductive emitter (TeraSED, GigaOptics GmbH) to generate a nearly single cycle THz pulse. Since the optical spot size and hence the THz generation area is much

smaller than the THz wavelength the emitted THz radiation is divergent and is collimated using an off-axis parabolic mirror, A, with focal length of 102 mm. The first of the two metal coated mirrors in between the parabolic mirrors directs the collimated THz beam onto the object while the second mirror directs a portion of the scattered THz onto a second parabolic mirror, B, of focal length 76 mm for electro-optic detection in ⟨110⟩ ZnTe crystal.

The probe pulse is reflected by a pellicle to propagate collinearly with the THz pulse. As the probe beam propagates through the ZnTe crystal, its polarization is rotated proportionally to the instantaneous THz electric field. ZnTe is favored for EOS because of its physical durability, its high second order nonlinearity $\chi^{(2)}$ coefficient and its excellent phase matching properties. The group velocity of the 800 nm probe pulse and the phase velocity of the THz field are approximately equal in ZnTe. The birefringence of ZnTe is modified by the external THz electric field and the probe beam polarization is rotated by the EO or Pockel's effect. The induced phase modulation of the probe pulse is converted into an intensity modulation and detected by a differential photodiode. By iteratively increasing the probe path length using a rapid delay scanner, the THz electric field at later times is measured and the temporal THz pulse profile is recorded.

The object that is to be imaged is mounted onto a rotation stage and rotated in increments of 1 degree for a full revolution. The resulting wavefield is then processed to generate a 2D image of the object's profile. In general, a reference waveform is obtained for system calibration and for use as the impulse response of the system. The reference waveform is obtained by placing a flat gold mirror at the pivot point of the rotation stage. An example of the reference waveform is shown in Fig. 4.2. The reference waveform clearly shows the presence of temporal ringing which will greatly clutter our reconstructed images if not properly taken into account.

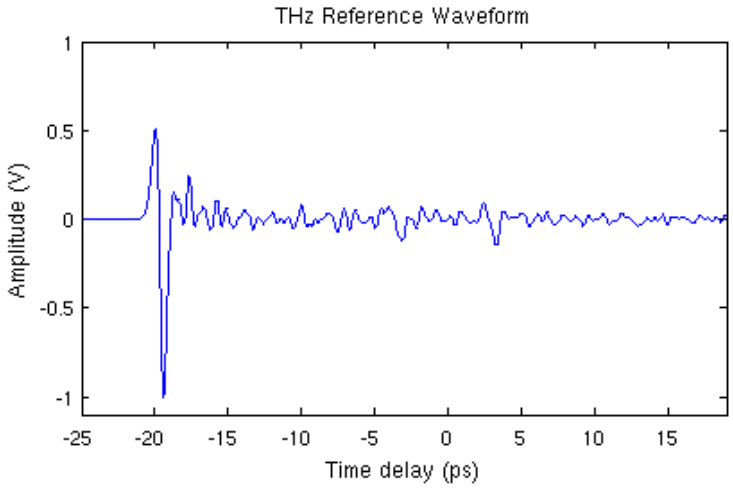


Figure 4.2. Illustrates the impulse response of the THz reflection mode Imaging System.

Problem Formulation

To develop a model-based image reconstruction algorithm for 2D THz imaging, we must first develop a system model that captures the forward imaging process. As discussed in the previous chapter, the forward imaging process can be described as a convolution between the THz impulse response and the object's spatial profile:

$$y_\theta(t) = \iint f(x,z)h(t-\tau_{\theta xz})\,dx\,dz \tag{4.1}$$

where $f(x,z)$ is the object's spatial profile and $h(t)$ is the impulse response of the THz system that is delayed by a known parameter $\tau_{\theta xz}$ that is calculated from the geometry of our setup. In image reconstruction, we are interested in the object's spatial function given a known kernel and the observed function. Since our measurement set is finite, $\{y_\theta\}_{\theta=1}^{\theta=N}$, there are uncountably many objects whose spatial profile fit our measurements and thus our image reconstruction problem is ill-posed. In order to address the ill-posedness of our problem and to develop a discrete model, we make the assumption that our object is band-limited spatially and thus can be represented by a set of basis. The simplest basis to work with is the set generated by spatially shifting $\delta(x,z)$

$$f(x,z) = \sum_{n=1}^{n=N}\sum_{m=1}^{m=M} f_{nm}\delta(x - n\Delta_x, z - m\Delta_z) \tag{4.2}$$

Inserting Eq. (4.2) into Eq. (4.1) and simplifying, we obtain the following discrete expression for Eq. (4.1).

$$y_\theta(t) \approx \sum_{m=1}^{m=M} \sum_{n=1}^{n=N} A_{\theta mn}(t) f_{nm}, \quad where \quad A_{\theta nm}(t) = h(t - \tau_{\theta mn}) \tag{4.3}$$

By parameterizing the object's spatial function using a comb function, we have essentially discretized our object space into a grid of possible reflectors and thus f_{nm} represents the reflectivity of the object. To obtain a matrix-vector expression of Eq. (4.3), we concatenate our measurements into a 1D vector of length $n_d \times 1$ and our 2D image, f_{nm}, into a 1D vector, f, of length $n_p \times 1$ using lexicographic ordering as discussed in the earlier chapter. The system matrix, $\boldsymbol{A}$, consists of shifted impulses, $h(t - \tau_{\theta mn})$, whose shift is determined by the geometry of our setup. The derivation for $\tau_{\theta mn}$ is with respect to the pivot point of the rotation stage and can be obtained by considering the following geometry.

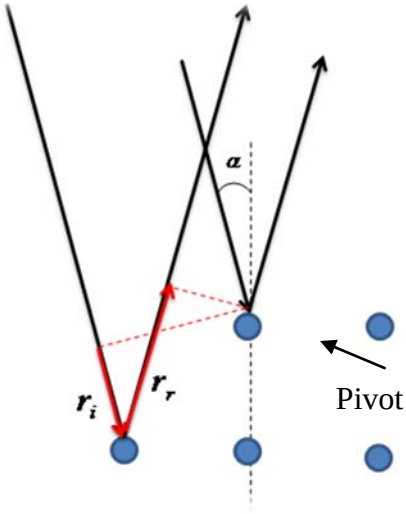


Figure 4.3. Illustrates the simple geometry used to derive an expression for $\tau_{\theta mn}$

As depicted in Fig. 4.3, the total path length difference, r_T, is shown in red. Hence, we have that

$$\tau_{\theta mn} = \frac{r_T}{c} = \frac{(r_i + r_r)}{c}$$

$$\tau_{\theta mn} = \frac{x * sin(\theta + \alpha) + z * sin(\theta + \alpha) + x * sin(\theta - \alpha) + z * cos(\theta - \alpha)}{c} \tag{4.4}$$

The expression for the time delay of the pulses is respect to the pivot point of the rotation stage. Hence it is very critical that the time delay of the pivot point is known.

Experimental Results

The expression given in Eq. (4.4) captures the rotating geometry of our imaging setup. However, it computes the relative time delays of pixels with respect to the pivot point. Hence, it is crucial that the location of the pivot point in our time window be known. Thus, to calibrate the 2D reflection mode THz imaging system, a steel post with a diameter of 12.66 mm was mounted onto a rotation stage and rotated in increments of 1 degree for a full revolution. We chose the steel post for calibration for its geometrical simplicity, anticipating that if the center of the post rotated about the pivot point, we should expect a flat wavefield plot as function of angle and thus the pivot point can be calculated easily. However, as shown in Fig 3.4a), the center of the post does not sit at the point of rotation and thus we must account for this offset either by shifting the coordinate system or by correcting for the location of the pivot point using this offset. In the following set of experiments, we opt for the latter and in later more complicated experiments; we actually shift the whole coordinate plane.

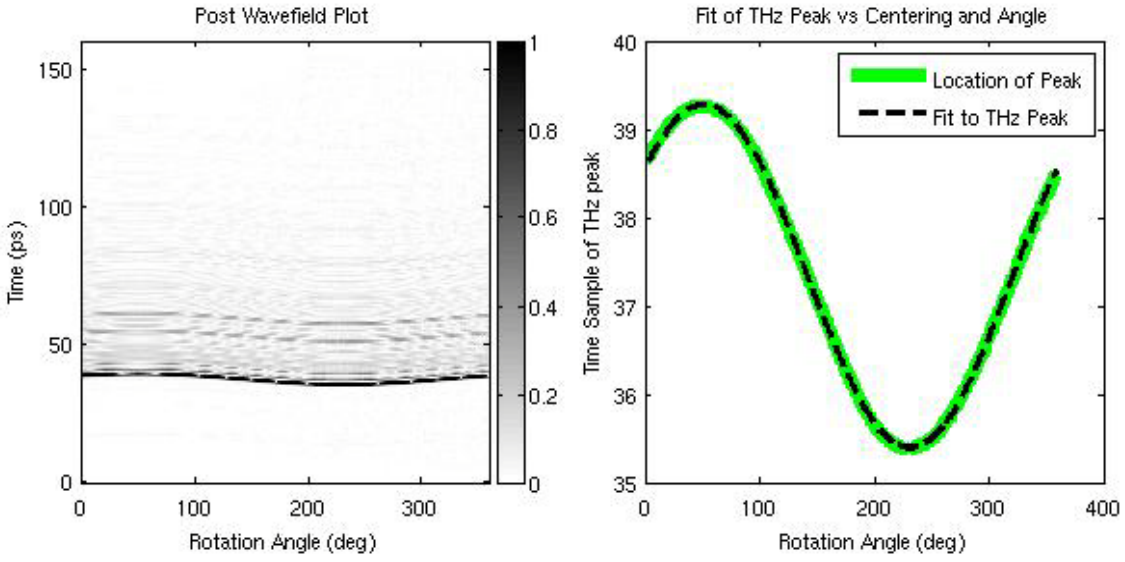


Figure 4.4. Plot of a) Post Wavefield. Illustration of b) the THz peak as a function of rotation angle (green line) and the fit to the THz peak as a function of angle (dashed black line).

To determine the location of the pivot point in our time window, we plot the THz peak reflected off the front surface of a 1/4 in. steel post as a function of angle (green line). The post is smooth and hence we can model the relative delay between the pivot point and the front surface of the post as a function of the post center and rotation angle accurately. We then seek the $x-$, and $z-$ values of the post center that will minimize the least-squares error between the actual graph and our time-delay model. Once the values are determined, the pivot point time location is easily calculated from the cylindrical geometry of the post.

Having determined the location of the pivot point in our time-window and using the expression for the relative time delay given in Eq. (4.4), we can model the forward imaging process by constructing a system matrix of time-shifted impulses. With the system model in place, we minimize via conjugate gradients the regularized least-squares cost function (QPLS) developed in the previous chapter. That is, we seek the minimum to the following expression:

$$\hat{f} = arg\min_{f}\left(\frac{1}{2}\|y - \boldsymbol{A}f\|^2 + \frac{\beta}{2}\|\boldsymbol{C}f\|^2\right) \tag{4.5}$$

The first term in the cost function is a measure of the error between the solution and the data and we would like this term to be small. The second term is a regularizing penalty term that penalizes an estimate according to how much it departs from our assumption of smoothness. Forcing the solution to be smooth will inevitably reduce the agreement between the solution and the data and hence introduce distortion. Thus, the regularized solution is a balance between the fidelity of the solution to the noisy data and the distortion introduced by the penalty term and this balance is controlled by the β term. As $\beta \longrightarrow \infty$, the original least squares solution is achieved whereas $\beta \longrightarrow 0$, a perfectly smooth solution is obtained. Thus, one must be careful in choosing the penalty parameter.

As discussed in the previous chapter, one must adjust β such that the desired spatial resolution is achieved. There are many computationally intensive ways for choosing β and it may be worth the effort to rigorously choose β according to such methods particularly if the system matrix does not change from one experiment to the next. In the 2D reflection THz imaging system, as long as the time window does not change the

system matrix will not change. However, we have found a simple and heuristic approach for determining β. We plot both the maximum amplitude of the point-spread function (psf) and the condition number for the Hessian of our cost function, $\kappa(\boldsymbol{H}) = (\boldsymbol{A}^T\boldsymbol{A} + \beta\boldsymbol{R})$, for a range of β as depicted in Fig. 4.5. From these two plots, we can extract a small range of useful β values to use in the model-based algorithm.

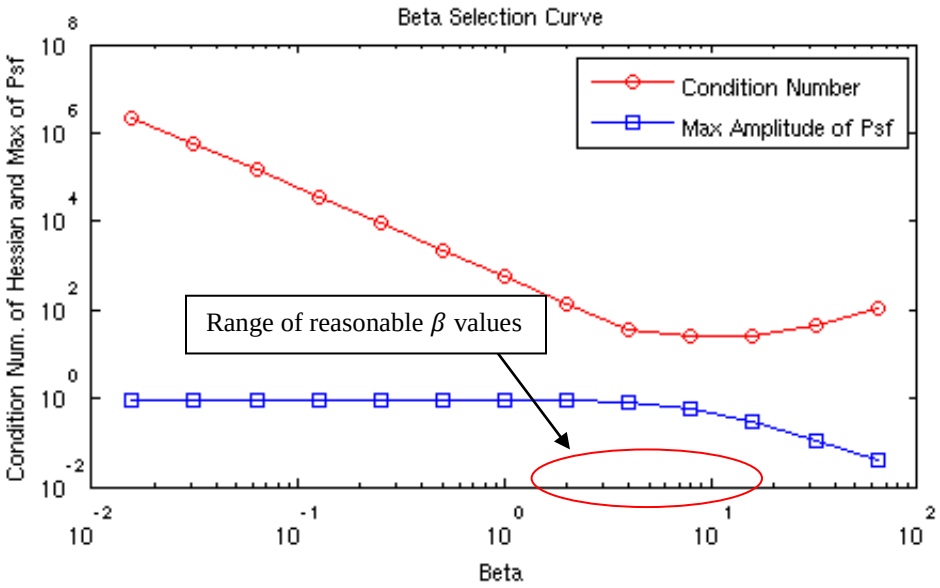


Figure 4.5. By plotting both the condition number of the Hessian and the maximum amplitude of the PSF we can extract a range of reasonable β values.

It is evident from the plot shown in Fig. 4.5 that by choosing a β value from the range shown in red, then we can better condition our overall system without overly smoothing our reconstructed images.

To better illustrate the smoothing effect of the β parameter, we plot the 2D PSF under two cases. In the first case, a plot of the PSF is shown in Fig. 4.6a) for a reasonable choice of 1 for β, in accordance with the curve in Fig. 4.5. The inset plot shown in Fig. 4.6a) is a horizontal line cut through the 2D plot of the PSF. It clearly shows that the PSF remains largely peaked at its pixel position.

