Parameter Selection in Non-quadratic Regularization-based SAR Imaging

by

Özge Batu

Submitted to the Graduate School of Engineering and Natural Sciences
in partial fulfillment of
the requirements for the degree of
Master of Science

Sabancı University
August 2008

PARAMETER SELECTION IN NON-QUADRATIC REGULARIZATION-BASED
SAR IMAGING

APPROVED BY

Assist. Prof. Dr. MÜJDAT ÇETİN   ....
(Thesis Supervisor)

Prof. Dr. AYTÜL ERÇİL   ....

Assist. Prof. Dr. HAKAN ERDOĞAN   ....

Assoc. Prof. Dr. MUSTAFA ÜNEL   ....

Assoc. Prof. Dr. Ş. İLKER BİRBİL   ....

DATE OF APPROVAL: ....

Acknowledgments

PARAMETER SELECTION IN NON-QUADRATIC REGULARIZATION-BASED SAR IMAGING

Özge Batu

EE, M.Sc. Thesis, 2008

Thesis Supervisor: Müjdat Çetin

Abstract

Many remote sensing applications such as weather forecasting and automatic target recognition (ATR) require high-resolution images. Synthetic Aperture Radar (SAR) has become an important imaging technology for these remote sensing tasks through its all-weather, day and night imaging capability. However the effectiveness of SAR imaging for a specific decision making task depends on the quality of certain features in the formed imagery. For example, in order to be able to successively use a SAR image in an ATR system, the SAR image should exhibit features of the objects in the scene that are relevant for ATR. Recently, advanced SAR image formation techniques have been developed to produce feature-enhanced SAR images.

In this thesis, we focus on one such technique, in particular a non-quadratic regularization-based approach which aims to produce so-called "point-enhanced SAR images". The idea behind this approach is to emphasize appropriate features by means of regularizing the solution. The stability of the solution is ensured through a scalar parameter, called the regularization parameter, balancing the contribution of the data and the *a priori* constraints on the formed image. Automatic selection of the regularization parameter is an important issue since SAR images are ideally aimed to be used in fully automated systems. However this issue has not been addressed in previous work.

To address the parameter selection problem in this image formation algorithm, we propose the use of Stein's unbiased risk estimation, generalized cross-validation, and L-curve techniques which have been mostly used in quadratic regularization methods previously. We have adapted these methods to the SAR imaging framework, and have developed a number of numerical tools to enable their usage. We demonstrate the effectiveness of the applied methods through experiments based on both synthetic as well as electromagnetically simulated realistic data.

KARESEL OLMAYAN DÜZENLİLEŞTİRMEYE BAĞLI SAR GÖRÜNTÜLEMEDE PARAMETRE SEÇİMİ

Özge Batu

EE, Master Tezi, 2008

Tez Danışmanı: Müjdat Çetin

Anahtar Kelimeler: parametre seçimi, sentetik açıklıklı radar, karesel olmayan düzenlileştirme, genelleştirilmiş çapraz geçerlilik sınaması, Stein'ın yansız risk kestiricisi, L-eğrisi

Özet

Hava tahmini, otomatik hedef tanıma (OHT) gibi birçok uzaktan algılama uygulaması yüksek çözünürlüklü görüntülere ihtiyaç duymaktadır. Sentetik açıklıklı radar (SAR), her türlü hava koşulunda, gündüz ve gece görüntüleme yeteneği sayesinde uzaktan algılama uygulamaları için önemli bir görüntüleme teknolojisi durumuna gelmiştir. Fakat SAR görüntülemenin belirli bir karar verme görevi için etkinliği oluşturulan görüntüdeki bazı özniteliklerin kalitesine bağlıdır. Örneğin, bir SAR görüntüsünü OHT sisteminde başarıyla kullanabilmek için, SAR görüntüsü sahnedeki nesnenin OHT ile ilgili özniteliklerini sergilemelidir. Son zamanlarda, öznitelikleri güçlendirilmiş SAR görüntüsü oluşturmak için ileri SAR görüntüleme teknikleri geliştirimiştir.

Bu tezde, özellikle "noktasal olarak güçlendirilmiş SAR görüntüsü" oluşturmayı amaçlayan karesel olmayan düzenlileştirmeye dayalı bir yaklaşıma odaklanıyoruz. Bu yaklaşımdaki fikir çözümü düzenlileştirerek uygun öznitelikleri güçlendirmektir. Çözümün kararlılığı düzenlileştirme parametresi olarak adlandırılan, ve verinin ve ön kısıtın oluşturulan görüntüye katkısını dengeleyen sayıl bir parametre ile sağlanmaktadır. SAR görüntülerinin idealde tamamen otomatik sistemlerde kullanılması hedeflendiğinden, düzenlileştirme parametresinin otomatik seçimi önemli bir sorundur. Fakat bu sorun henüz çözülmemiştir.

Bu görüntü oluşturma algoritmasındaki parametre seçme problemi için, daha önce çoğunlukla karesel düzenlileştirme yöntemlerinde kullanılmış Stein'ın yansız risk kestirimi, genelleştirilmiş çapraz geçerlilik sınaması ve L-eğrisi yöntemlerinin kullanımını öneriyoruz. Bu yöntemleri SAR görüntüleme problemine uyarladık ve kullanımlarına olanak tanımak için bir takım sayısal araçlar geliştirdik. Uygulanan yöntemlerin etkinliğini hem sentetik hem de elektromanyetik benzetimlerle elde edilmis gerçekçi veriler üzerindeki deneylerimizle gösteriyoruz.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis addresses the parameter selection problem in non-quadratic regularization based synthetic aperture radar (SAR) image reconstruction. SAR is a form of radar which is used in a variety of remote sensing applications and has the capability of imaging at high resolutions even in bad-weather or during night as well as day. The purpose of this chapter is to introduce the problem addressed in this thesis, discuss the needs for automatic parameter selection techniques, point out the main contributions, and present an outline of the thesis.

1.1   Synthetic Aperture Radar Imaging

The increased use of imaging radars in remote sensing technologies is mainly based upon several principles [2]:

1. A radar carries its own illumination, so it works in dark as well.

2. Radar systems usually emit radio waves. Waves at these frequencies pass through clouds and precipitation with little or no deterioration.

3. The radar energy scatters off materials differently from optical energy, providing a complementary and sometimes better discrimination of surface features than optical sensors.

Thus, taking advantage of these properties, synthetic aperture radar comes into favour in several military and civilian remote sensing technologies.

SAR systems usually emit microwaves or radio waves. This frequency band does not cause cloud, fog and rain effects to be observable in the radar image.
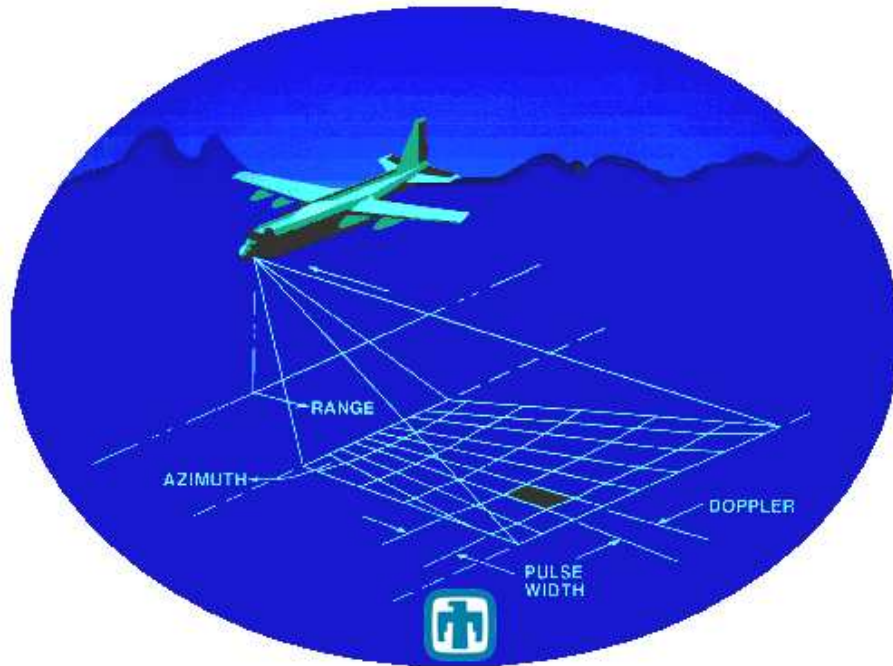
Figure 1.1: Simple illustration of data collection by synthetic aperture radar. (Image obtained from the web site of Sandia National Laboratories.)

All-weather, day and night imaging capability of the radar is the primary reason for radar to become popular for earth surface imaging tasks. Figure 1.1 shows an airborne imaging radar illuminating an area in the ground. A SAR imaging system consists of a transmit and receive antenna attached to an aircraft. While the aircraft moves along its path, it transmits microwave pulses towards the area that is to be imaged. At the earth surface, the radar pulses are scattered in all directions and the backscattered waves are received by the sensor and demodulated. These data are essentially a slice of the spatial Fourier transform of the electromagnetic reflectivity of the ground and must be inverse transformed to form an image.

In most current SAR systems, image formation is achieved through a 2-D fast Fourier transform (FFT) algorithm. This technique, which is also called *conventional* method throughout this thesis, supposes that clean and complete data can be obtained, however most of the time only reduced data are available. Furthermore, the conventional technique does not make allowance for any prior information. The conventional image formation method is basically only data driven and hence the output image has limited quality and does not incorporate any constraints regarding prior information about the scene or the objectives for the automated decision task.

(a)   (b)
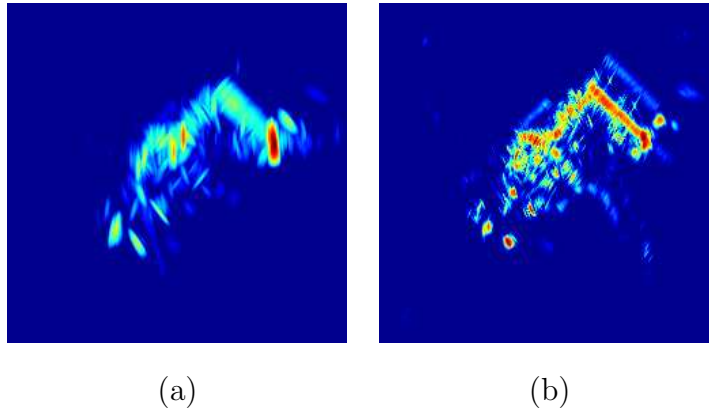
Figure 1.2: SAR image of the backhoe. (a) Conventional image. (b) Point-enhanced image.

1.2   Non-Quadratic Regularization-Based SAR Imaging

Recently developed processing techniques [3, 4, 1] have the potential to enable SAR systems to produce high-quality images suitable for automated decision making systems even when the data are limited or corrupted. Among these techniques, we consider a non-quadratic regularization-based approach which aims to produce so-called "point-enhanced SAR images" [1].

Point-enhanced SAR image formation technique mainly suggests that the reflectivities in the scene have a sparse representation and incorporates this prior belief into the solution through a non-quadratic regularizer. With the inclusion of the sparsity constraint, only the most dominant scatterers are represented by non-zero components in the image. Figure 1.2 shows the image of the backhoe model obtained using the conventional method and the point-enhanced image formation algorithm. In Figure 1.2 (b), the dominant scatterers are enhanced and the components of the backhoe model are resolved better. The conventional image is actually a smoothed or unfocused version of the underlying scene and hence enhancing the point-based features reveals the details which are not observable in the conventional image.