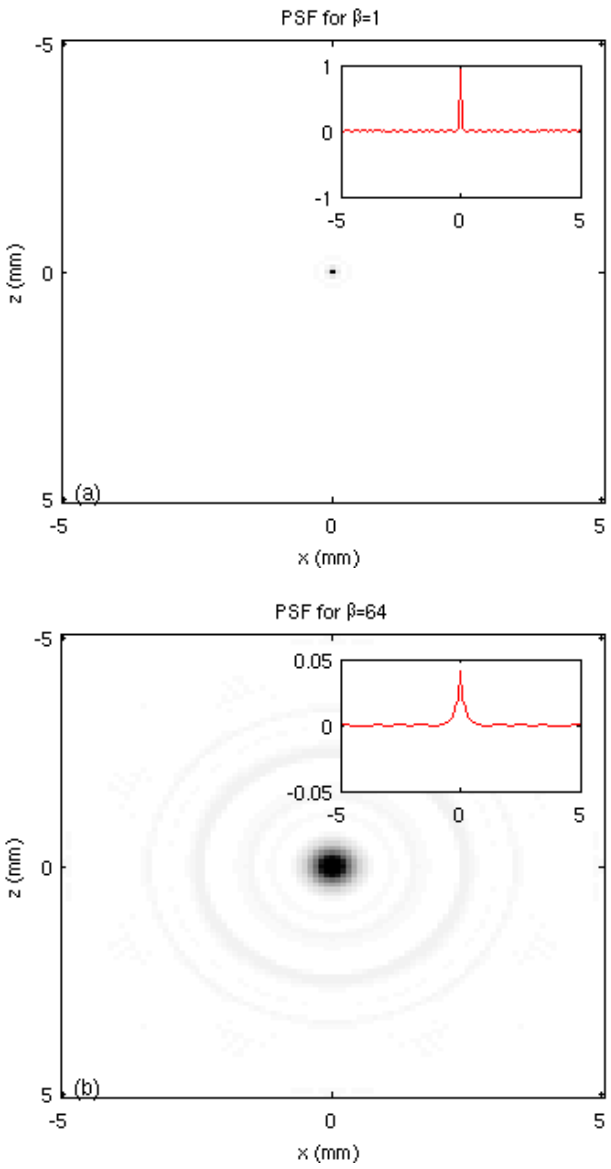


Figure 4.6. 2D plot of the PSF for a) $\beta = 1$ and b) $\beta = 64$. The insets are horizontal line cuts of the 2D PSF plots.

In Figure 4.6b) we consider the case in which the β parameter is large. As evident in the 2D PSF for $\beta = 64$, the energy in the PSF has been spread to many pixels which has resulted in a reduction in the peak amplitude of the PSF, as shown in the inset. Clearly, we must choose a large enough value for β such that it stabilizes the Hessian matrix but not large enough such that it smears the energy of a signal over a large number of pixels.

To assess the performance of the model-based algorithm via regularized least-squares, we compared the resulting reconstructed images to those obtained via the time-reversal

algorithm. The first objected to be reconstructed was the 1/4 in steel post whose wavefield plot is given in Fig3.4a).

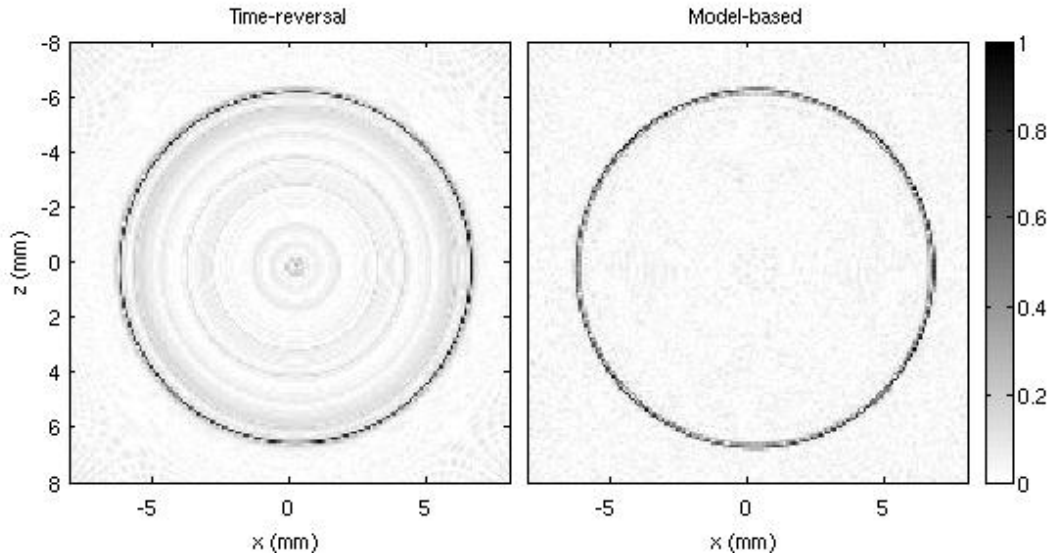


Figure 4.7. Reconstruction of post wavefield data via a) Time-reversal and via b) Model-based.

After five iterations, the model-based algorithm has achieved a better reconstruction of the steel post than did the time reversal. As evidenced by the reconstructed images shown in Fig. 4.7a) and 4.7b), normalized to their respective peaks, we see that both algorithms reconstructed the cross section of the steel post with the right dimensions of 12.7 mm. However, the model-based image clearly shows that the grid points corresponding to points inside the steel post are very close to zero. However, the time reversal algorithm clearly has signal inside the perimeter of the post which arises because the time-reversal algorithm does not take into account the temporal ringing in the impulse response. Thus the image produced by the model-based algorithm is qualitatively more consistent with the object than the time-reversal. We can quantify the improvement of the model-based algorithm versus the time-reversal algorithm by integrating the reconstructed images azimuthally and plotting the resulting signals as a function of radius as shown in Fig. 4.8a) and 4.8d). From the plots we calculated a peak SNR (in amplitude) for the model-based algorithm of 80.0 and a peak SNR of 16.9 for the time-reversal algorithm. Hence, the model-based algorithm has resulted in a factor of approximately 5 (14 dB) improvement in SNR over the time-reversal algorithm.

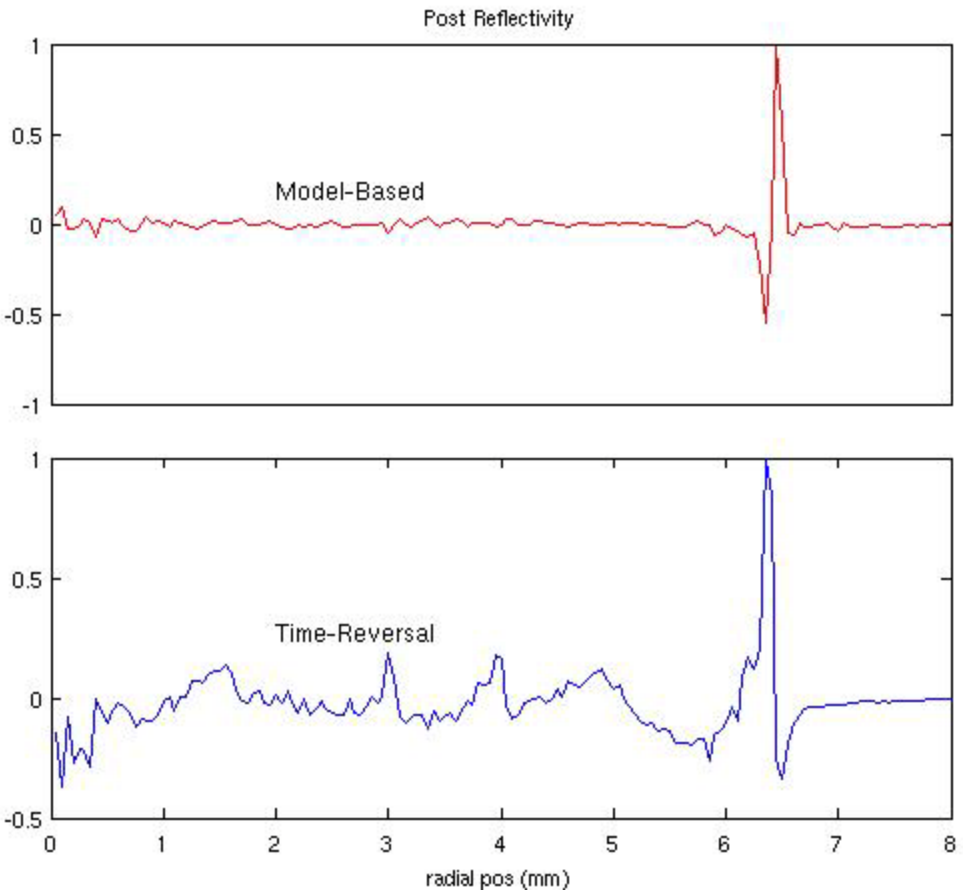


Figure 4.8. Illustrates the resulting 1D signals from azimuthally integrated the reconstructed images in Fig. 4.5 as a function of radius via a) time-reversal and b) via model-based.

In the second experiment, we imaged two wires with diameters of 2 mm and a spacing of 8.5 mm. Since the time window during this experiment is different than the first experiment with the post, we had to calibrate the imaging system in the same manner outline above to find the new time location of the pivot point. Furthermore, we created a new β-curve for the new experiment in order to extract a reasonable value for β. However, the curve was nearly identical to the one shown in Fig. 4.5 so we used the same value for β as the post experiment. We measured the scattered field off our-two wire object in increments of 1 degree for a full revolution.

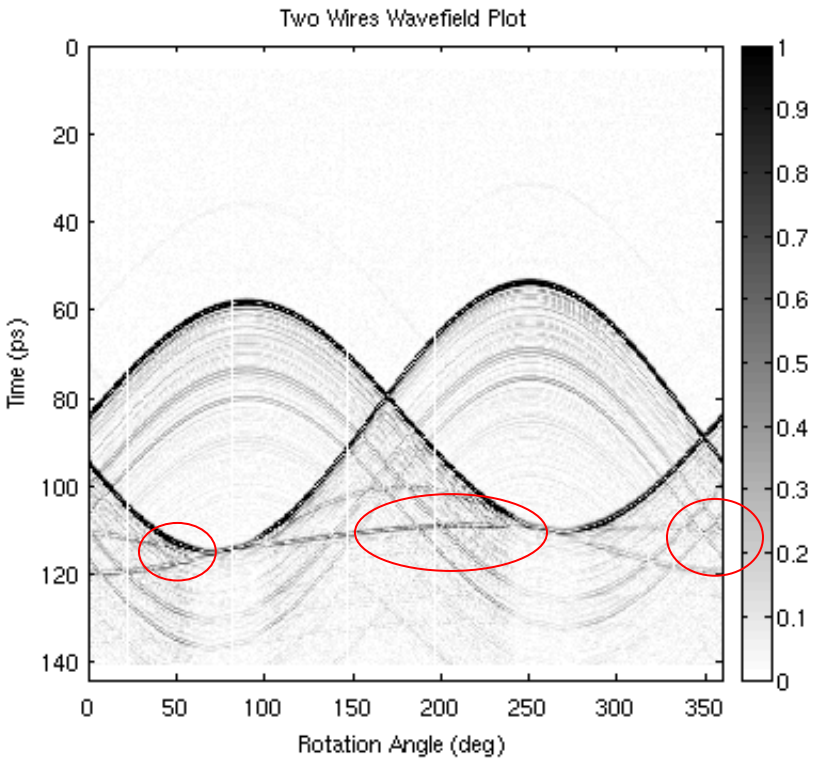


Figure 4.9. Illustrates the wavefield plot for the two wires

The wavefield plot shown in Fig. 4.9 illustrates that one of the wires can at certain angles be completely shadowed by the other wire. The second wire is in the shadow of the first wire over the range of angles between 65-95 degrees and the first wire is in the shadow of the second wire over the range of angles between 245-275 degrees. This shadowing will not however present a problem to both algorithms because of the strong scattering from the two wires over a significantly large range of angles. In addition to the main scattering off the wires, the wavefield plot clearly shows the presence of weak scattering between the two wires, shown in red. This scattering will present a challenge to the time-reversal algorithm because of the lack of a system model and it will also present a challenge to the model-based reconstruction method unless we alter the system model to take into account the weak scattering between the two wires. However, instead of modeling the complex scattering process, it was easier to choose a cost function that enforces the sparsity of the reconstructed images. Prior to discussing sparsity regularization, we show the reconstructed images from the time-reversal algorithm and

also the model-based algorithm using a quadratically penalized least-squares (QPLS) cost function.

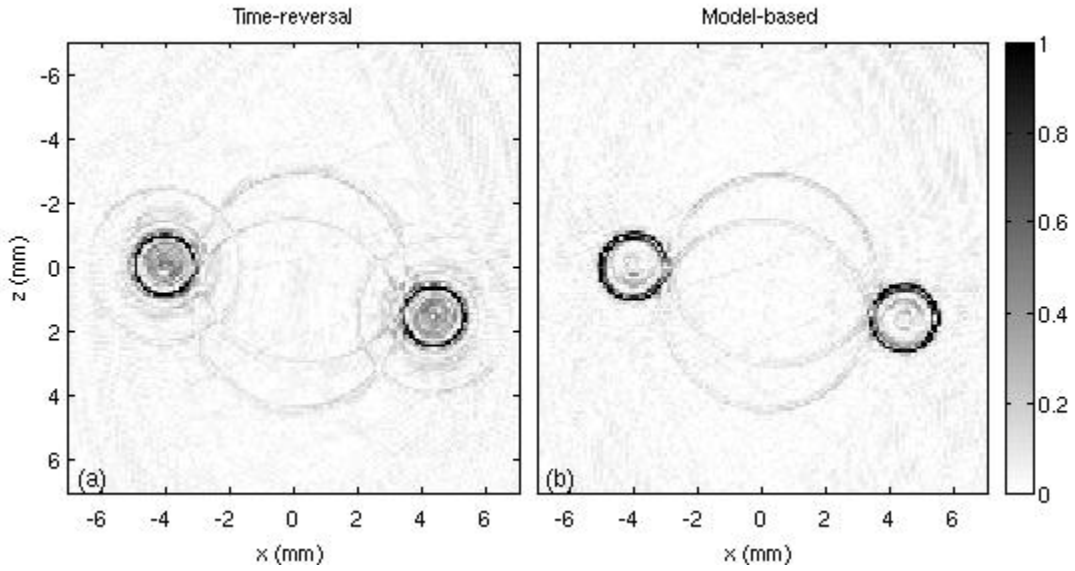


Figure 4.10. Reconstructed image of two wires using a) time-reversal algorithm and b) model-based with QPLS.

The reconstructed image via the QPLS reconstructed algorithm is qualitatively better than the one generated via the time-reversal algorithm. The ringing around the wires has been suppressed in the QPLS reconstructed image whereas it is apparent in the time-reversal reconstructed image. Furthermore, if we were to measure the diameter of the wires in the reconstructed images, we obtain a value of 2 mm for the QPLS image and a value of 1.8 mm for the time-reversal image. Hence, we see the QPLS reconstruction of the wires is more consistent with the true object than the time-reversal algorithm.

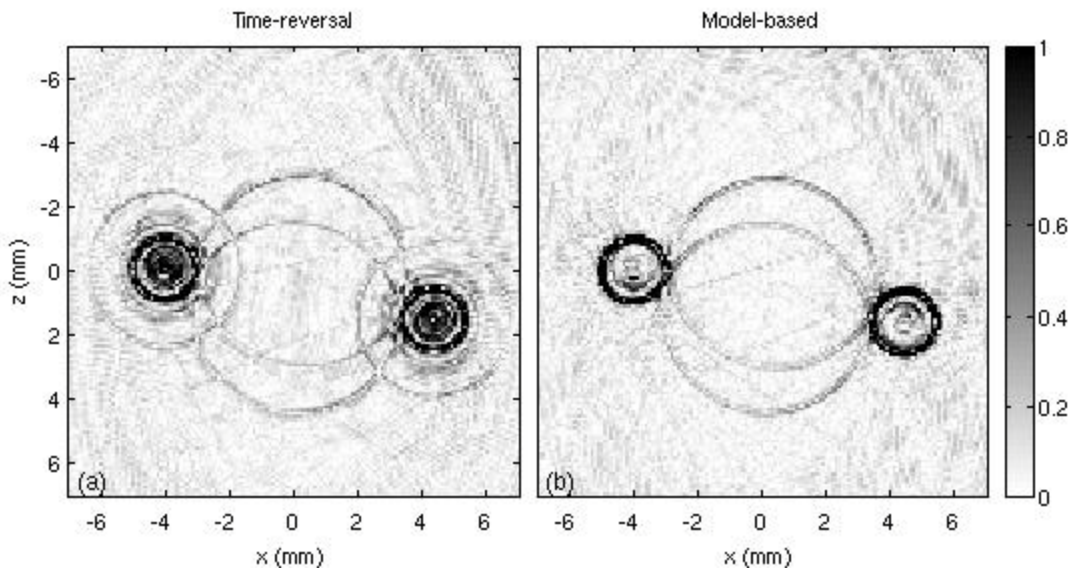


Figure 4.11. Saturated image of two wires via a) time reversal via b) model-based with QPLS.

In Fig. 4.11 we saturate the reconstructed images shown in Fig. 4.10 in order to raise the background and display the better performance of the model-based algorithm as compared to the time reversal. Again we see the strong signal inside the two wires in Fig. 4.11a) which is attributed to the temporal ringing of THz impulse. We also have temporal ringing around the wires in Fig. 4.11a). However, in Fig. 4.11b), we see that much of these artifacts have been suppressed. We also see that the reconstructed image in Fig. 4.11b) is much sharper than the image in Fig. 4.11a).

In both reconstructed images, however, we see artifacts due to the presence of the weakly scattered signals in the wavefield data. There is not much that can be done with the time-reversal algorithm that can mitigate the effect of the weak scattering that is present in the measurement data. However, in the model based algorithm, we have two options as stated earlier. We can either improve on the system model which would lead to an improvement in the qualitative accuracy of the reconstructed images. Modeling the complex scattering process though can be difficult and thus we choose the alternate option of using a different cost function. We know prior hand that the reconstructed images will be sparse spatially. Even if the images were not sparse spatially they can be sparse in another domain such as the wavelet domain. That is the transform of the image is sparse. However, in our case, the reconstructed images are spatially sparse to begin with. Thus we can construct a cost function that consists, as usual, of a data mismatch term and also a penalty term that enforces the sparsity of our reconstructed images. One

could use the $\mathcal{L}_0$ norm. The zero norm of f is simply the number of non-zero elements of f and it derives its name as being the limit of p-norms as p approaches 0. That is:

$$\|f\|_0 = \lim_{p\to 0}\|f\|_p^p = \sum_j 1_{\{f_j\neq 0\}} \tag{4.6}$$

Despite its name, the $\mathcal{L}_0$ norm is not really a norm because it violates the triangle inequality. Thus we replace it with the $\mathcal{L}_1$ norm which is defined as:

$$\|f\|_1 = \sum_{j=1}^{np}|f_j| \tag{4.7}$$

Thus our new cost function can be represented as follows:

$$\Psi(f) = L(f) + \beta R(f) = \frac{1}{2}\|y - \boldsymbol{A}f\|_2^2 + \frac{B}{2}\|f\|_1 \tag{4.8}$$

and thus we seek:

$$\hat{f} = arg\min_f\left(\frac{1}{2}\|y - \boldsymbol{A}f\|_2^2 + \frac{B}{2}\|f\|_1\right) \tag{4.9}$$

It is clear that this remains an unconstrained convex optimization problem in terms of f. However, this problem is non-differentiable when $f_j = 0$ for any f_j. Thus, we can't obtain a closed form solution for the global minimum in the same way that was done with the QPLS cost function. This drawback has led to the recent introduction of a multitude of techniques for dealing with cost functions that are convex but not differentiable. One such approach is optimization transfer.

The basic idea of optimization transfer is that when one is faced with a cost function $\Psi(f)$ that is difficult to minimize, at the nth iteration replace $\Psi(f)$ with a surrogate function $\phi^{(n)}(f)$ that is easier to minimize. Usually, minimizing $\phi^{(n)}(f)$ will not yield the global minimizer $\hat{f}$ of $\Psi(f)$ in one step so we must repeat the process. We alternate between the "S-step:" choosing a surrogate function $\phi^{(n)}(f)$ and the "M-Step:" finding the minimizer of $\phi^{(n)}(f)$. Thus if we choose the sequence of surrogate functions $\phi^{(n)}(f)$ properly then the sequence of iterates $\{f^{(n)}\}$ will converge to the global minimizer $\hat{f}$ provided that the following three montonicity conditions are upheld

- $\phi^{(n)}\left(f^{(n)}\right) = \Psi\left(f^{(n)}\right)$ Matched current value
- $\nabla\phi^{(n)}(f)|_{f=f^{(n)}} = \nabla\Psi(f)|_{f=f^{(n)}}$ Matched Gradient
- $\phi^{(n)}(f) \geq \Psi(x) \quad \forall x \geq 0$ Lies above

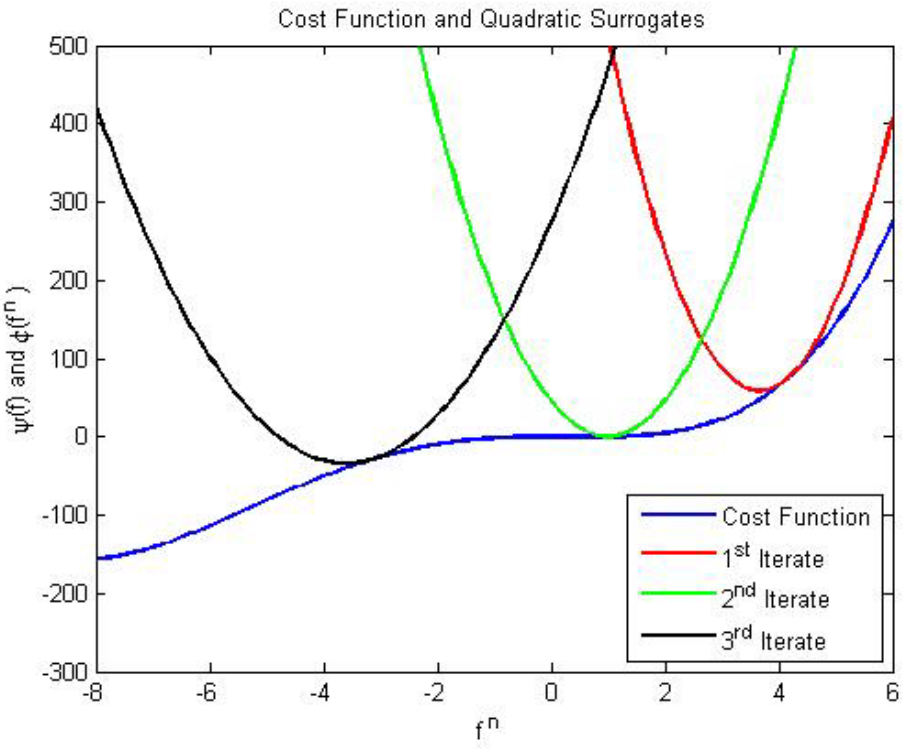


Figure 4.12. Illustration of 1D cost function $\Psi(f)$ and quadratic surrogates at 3 iterations.

In Fig. 4.12, we illustrate the basic idea of optimization transfer. The parabolic surrogate function $\phi\left(f^{(n)}\right)$ has the same value as the cost function $\Psi\left(f^{(n)}\right)$ at the current iterate $f^{(n)}$ and has the same slope at that point. In addition, the parabolic function lies above $\Psi\left(f^{(n)}\right)$ and this is the key to having a monotonic algorithm.

To minimize the cost function $\Psi(f)$ given in Eq. (4.8), we note the penalty function, $R(f)$, like the energy penalty discussed in the last chapter is a separable function, as shown in Eq. (4.7). However, the likelihood term, or the data mismatch term, $L(f)$ is a non-separable function due to the presence of the system matrix. Hence we seek separable surrogates for $L(f)$. Expanding the non-separable quadratic $L(f)$ exactly in terms of a second order Taylor series, we obtain the following expression

$$L(f) = L\left(f^{(n)}\right) + \left(f - f^{(n)}\right)^T \nabla L\left(f^{(n)}\right) + \frac{1}{2}\left(f - f^{(n)}\right)^T \boldsymbol{A^T A}\left(f - f^{(n)}\right) \tag{4.10}$$

A separable quadratic surrogate can be expressed as:

$$\phi_L^{(n)}(f) = L\left(f^{(n)}\right) + \left(f - f^{(n)}\right)^T \nabla L\left(f^{(n)}\right) + \frac{1}{2}\left(f - f^{(n)}\right)^T \check{c}\boldsymbol{I}\left(f - f^{(n)}\right) \tag{4.11}$$

where $\check{c}$ is a scalar value and $\boldsymbol{I}$ is the identity matrix with the same dimension as $\boldsymbol{A^T A}$. The scalar $\check{c}$ defines the curvature of the surrogate. If the surrogate function $\phi_L^{(n)}$ has low curvature, then it will appear "broad" and the algorithm will take large steps and reach the minimizer quickly. Conversely, if the surrogate function has high curvature, then it appears as a "skinny" graph, the steps are small, slowing convergence. In general we would like to find low-curvature surrogate function with the caveat that the surrogate function is to lie above $\Psi(f)$. Thus the following must be true to ensure montonicity:

$$\check{c}\boldsymbol{I} \geq \boldsymbol{A^T A} \tag{4.12}$$

Thus one option is to choose the curvature value to be the spectral radius of $\boldsymbol{A^T A}$, $\check{c} = \rho\left(\boldsymbol{A^T A}\right)$. Using the spectral radius may not give the tightest upper bound but it will likely give a decent upper bound to use in the iterative algorithm. The overall quadratic surrogate for the original cost function $\Psi(f)$ is

$$\phi^{(n)}(f) = \phi_L^{(n)}\left(f^{(n)}\right) + \frac{B}{2}\|f\|_1 \tag{4.13}$$

Thus our update expression should have the following form:

$$f^{(n+1)} = \underset{f}{\operatorname{argmin}}\left(\phi^{(n)}(f)\right) \tag{4.14}$$

By completing the square, we can express the likelihood surrogate $\phi_L^{(n)}(f)$ as

$$\phi_L^{(n)}(f) = constant + \frac{\check{c}}{2}\left\|f - f^{(n)} + \frac{1}{\check{c}}\nabla L\left(f^{(n)}\right)\right\|_2^2 \tag{4.15}$$

Dropping the constant term and factoring out a negative sign, we can write the surrogate for the cost function $\phi^{(n)}(f)$ as

$$\phi^{(n)}(f) = \frac{\check{c}}{2}\left\|f^{(n)} - f - \frac{1}{\check{c}}\nabla L\left(f^{(n)}\right)\right\|_2^2 + \beta\|f\|_1 \tag{4.16}$$

Since this surrogate is separable, we can expand it in terms of its unknown parameters,

$$\phi^{(n)}(f) = \frac{\check{c}}{2}\sum_{j=1}^{n_p}\left(\left|f_j^{(n)} - f_j - \frac{1}{\check{c}}\nabla L\left(f_j^{(n)}\right)\right|^2 + \beta|f_j|\right) \tag{4.17}$$

Thus to minimize the expression in Eq. (4.17), we have to minimize each of the individual terms separately. It can be shown that the soft threshold function minimizes a function with the following form:

$$soft(t, \alpha) = \min_{s}(|t - s|^2 + \alpha|s|) \tag{4.18}$$

The soft thresholding function is a nonlinear function that sets all points between $-\alpha$ and α to 0. The following graph in Fig. 4.13 illustrates the soft thresholding function for $\alpha = 2$.

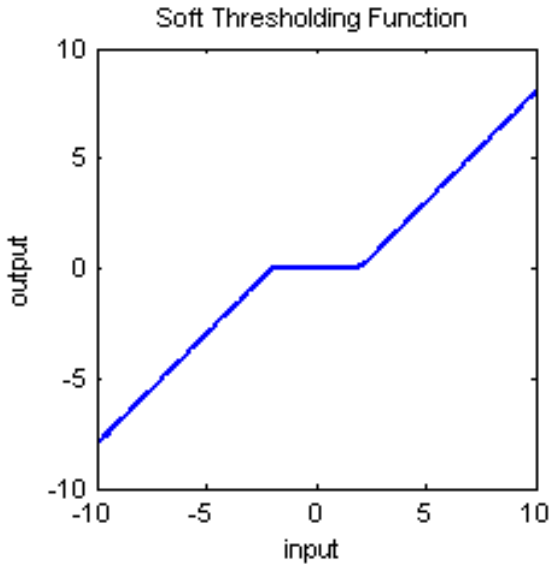


Figure 4.13. Illustrates the soft thresholding function.

Hence the expression for the minimizer can be written in terms of the soft thresholding function as:

$$f^{(n+1)} = soft\left(f^{(n)} - \frac{1}{\check{c}}\nabla L\left(f^{(n)}\right), \frac{\beta}{\check{c}}\right) \tag{4.19}$$

where $\check{c}$ has been moved underneath β because multiplying a cost function by constant does not change the location of the minimizer. To simplify the expression given in Eq. (4.19), we can write the gradient of likelihood as:

$$\nabla L\left(f^{(n)}\right) = -\boldsymbol{A^T}\left(y - \boldsymbol{A}f^{(n)}\right) \tag{4.20}$$

Thus Eq. (4.19) can be expressed as

$$f^{(n+1)} = soft\left(\underbrace{f^{(n)} + \frac{1}{\check{c}}\boldsymbol{A^T}\left(y - \boldsymbol{A}f^{(n)}\right)}_{\text{Preconditioned Gradient Descent}}, \underbrace{\frac{\beta}{\check{c}}}_{\text{Threshold pixel values}}\right) \tag{4.21}$$

The algorithm in Eq. (4.21) uses preconditioned gradient descent for the update step and then thresholds the resulting pixel values according to the soft thresholding function after every iteration. Heuristically, one could have obtained this algorithm by deciding to threshold the values after the update step based on precondition gradient descent. However, it is unlikely a suitable step size would have been chosen that would have guaranteed a monotonic decrease of the cost function.

The algorithm given in Eq. 4.21 is termed Iterative Soft-Thresholding (IST) and we have applied this algorithm to the data shown in Fig. 4.9 to see if we can obtain better reconstructions than the ones given in Fig. 4.10. Values for both $\check{c}$ and β have to be chosen. We choose the same value for β in the IST algorithm as the value used in the QPLS algorithm. In general, for the curvature, $\check{c}$, a suitable choice for the curvature parameter, $\check{c}$, is spectral radius of $\boldsymbol{A}^T\boldsymbol{A}$. However, finding the largest eigenvalue of $\boldsymbol{A}^T\boldsymbol{A}$ is not practical due to the dimensions of $\boldsymbol{A}$. Thus we simply create the system matrix, $|\boldsymbol{A}|$, using the absolute value of the impulse function, sum the columns of $|\boldsymbol{A}|^T|\boldsymbol{A}|$ and choose the maximum element in resulting column. That is, we choose $\check{c}$ as follows:

$$\check{c} = \max\big(\text{sum}(|\boldsymbol{A}|^T|\boldsymbol{A}|)\big) \tag{4.22}$$

This choice for the curvature will be a very conservative upper bound on $|\boldsymbol{A}|^T|\boldsymbol{A}|$ but will nonetheless guarantee a monotonic decrease of the cost function.