1.3   Parameter Selection Problem

As mentioned before, point-enhanced SAR imaging is a non-quadratic regularization-based approach. As in other regularization-based image reconstruction problems, it requires selection of the regularization parameter. This parameter balances the

3

contribution of the data and the *a priori* constraints on the formed image. Small parameter provides that the estimate is well fitted to the data whereas large parameter ensures the dominance of the prior knowledge or constraints. Particularly in point-enhanced SAR imaging framework, the image obtained with a small regularization parameter would look like the conventional image. On the other hand, use of a large parameter would produce an image with a very few number of dominant scatterers. For this reason, the choice of the regularization parameter is very crucial. However this issue has not been addressed in previous work.

Selection of the regularization parameter has a great importance not only in point-enhanced SAR imaging but also in other regularization-based image reconstruction problems. Most parameter choice methods have been proposed to choose the regularization parameter in the Tikhonov method [5] which is a well-known and widely used quadratic regularization approach. There exist some methods which are based on statistical considerations such as the discrepancy principle [6], Stein's unbiased risk estimator (SURE) [7], the generalized cross-validation (GCV) [8, 9] and Bayesian methods [10], as well as the graphical tools such as the L-curve [11]. All these methods were initially developed to serve to the Tikhonov regularization. The quadratic form of the Tikhonov solution yields a linear optimization problem which simplifies the computation of the regularized solution and the regularization parameter.

As the advantage of sparse representation has been discovered in many different fields such as optical flow estimation [12], compressed sensing [13] and functional regression [14], constraints which impose sparsity have became more prevalent. It has been shown that a non-quadratic regularizer promotes sparsity in the solution [15]. However, inclusion of such a non-quadratic constraint yields a non-linear optimization problem for image formation. Unlike quadratic methods, in this case the selection of the regularization parameter is not straightforward. For the parameter choice in non-quadratic regularization-based techniques, the application of SURE, GCV and L-curve is limited [16, 17, 10]. Especially for the form of our problem which considers an ℓ_p-norm penalty with $p \leq 1$ for complex-valued inverse problems, the use and effectiveness of these methods have not been explored yet.

4

1.4   Contribution of this Thesis

The first contribution of this thesis is the formulation of the parameter selection methods SURE and GCV for the point-enhanced SAR imaging framework. This framework provides a generalized form of previously developed parameter estimation methods. It is also an extension for complex-valued image reconstruction problems. We have used SURE, GCV and L-curve for the selection of the regularization parameter in point-enhanced SAR imaging. The second contribution of this thesis is the use of numerical tools for efficient implementation of the methods considered. We have used randomized trace estimation to compute the SURE and GCV function and golden section search to minimize them. These tools enable the application of memory-intensive methods to large-scale problems in SAR imaging.

1.5   Organization

This dissertation is organized as follows. In Chapter 2, we review the principles of SAR, regularization-based imaging, and parameter selection methods in general. We give a brief summary of quadratic and non-quadratic regularization methods, their advantages and shortcomings. We also explain several standard methods for parameter selection in general regularization problems. In Chapter 3, we formulate parameter selection methods for complex-valued, non-quadratic regularization-based SAR imaging. This chapter also explains the numerical optimization tools that we use for implementation of the applied methods. Experimental results are also provided in this chapter. Finally, Chapter 4 summarizes the results we have obtained, and suggests a number of topics to be considered in future research.

Chapter 2

Background

In this chapter, we give background information on SAR and feature-enhanced SAR imaging upon which this thesis is built. To gain a deeper understanding of point-enhanced imaging, regularization based image restoration and reconstruction is revised. Already existing methods for regularization parameter selection are also presented.

2.1 Synthetic Aperture Radar Imaging

Radar was originally developed for military purposes to measure the range to a target, i.e. scattering field. It was used to determine the locations of military vehicles such as ships, aircrafts, tanks, as well as track them. A radar system usually consists of an active sensor, i.e. radar antenna, operating in the radio or microwave region of the electromagnetic spectrum. This frequency band does not allow cloud, fog and rain effects to be observable in the radar image. All-weather, day and night imaging capability of radar is the reason of primary importance for radar to become popular for earth surface imaging tasks. However there is also the disadvantage that the resolution achievable with the operating wavelength is very poor. It is known that a better vision, i.e. better resolution, can be attained with a larger radar antenna. Unfortunately, constructing such a physically large antenna is not feasible. Similar to optical systems, resolution is proportional to the ratio between the radiation wavelength and the sensor antenna dimension. In particular, the cross-range resolution for radars is given by

$$\Delta x = \frac{R\lambda}{D} \tag{2.1}$$

where λ is the operating wavelength, R is the target range and D is the size of the antenna aperture. Let us assume a $D = 1$-meter diameter radar antenna operating at $\lambda = 1$ meter, and attached to an aircraft at a range $R = 1000$ meter. The cross-range resolution for this antenna is $\Delta x = 1000$ meter which is very poor. To reach a reasonable level of resolution like $\Delta x = 1$ meter; such that, e.g. vehicles in the scene can be distinguished from one another, a $D = 1000$-meter antenna would be required. Building an antenna aperture of this size is physically infeasible, but in fact it can be reachable by means of a synthetic aperture. This kind of a system is called *synthetic aperture radar* (SAR).

The principle idea behind SAR is to synthesize the effect of a large-aperture physical radar using multiple observations from a small antenna. Both amplitude and phase of the received signal must be recorded to synthesize the receiving antenna, however, in fact the measured data are quite unfocused and seems much like a random noise. These data are called *phase history* since the essential information is hidden in the phase of the received signal and phase-sensitive processing is needed to focus the image. In the early years of the SAR technologies, optical SAR processors were used to produce a well-focused image but optical processing was so demanding that digital SAR processors have emerged and replaced optical ones. By means of developing digital SAR processor algorithms SAR became prominent with its ability to produce high resolution images.

A SAR imaging system can operate mainly in two distinct modes: stripmap-mode SAR and spotlight-mode SAR. In stripmap-mode SAR, the pointing direction of the antenna is fixed so that a continous strip of ground is imaged. On the other hand, the antenna is steered and illuminates a single spot on the ground in the spotlight-mode SAR. Spotlight-mode SAR is more useful if a high resolution image of a limited area is desirable since it simulates a wider antenna. We will be focusing on this type of SAR throughout this thesis.

The data collection geometry for spotlight-mode SAR is demonstrated in Figure 2.1. The $x - y$ coordinate system (denoting range and azimuth coordinates respectively) is centered on a relatively small patch of ground illuminated by a narrow RF beam from the moving radar. As the aircraft moves along the synthetic aperture, the radar beam is steered such that the spotlighted target area is fixed as shown
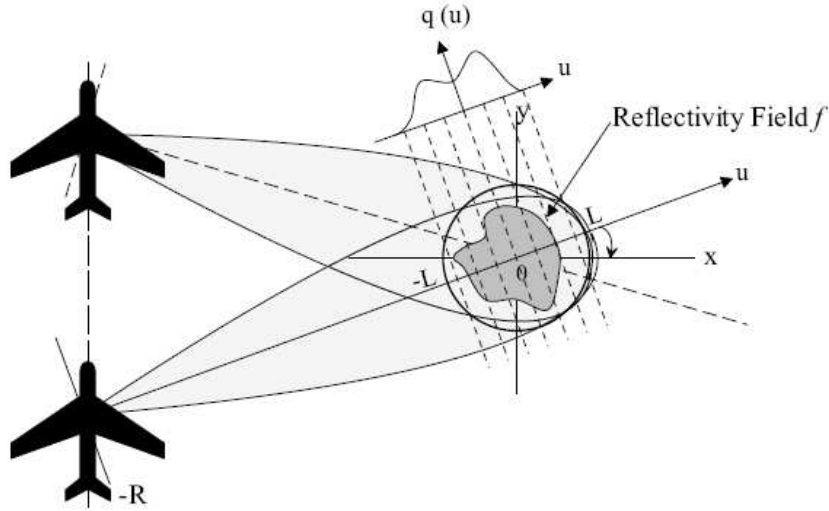
Figure 2.1: Ground-plane geometry for data collection in spotlight-mode SAR. (This figure is taken from [1].)

in Figure 2.1. At points corresponding to equal increments of θ (the angle between the x-axis and u-axis in Figure 2.1), high-bandwidth pulses are transmitted to the ground patch and echoes are then received and processed. At each observation point, received SAR signals are demodulated. After some pre-processing and certain approximations, the observed signal is related to a particular projectional view of the underlying scene. Then the relationship between the field $f(x, y)$ and the observed signal $G_\theta(t)$ is given by:

$$G_\theta(t) = \int\int_{x^2+y^2 \leq L^2} f(x, y) \, exp\left\{-j\Omega(t)(xcos\theta + ysin\theta)\right\} dxdy \qquad (2.2)$$

where L is the radius of the ground region of interest, as shown in Figure 2.1, and Ω denotes the radial spatial frequency. (The structure of Ω and other details on the pre-processing of the radar signals are given in [1].) The observed signal $G_\theta(t)$ can be interpreted as a slice through the 2-D Fourier transform of the scene $f(x, y)$. As a consequence of this, the standard image formation algorithm is the polar format algorithm [18] based on the two-dimensional fast Fourier transform (FFT). It is not convenient to compute approximate samples of $f(x, y)$ directly from polar samples of its Fourier transform. Therefore, the known data samples are first interpolated to a Cartesian grid, assuming unknown samples to be zero. After interpolation, an inverse 2-D FFT is applied and the magnitude of the reconstructed complex

8

image is displayed for viewing. Although this system is commonly used in SAR systems, it also involves some drawbacks. First of all, polar format algorithm is not robust to noisy or limited data. In addition, it does not increase the resolution or exploit the features or structures which we favour to see in a SAR image. To address these issues, a feature-enhanced synthetic aperture radar imaging technique has been developed [1]. This is a model-based and regularized image reconstruction technique which enables high-resolution SAR images with reduced artifacts. Before going into details of this approach, we will briefly repeat the notions of regularization and regularization-based image reconstruction.

2.2   Regularization-Based Image Reconstruction

The problem of image restoration and reconstruction is to recover a 2-D unknown field $f(x, y)$ given an indirect observation $g(x, y)$ of it. In many engineering problems, these observations can be accurately modeled by a linear transformation of the field of interest. Then, the relation between the data and the unknown field, i.e. the distortion model, can be expressed by a linear integral equation, in particular a Fredholm integral equation of the first kind:

$$g(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y; x', y') f(x', y') \, dx' dy' \tag{2.3}$$

where $h(x, y; x', y')$ is the point spread function (PSF) of the distortion model and assumed to be known in image restoration and reconstruction framework. We focus here on the discrete representation of this kind of relation in the presence of noise which is given by:

$$g = Hf + w \tag{2.4}$$

where g and f are the vectors representing the observation, and the unknown field, respectively, w is the noise and H is a matrix modelling the distortion. The problem that we consider in general is called a discrete ill-posed problem since it is the discretization of the Fredholm integral of the first kind in (2.3), which itself involves an ill-posed problem.

The notion of ill-posedness was first introduced by Hadamard [19]. According to his definition a problem is ill-posed if it violates any of the following:

1. for each g there exist a solution f;

2. the solution f is unique; and

3. the solution is stable with respect to perturbations in g.

The existence of the solution can be ensured through least squares solution:

$$\hat{f}_{ls} = \text{argmin}_f \|g - Hf\|_2^2 \tag{2.5}$$

where $\|.\|_2$ represents the ℓ_2-norm. Although least squares approach guarantees that the first condition is satisfied, it does not necessarily end up with a unique solution when the null space of H is nonempty, i.e. $\text{N}(H) \neq \emptyset$. To obtain a unique solution it is common to choose the one with the minimum size among all possible solutions. Generally, ℓ_2-norm is used to measure the *size* of the solution. Then, the minimum norm solution $\hat{f}_{mn}$ is defined as:

$$\hat{f}_{mn} = \text{argmin}_f \|f\|_2^2 \text{ s.t. } \min \|g - Hf\|_2^2 \tag{2.6}$$

When the data are noisy, the generalized solution tries to fit the solution to these noisy observations and since H is ill-conditioned the noise components of the data are extremely amplified in the generalized solution. To be able to obtain a stable solution, one common way is to employ *regularization*. The purpose of regularization is to single out a useful and stable solution by incorporating prior information about the desired solution. This prior information is imposed in different forms depending what we have or believe a priori and leads to different regularization methods. In the following sections, we will explain the details of Tikhonov regularization and non-quadratic regularization.