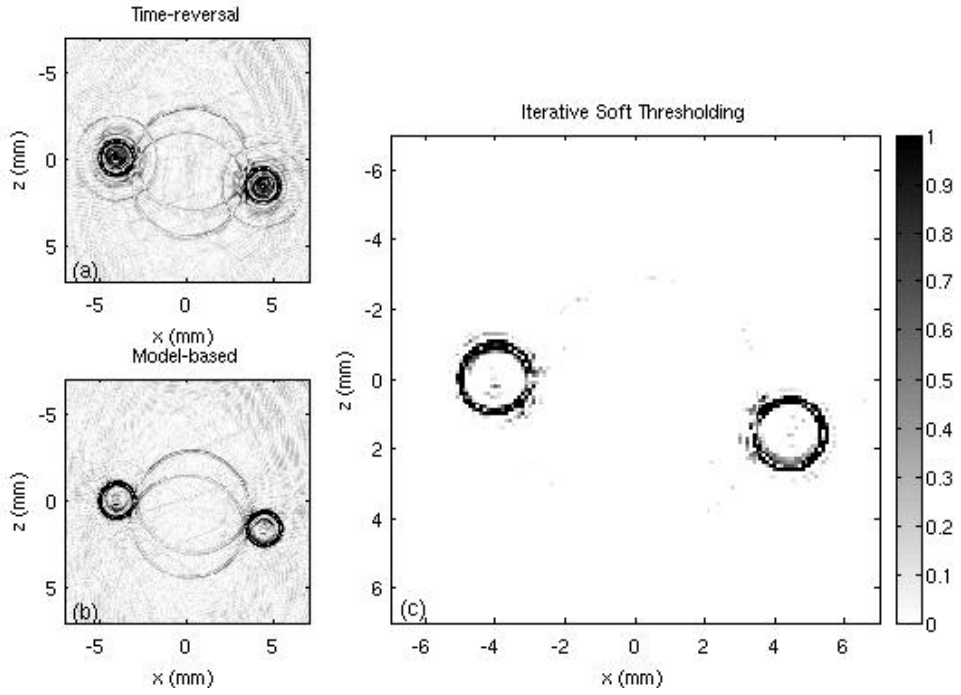


Figure 4.14. Time-reversal reconstruction of two wires a). Model-based reconstruction of two wires using QPLS b). Iterative Soft Thresholding reconstruction of two wires c).

For easier comparison, we have included the reconstructed images using the time-reversal and the QPLS algorithms again to illustrate the better performance of the Iterative Soft Thresholding algorithm as shown in Fig.4.14. Clearly, the IST algorithm has removed the phantom rings due to scattering between the wires while retaining a sharp image of the wires. It has selectively threshold the background while maintaining a quality reconstruction of the wires. One may argue that it possible to achieve the same results via thresholding the background of the QPLS reconstructed image. Although, this may be true, the justification for using the soft-thresholding function in the IST algorithm has stemmed from incorporating a sparsity penalty term in the cost function and then using the optimization transfer technique to minimize the resulting cost function. In more complicated settings or for more complex objects, one may not know what ad-hoc steps one must take in order to post-process the resulting images to achieve the same results as can be achieved via tailored cost functions and optimization transfer. However for the sake of being complete, we demonstrate the effect of wavelet denoising, Fig. 4.15b), of the QPLS reconstructed image and also the effect of simply soft-thresholding background

of the QPLS reconstructed image, Fig. 4.15c). The wavelet denoising was implemented using 4 levels of decomposition, the symlets wavelet of order 4, and soft-thresholding. The wavelet denoising technique was somewhat successful at removing the artifact in the image but not as well as the iterative soft thresholding technique. Since the QPLS reconstructed image is sparse, it more straightforward to threshold the pixel values in the spatial domain than to transform to the wavelet domain and so we illustrate the effect of simply soft-thresholding the image. The performance of this technique is comparable to that of the IST algorithm as shown in Fig. 4.15c).

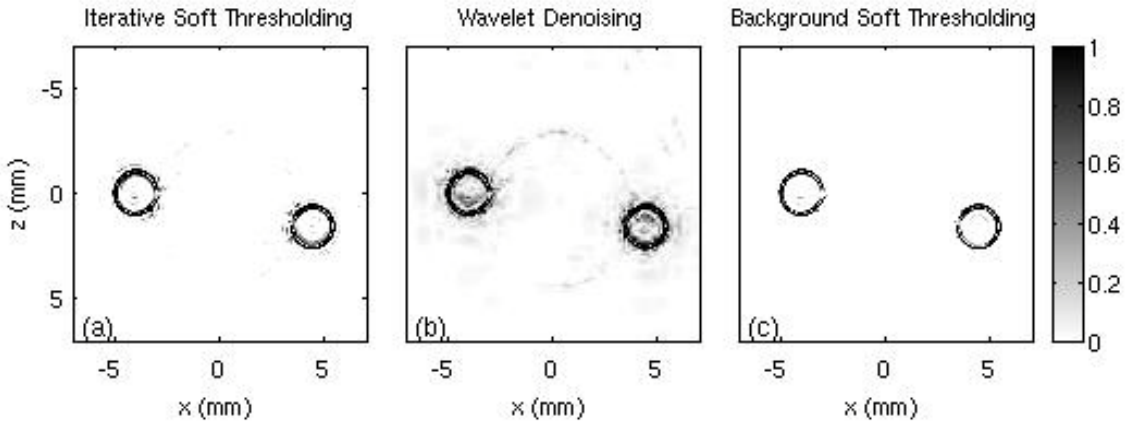


Figure 4.15. Comparison of Iterative Soft Thresholding a) to post processing of image via b) wavelet denoising and c) background soft thresholding.

In the previous two experimental studies, the geometry was relatively easy to model because the objects were metallic. For dielectrics, however, the modeling of the geometry needs to take into account other parameters for accurate timing of the arrival of the pulses at the detector plane. In the first of two dielectric experiments, we image a pencil with a wooden outer shell and a graphite core. The pencil was rotated in increments of 1 degree for a full revolution and the scattered THz field was collected. The wavefield plot for the pencil is shown in Fig. 4.14.

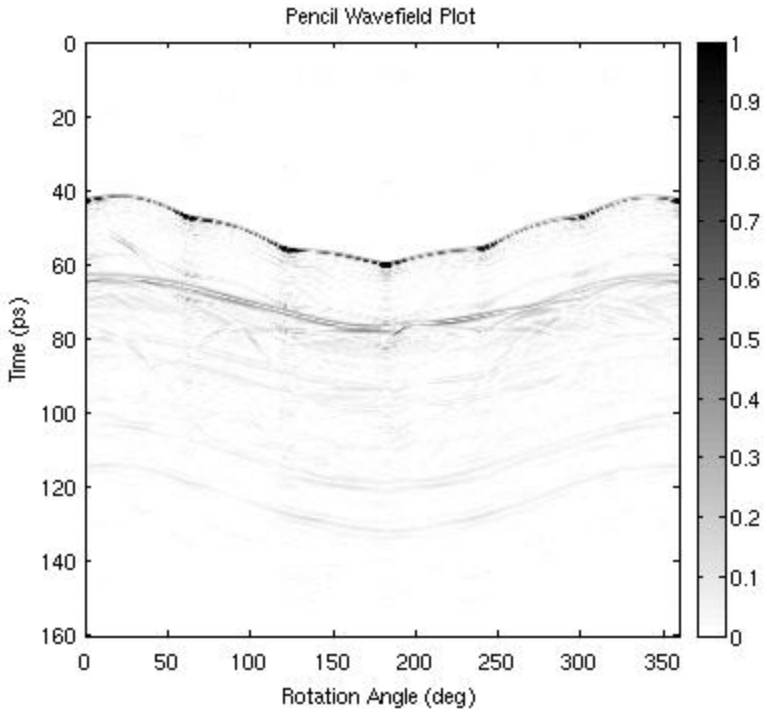


Figure 4.16. Illustrates the wavefield plot for the pencil.

The first signal at about 40 ps in Fig. 4.16 corresponds to reflections off the six faces of the pencil as it is rotated. The second fainter signal at about 60 ps corresponds to reflections off the graphite core. Thus we expect to see in the reconstruction of the data a hexagonal cross-section corresponding to the wooden outer shell of the pencil and a circular cross-section of the graphite core.

We reconstructed the above wavefield data using both the time-reversal algorithm and the model-based algorithm with a quadratic penalty. Since the time-window for this experiment is the same as the previous experiment with two wires, we kept the β value the same. The reconstructed image of the pencil via time-reversal is shown in Fig. 4.17a) and via the model-based algorithm is shown in Fig. 4.17b). Clearly, the model-based algorithm has produced a qualitatively more accurate reconstruction of the pencil than the time-reversal method particularly inside the hexagonal cross section of the pencil. We see that in both images the cross-section of the graphite core did not fully reconstruct and this is due to the fact that at certain angles the scattering off the graphite core is very weak (angles 300-360) as shown in the wavefield plot of Fig. 4.16. However, the model-based reconstruction of the core is still significantly better than the time-reversal. We measured the graphite core to be 2 mm in diameter but the diameter according the model-

based image is only 1.1 mm. Having neglected the index of refraction in the model-based algorithm has resulted in the wrong dimension for the diameter of the graphite core. Thus, to get more meaningful reconstructions, we must include the index of refraction.

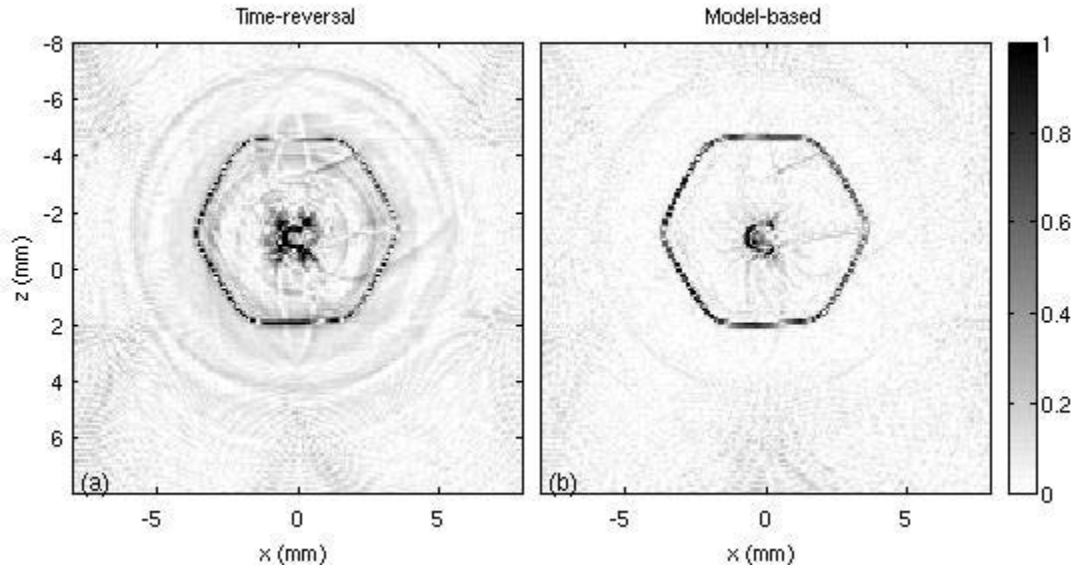


Figure 4.17. Illustrates the a) time-reversal reconstruction and b) model-based reconstruction of pencil.

The last experiment, in which we attempt to image a plastic cylinder with a defect in it, was conducted at Picometrix. As we shall see, we must take into account refraction in the geometrical modeling of the imaging process in order to generate meaningful images. The object's dimensions as well as the dimensions of the defect are shown in the following figure:

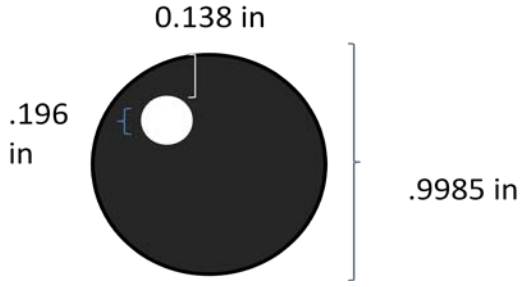


Figure 4.18. Illustrates the plastic cylinder and the dimensions of the defect in it.

The data was collected by rotating cylinder in increments of 1 degree for a full revolution. The data is shown in Figure 3.17.

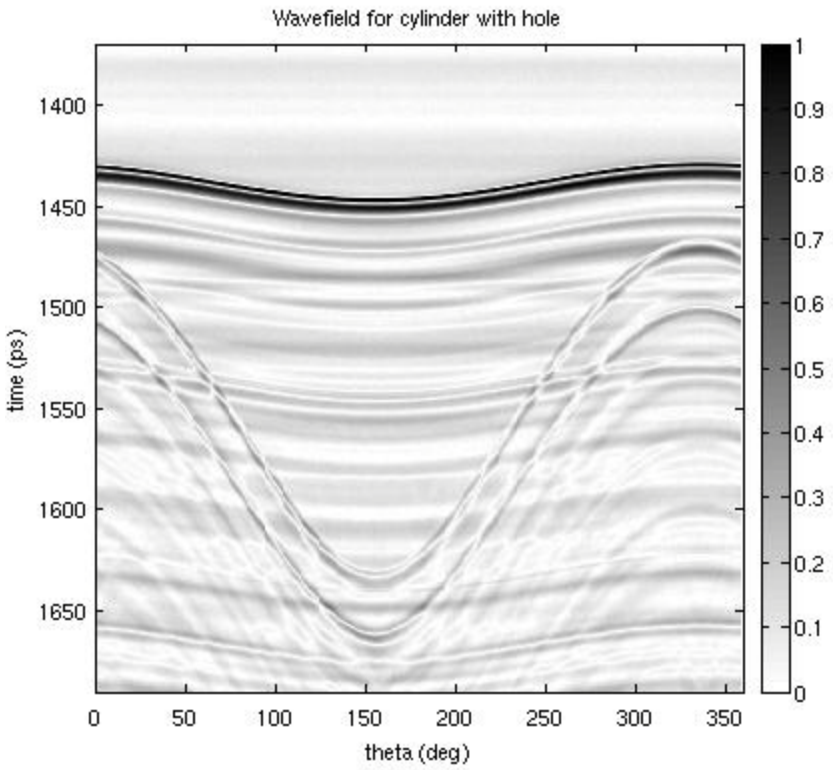


Figure 4.19. Illustrates the wavefield data for plastic cylinder with defect.

The signal shown in Fig. 4.19 between 1400 and 1450 ps corresponds to reflections off the plastic cylinder as it is rotated about the pivot point. The fainter signal corresponds to the reflection off the defect. We can easily make out the signal due to the defect in the wavefield plot and thus we should expect to reconstruct the defect.

We can naively attempt a reconstruction of the wavefield data by simply including the index of refraction, n, for points inside the cylinder. We were not given the value for n, thus we had to estimate it from the wavefield data. We estimated a value of $n = 1.647$ based on the time of flight difference between the arrival of the first surface pulse and the arrival of the second pulse due to the defect. We also had to calibrate this system in order to find the location of the pivot sample and a reasonable value for β. We used the estimated value of n in both the time-reversal algorithm and the model-based algorithm to generate the following reconstruction of the cylinder with defect:

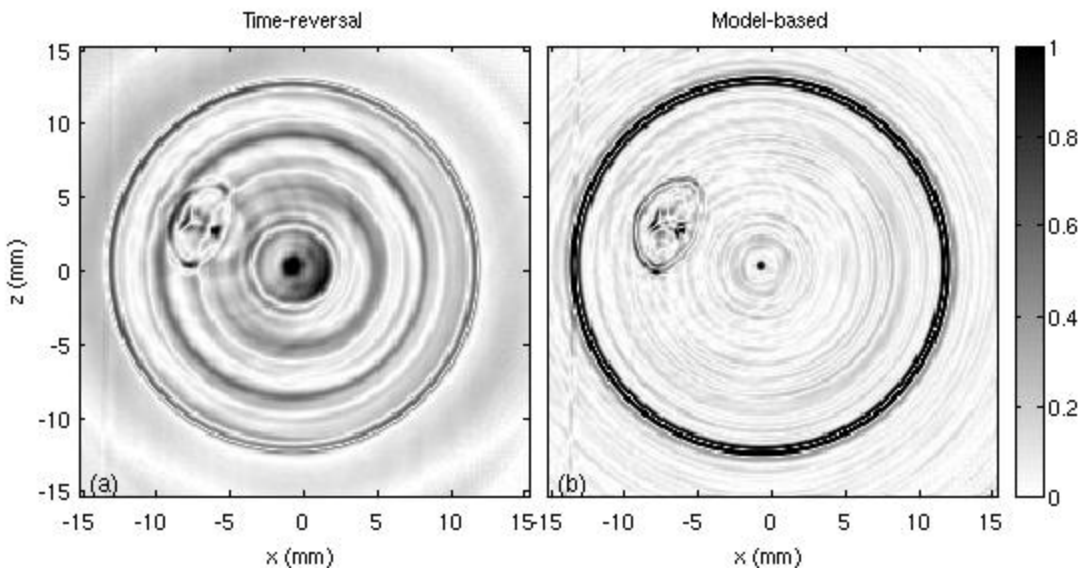


Figure 4.20. Illustrates the reconstruction of the plastic cylinder with defect via a) time reversal and b) via model-based.

The reconstructed images in Fig. 4.20 clearly show the presence a defect within the cylinder. However the shape of the defect and the size of it are not correct. Thus to get more accurate reconstructions we must include refraction in the system model. For a detailed explanation of the system model derivation, refer to appendix c. It suffices to note that the geometry of the setup is as follows.

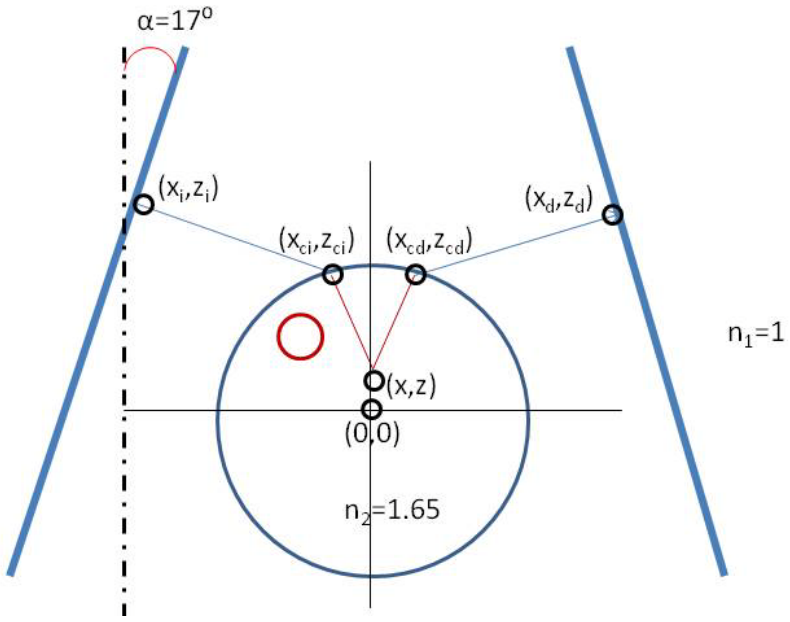


Figure 4.21. Illustrates geometrical setup of the experiment.

In the above geometrical setup, there are three regions to consider. The first region corresponds to points that fall in the shadow of the cylinder. Those points do not contribute to the wavefield data. The second region corresponds to points outside of the region, and the optical path length and hence the time of flight is given by Eq. (4.4). The third region corresponds to points that fall within escapable areas of the cylinder. The optical path length for a point that falls in the third region is given by the following expression:

$$\begin{aligned} OPL = {} & n_1\sqrt{(x_i - x_{ci})^2 + (z_i - z_{ci})^2} + n_2\sqrt{(x_{ci} - x)^2 + (z_{ci} - z)^2} \\ & + n_2\sqrt{(x - x_{cd})^2 + (z - z_{cd})^2} \\ & + n_1\sqrt{(x_{cd} - x_d)^2 + (z_{cd} - z_c)^2} \end{aligned} \tag{4.23}$$

According to the expression given in Eq. (4.23), we must find the two points on the surface of the cylinder (x_{ci}, z_{ci}) and (x_{cd}, z_{cd}) that the grid point (x, z) refract to for both illumination and detection respectively. However, in the current geometrical setup, it is difficult to find these two points because to solve the resulting equation requires good initial guess for the location of these two points. Thus, it is much simpler to split the problem into an illumination and detection phase and then to apply a coordinate rotation such that our initial guess will always be correct. The following figure illustrates the splitting of our geometrical setup shown in Fig. 4.21 into two an illumination phase and a detection phase.

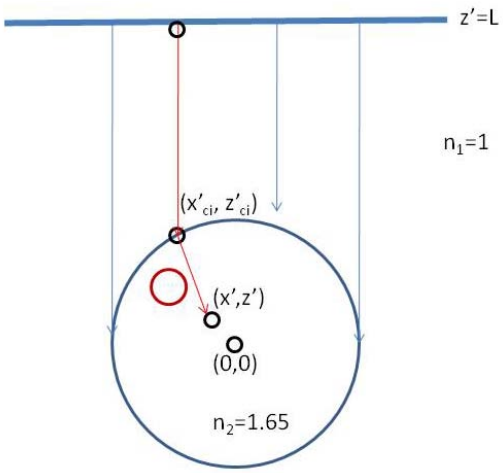


Figure 4.22. Illustrates the illumination phase of the setup.

In the original coordinate system, the illuminating rays are at 107^0 with respect to the horizontal, so we rotate the coordinate plane by -17^0 so that the rays are parallel to the z-axis. Since the illumination OPL calculation is the most complicated for points that fall within the escapable regions of the cylinder, the following analysis is for points in that region. A grid point, (x, z) is first transformed into the grid point (x', z') by a coordinate transformation. For the transformed grid point, we must find the corresponding refracted point. To find the refracted point, (x'_{ci}, z'_{ci}), we must solve the following transcendental function for the normal angle, ϕ, given the transformed grid point (x', z'):

$$f(\phi) = \sin(\psi - \phi) - n * \sin\left(\tan^{-1}\left[\frac{r*\sin(\phi)-z'}{r*\cos\phi - x'}\right] - \phi\right) = 0 \tag{4.24}$$

where $\psi = \frac{\pi}{2}$ because of the rotation and r is the radius of the cylinder. Once we have found the value for the normal angle ϕ, then we can easily solve for (x'_{ci}, z'_{ci}) by using the following trignometric identities:

$$x'_{ci} = r * \cos\phi$$
$$z'_{ci} = r * \sin\phi$$

Hence, the illumination OPL is given by the following simpler expression:

$$OPL_{ill} = n_1(L - z'_{ci}) + n_2\sqrt{\left(x'_{ci} - x'\right)^2 + \left(z'_{ci} - z'\right)^2} \tag{4.25}$$

where L is the distance to the illumination reference plane.

We can proceed in a similar manner for the detection phase. In the original coordinate plane, the detected rays are at 73^0 with respect to the horizontal so we rotate the coordinate grid by 17^0 so that the rays are parallel to the z-axis.

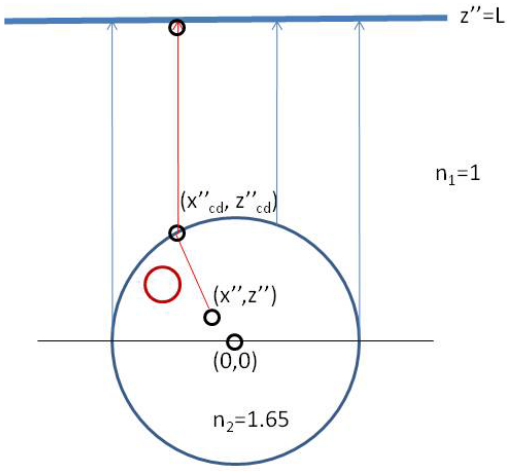


Figure 4.23. Illustrates the detection phase of the setup.

A grid point (x, z) within the escapable region of the cylinder is transformed into the point (x'', z''). Given the transformed gird point, we locate the point (x''_{cd}, z''_{cd}) to where it refracts to on the surface of the cylinder. Again, solving for the point (x''_{cd}, z''_{cd}), requires solving the transcendental equation given in Eq. (4.24) for the normal angle, ϕ. Using the computed value for ϕ we can solve for (x''_{cd}, z''_{cd}) using the above trig. identities. The final expression for the detected OPL is given by:

$$OPL_{det} = n_1(L - z''_{cd}) + n_2\sqrt{(x''_{cd} - x'')^2 + (z''_{cd} - z'')^2} \tag{4.26}$$

To test the analysis presented here, we simulated a cylinder with the same dimensions as our cylinder for a rotation angle $\Theta = 0$ as shown in Fig. 4.24. The left half of Fig. 4.24 shows the three regions discussed earlier for illumination, detection, and combined phases. The right half of Fig. 4.24 shows the computed OPL for all three phases.

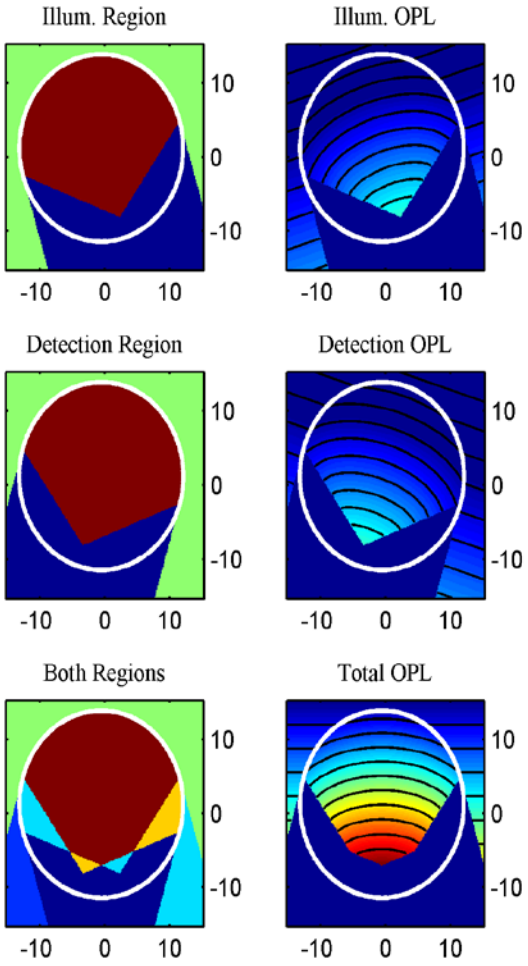


Figure 4.24. Simulation of illumination, detection, and total OPL for cylinder.

The analysis presented earlier for calculating the correct time of flight information by including refraction assumed that the center of the object was at the pivot point. Thus, we correct for the off-centering by first locating the center of plastic cylinder and then shifting the center and hence the coordinate grid to the origin, which we have designated as our pivot point. We have incorporated the correct time of flight information based on the new OPL calculation for both the time-reversal algorithm and the model-based algorithm. We then reconstructed the wavefield data for the cylinder with the defect. The reconstructed images via both algorithms are shown in Fig. 4.23.

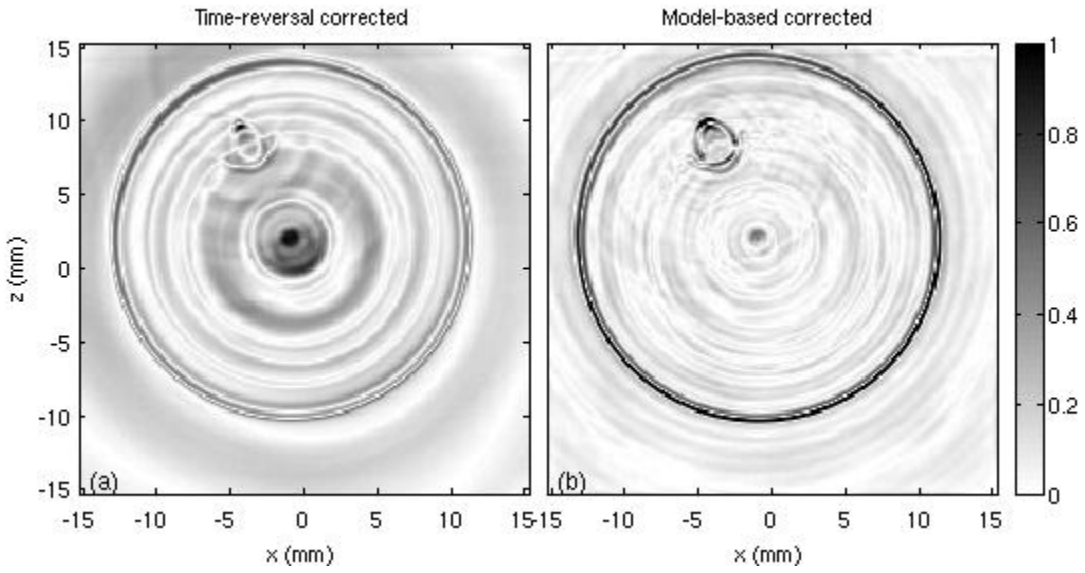


Figure 4.25. Illustrates the reconstructed images with correct time of flight via a) time-reversal and via b) model-based.

The correct time of flight information has resulted in a better reconstruction of the defect via the model-based algorithm, Fig. 4.25b), than via the time reversal method, Fig. 4.25a). Accounting for refraction did very little to improve the reconstruction of the defect via time-reversal but has significantly aided in the reconstruction of the defect via the model-based algorithm. Furthermore, the location and size of the defect shown in Fig. 4.25b) is more comparable to the true object's defect than the reconstructed defect in the uncorrected case, Fig. 4.20. In the corrected case, the reconstructed diameter of the defect is approximately 3.5 mm whereas the true diameter size of defect is 5.0 mm. The reconstructed defect's distance from the outer surface of the object is 3.5 mm and the true distance is 4.0 mm. It's plausible that much of the deviations in the reconstructed

defect's dimension and location stems from the fact our estimate of the index of refraction, n, might not be correct. As stated earlier, this value was not supplied to us by PicoMetrix and thus had to be estimated.

Conclusion

In this chapter, we have applied the principles of model-based imaging, developed in the previous chapter, to a 2D THz system operating in reflection mode to show fast reconstruction of metallic and dielectric objects. We have compared the performance of this new inverse imaging technique to the time-reversal method in order to show that we can obtain better quality images without an increase in the acquisition speed of the THz imaging system. Initially, we presented a brief discussion on the need to calibrate our 2D THz system using a well machined post. We then applied the model-based algorithm with a simple quadratic penalty to the post wavefield data and showed the better performance achieved via this algorithm as compared to the time-reversal algorithm. We then showed the limitation of the QPLS algorithm when applied to a two wire object because of its inability to suppress the weakly scattered signals in the data. Although, we could have changed the system model to take into account the weak scattering process, we opted to go in a different route and develop the Iterative Soft Thresholding algorithm based on optimization transfer to suppress the weakly scattered signals. We then characterized the performance of this algorithm as compared to the time-reversal and the QPLS algorithms. Lastly, we discussed imaging dielectric objects and the obstacles associated with such an endeavor. We began with imaging a pencil and compared the reconstruction of the pencil wavefield data via the time-reversal and the model-based algorithm based on QPLS. Although, qualitatively the model-based algorithm was superior in reconstructing a better image of the pencil than the time-reversal, the reconstructed graphite core did not have the correct diameter because we neglected the index of refraction of the pencil. We then naively attempted to reconstruct the wavefield data belonging to a plastic cylinder with a 5 mm defect without taking into account refraction. The reconstruction via both algorithms showed the presence of a defect in the cylinder. However, neither the size nor the location of the defect was correct. We then modeled the refraction geometry to obtain better timing information and then incorporated that information in both algorithms. In the model-based algorithm, the

correct time greatly improved the reconstruction of the defect but did very little to help the time-reversal algorithm. The experimental studies in this chapter clearly show that the theory presented in the previous chapter is very applicable to THz imaging.