2.2.1 Tikhonov Regularization

Tikhonov regularization [20, 5] is probably the most widely used regularization method. The key idea in Tikhonov method is to incorporate a priori assumptions about the size and smoothness of the desired solution, in the form of the (semi)norm $\|Lf\|_2^2$. Tikhonov regularization in general form leads to the minimization problem:

$$\hat{f}_{Tikh} = \text{argmin}_f \|g - Hf\|_2^2 + \lambda \|Lf\|_2^2 \tag{2.7}$$

where the regularization parameter λ controls the weight given to minimization of the regularization term, relative to the minimization of the residual norm.

The simplest choice for L is the identity matrix. Then, the second term in (2.7) simply becomes the ℓ_2-norm of the solution f. Such a regularizer puts penalty on the magnitude of the solution and prevents the size of the solution from being too large. However this may not be the best choice in many applications. Most of the time, it is more useful to choose the regularization operator L as an approximation to the 2-D derivative (gradient) operator so that the regularizer is the measure of the smoothness of the solution. In such a case, the gradient of the solution is penalized instead of the solution itself and thus, large gradients, i.e. brightness changes in the image, are suppressed. These large gradients may result either from the noise components or the edges in the image. Therefore, while suppressing the effect of noise, this method also reduce the gradients which originally exist in the image and blur the edges.

To minimize the Tikhonov cost, we take the gradient of (2.7) with respect to f and set it equal to zero; and thus we obtain the following set of linear equations for the Tikhonov solution $\hat{f}_{tikh}$:

$$\left(H^T H + \lambda L^T L\right) \hat{f}_{Tikh} = H^T g \tag{2.8}$$

If the null spaces of H and L intersect trivally, i.e. if $N(H) \cap N(L) = \{0\}$, then the Tikhonov solution $\hat{f}_{Tikh}$ is unique. Note that here for $\lambda = 0$, the Tikhonov solution is nothing but the least squares solution.

2.2.2 Non-Quadratic Regularization

The aim of the standard Tikhonov method was to include a quadratic side constraint $\|Lf\|_2^2$ to the least squares criterion. Using such a quadratic regularization term leads to the linear problem (2.8) and thus the solution becomes a linear function of the data and can be computed easily. Despite its computational advantage, quadratic regularizers are limiting in the sense that they do not take into account the outliers in the true image while suppressing noise components. To be able to obtain more effective results, non-quadratic constraints can be incorporated. For this reason, we consider more general problems of the following form:

$$\hat{f}_{NQ} = \operatorname{argmin}_f \|g - Hf\|_2^2 + \lambda \sum_{i=1}^{n} \Psi\left((Lf)_i\right) \tag{2.9}$$

11

where n is the length of the vector Lf, and $(Lf)_i$ denotes its i-th element.

The form of the employed regularizer depends on the characteristics of the field of interest and the purpose of imaging. Regarding the regularization term, there is a special attention on the ℓ_1-norm in some applications, such as image deblurring, where this norm preserves the edges. Sparsity is an important feature that occurs naturally in some image types, e.g., molecules and star fields. The emerging field of compressed sensing [13] uses sparsity in the parameter vector and similarly it has been used in optical flow estimation [12]. The ℓ_p-norm penalty, where $p < 2$ on the image values is known to increase sparsity in the estimate. As the value of p gets smaller, the relative penalty on large values of the function reduces. Comparing to quadratic constraints, ℓ_1-norm puts a smaller penalty on large values of the components of the vector whose norm is being computed. As a results of this behavior, it has the capability of producing sparse images if L is chosen as identity [14, 15, 21], or images with preserved edges if L is a derivative operator [22]. Point-enhanced SAR imaging is based on a similar reasoning, therefore we focus here on the problem with ℓ_p-norm regularizer:

$$\hat{f} = \text{argmin}_f \, \|g - Hf\|_2^2 + \lambda \sum_{i=1}^{n} ((Lf)_i)^p \tag{2.10}$$

where $0 < p < 1$.

While taking the gradient of the above cost function, we also have to consider the non-differentiability of the ℓ_p-norm around the origin. One way to deal with this issue is to smooth the function such that the regularizer has the form

$$\|Lf\|_p^p \approx \sum_{i=1}^{n} \left(|(Lf)_i|^2 + \beta \right)^{p/2} \tag{2.11}$$

where β is a small smoothing constant. Then, the estimate is the solution of the following set of equations:

$$\left(2H^T H + \lambda L^T W \left(\hat{f}, \beta \right) L \right) \hat{f} = 2H^T g \tag{2.12}$$

where the diagonal weight matrix $W \left(\hat{f}, \beta \right)$ is given as:

$$W \left(\hat{f}, \beta \right) = \text{diag} \left\{ \frac{p}{\left[|(Lf)_i|^2 + \beta \right]^{1 - \frac{p}{2}}} \right\}. \tag{2.13}$$

where diag $\{.\}$ is a diagonal matrix whose i-th diagonal element is given by the expression inside the brackets. Note that the only difference between (2.8) and (2.12) is the weight matrix $W\left(\hat{f},\beta\right)$ which makes the non-quadratic estimate adaptive to the underlying field. When $|(Lf)_i|^2$ is small, the weight goes to a large value, imposing greater penalty to the solution in these regions. When $|(Lf)_i|^2$ is large, the weight goes to a small value, allowing large values or large gradients at those points [23]. On the other hand, Tikhonov regularization does not consider such weighting and adds just a spatially invariant penalty term to the cost function.

Although non-quadratic regularization is a much more powerful method, it requires extra effort to compute the solution because of the non-linearity in (2.12). Fortunately, through a fixed point iteration [24, 25], each iteration of the non-linear problem turns out to be a linear problem:

$$\hat{f}^{(k+1)} = \left(2H^T H + \lambda L^T W\left(\hat{f}^{(k)},\beta\right) L\right)^{-1} 2H^T g \qquad (2.14)$$

where $\hat{f}^{(k)}$ is the solution obtained in the kth iteration.

In this section, we tried to highlight the superiority of non-quadratic methods compared to quadratic ones. In the following section, we explain how this kind of non-quadratic regularization methods can be used to produce point-enhanced SAR images.

2.3   Point-Enhanced SAR Imaging

In Section 2.1, we briefly explained the SAR observation geometry and the pre-processing of the collected SAR data. Now, the problem is to reconstruct a SAR image from this pre-processed SAR data. In this section, we focus on a regularization-based SAR imaging framework proposed by [1].

Conventional SAR image formation methods mostly do not have an explicit dependence on the SAR observation relation in (2.2). However, an explicit discrete model of the SAR sensor and observation geometry can provide some advantages given in [1]:

- It lets us handle limitations in data quantity more effectively. Examples of such limitations are angular diversity limitations (e.g. due to sensor re-tasking), resolution limitations, and missing observations.

- The model-based approach ties readily into the statistical processing methods. This lets us handle limitations in data quality more effectively, through a noisy observation model.

- When the particular data relationship deviates significantly from the Fourier model, a model-based approach can readily take this into account.

Starting from the expression in (2.2), if we discretize and interpolate the observed data, then we can represent this discrete model through the relation :

$$G = \tilde{H} F \tag{2.15}$$

where F represent the 2-D Fourier transform of f and $\tilde{H}$ defines a rectangular window in Fourier domain to which the polar data is interpolated. By using the property of Fourier transform which suggests that the transform of a product is equal to the convolution of the transform of the individual terms, the image produced by computing a 2-D inverse Fourier transform of the observed data can be written as:

$$g = h * f \tag{2.16}$$

where $*$ denotes two dimensional convolution, g and h is the 2-D inverse Fourier transform of G and $\tilde{H}$ respectively. Here, we call h the point spread function (PSF) of the system. To be consistent with the discussion in Section 2.2.2, we write this relation in vector matrix notation in the presence of noise w:

$$g = Hf + w \tag{2.17}$$

where H is a convolution matrix for the observation kernel h.

We have explained model-based, regularized image reconstruction in the previous section. With the same motivation, the SAR image reconstruction problem is defined as:

$$\hat{f} = \mathrm{argmin}_f J(f) \tag{2.18}$$

where $J(f)$ has the following form:

$$J(f) = \|g - Hf\|_2^2 + \lambda \Psi(f) \tag{2.19}$$

where $\Psi(f)$ is the side constraint. The first term in (2.19) is the data fidelity term and measures the squared error between the actual observations and the observations that would be obtained by passing the solution of the reconstruction problem

through the linear forward model. The second term $\Psi(f)$ brings in the prior information we would like to impose. The side constraint should be determined in such a way that it causes the artifacts to be suppressed and the desired features to be enhanced. We have mentioned that the standard Tikhonov solution cannot satisfy such objectives in general. Therefore, in regularized SAR imaging, constraints in a non-quadratic form are considered. One favoured choice for $\Psi(f)$ is $\|f\|_p^p$, where $\|.\|_p$ denotes the ℓ_p-norm. This regularizer is aimed at enhancing point-based features by imposing a constraint on the energy of the solution. The outcome of the use of this term is to suppress the artifacts and increase the resolvability of scatterers. A smaller value of p puts a smaller penalty on large pixel values as compared to a larger p, and thus produces a field with a smaller number of dominant scatterers, and results in better preservation of the scatterer magnitudes.

As in other regularization-based image reconstruction problems, the choice of the regularization parameter λ is again crucial. When the value of λ is too large, very few scatterers would be observable. Although the artifacts are mostly suppressed in such a result we cannot consider the reconstruction as satisfactory since a number of scatterers is most probably not observable. On the other hand, for a very small λ the solution would be dominated by the artifacts and the noise components which are present in the observed data would be amplified in the solution. To address the issue of parameter selection in feature-enhanced SAR imaging, we investigate some parameter selection methods and apply them to our problem.

2.4   Parameter Selection in Regularization-Based Imaging

As discussed in Section 2.2.1 and Section 2.2.2, the regularized image reconstruction framework involves a scalar parameter λ which is called a regularization parameter. Let us consider the general ℓ_p-norm solution in (2.10), which we repeat here for convenience:

$$\hat{f} = \text{argmin}_f \|g - Hf\|_2^2 + \lambda \sum_{i=1}^{n} ((Lf)_i)^p. \tag{2.20}$$

The first term in (2.20) is usually refered as the data fidelity term, whereas the second term imposes the prior knowledge about the unknown field. The parameter λ controls the tradeoff between these two terms. Small λ makes the residual norm $\|g - Hf\|_2^2$ become more dominant in the optimization problem (2.20), and therefore

the estimate is well fitted to the data. However, it is obvious that this is not a good idea when the noise level is high. On the other hand, when λ is large, the side constraint $\sum_{i=1}^{n} \left((Lf)_i \right)^p$ is dominant. If the data are relatively less noisy, then choosing a large λ means mostly relying on the prior information and sacrificing some useful data. In summary, the choice of λ is very crucial to obtain a meaningful estimate.

In this section, some methods for choosing the regularization parameter will be discussed: Stein's unbiased risk estimator and generalized cross-validation, both based on prediction error; and the L-curve, based on a plot of the residual norm versus the side constraint norm.

2.4.1 Stein's Unbiased Risk Estimation

The unbiased risk estimator method was concurrently and independently developed by Stein [7] and Mallow [26] for model selection in linear regression and thereafter adapted for the solution of inverse problems. It is usually called Stein's unbiased risk estimator (SURE) in the literature.