Inverse imaging architectures such as the model-based THz imaging have the potential to reconstruct quality images from scattered fields without compromising the acquisition speed of the system. The iterative algorithms developed for demonstrating model-based THz imaging were not optimized for speed and were meant for demonstration purposes. To obtain a speed up of the algorithms, one would need to migrate the algorithms to a compiled language like C and not rely on the Matlab (Mathworks) environment. A further speed of these algorithms can be obtained via parallelization of the algorithms. Thus the acquisition speed of a model-based THz imaging system much like the time-reversal imaging system is largely dependent on the hardware of the system and the number of measurements and not so much on the algorithms.

Chapter 5
Conclusion

THz imaging is an emergent technology that has found relevancy in many diverse fields. Despite it promising potential, there are many obstacles that stand in the way of its large scale industrial introduction. There are a number of key research areas that promise significant continuing advances in THz technology. Current efforts in THz hardware are vital for THz to make its transition from the laboratory to industrial settings. Equally important is the development of imaging architectures and algorithms that accurately and quickly process THz data into images. Direct imaging architectures are the most established and probably the most commonly used THz imaging techniques. However, these systems are limited either by speed or by SNR. For example, the traditional scanned THz imaging system is very popular because of its simple setup and its impressive SNR. The need to raster scan the object however leads to a long acquisition time. Attempts to increase the speed of the system via 2D free space electro-optic sampling or via chirping the probe pulse do so but at the expense of SNR. However, by using an inverse architecture one can still use a single point detection scheme to benefit from the high SNR while collecting a fraction of the number of measurements need by a scanning THz system.

The goal of this thesis has largely been on advancing the field of THz imaging via inverse imaging methods. The inspiration for this thesis has sprung from Ruffin's work in time-reversal THz imaging. Our goal in chapter 2 had been to improve on the time-reversal method via two approaches. The first method was along the lines of a hardware improvement. We wanted to improve the numerical aperture of our THz imaging system and so we adapted a waveguide approach first pioneered in ultrasound to reach that goal. The waveguide approach was a success because of the 2.6 x improvement in intensity and the approximate 30% improvement in resolution. The second method presented an algorithm improvement to the time-reversal technique. The time-reversal method simply

back-propagates the diffracted field so that it focuses at the object plane. It does not take into account the impulse response of the THz system which has significant ringing. The ringing degrades the quality of the reconstructed image. In chapter 2 we develop a model-based algorithm that takes into account the impulse response of the THz system, and attempts to recover some underlying function that describes the object from the collected data in a "best fit" manner without overly fitting the noise. The 1D RLS algorithm is developed in chapter 2 without a rigorous discussion of the theory behind model-based image reconstruction, which has been deferred till chapter3. The algorithm was implemented in the presence and absence of the waveguide. In absence of the waveguide, the algorithm was able to achieve a peak SNR ratio improvement of 2.2 and an approximate improvement of 25% in the resolution. In the case of the waveguide, the improvement was not as dramatic as one would have hoped and this largely stems from the fact that we could not accurately model the waveguide. In hind sight, we should have fitted the wavefield data to obtain the distance and pitch of the waveguide instead of relying on the approximate measurements from the lab. Nonetheless the algorithm did result in a peak SNR ratio improvement of 1.47. The improvement of resolution though was negligible.

In chapter 3, we focused exclusively on the mathematical formulism of model-based image reconstruction as means for deriving algorithms that solve inverse problems. The model-based approach through regularization provides a stable method for inverting a forward imaging model that incorporates the actual impulse response of the imaging system. We discussed the five steps to model-based image reconstruction which are:

- Parameterization of the object using a parametric model
- Development of a system model that relates the unknown image to the expectation of each detector measurement
- Development of a statistical model for how the detector measurements vary about their expectations
- The selection of a cost function with an appropriate regularization term based on the tools of penalized-likelihood estimation

- The selection of an interative algorithm for the minimization of the regularized cost function

We provided a thorough discussion of each of the bulleted points in preparation for the experimental studies presented in chapter 4.

In Chapter 4, we apply the principles of model-based image reconstruction developed in chapter 3 to a 2D THz imaging system operating in reflection to show fast reconstruction of metallic and dielectric objects. We then compared the performance of this new inverse imaging technique to the time-reversal method to show that we can obtain better quality images without compromising the acquisition speed of the THz imaging system. The first experiment consisted of reconstructing the wavefield data for a well machined post. Both the model-based and time-reversal method reconstructed the post. However the artifacts caused by the temporal ringing on the THz pulse have been mitigated by the model-based algorithm and has resulted in a 14 dB improvement in the SNR over the time-reversal algorithm. In the next experiment we imaged two metal wires. The proximity of the two wires led to weak scattering between them and this was very apparent in the wavefield data. The reconstructed image via the model-based algorithm based on QPLS was better than the time-reversal because much of the ringing has been suppressed and the image looks sharper than the time-reversal image. However, in both images the weak scattering that was present in the wavefield data has reconstructed to a double ring artifact in the image. This is attributed to the fact that we did not model the weak scattering. One option to suppress the double ring artifact that we pursued was to use a penalty function that enforced the sparsity of our image. However, the use of this penalty function has resulted in a cost function that is difficult to minimize and thus we introduced the concept of optimization transfer which enabled us to use surrogate functions to minimize our cost function. The resulting Iterative Soft Thresholding algorithm was able to suppress the double ring artifact while preserving the reconstruction of the wire. Thus, optimization transfer allows one to minimize custom cost functions that in general may not be amiable to general purpose minimization techniques such as conjugate gradients.

We showed the applicability of model-based image reconstruction in the context of imaging dielectrics. We showed that the reconstruction of the pencil via model-based

was better than the time-reversal algorithm because of the suppression of the artifacts due to the temporal ringing on the impulse response of the system. However, since we neglected the index of refraction of the pencil, the dimension of the reconstructed graphite core was not accurate. We showed that when accounting for the proper time delays we were able to obtain a reconstruction of a defect inside a plastic cylinder via the model-based algorithm. However, the ringing in the reconstructed image via time-reversal obscured the defect.

Future Work

There are a vast number of open questions and promising future research problems associated with any rapidly emerging technology and THz imaging is no exception. This section highlights promising extensions of the work presented in this thesis. Development of higher power THz sources and higher sensitivity detectors remains a fundamental area of research. However, engineering considerations such as the development of high speed imaging systems and algorithms for accurately processing THz measurements are equally important for large scale introduction of THz systems.

In chapter 2, we developed the waveguide technique as way to effectively increase the numerical aperture without compromising acquisition speed. A promising avenue to pursue is to develop a better model for the waveguide technique and to extend it to 2-dimensions. In chapter 3, model-based image reconstruction in context of THz imaging systems was developed as an alternative to time-reversal THz imaging. The model-based algorithms provide a better approach than the time-reversal technique for imaging point scatters in homogenous backgrounds and this may find application in identifying defects such as voids in foam insulation. However, for more general targets, nonlinear estimation techniques need to be explored and this can serve as a future research endeavor. Furthermore, in the context of imaging dielectrics better modeling of dielectric properties are needed. In chapter 4, several experimental case studies are conducted to verify the performance of the model-based algorithms. Future extensions to this chapter could include demonstrating the time-reversal technique with negative signals, developing a better statistical model for a “non quiet laser”, and incorporating a sign reversal for the pencil data. . Another avenue of research is to develop nonlinear

estimation algorithms for general targets and to experimentally assess the performance of these algorithms.

Appendix A
Convergence Analysis of Steepest Descent Algorithm

In order to analyze the converges of steepest descent algorithm for a quadratic cost function given by,

$$\Psi(x) = \frac{1}{2}x^T Ax - bx + c \tag{1.1}$$

we introduce a few definitions: The error $e_{(i)} = x_{(i)} - \hat{x}$ is a vector that indicates how far we are from the solution. The residual $r_{(i)} = y - Ax_{(i)}$ indicates how far we are from the correct value of y. It is important to note that the residual is the error transformed by A into the same space as y so that that $r_{(i)} = -Ae_{(i)}$ as well as being the direction of steepest descent $r_{(i)} = -\Psi'(x_i)$.

It can be shown that if A is symmetric, there exists a set of n orthogonal eigenvectors of A. Furthermore these eigenvectors can be scaled to have unit length.

$$v_j^T v_k = \begin{cases} 1 & j = k \\ 0 & j \neq k \end{cases} \tag{1.2}$$

To begin with, we express the error vector $e_{(i)}$ as a linear combination of eigenvectors.

$$e_{(i)} = \sum_{j=1}^{n} \xi_j v_j \tag{1.3}$$

Where ξ_j is the length of each component of $e_{(i)}$. Form the above two equations, we have the following identities:

$$r_{(i)} = -Ae_{(i)} = -\sum_{j=1}^{n} \xi_j \lambda_j v_j \tag{1.4}$$

$$\left\|e_{(i)}\right\|^2 = {e_{(i)}}^T e_{(i)} = \left(\sum_{j=1}^{n} \xi_j {v_j}^T\right)\left(\sum_{m=1}^{n} \xi_m v_m\right)$$

$$\left\|e_{(i)}\right\|^2 = \sum_{j=1}^{n}\sum_{m=1}^{n} \xi_j \xi_m v_j^T v_m = \sum_{j=1}^{n}\sum_{m=1}^{n} \xi_j \xi_m \delta(j-m) \tag{1.5}$$

$$\left\|e_{(i)}\right\|^2 = \sum_{j=1}^{n} \xi_j^2$$

$$\left\|e_{(i)}\right\|_A^2 = {e_{(i)}}^T A e_{(i)}$$

$$= \left(\sum_{j=1}^{n} \xi_j {v_j}^T\right) A \left(\sum_{m=1}^{n} \xi_m v_m\right)$$

$$\left\|e_{(i)}\right\|_A^2 = \left(\sum_{j=1}^{n} \xi_j {v_j}^T\right)\left(\sum_{m=1}^{n} \xi_m A v_m\right)$$

$$= \left(\sum_{j=1}^{n} \xi_j {v_j}^T\right)\left(\sum_{m=1}^{n} \xi_m \lambda_m v_m\right) \tag{1.6}$$

$$\left\|e_{(i)}\right\|_A^2 = \sum_{j=1}^{n}\sum_{m=1}^{n} \xi_j \xi_m \lambda_m \delta(j-m) = \sum_{j=1}^{n} \xi_j^2 \lambda_j$$

Likewise for the residuals we have the following identities without proofs:

$$\left\|r_{(i)}\right\|^2 = {r_{(i)}}^T r_{(i)} = \sum_{j=1}^{n} \xi_j^2 {\lambda_j}^2 \tag{1.7}$$

and

$$\left\|r_{(i)}\right\|_A^2 = {r_{(i)}}^T A r_{(i)} = \sum_{j=1}^{n} \xi_j^2 {\lambda_j}^3 \tag{1.8}$$

To bound the convergence of steepest descent in the general case, we use the energy norm instead of Euclidean norm. We first note that:

$$e_{(i+1)} = x_{(i+1)} - \hat{x} = x_{(i)} + \alpha_i r_i - \hat{x} = e_{(i)} + +\alpha_i r_i \tag{1.9}$$

and that the step length for the steepest descent algorithm is given by

$$\alpha_i = \frac{{r_{(i)}}^T r_{(i)}}{{r_{(i)}}^T A r_{(i)}} \tag{1.10}$$

So that we can define the following

$$\left\|e_{(i+1)}\right\|_A^2 = {e_{(i+1)}}^T A e_{(i+1)} = \left(e_{(i)} + +\alpha_i r_{(i)}\right)^T A\left(e_{(i)} + +\alpha_i r_{(i)}\right)$$

$$\left\|e_{(i+1)}\right\|_A^2 = {e_{(i)}}^T A e_{(i)} + 2\alpha_i {r_{(i)}}^T A e_{(i)} + {\alpha_i}^2 {r_{(i)}}^T A r_{(i)}$$

$$\left\|e_{(i+1)}\right\|_A^2 = \left\|e_{(i)}\right\|_A^2 + 2\frac{{r_{(i)}}^T r_{(i)}}{{r_{(i)}}^T A r_{(i)}}\left(-{r_{(i)}}^T r_{(i)}\right) + \left(\frac{{r_{(i)}}^T r_{(i)}}{{r_{(i)}}^T A r_{(i)}}\right)^2 {r_{(i)}}^T A r_{(i)} \tag{1.11}$$

$$\left\|e_{(i+1)}\right\|_A^2 = \left\|e_{(i)}\right\|_A^2 - \frac{\left({r_{(i)}}^T r_{(i)}\right)^2}{\left({r_{(i)}}^T A r_{(i)}\right)}$$

We can further simplify this expression as follows:

$$\left\|e_{(i+1)}\right\|_A^2 = \left\|e_{(i)}\right\|_A^2 \left(1 - \frac{\left({r_{(i)}}^T r_{(i)}\right)^2}{\left({r_{(i)}}^T A r_{(i)}\right)\left({e_{(i)}}^T A e_{(i)}\right)}\right) \tag{1.12}$$

Substituting into eq. 1.12 the expressions given in eq. 1.5, 1.7, and 1.8, we get the following expression

$$\left\|e_{(i+1)}\right\|_A^2 = \left\|e_{(i)}\right\|_A^2 \left(1 - \frac{\left(\sum_{j=1}^n \xi_j^2 {\lambda_j}^2\right)^2}{\left(\sum_{j=1}^n \xi_j^2 {\lambda_j}^3\right)\left(\sum_{j=1}^n \xi_j^2 \lambda_j\right)}\right) \tag{1.13}$$

which can be rewritten as:

$$\left\|e_{(i+1)}\right\|_A^2 = \left\|e_{(i)}\right\|_A^2 w^2, \quad w^2 = \left(1 - \frac{\left(\sum_{j=1}^n \xi_j^2 {\lambda_j}^2\right)^2}{\left(\sum_{j=1}^n \xi_j^2 {\lambda_j}^3\right)\left(\sum_{j=1}^n \xi_j^2 \lambda_j\right)}\right) \tag{1.14}$$

The convergence analysis now hinges on finding an upper bound for w which without proof is shown to be

$$w \leq \frac{\kappa - 1}{\kappa + 1} \tag{1.15}$$

where κ, the condition number of A, is the ratio of the largest eigenvalue of A to the smallest eigenvalue of A. Hence, the convergence result for steepest descent is

$$\left\|e_{(i+1)}\right\|_A^2 \leq \left\|e_{(i)}\right\|_A^2 \left(\frac{\kappa - 1}{\kappa + 1}\right) \tag{1.16}$$

Appendix B
Derivation of CG Update Expression

The conjugate-gradient algorithm is an iterative method for unconstrained minimization that produces a better approximation to the minimum of a general unconstrained nonlinear equation of N variables, $x_1, x_2, \ldots, x_N$ with each iteration [58]. Within a given iteration an estimate is made of the best way to change each component of the vector $\boldsymbol{x}$ so as to produce the maximum reduction of the function by finding the gradient fo the function with respect to the variables and combining this gradient with information from previous iterations to produce a search direction. The search direction is an estimate of the relative change in each component of the vector $\boldsymbol{x}$ to produce the maximum reduction in the function F. To find the magnitude of the changes along the search direction, an optimal step size must be estimated. The new vector after an iteration of the conjugate gradient $\boldsymbol{x_{k+1}}$ is given by the previous vector $\boldsymbol{x_{k+1}}$ plus an optimal step size times the search direction.