SURE mainly aims to minimize the following predictive risk, i.e. predictive mean squared error:

$$\left\| H\hat{f}_\lambda - Hf_{true} \right\|_2^2. \tag{2.21}$$

Here, $\hat{f}_\lambda$ denotes the solution obtained by using λ and f_{true} is the true, unknown field. Obviously, the predictive risk cannot be calculated exactly since it depends on f_{true} which is the unknown of the problem. However, Stein's method achieves an unbiased estimate of the predictive risk.

To explain Stein's method, let $\mu = Hf$ and consider the following observation model:

$$g = \mu + w \tag{2.22}$$

Let $\hat{\mu}$ be a nearly arbitrary estimate of μ and assume that it has the form $\hat{\mu} = g + e(g)$, where $e(.)$ is a weakly differentiable function. To measure the mismatch between $\hat{\mu}$ and μ, introduce the mean squared risk of $\hat{\mu}$:

$$R_\lambda = \|\hat{\mu} - \mu\|_2^2. \tag{2.23}$$

SURE provides the expected value of this risk as:

$$\hat{R}_\lambda = n\sigma^2 + \|e(g)\|_2^2 + 2\sigma^2 \nabla e(g) \qquad (2.24)$$

where $\nabla e(g) = \sum \partial e_i(g)/\partial g_i$. Thus, $\hat{R}_\lambda$ is an unbiased estimate of the risk of the nearly arbitrary estimate $g + e(g)$. Here, $e(g)$ is a measure for the fitness of the estimate $\hat{\mu}$ to the observation g, and usually called residual.

For standard Tikhonov solution, the computation of the gradient in (2.24) is straightforward since the regularized solution is linearly dependent on the data as is given in (2.8). However, when non-quadratic regularization methods are considered, a nonlinear relation arises between $\hat{f}$ and g. Holding for all cases, also for non-quadratic methods, it has been shown in [16] that the risk estimate takes the following form:

$$\hat{R}_\lambda = -n\sigma^2 + \|e\|_2^2 + 2\sigma^2 \mathrm{trace}\left(HJ_{\hat{f}\hat{f}}^{-1}J_{\hat{f}g}\right) \qquad (2.25)$$

where J is the objective function, $J_{\hat{f}\hat{f}}$ is the Hessian and $J_{\hat{f}g} = \partial^2 J/\partial\hat{f}\partial g^T$. Then, as far as σ^2 is known or accurately estimated, SURE parameter selection method is to pick:

$$\lambda_{sure} = \mathrm{argmin}_\lambda \hat{R}_\lambda. \qquad (2.26)$$

It is actually not clear that the λ which minimizes the predictive risk in (2.21) will yield a small value for the estimation error $\left\|\hat{f}_\lambda - f_{true}\right\|_2^2$. We can just make some interpretations such that minimizing the predictive risk minimizes the upper bound on the estimation risk, if the the columns of H are linearly independent.

2.4.2 Generalized Cross-Validation

As mentioned in the previous section, SURE requires the prior knowledge of the variance σ^2. On the other hand, the method of generalized cross-validation (GCV) [8, 9] provides an estimate for the λ which approximately minimizes the expected value of the predictive risk, without needing the variance σ^2.

This method, in fact, is the weighted version of the ordinary cross-validation [27, 28]. The ordinary cross-validation has some shortcomings in estimating λ, and thus GCV has been developed, see [9]. The idea of this method is as follows. Let $\hat{f}_\lambda^{(k)}$ be the estimate of f with the kth data point g_k, omitted. If λ is a good choice, then

the k-th component $\left[H\hat{f}_\lambda^{(k)}\right]_k$ of $H\hat{f}_\lambda^{(k)}$ should be a good predictor of g_k. Therefore, cross-validation estimate of λ is the minimizer of:

$$P_\lambda = \frac{1}{n}\sum_{k=1}^{n}\left(\left[H\hat{f}_\lambda^{(k)}\right]_k - g_k\right)^2. \tag{2.27}$$

The ordinary cross-validation function has the form

$$OCV_\lambda = \frac{1}{n}\sum_{k=1}^{n}\left(\frac{\left[H\hat{f}_\lambda\right]_k - g_k}{1 - a_{kk}}\right)^2. \tag{2.28}$$

where a_{ii} denotes the i-th diagonal element of the influence matrix A_λ which is defined to be:

$$H\hat{f}_\lambda = A_\lambda g. \tag{2.29}$$

Then, the weighted form of $P(\lambda)$ is obtained by replacing a_{ii} in (2.28) by the average of all diagonal elements:

$$V_\lambda = \frac{\frac{1}{n}\|e_\lambda\|_2^2}{\left[\frac{1}{n}\text{trace}(I - A_\lambda)\right]^2} \tag{2.30}$$

Note that GCV is a rotationally invariant extension of OCV.

The GCV method was originally designed for problems in which A_λ is independent of g. For more general regularization methods it was proposed in [29] to replace the denominator in (2.30) by $\left[\frac{1}{n}\text{trace}\left(I - HJ\left(\hat{f}\right)\right)\right]^2$, where $J\left(\hat{f}\right)$ is the Jacobian of $\hat{f}$ with respect to g. Note that $HJ\left(\hat{f}\right) = A_\lambda$ if $\hat{f}$ is linearly dependent on g. The GCV estimate for λ is:

$$\lambda_{gcv} = \text{argmin}_\lambda V_\lambda. \tag{2.31}$$

2.4.3 L-curve

L-curve was first defined as a parametric plot of the (semi)norm $\left\|L\hat{f}_\lambda\right\|_2$, versus the corresponding residual norm $\left\|H\hat{f}_\lambda - g\right\|_2$, with the regularization parameter λ as the parameter [11, 30]. The L-shaped corner of the L-curve appears for regularization parameters close to the optimal parameter that balances the regularization errors and perturbation errors in $\hat{f}_\lambda$. The L-curve criterion for choosing the regularization parameter is based on this feature.

The L-curve basically consists of a vertical part and an adjacent horizontal part as in Figure 2.2. It is important to plot the L-curve in $\log - \log$ scale in order to
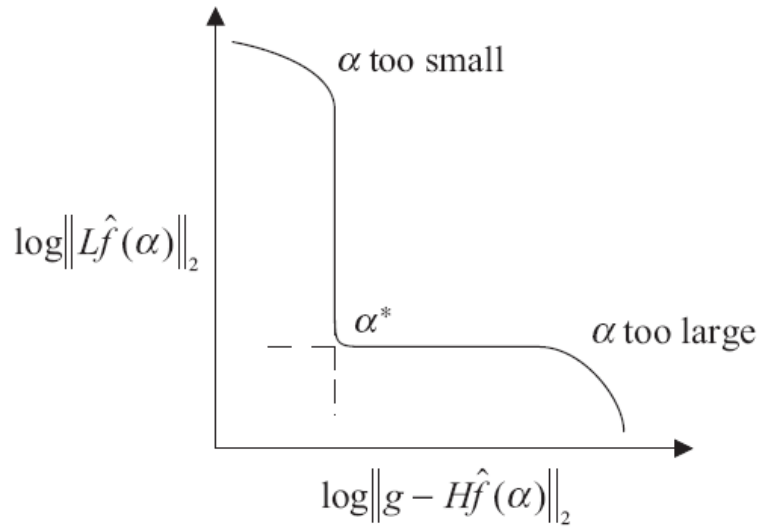
Figure 2.2: The generic form of the L-curve.

emphasize the two different parts of the curve. The horizontal part corresponds to oversmoothed solutions where the regularization parameter is too large and the solution is dominated by regularization errors. The vertical part corresponds to underregularized solutions where the regularization parameter is too small and the solution is dominated by perturbation errors. This behavior does not rely on any additional properties of the problem, such as the distortion model or the statistical distribution of the errors. Besides, L-curve does not need any prior knowledge of the noise.

As the regularization parameter λ increases, the solution norm corresponding the vertical axis becomes smaller and the residual norm corresponding the horizontal axis becomes larger. For relatively small λ values, the decrease in the solution norm is very steep however the increase in the residual is quite slight. On the other hand, for larger λ values, the solution norm decreases very small amount and the residual norm increases significantly. The corner of the L-curve should satisfy the balance between these two extremes. Although this intuition is natural and quite simple, computing the corner of the L-curve may not be so easy. Several ideas are applied to determine the corner including the point of maximum curvature, the point closest to a reference location, such as the origin, and the point of tangency with a line of slope -1.

As mentioned before, L-curve was first developed for choosing the regularization parameter in Tikhonov regularization. Then, it was extended to the case of general functions [31]. L-curve for general functions is nothing but the $\log - \log$ plot of those functions. This is quite natural since L-curve tries to balance the two terms involved in the regularization problem regardless of their structures.

Chapter 3

Parameter Selection in Point-Enhanced SAR Imaging

In this chapter, we expand all techniques mentioned in Section 2.4 to SAR imaging framework. Unlike usual image reconstruction problems, SAR involves complex-valued and random-phase reflectivities. For this reason, standard parameter selection methods should be modified and validated for this particular problem. In SAR imaging, to achieve the desired resolution for a wide observation range, one has to deal with large size images. For example, if we image a 10 m by 10 m field with a resolution of 0.3 m, we obtain a 256×256 representation. This size may seem reasonable for many image processing problems, however it is quite problematic for our task. Problem size is an important limitation especially for SURE and GCV. To be more explicit, for the same example, the observation is modelled with a 65536×65536 convolution matrix. To find the regularized solution and the optimum regularization parameter, we need to execute some operations including taking the inverse for matrices of that size. Most of the time, these operations are held by means of singular value decomposition (SVD). For large-scale problems, however, neither to construct the system matrix nor to find its SVD is easy . To avoid these obstacles, all the involved matrix vector products are actually carried out by convolution operations (in the Fourier domain) such that there is no need to construct the convolution matrix and deal with memory-intensive matrix operations. Unfortunately, this strategy brings another difficulty when computing the trace term in SURE and GCV.

In this chapter, we first formulate SURE and GCV functions for point-enhanced SAR imaging problem. Then, we explain the numerical tools used for the evaluation and minimization of the parameter selection methods. Finally, we present exper-

imental results based on synthetic and real SAR data to demonstrate and discuss the effectiveness of these methods.

Throughout this chapter, we consider the objective function in (2.10) with $L = I$, which we repeat here for convenience:

$$J = \|g - Hf\|_2^2 + \lambda \sum_{i=1}^{n} \left(|f_i|^2 + \beta\right)^{p/2} \tag{3.1}$$

where $p \leq 1$.

3.1   SURE for Point-Enhanced SAR Imaging

Starting point that we use is the SURE function in (2.25). Evaluation of the first two terms is straightforward, therefore we directly focus on the trace term, i.e. $\text{trace}\left(HJ_{\hat{f}\hat{f}}^{-1}2H^T\right)$. From now on, let us call $HJ_{\hat{f}\hat{f}}^{-1}2H^T = T_\lambda$. The Hessian of (3.1) with respect to f is:

$$J_{\hat{f}\hat{f}} = 2H^T H + \lambda K\left(\hat{f}_\lambda, \beta\right) \tag{3.2}$$

where $K\left(\hat{f}_\lambda, \beta\right)$ is given by:

$$K\left(\hat{f}_\lambda, \beta\right) = \text{diag}\left[p\left(|f_i|^2 + \beta\right)^{\frac{p}{2}-2}\left((p-1)|f_i|^2 + \beta\right)\right]. \tag{3.3}$$

Note that for $p = 2$, $T_\lambda g = H\hat{f}_\lambda$, i.e. T_λ is equal to the influence matrix A_λ mentioned in Section 2.4.2. However, for any other choice of p, T_λ has a different structure. When we substitute the above expression in (2.25), we obtain the SURE function as:

$$\hat{R}_\lambda = -n\sigma^2 + \left\|H\hat{f}_\lambda - g\right\|_2^2 + 2\sigma^2\text{trace}\left(T_\lambda\right) \tag{3.4}$$

where

$$T_\lambda = H\left(2H^T H + \lambda K\left(\hat{f}_\lambda, \beta\right)\right)^{-1}2H^T. \tag{3.5}$$

Now we need to calculate the trace term. As we have mentioned before, we define the PSF of the system and obtain the regularized solution by means of convolution. Since we do not really construct T_λ, it does not seem possible to find its trace directly.

At this point, we seek help from another technique called *randomized trace estimation* to be able to estimate the trace of T_λ. We use the trace estimation approach

also in GCV computation, and therefore the details of this technique will be explained in Section 3.4.1, after we discuss the GCV method for point-enhanced SAR imaging problem.

3.2 GCV for Point-Enhanced SAR Imaging

In Section 2.4.2, we mentioned that the Jacobian of the regularized solution $\hat{f}_\lambda$ with respect to g can be employed instead of A_λ when A_λ is not independent of g. This modification leads up to the same expression that we derived for SURE. In other words, we again replace $\hat{f}_g$ by $J_{\hat{f}\hat{f}}^{-1} J_{\hat{f}g}$ (Here, we drop the λ subscript of $\hat{f}$ for the sake of a simpler notation). This produces the same T_λ stated in (3.5). Then, the GCV function can be expressed as follows:

$$V_\lambda = \frac{\frac{1}{n} \left\| H\hat{f}_\lambda - g \right\|_2^2}{\left[\frac{1}{n} trace(I - T_\lambda) \right]^2}. \tag{3.6}$$

3.3 L-curve for Point-Enhanced SAR Imaging

We can directly use the standard L-curve method in our problem, since it does not require computation of any additional quantities. We just constitute the log-log plot of the data fidelity term and the regularizer in (3.1). It is quite natural to use L-curve for any functions regardless of their structures because it simply tries to set a balance between those functions. On the other hand, there is still the issue of locating the corner of the L-curve. In our experiments, we choose the point with a tangent line of slope -1 and positive curvature as the corner of the L-curve.

3.4 Numerical Optimization Tools

3.4.1 Randomized Trace Estimation

As mentioned in Section 3.1 and 3.2, SURE and GCV methods require the trace of the matrix T_λ. For large scale problems, T_λ can not be easily constructed due to the memory limitations of computers. In such cases, it is more convenient to find an estimate of the trace instead of making the exact calculation.

This method has been developed to calculate an estimate of the trace of the influence matrix involved in regularization of linear equations and aimed to enable the

use of the GCV method for choosing λ in large-scale problems [32]. The influence matrix A_λ for standard Tikhonov solution is defined in (2.29). In quadratic regularization methods, the solution depends linearly on the data, and hence the influence matrix is independent of the solution. Since this is not the case for non-quadratic regularization problems, an approximation of the influence matrix is used as given in (3.5). Here we compute the estimate of this approximated influence matrix T_λ. To use trace estimation, the matrix has to be symmetric, and this assumption holds for T_λ.

If Q is a white noise vector with zero mean and unit variance, then an unbiased estimator for trace (T_λ) is the random variable:

$$t(\lambda) = Q^T T_\lambda Q. \tag{3.7}$$

The method can be applied through the following algorithm:

- generate K independent realizations of q_i of Q,

- compute $t_i(\lambda) = q_i^T T_\lambda q_i$, and then

- take the sample mean $\bar{t}(\lambda) = \sum_{1=1}^{K} t_i(\lambda)/K$ to be the trace estimate.

The accuracy of this estimate depends on the variability of the $t_i(\lambda)$'s, and this variability can be quatified in terms of the variance of $t(\lambda)$. It is shown that this variance is minimized by taking Q to be a random vector whose components are independent and take values +1 and -1, with probability 1/2 [33].

To simulate such a random vector, we first generate a realization z of a random vector Z whose components are independent and uniformly distributed on the interval $[0, 1]$. Then we take q to have components $q_i = +1$ if $z_i \geq 1/2$, and $q_i = -1$ if $z_i \leq 1/2$.

This algorithm has an explicit dependence to the matrix T_λ. However, here we do not construct T_λ. Note that the matrix H in (3.5) is a convolution operator. Thus, in practice, whenever a vector matrix product appears, we perform convolution. In Section 3.5, we will investigate further properties of this method utilizing our experimental results.
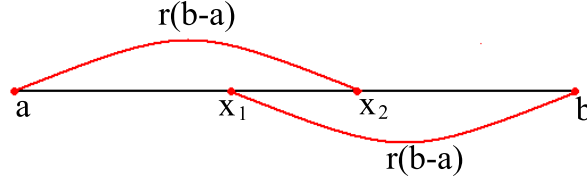
Figure 3.1: Location of x_1 and x_2 for Golden section search.

3.4.2 Golden Section Search

Golden section search is a derivative-free optimization method for unimodal functions [34]. If the function $f(x)$ is unimodal in an interval $[a, b]$, then $f(x)$ will have only one local minimum on $[a, b]$. As we know that the optimal solution is in the interval $[a, b]$, we can reduce the size of the interval by evaluating $f(x)$ at two points on $[a, b]$. The search goes as follows:

Begin with two test points x_1 and x_2 on $[a, b]$ such that $x_1 < x_2$. Evaluate $f(x_1)$ and $f(x_2)$ and shrink the interval according to the following rules:

- If $f(x_1) \leq f(x_2)$, then the new interval becomes $[a, x_2]$.

- If $f(x_1) > f(x_2)$, then the new interval becomes $[x_1, b]$.

Determine two new test points on the new interval and continue to shrink interval until it is sufficiently small. The new test points are set so that they divide the interval into the Golden section. Golden section requires:

$$\frac{\text{length of the whole interval}}{\text{length of the larger part of the interval}} = \frac{\text{length of the larger part of the interval}}{\text{length of the smaller part of the interval}}.$$

This relation results in the quadratic equation $r^2 + r = 1$. The positive root of this equation is $r = 0.618$ and used to determine the new test points x_1 and x_2, as shown in Figure 3.1.

SURE and GCV functions obtained for point-enhanced SAR imaging are unimodal in general. However we have also observed that for very small regularization parameters SURE cost has some local minima around that region. These minima probably arise due to the large noise amplification in the regularized solution for

25

very small parameters. We note that these minima occur at very small parameters which cannot produce a reasonable solution. Therefore, if the initial interval is selected properly, then the SURE function will be unimodal in that interval and the assumption will hold for the golden section search. In our experiments, we use the interval $[10^{-4}, 10]$ for the regularization parameter λ which covers all the solutions ranging from under-regularized to over-regularized, and hence definitely contains the global minimum.

3.5 Experimental Results

We demonstrate the effectiveness of the applied methods on synthetic and real SAR scenes. We present point-enhanced SAR images with selected parameters and compare these results to different parameter choices and conventional reconstructions. The real SAR data used in the following sections provide formed imagery only however we need to know the true scene to validate the performances of the applied methods. Thus, we generate a synthetic scene to test the methods first.

3.5.1 Synthetic Scene Reconstructions

The synthetic scene consists of 15 point scatterers as shown in Figure 3.2(a). Figure 3.2(b) and (c) shows the PSF of the SAR imaging system defined in (2.16) and the conventional SAR image of the synthetic scene. The PSF is a 2-D sinc function which comes from the 2-D Fourier transform of the rectangular window and defined as h in 2.16. Thus, the conventional image is a filtered or smoothed version of the true scene. We perform the experiments for two different noise level. We add complex Gaussian noise to the simulated SAR image, so that the signal-to-noise ratio (SNR) is 30, 20 and 10 decibels (dB). We take the SNR[1] to be the variance ratio of the noise-free data to noise in dB. Throughout our work, we display the magnitude (in dB) of the complex-valued reflectivities.

We first show some experimental results to measure the accuracy of the randomized trace estimation method. To validate this method we need to know the true trace value of T_λ and therefore we have to construct the approximated influence

---

[1]$\mathrm{SNR}\,(\mathrm{dB}) = 10 \log_{10} \left[\mathrm{Var}(Hf)/\mathrm{Var}(w) \right]$

(a)   (b)   (c)
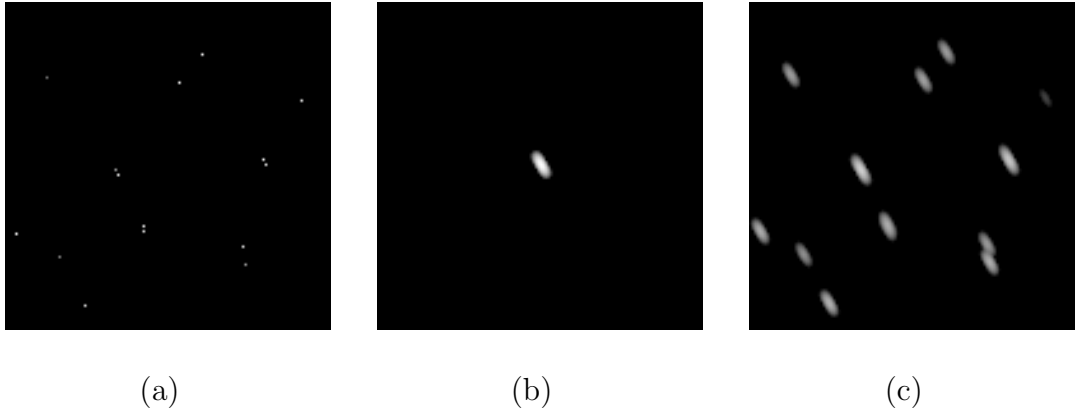
Figure 3.2: The grayscale plot of the magnitude of the (a) 128×128 synthetic scene, (b) PSF and (c) conventional SAR image.

matrix T_λ. However it is not easy to construct T_λ for large-scale problems. For this reason, we generate a 32×32 synthetic scene which is reasonably small. We first construct the convolution matrix for the PSF and then the influence matrix T_λ as is given by (3.5). Trace of this matrix is the true trace value. On the other hand, we use the randomized trace estimation method and compute the estimated trace value for the same influence matrix. Figure 3.3(a) and (b) show the estimated and true trace values for $p = 1$ and $p = 0.8$, respectively.

As mentioned in Section 3.4.1, randomized trace estimator is defined as the mean of the random variable $t(\lambda)$. To find a reliable estimate with small variance we should consider sufficient number of realizations. The ratio between the standard deviation and the mean of the random variable $t(\lambda)$ for different number of realizations of Q is given in Figure 3.4. It is shown that running K independent realizations reduces the variance by a factor of K^{-1} [33]. Making use of this graph and as a result of our experiments we conclude that 10 realization is a reasonable number to compute and sufficient to serve our purpose.