The conjugate gradient falls in the family of conjugate-direction method and was developed by Hesstenes and Stieffel (1952) [58]. To understand CG or CD, one must understand the concept of conjugacy. Suppose we consider the following quadratic function given by:

$$F(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}'\boldsymbol{G}\boldsymbol{x} + \boldsymbol{b}'\boldsymbol{x} + c \tag{2.1}$$

where $\boldsymbol{G}$ is a positive definite symmetric matrix, $\boldsymbol{b}$ a vector, and c a scalar. Then the directions directions respresented by two vectors $\boldsymbol{u} \neq 0$ and $\boldsymbol{v} \neq 0$ are conjugate with respect to $\boldsymbol{G}$ if

$$\boldsymbol{u}'\boldsymbol{G}\boldsymbol{v} = 0 \tag{2.2}$$

Theorem 1. *If the vectors $\boldsymbol{d}_j$ are mutually conjugate (i.e. $\boldsymbol{d}_i'\boldsymbol{G}\boldsymbol{d}_j = 0$ $for\ i \neq j, for\ all\ i\ and\ j$), then they are linearly independent.*

It follows that there exists at least one set of n independent vectors mutally conjugate with respect to the matrix $\boldsymbol{G}$; the set of eigenvectors of **G** forms such a set. We now look at minimizing $F(x)$ subject to $\boldsymbol{x} \in \mathbb{R}^N$.

Theorem 2. *Let $\boldsymbol{x_{k+1}}$and $\boldsymbol{x_k}$ be consecutive current points in a minimization of $F(\boldsymbol{x})$. If (i) $\boldsymbol{x_k}$ minimizes $F(\boldsymbol{x})$ in direction $\boldsymbol{d_l}$, (ii) $\boldsymbol{x_{k+1}}$ minimizes $F(\boldsymbol{x})$ in direction $\boldsymbol{d_m}$, (iii) $\boldsymbol{d_l}$ and $\boldsymbol{d_m}$ are conjugate directions, then $\boldsymbol{x_{k+1}}$ also minimizes $F(\boldsymbol{x})$ in direction $\boldsymbol{d_l}$.*

If we denote gradient evaluated at the point $\boldsymbol{x_k}$ by $\boldsymbol{g}_k = \nabla F(\boldsymbol{x_k})$, then condition (i) and condition (iii) imply that $\boldsymbol{d_{k-1}^T g}_l = 0 \quad for\ each\ k = 0, \ldots, l$

Starting with Eq. 1, the gradient of $F(\boldsymbol{x})$ is

$$\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{b} + \boldsymbol{G}\boldsymbol{x} \tag{2.3}$$

and for another point $\boldsymbol{x_k}$, the gradient is equal to the following:

$$\boldsymbol{g}(\boldsymbol{x_k}) = \boldsymbol{b} + \boldsymbol{G}\boldsymbol{x_k} \tag{2.4}$$

The line minimization is given as:

$$\frac{dF(\boldsymbol{x_k} + \alpha \boldsymbol{d_k})}{d\alpha} = 0 \tag{2.5}$$

Taking a quadratic approximation about point $\boldsymbol{x_k}$ using Taylor series:

$$F(\boldsymbol{x}) \approx F(\boldsymbol{x_k}) + \nabla^T F(\boldsymbol{x_k})(\boldsymbol{x} - \boldsymbol{x_k}) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x_k})^T \nabla^2 F(\boldsymbol{x_k})(\boldsymbol{x} - \boldsymbol{x_k})$$

Taking **G**=$\nabla^2 F(\boldsymbol{x_k})$, the Hessian of F, and setting $\boldsymbol{x} = \boldsymbol{x_k} + \alpha \boldsymbol{d_k}$ yielding

$$F(\boldsymbol{x_k} + \alpha \boldsymbol{d_k}) \approx F(\boldsymbol{x_k}) + \alpha \nabla^T F(\boldsymbol{x_k}) \boldsymbol{d_k} + \frac{1}{2}\alpha^2 (\boldsymbol{d_k})^T \boldsymbol{G} \boldsymbol{d_k}$$

$$\frac{dF(\boldsymbol{x_k} + \alpha \boldsymbol{d_k})}{d\alpha} = \nabla^T F(\boldsymbol{x_k}) \boldsymbol{d_k} + \alpha (\boldsymbol{d_k})^T \boldsymbol{G} \boldsymbol{d_k} = 0$$

Solving for α yields

$$\alpha_k = \frac{-\nabla^T F(\boldsymbol{x_k}) \boldsymbol{d_k}}{(\boldsymbol{d_k})^T \boldsymbol{G} \boldsymbol{d_k}} = \frac{-\boldsymbol{g}(\boldsymbol{x_k})^T \boldsymbol{d_k}}{(\boldsymbol{d_k})^T \boldsymbol{G} \boldsymbol{d_k}} = \frac{-\boldsymbol{g_k}^T \boldsymbol{d_k}}{(\boldsymbol{d_k})^T \boldsymbol{G} \boldsymbol{d_k}} \tag{2.6}$$

We also have by expanding the gradient term and introducing $\boldsymbol{x}_{k+1} = \boldsymbol{x_k} + \alpha_k \boldsymbol{d_k}$ that

$$(\boldsymbol{b} + \boldsymbol{G}\boldsymbol{x_k})^T \boldsymbol{d_k} + \alpha_k (\boldsymbol{d_k})^T \boldsymbol{G} \boldsymbol{d_k} = \boldsymbol{0}$$

$$\boldsymbol{d_k}^T (\boldsymbol{b} + \boldsymbol{G}\boldsymbol{x_k}) + (\boldsymbol{d_k})^T \boldsymbol{G}(\boldsymbol{x}_{k+1} - \boldsymbol{x_k}) = \boldsymbol{0}$$

$$\boldsymbol{d_k}^T (\boldsymbol{b} + \boldsymbol{G}\boldsymbol{x}_{k+1}) = 0$$

$$\boldsymbol{d_k}^T \boldsymbol{g_{k+1}} = \boldsymbol{0} \tag{2.7}$$

This equation shows that the gradient at the point $\boldsymbol{x}_{k+1}$ is orthogonal to the previous search direction $\boldsymbol{d}_{\boldsymbol{k}}$. To prove now that $\boldsymbol{d}_{\boldsymbol{k-1}}^{T}\boldsymbol{g}_{l} = 0 \quad for\ each\ k = 0, \ldots, l$, we first note the recursion relationship:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_{\boldsymbol{k}} + \alpha_{k}\boldsymbol{d}_{\boldsymbol{k}}$$

$$\boldsymbol{x}_{k+3} = \boldsymbol{x}_{\boldsymbol{k}} + \alpha_{k}\boldsymbol{d}_{\boldsymbol{k}} + \alpha_{k+1}\boldsymbol{d}_{\boldsymbol{k+1}} + \alpha_{k+2}\boldsymbol{d}_{\boldsymbol{k+2}} = \boldsymbol{x}_{\boldsymbol{k}} + \sum_{\boldsymbol{i=0}}^{\boldsymbol{2}} \alpha_{i}\boldsymbol{d}_{\boldsymbol{i}}$$

Hence in general we have the following recursive formula

$$\boldsymbol{x}_{l} = \boldsymbol{x}_{\boldsymbol{k}} + \sum_{\boldsymbol{i=k}}^{\boldsymbol{l-1}} \alpha_{i}\boldsymbol{d}_{\boldsymbol{i}} \tag{2.8}$$

We start from the expression for the gradient

$$\boldsymbol{g}_{l} = \boldsymbol{b} + \boldsymbol{G}\boldsymbol{x}_{\boldsymbol{l}} = \boldsymbol{b} + \boldsymbol{G}\left(\boldsymbol{x}_{\boldsymbol{k}} + \sum_{\boldsymbol{i=k}}^{\boldsymbol{l-1}} \alpha_{i}\boldsymbol{d}_{\boldsymbol{i}} \right) = \boldsymbol{b} + \boldsymbol{G}\boldsymbol{x}_{\boldsymbol{k}} + \sum_{\boldsymbol{i=k}}^{\boldsymbol{l-1}} \alpha_{i}\boldsymbol{G}\boldsymbol{d}_{\boldsymbol{i}}$$

This expression further simplifies to

$$\boldsymbol{g}_{l} = \boldsymbol{g}_{\boldsymbol{k}} + \sum_{\boldsymbol{i=k}}^{\boldsymbol{l-1}} \alpha_{i}\boldsymbol{G}\boldsymbol{d}_{\boldsymbol{i}}$$

Premultiplication of this equation by $\boldsymbol{d}_{k-1}^{T}$ yields

$$\boldsymbol{d}_{k-1}^{T}\boldsymbol{g}_{l} = \boldsymbol{d}_{k-1}^{T}\boldsymbol{g}_{k} + \sum_{\boldsymbol{i=k}}^{\boldsymbol{l-1}} \alpha_{i}\boldsymbol{d}_{k-1}^{T}\boldsymbol{G}\boldsymbol{d}_{\boldsymbol{i}}$$

The first term on the right hand side of the above equation vanishes since we have already proven that the gradient at a point is orthogonal to the previous search direction if the quadratic function is minimized in that search direction. All of the terms in the sum of the second term on the right hand side of the Eq. () vanishes because of conjugacy. Hence,

$$\boldsymbol{d}_{k-1}^{T}\boldsymbol{g}_{l} = 0, \quad for \quad 0 \leq k \leq l \tag{2.9}$$

Theorem 3: *Let* $\boldsymbol{d}_{i}, \quad i = 1, \ldots, m\ (m \leq n)$ *be mutually conjugate directions. Then the global minimum of* $F(\boldsymbol{x})$ *can be found from an arbitrary starting point* $\boldsymbol{x}^{0}$ *by a finite descent computation in which each of the* $\boldsymbol{d}_{i}$ *is used as a descent direction only once.*

Construction of a set of mutually conjugate directions

Given a set of linearly independent vectors $\boldsymbol{v_0}, \dots, \boldsymbol{v_{n-1}}$, one can construct a set of mutually **G**-conjugate directions $\boldsymbol{d}_0, \dots, \boldsymbol{d}_{n-1}$ by the following procedure [58]. Set

$$\boldsymbol{d}_0 = \boldsymbol{v_0}$$

And then for $i = 1, \dots n-1$ successively define

$$\boldsymbol{d}_i = \boldsymbol{v_i} + \sum_{\boldsymbol{j=0}}^{\boldsymbol{i-1}} a_{ij}\boldsymbol{d_j}$$

where a_{ij} are coefficients chosen so that $\boldsymbol{d}_i$ is **G**-conjugate to the previous directions $\boldsymbol{d}_{i-1}, \boldsymbol{d}_{i-2}, \dots, \boldsymbol{d}_0$. This is possible if, for $l = 0, \dots, i-1$,

$$\boldsymbol{d}_i^T \boldsymbol{G} \boldsymbol{d_l} = \boldsymbol{v}_i^T \boldsymbol{G} \boldsymbol{d}_l + \sum_{\boldsymbol{j=0}}^{\boldsymbol{i-1}} a_{ij} \boldsymbol{d_j}^T \boldsymbol{G} \boldsymbol{d}_l = 0$$

If previous coefficients a_{ij} were chosen so that $\boldsymbol{d}_0, \dots, \boldsymbol{d}_{i-1}$ are **G**-conjugate, then we have that

$$\boldsymbol{d_j}^T \boldsymbol{G} \boldsymbol{d}_l = \begin{cases} 1 & j = l \\ 0 & j \neq l \end{cases}$$

Thus only one term in the summation is present and thus we have that:

$$a_{ij} = \frac{\boldsymbol{v}_i^T \boldsymbol{G} \boldsymbol{d}_l}{\boldsymbol{d_j}^T \boldsymbol{G} \boldsymbol{d}_l} \quad for\ all\ i = 1, \dots n-1\ and\ j = 0, \dots, i-1 \tag{2.10}$$

Thus the generated set of directions $\boldsymbol{d}_0, \dots, \boldsymbol{d}_{n-1}$ is **G**-conjugate and the subspaces spanned by $\boldsymbol{d}_0, \dots, \boldsymbol{d}_i$ is equal to the subspace spanned by $\boldsymbol{v}_0, \dots, \boldsymbol{v}_i$.

If in the construction process described above we take the initial step is in the direction of steepest descent

$$\boldsymbol{v}_0 = -\boldsymbol{g}_0$$

And the rest of the vectors are $\boldsymbol{v}_1 = -\boldsymbol{g}_1, \dots, \boldsymbol{v}_{n-1} = -\boldsymbol{g}_{n-1}$, we then find by a line search the point

$$\boldsymbol{x}_1 = \boldsymbol{x}_0 + \alpha_0 \boldsymbol{d}_0$$

The second direction $\boldsymbol{d}_1$ is found using Eq. 11 with $\boldsymbol{v}_0 = -\boldsymbol{g}_0$ and $\boldsymbol{v}_1 = -\boldsymbol{g}_1$. This gives

$$\boldsymbol{d}_1 = -\boldsymbol{g}_1 + \frac{\boldsymbol{g}_1^T \boldsymbol{G} \boldsymbol{d_0}}{\boldsymbol{d}_0^T \boldsymbol{G} \boldsymbol{d_0}} \boldsymbol{d_0}$$

We also have that $\boldsymbol{g}_1 - \boldsymbol{g}_0 = \boldsymbol{G}(\boldsymbol{x}_1 - \boldsymbol{x}_0) = \boldsymbol{G}(\boldsymbol{x}_0 + \alpha_0 \boldsymbol{d}_0 - \boldsymbol{x}_0) = \alpha_0 \boldsymbol{G}\boldsymbol{d}_0$ and so we can write

$$\boldsymbol{d}_1 = -\boldsymbol{g}_1 + \frac{\boldsymbol{g}_1^T(\boldsymbol{g}_1 - \boldsymbol{g}_0)}{\boldsymbol{d}_0^T(\boldsymbol{g}_1 - \boldsymbol{g}_0)}\boldsymbol{d_0}$$

Repeating the procedure with $\boldsymbol{v}_0 = -\boldsymbol{g}_0,\ \boldsymbol{v}_1 = -\boldsymbol{g}_1$, and $\boldsymbol{v}_k = -\boldsymbol{g}_k$, we obtain

$$\boldsymbol{d}_k = -\boldsymbol{g}_k + \sum_{j=0}^{k-1} \frac{\boldsymbol{g}_k^T(\boldsymbol{g}_{j+1} - \boldsymbol{g}_j)}{\boldsymbol{d}_j^T(\boldsymbol{g}_{j+1} - \boldsymbol{g}_j)}\boldsymbol{d_j}$$

Using the fact that $\boldsymbol{g}_k^T \boldsymbol{d}_i = 0, \quad for \quad 0 \le i \le k-1$, we obtain the following

$$\boldsymbol{g}_k^T \boldsymbol{d}_i = -\boldsymbol{g}_k^T \boldsymbol{g}_i + \sum_{j=0}^{k-1} \frac{\boldsymbol{g}_k^T(\boldsymbol{g}_{j+1} - \boldsymbol{g}_j)}{\boldsymbol{d}_j^T(\boldsymbol{g}_{j+1} - \boldsymbol{g}_j)}\boldsymbol{g}_k^T \boldsymbol{d_i} = \boldsymbol{0}$$

Hence we obtain that

$$\boldsymbol{g}_k^T \boldsymbol{g}_i = 0 \qquad for \quad 0 \le i \le k-1$$

We can simplify Eq. 20 to the simpler formula

$$\boldsymbol{d}_k = -\boldsymbol{g}_k + \beta_{k-1}\boldsymbol{d}_{k-1}$$

With

$$\beta_{k-1} = \frac{\boldsymbol{g}_k^T(\boldsymbol{g}_k - \boldsymbol{g}_k)}{\boldsymbol{d}_{k-1}^T(\boldsymbol{g}_k - \boldsymbol{g}_{k-1})} = \frac{\boldsymbol{g}_k^T \boldsymbol{g}_k}{\boldsymbol{d}_{k-1}^T(-\boldsymbol{g}_{k-1})}$$

We know that $\boldsymbol{d}_{k-1} = -\boldsymbol{g}_{k-1} + \beta_{k-1}\boldsymbol{d}_{k-2}$ and so β_k further reduces to

$$\begin{aligned} \beta_{k-1} = \frac{\boldsymbol{g}_k^T \boldsymbol{g}_k}{\boldsymbol{d}_{k-1}^T(-\boldsymbol{g}_{k-1})} &= \frac{\boldsymbol{g}_k^T \boldsymbol{g}_k}{(-\boldsymbol{g}_{k-1} + \beta_{k-1}\boldsymbol{d}_{k-2})(-\boldsymbol{g}_{k-1})} \\ &= \frac{\boldsymbol{g}_k^T \boldsymbol{g}_k}{\boldsymbol{g}_{k-1}\boldsymbol{g}_{k-1}} \end{aligned} \tag{2.11}$$

The important conclusion is that in order to generate the direction $\boldsymbol{d}_k$ we need to only know current and previous gradients $\boldsymbol{g}_k$ and $\boldsymbol{g}_{k-1}$ and the previous search direction $\boldsymbol{d}_k$.