Now we present the parameter selection results for this synthetic problem. Since we know the underlying scene, we can compare SURE and GCV estimates to the true value of the predictive risk given by (2.21). We first consider conventional SAR image with 30 dB SNR. There are two determining terms for SURE and GCV costs: the regularized residual and the trace of the influence matrix. Figure 3.5 shows the trace and the residual terms for varying regularization parameter λ. As λ increases,

27
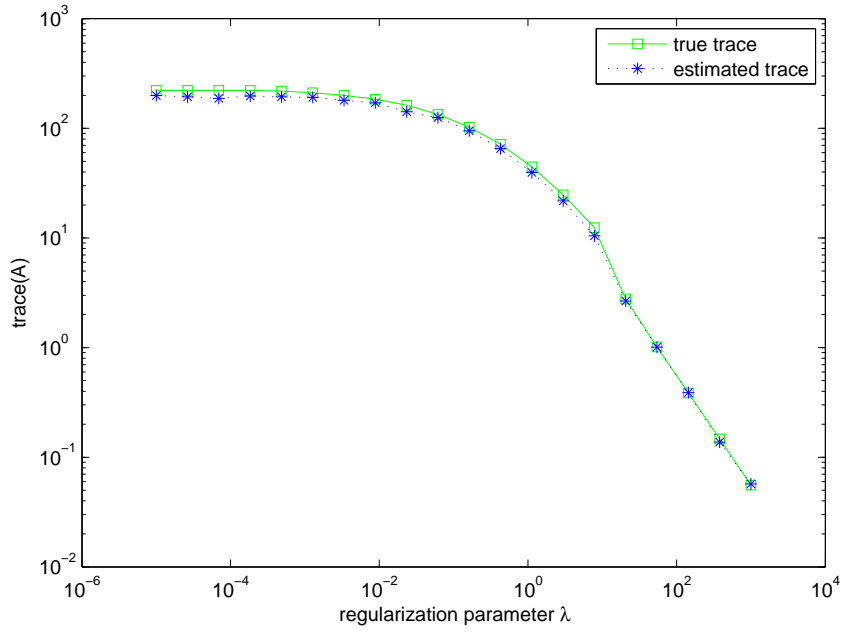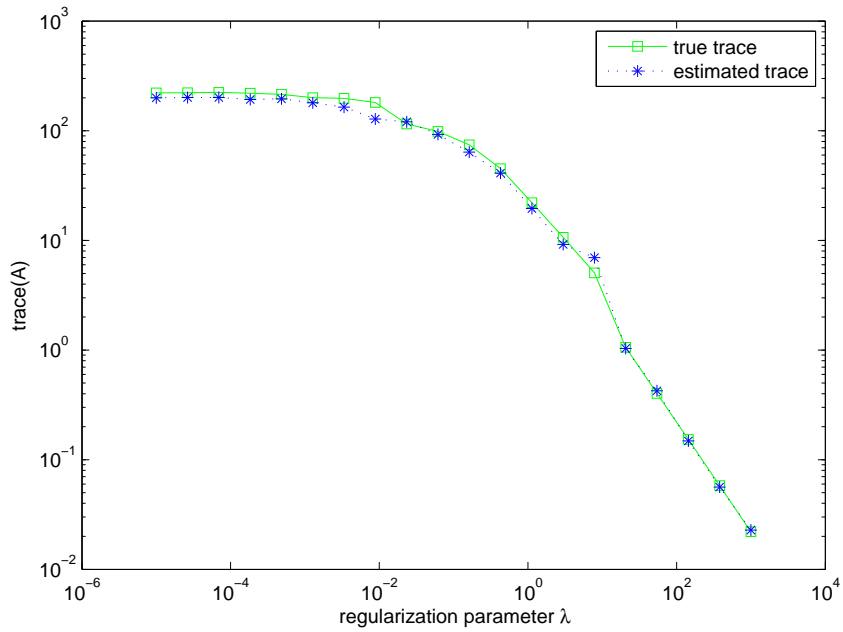
Figure 3.3: True and estimated trace values of T_λ. (a) p=1. (b) p=0.8.
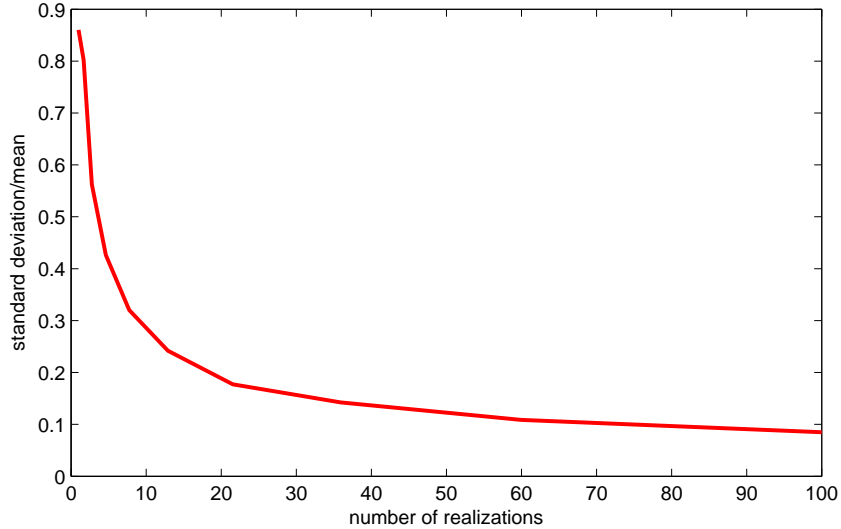
Figure 3.4: The ratio between the standard deviation and the mean of trace (A) for different number of realizations.

the residual become larger since the solution fits less to the data, and the value of the trace term decreases since λ and the solution norm get larger.

In our experiments, we consider the interval $[10^{-4}, 10^1]$ for λ and divide this interval in 12 constant grids in logarithmic scale . Then, we calculate SURE and GCV costs for each of these λ values. Figure 3.6 shows the true predictive risk, SURE and GCV costs in this interval. As seen from the figure, both SURE and GCV have their minima at the same λ which minimizes the true predictive risk. Stein's estimator is an unbiased estimator, GCV function, on the other hand, takes distinct values and very flat around its minimum. Still, all of them attain the same minimizers. Now, let us have a closer look around the minima. In Figure 3.7, we consider a smaller interval to see whether they have the same minima in a larger scale. We observe that SURE and GCV have their minima at a smaller λ than the true predictive risk. This leads an under-regularized solution but the difference between these solutions is not realised visually. Despite this slight under-regularization effect, SURE and GCV work well for this complex-valued, non-quadratic regularization-based imaging problem. To see whether the selected λ produces a visually satisfying reconstruction, we compare it to the solutions obtained by using other choices of regularization parameters. These results are presented in Figure 3.8.

SURE and GCV functions are minimized at $\lambda = 1.9 \times 10^{-2}$. This λ value achieves
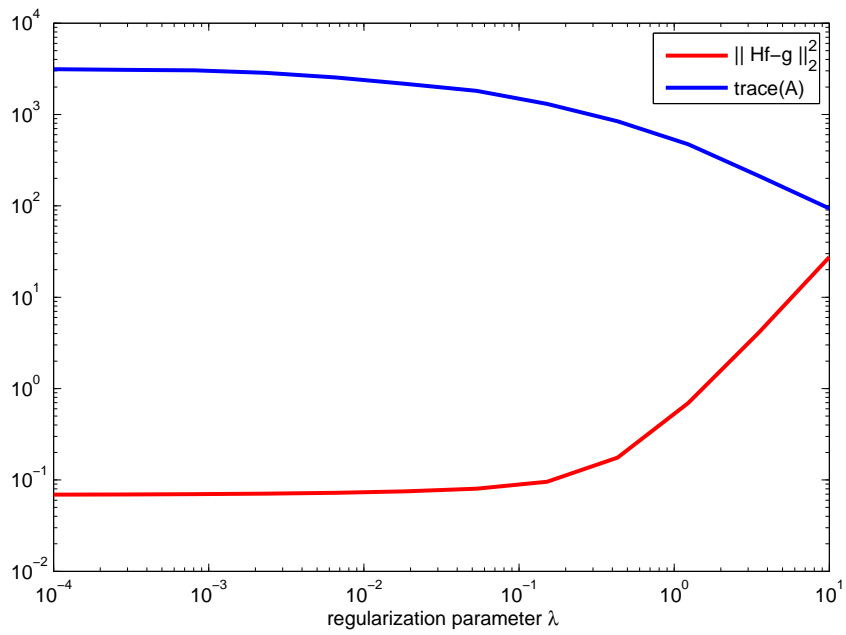
Figure 3.5: The residual $\left\|H\hat{f}_\lambda - g\right\|_2^2$ and trace(A) for 30 dB SNR data and p=1.

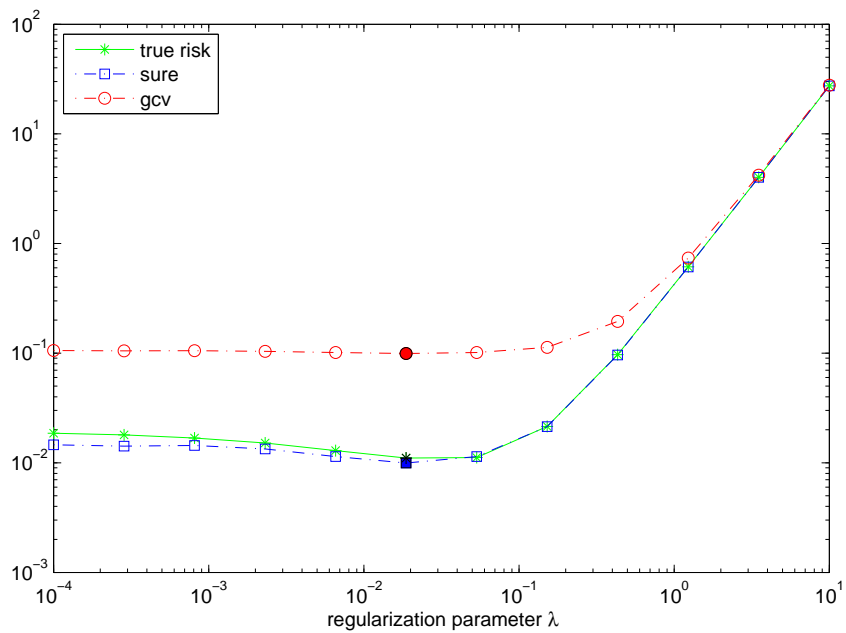

Figure 3.6: True predictive error, SURE and GCV results for 30 dB SNR data and p=1. The minimum of the true predictive risk is specified with the black asteriks and the minimum values of the SURE and GCV functions are marked with full markers.
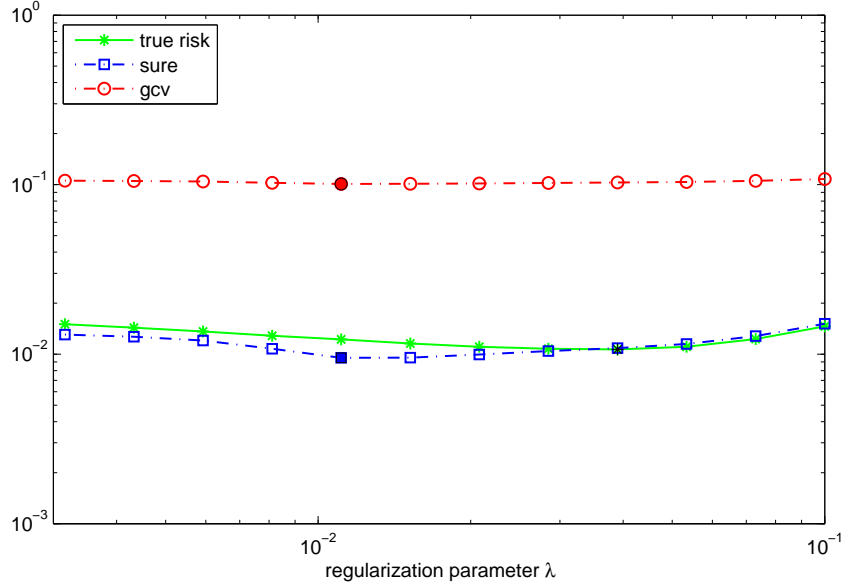
Figure 3.7: A closer look to the true predictive error, SURE and GCV results for 30 dB SNR data and p=1.

to produce a good point-enhanced image with very little artifacts. However, in fact some of the reconstructions for larger λ values seem even better. This case probably arises because both methods are trying to minimize the predictive risk instead of the solution error, i.e $\left\|\hat{f}_\lambda - f_{true}\right\|_2^2$. For this example, since the true scene is known, we can compute the solution error and compare it to the predictive risk. As it can be seen in Figure 3.9, the predictive risk and the solution error exhibit different behaviour. The curves are quite different and not minimized at the same λ. For this reason, SURE and GCV methods cannot guarantee that the selected parameter would produce the minimum solution error even if the predictive risk is estimated precisely. In this particular example, $\lambda = 1.5 \times 10^{-1}$ is the best solution regarding the solution error and visual inspection, however SURE and GCV select the optimum regularization parameter as $\lambda = 1.9 \times 10^{-2}$ since they aim to estimate the predictive risk.