The conjugate gradient method is used as a method of solving a set of positive definite symmetric linear equations. The linear conjugate-gradient method only uses products of a matrix with a vector and does not require the elements of the matrix explicitly. One can also solve for a positive definite matrix $\boldsymbol{R}$

Appendix C
Calculation of Refracted Point on Cylinder

Due to the symmetry of the problem, we opt to choose the detection phase in order to derive the transcendental equation for the normal angle that allows the calculation of the refracted point on the cylinder. The following figure illustrates the geometry for the derivation of the transcendental equation.

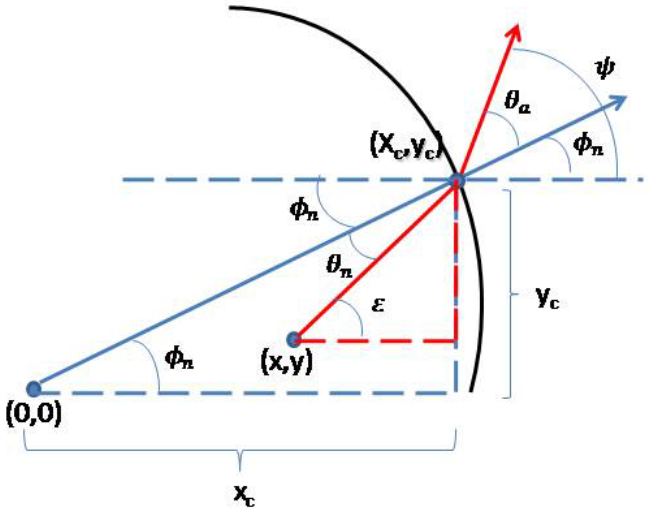


Figure C.1. Illustrates the geometry for the derivation of transcendental equation.

We derive the transcendental equation for an arbitrary orientation of the detection plane. In the actual experiment of chapter 4, we rotated our coordinate system so that the angle of the detector is at 90^0 with respect to the vertical. This simplified our computation of the temporal delays. However, here we assume that the detector plane is at ψ with respect to the vertical. Assuming that our grid point is at (x, y), we would like to find the refracted point (x_c, y_c). Hence the following is true:

$$\psi = \theta_a + \phi_n \tag{3.1}$$

but

$$\phi_n = \tan^{-1}\left(\frac{y_c}{x_c}\right) \tag{3.2}$$

By Snell's law, we have the following expression for θ_a

$$\theta_a = \sin^{-1}(n * \sin(\theta_n)) \tag{3.3}$$

and according to Fig. C.1, we have the following express for θ_n

$$\begin{aligned} \theta_n &= \varepsilon - \phi_n \\ \theta_n &= \tan^{-1}\left(\frac{y_c - y}{x_c - x}\right) - \phi_n \end{aligned} \tag{3.4}$$

We also rewrite x_c and y_c in terms of polar coordinates so that we obtain a final expression for the normal angle.

$$\begin{aligned} y_c &= r\sin(\phi_n) \\ x_c &= r\cos(\phi_n) \end{aligned} \tag{3.5}$$

Thus we have the following expression for θ_n

$$\theta_n = \tan^{-1}\left(\frac{r\sin(\phi_n) - y}{r\cos(\phi_n) - x}\right) - \phi_n \tag{3.6}$$

Substituting Eq. (3.6) and Eq. (3.3) into Eq. (3.1), we obtain the following:

$$\psi = \sin^{-1}\left(n * \sin\left(\tan^{-1}\left(\frac{r\sin(\phi_n) - y}{r\cos(\phi_n) - x}\right) - \phi_n\right)\right) + \phi_n \tag{3.7}$$

We can simplify further by moving ϕ_n to the left hand-side of Eq. (3.7) and taking the sin of both sides.

$$\sin(\psi - \phi_n) = n * \sin\left(\tan^{-1}\left(\frac{r\sin(\phi_n) - y}{r\cos(\phi_n) - x}\right) - \phi_n\right) \tag{3.8}$$

We now have an expression that is a function of ϕ_n and we would like to find the zero of it. Thus we seek the zero of the following function

$$f(\phi_n) = \sin(\psi - \phi_n) - n * \sin\left(\tan^{-1}\left(\frac{r\sin(\phi_n) - y}{r\cos(\phi_n) - x}\right) - \phi_n\right) \tag{3.9}$$

Bibliography

1. NASA. Space Shuttle Columbia and Her Crew; Available from: http://www.nasa.gov/columbia/home/index.html.
2. NASA. Columbia Accident Investigation Board: *A History of FOAM Anomalies.* Report Volume I August 2003; Available from: http://caib.nasa.gov/news/report/pdf/vol1/chapters/chapter6.pdf.
3. Jansen, C., et al., *Terrahertz imaging: application and perspectives.* Applied Optics. 2010. **49**(19): p E48-E57.
4. Mittleman, D., et al., *Recent advances in terahertz imaging.* Applied Physics B: Lasers and Optics, **68**(6): p. 1085-1094.
5. Mittleman, D., et al., *T-ray imaging.* IEEE Journal of Selected Topics in Quantum Electronics. 1996. **2**(3): p. 679-692.
6. Smye, S., et al., *The interaction between terahertz radiation and biological tissue.* Physics in Medicine and Biology. 2001. **46**(9): p. R101-R112.
7. Menikh, A., et al., *Terahertz biosensing technology: Frontiers and progress.* 2002. **3**(8): p. 655-658.
8. Mickan, S., et al., *Analysis of system trade-offs for terahertz imaging.* Microelectronics Journal. 2000. **31**(7): p. 503-514.
9. Ohtake, H., et al. *Intense THz-radiation generation from an intracavity saturable Bragg reflector in a magnetic field*. in *Lasers and Electro-Optics, 1998. CLEO 98. Technical Digest. Summaries of papers presented at the Conference on*. 1998.
10. You, D., et al., *Generation of high-power sub-single-cycle 500-fs electromagnetic pulses.* Opt. Lett., 1993. **18**(4): p. 290.
11. Smith, P., et al., *Subpicosecond photoconducting dipole antennas.* IEEE Journal of Quantum Electronics. 1988. **24**(2): p. 255-260.
12. Ralph, S., et al., *Terahertz spectroscopy of optically thick multilayered semiconductor structures.* Journal of the Optical Society of America B. 1994. **11**(12): p. 2528-2532.
13. Koch, M., et al., *Terahertz Applications and Techniques, 2007. OFC 2007.*
14. Jepsen, P., et al., *Generation and detection of terahertz pulses from biased semiconductor antennas.* Journal of the Optical Society of America B. 1996. **13**(11): p. 2424-2436.
15. Hu, B., et al., *Imaging with terahertz waves.* Optics Letters. 1995. **20**(16): p. 1716-1718.
16. Wu, Q., et al., *Dynamic range of an electro-optic field sensor and its imaging applications.* Applied Physics Letters. 1996. **68**(23): p. 3224-3226.
17. Kac, M., *Can one hear the shape of a drum.* American Math Monthly. 1966. **73**(4): p. 1-23.

18. Mourou, G., et al., *Picosecond microwave pulses generated with a subpicosecond laser driven semiconductor switch.* Applied Physics Letters. 1981. **39**(4): p. 295-296.
19. Darrow, J., et al., *Power Scaling of large aperture photoconducting antennas.* Applied Physics Letters. **58**(1): p. 25-27.
20. Darrow, J., et al., *Saturation properties of large aperture photoconducting antennas.* IEEE Journal of Quantum Electronics. 1992. **28**(6): p. 1607-1616.
21. Franken, P.A., et al., *Generation of Optical Harmonics.* Physical Review Letters, 1961. **7**(4): p. 118.
22. Bass, M., et al., *Optical Rectification.* Physical Review Letters, 1962. **9**(11): p. 446.
23. Zhang, X., et al., *Coherent measurement of THz optical rectification from electro-optic crystals.* Applied Physics Letters. 1992. **61**(23): p. 2764-2766.
24. Löffler, T., et al., *Large-area electro-optic ZnTe terahertz emitters.* Opt. Express, 2005. **13**(14): p. 5353-5362.
25. Verghese, S., et al., *Highly tunable fiber-coupled photomixers with coherent terahertz output power.* Microwave Theory and Techniques, IEEE Transactions on, 1997. **45**(8): p. 1301-1309.
26. Stepanov, A., et al., *Scaling up the energy of THz pulses created by optical rectification.* Opt. Express, 2005. **13**(15): p. 5762-5768.
27. Yeh, K.L., et al., *Generation of 10 uJ ultrashort terahertz pulses by optical rectification.* Applied Physics Letters. **90**: p. 171121.
28. Blanchard, F., et al., *Generation of 1.5 ÂµJ single-cycle terahertz pulses by optical rectification from a large aperture ZnTe crystal.* Opt. Express, 2007. **15**(20): p. 13212-13220.
29. Chang, G., et al., *Power scalable compact THz system based on an ultrafast Yb-doped fiber amplifier.* Optics Express, 2006. **14**(17): p. 7909-7913.
30. Hebling, J.n., et al., *Generation of high-power terahertz pulses by tilted-pulse-front excitation and their application possibilities.* J. Opt. Soc. Am. B, 2008. **25**(7): p. B6-B19.
31. Chang, G., et al., *Broadband THz generation pumped by Yb-doped fiber lasers.* Optics Express. 2007. **15**(25): p. 16308-16315.
32. Wu, Q., et al., *Two-dimensional electro-optic imaging of terahertz beams.* Applied Physics Letters. 1996. **68**(8): p. 1026-1028.
33. Herman, G., *Image Reconstruction from Projections-The Fundamentals of Computerized Tomography.* Academic Press, New York.
34. Markel, A., et al., *Inverse Scattering for the diffusion equation with general boundary conditions.* Physical Review E. 2001. **64**(3): p. 035601.
35. Jiang, Z., et al., *Electro-optic sampling near zero optical transmission point.* Applied Physics Letters. 2000. **76**(22): p. 3221-3223.
36. Jiang, Z., et al., *Improvement of terahertz imaging with a dynamic subtraction technique.* Applied Optics. 2000. **74**(9): p. 1191-1193.
37. Jiang, Z., et al., *Electro-optic measurement of THz field pulses with a chirped optical beam.* Applied Physics Letters. 1998. **72**(16): p. 1945-1947.

38. Jiang, Z., et al., *Measurement of spatio-temporal terahertz field distribution by using chirped pulse technology.* IEEE Journal of Quantum Electronics. 2000. **36**(10): p. 3221-3223.
39. Choi, M., et al., *Potential for detection of explosive and biological hazards with electronic terahertz impulse ranging.* Philosophical Transactions of the Royal Society of London. 2004. **362**(1815): p. 337-349.
40. Ferguson, B. "Three dimensional T-ray inspection systems". Ph.D. dissertation, University of Adelaide, 2004.
41. Ruffin, A., et al., *Time reversal terahertz imaging.* Optics Letters. 2001. **26**(10): p. 681-683.
42. Ruffin, A., et al., *Time reversal terahertz imaging.* IEEE Journal of Quantum Electronics. 2002. **38**(8): p. 1110-1119.
43. Buma, T., et al., *Time reversal three-dimensional imaging using single-cycle terahertz pulses.* Applied Physics Letters. 2004. **84**(12): p. 2196-2198.
44. Planken, P.C.M., et al., *Measurement and calculation of the orientation dependence of terahertz pulse detection in ZnTe.* Journal of the Optical Society of America B: Optical Physics, Volume 18, Issue 3, March 2001, pp.313-317, 2001. **18**: p. 313-317.
45. Born, M., and Wolf, E., *Principles of Optics,* 7th ed., Cambridge University Press, Cambridge U.K.
46. Guenther, D., *Modern Optics,* 2nd ed., New York: John Wiley and Sons, 1990.
47. Fink, M,. et al., *Acoustic time-reversal mirrors.* Inverse Problems. 2001. **17**(1): p. R1-R38.
48. Roux, P., et al., *Time reversal in an ultrasonic waveguide.* Applied Physics Letters. 1997. **70**(17): p. 1811-1813.
49. Wu, Q., et al., *Free-space electro-optic sampling of terahertz beams.* Applied Physics Letters. 1995. **67**(24): p. 3523-3525.
50. Fessler, A., *Penalized Weighted Least-Squares Image Reconstruction for Positron Emission Tomography.* IEEE Trans. Medical Imaging. 1994. **13**(2): p. 290-300.
51. Bertero, M., et al., *Ill-Posed problems in early vision.* Proceedings of the IEEE. 1988. **76**(8): p. 869-889.
52. Groetsch, C., *The theory of Tikhonov regularization for Fredholm equations of the first kind, Research notes in Mathematics,* vol. 105. Boston, MA: Pitman, 1984.
53. Hadamard, J., *Sur les problemes aux derivees partielles et leur signification physicque, University Bulletin,* vol. 13, 1902.
54. Fessler J., Department of Electrical and Computer Engineering, University of Michigan, 1301 Beal Avenue, Ann Arbor, Michigan 48109-2122, USA, is preparing a book to be titled *Image Reconstruction: Algorithms and Analysis. Image Restoration Chapter*.
55. Fessler, J., et al., *Spatial resolution properties of penalized-likelihood image reconstruction methods: Space invariant tomographs.* IEEE Trans. Image Proccess. 1996. **5**: p. 1346-1358.
56. Fessler J., Department of Electrical and Computer Engineering, University of Michigan, 1301 Beal Avenue, Ann Arbor, Michigan 48109-2122, USA, is

preparing a book to be titled *Image Reconstruction: Algorithms and Analysis. Regularization Chapter*.

57. Fessler J., Department of Electrical and Computer Engineering, University of Michigan, 1301 Beal Avenue, Ann Arbor, Michigan 48109-2122, USA, is preparing a book to be titled *Image Reconstruction: Algorithms and Analysis. General-purpose optimiation Chapter*.
58. Shewchuk, J., *An introduction to the conjugate gradient method without the agonizing pain.* 1994. Available from: http://www.cs.cmu.edu/~quake-papers/painless-conjugate-gradient.pdf.
59. Fessler J., Department of Electrical and Computer Engineering, University of Michigan, 1301 Beal Avenue, Ann Arbor, Michigan 48109-2122, USA, is preparing a book to be titled *Image Reconstruction: Algorithms and Analysis. Optimization transfer method*.