We now show the results of another parameter selection method L-curve in Figure 3.10. To be able to compare to the solution error and other methods, we consider the same interval and grids for L-curve. The corner of the L-curve is selected as the point of tangency with a line of slope closest to -1. To do this, we first compute the numerical derivative of the L-curve at each λ and select the one with slope -1 and positive curvature as the corner of the L-curve. The corresponding λ equals to
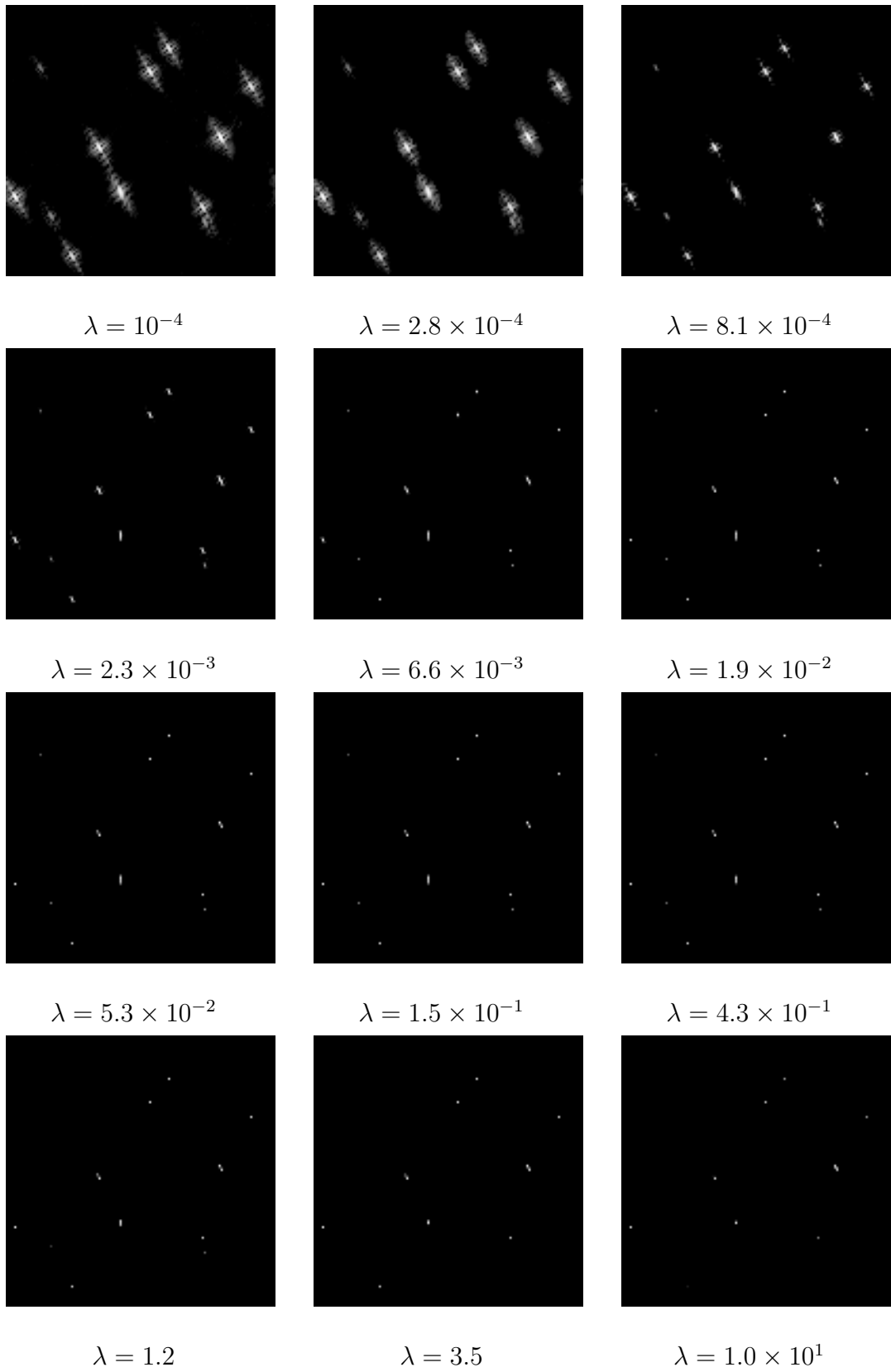
Figure 3.8: Point-enhanced reconstructions from the conventional image with 30 dB SNR and for p=1
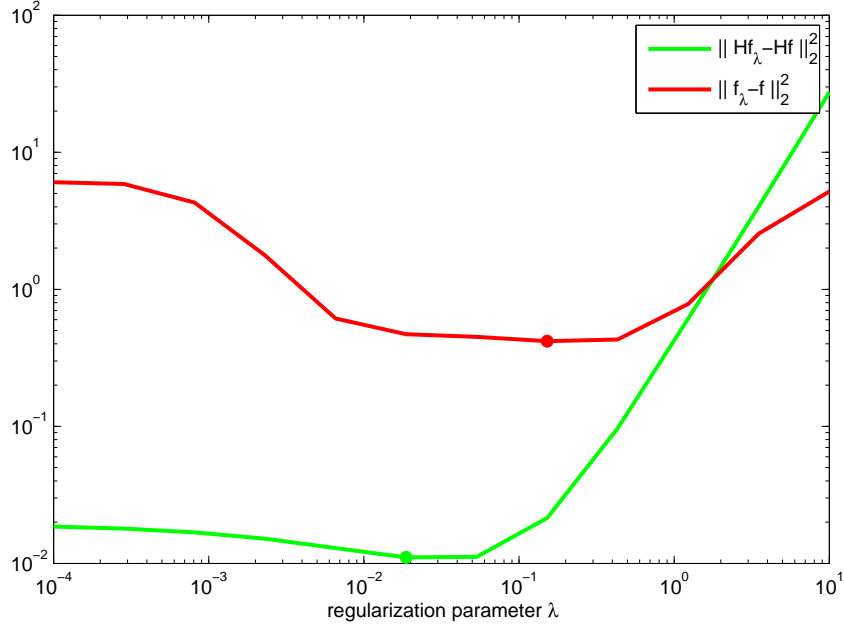
Figure 3.9: Mean squared error and the predictive risk of the solution for the reconstruction in Figure (3.8).

	SURE	GCV	L-curve
30 dB	0.0107	0.0107	0.1270
20 dB	0.1620	0.1620	0.3511
10 dB	0.5844	0.4574	1.3879

Table 3.1: Selected parameters for the synthetic scene when p=1.

5.3×10^{-2}, which in this case is the parameter value that minimizes the solution error.

Up to now, we showed the general behaviour of the SURE, GCV and L-curve in a particular interval. However, bruteforce searching with a reasonably small number of λ values provides just a rough parameter choice since the interval is quite large. For example, to choose the minimizer with an error of ± 0.1 in logarithmic scale, 50 reconstructions should be computed. On the other hand, when we use the Golden section search, we observe that this number is not more than 20 to obtain the same sensitivity. For these reasons, we give the minimizing λ values for 20 dB and 10 dB SNR data in Table 3.1 and 3.2. When the noise level is higher, then SURE, GCV
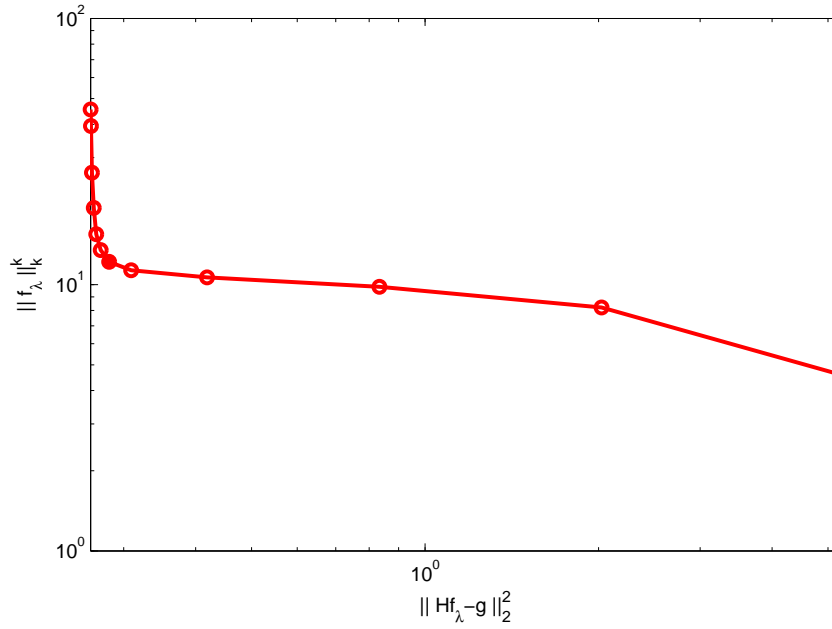
Figure 3.10: The L-curve for the reconstructions in Figure (3.8).

	SURE	GCV	L-curve
30 dB	0.0449	0.0630	0.1520
20 dB	0.1620	0.1620	0.3545
10 dB	0.3379	0.4574	1.0560

Table 3.2: Selected parameters for the synthetic scene when p=0.8.

Figure 3.11: Conventional image for Slicy target.

	SURE	GCV	L-curve
30 dB	0.0008	0.0015	0.0045
20 dB	0.0107	0.0117	0.0480
10 dB	0.0733	0.1196	0.3560

Table 3.3: Selected parameters for Slicy target when p=1.

and L-curve select a larger parameter, as expected. We note that SURE and GCV give almost the same results while L-curve tends to choose a larger λ in general.

3.5.2 MSTAR Data Reconstructions

We now show results on data from the Moving and Stationary Target Acquisition and Recognition (MSTAR) public target data set [35]. The *slicy* target of the MSTAR data set is a precisely designed and machined engineering test target containing standard radar reflector primitive shapes such as flat plates, dihedrals, trihedrals, and top hats. The conventional image of the slicy target is shown in Figure 3.11. We add complex Gaussian noise such that SNR is 30 dB, 20 dB and 10 dB.

Similar to the synthetic example, the selected parameters and the correspond-

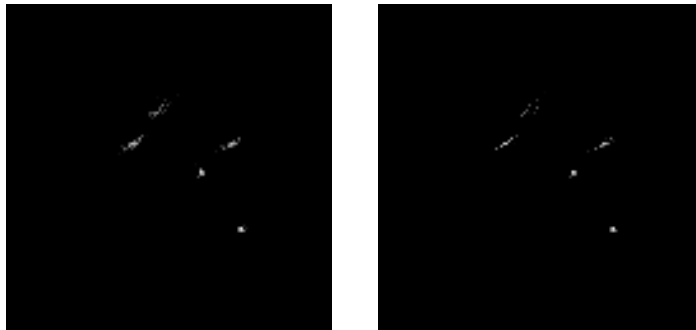|	SURE	GCV	L-curve
30 dB	0.0006	0.0008	0.0038
20 dB	0.0042	0.0048	0.0355
10 dB	0.0574	0.0574	0.3480

Table 3.4: Selected parameters for Slicy target when p=0.8.
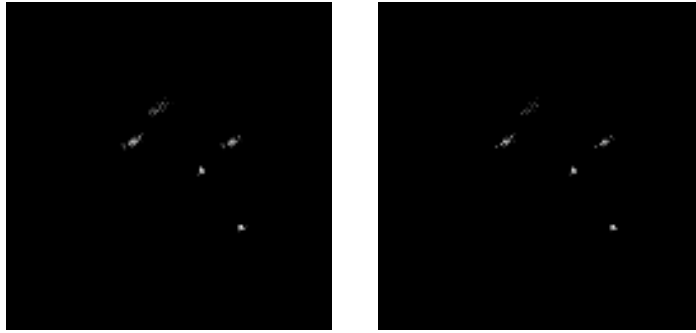
(a)   (b)

Figure 3.12: Point-enhanced images for Slicy target when p=1. (a) Parameter selected by SURE and GCV. (b) Parameter selected by L-curve.

(a)   (b)

Figure 3.13: Point-enhanced images for Slicy target when p=0.8. (a) Parameter selected by SURE and GCV. (b) Parameter selected by L-curve.

ing images are shown in Figure 3.12-3.13 and in Table 3.3-3.4. L-curve selects the regularization parameter larger than the SURE and GCV choice. We cannot exactly conclude that L-curve performs better, however the experiments and visual inspection indicate this. We can observe that a smaller p value yield a smaller regularization parameter in general. Since smaller p favors that less dominant scatterers are observable in the image, a smaller regularization parameter may be sufficient to produce images with resolved point scatterers. Although the reconstructed images look visually indistinguishable even when the selected parameter values are distinct, this is probably not the case when we consider an ATR system.

Figure 3.14: Backhoe model used in Xpatch scattering predictions. The view to the right corresponds approximately to the images in our experiments.

3.5.3 Backhoe Data Reconstructions

We now present 2D image reconstruction experiments based on the Air Force Research Laboratory (AFRL) Backhoe Data Dome, which consists of simulated wideband (7-13 GHz), full polarization, complex backscatter data from a backhoe vehicle in free space [36]. The backscatter data are available over a full upper 2π steradian viewing hemisphere. In our experiments, we use VV polarization data, centered at 10 GHz, and with an azimuthal span of $110°$ (centered at $45°$). The backhoe model is given in Figure 3.14.

Advanced imaging strategies such as point-enhanced imaging have enabled resolution-enhanced wide angle SAR imaging. We consider the point-enhanced composite imaging technique [37] and show experimental results in this framework. In this framework, the whole angle is divided into subapertures and a seperate image is formed for each subaperture. Then, the dominant components of each subaperture image are combined and one final image is obtained for a wide observation angle. For composite imaging, we use 19 subapertures, with azimuth centers at $0°, 5°, \ldots, 90°$, and each with an azimuthal width of $20°$. We consider two different bandwidths: 500 MHz and 1 GHz. For each of these bandwidths, we consider data with two different signal-to-noise ratios: 30 dB and 20 dB.

Figure 3.15 shows one example of the SURE and GCV costs and the L-curve for backhoe image for 20 dB data with 500 MHz bandwidth. SURE and GCV again selects the same parameter and it is smaller than the one selected by L-curve.

37
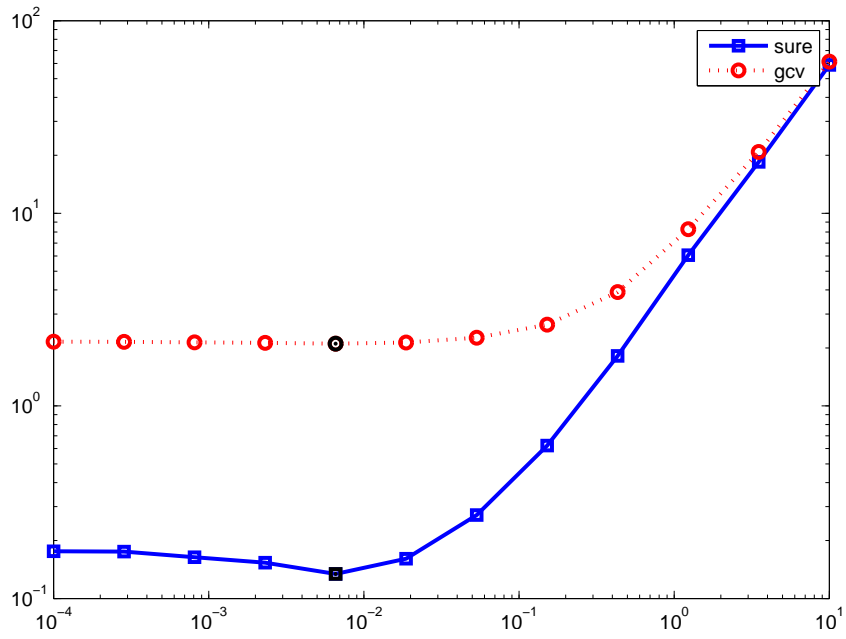
(a)
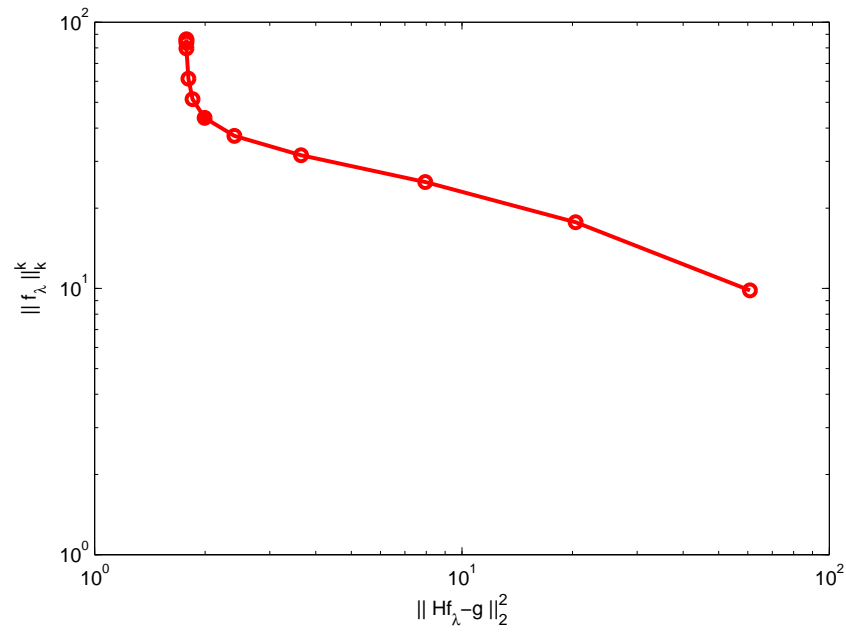
(b)

Figure 3.15: Parameter selection of the backhoe image for $p = 0.8$, SNR=20 dB and BW=500 MHz. (a) SURE and GCV. (b) L-curve.

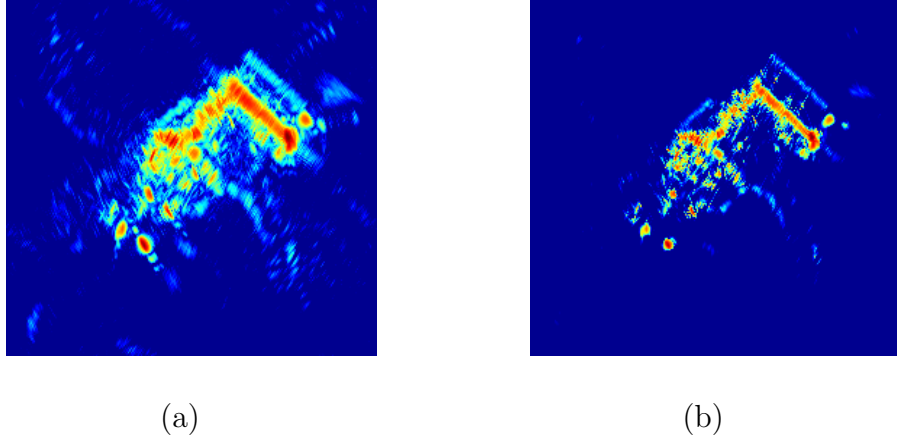|	
(a)	(b)

Figure 3.16: Point-enhanced backhoe images for $p = 0.8$, SNR=20 dB and BW=500 MHz. (a) Parameter selected by SURE and GCV. (b) Parameter selected by L-curve.

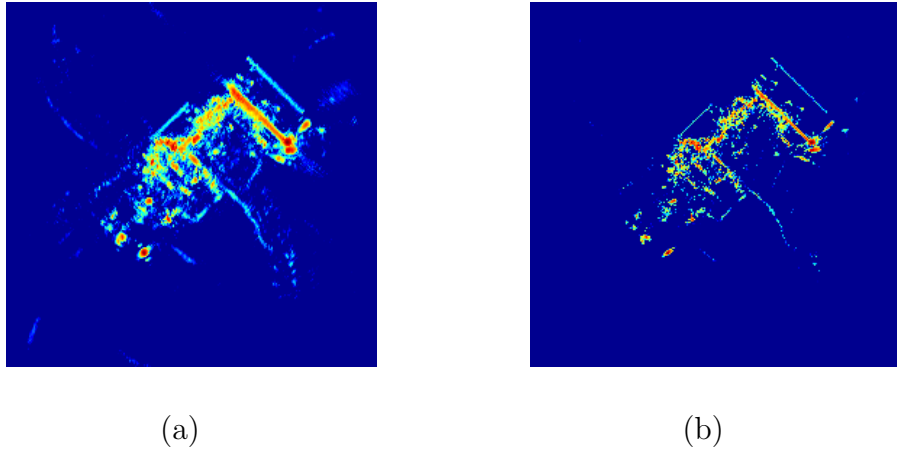|	
(a)	(b)

Figure 3.17: Point-enhanced backhoe images for $p = 1$, SNR=20 dB and BW=1 GHz. (a) Parameter selected by SURE and GCV. (b) Parameter selected by L-curve.

The reconstructed backhoe images are shown in Figure 3.17. L-curve choice gives a visually better solution; the artifacts are minimized successfully. Although the point scatterers are enhanced in SURE-GCV result, it suffers from side artifacts.

In Figure 3.14, the backhoe reconstructions with a bandwidth of 1 GHz and $p = 1$ are shown. In this example the point scatterers are resolved better. Generally, the choice of p and λ is very sensitive in the sense that the choice of λ directly effects the result for p. Therefore, the choice for p and λ mutually affects the overall performance.

To give a general idea, we also display the reconstructed images with different regularization parameters in Figure 3.18. Here we present how the reconstructions
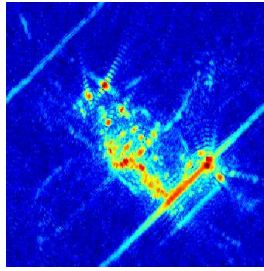
(a)
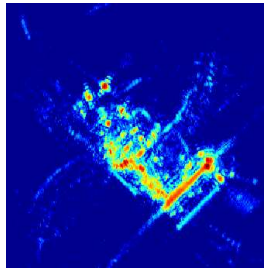
(b)
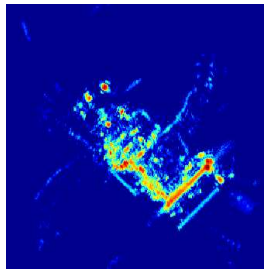
(c)
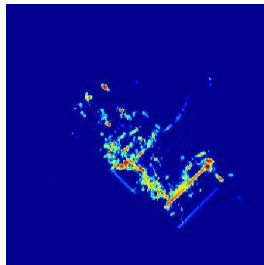
(d)
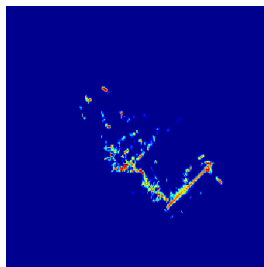
(e)

Figure 3.18: Point-enhanced composite images using different λ's, BW=1 GHz and SNR=20 dB. λ_{GCV} denotes the parameter selected by GCV. (a) $\lambda = 10^{-2}\lambda_{GCV}$. (b) $\lambda = 10^{-1}\lambda_{GCV}$. (c) $\lambda = \lambda_{GCV}$. (d) $\lambda = 10^{1}\lambda_{GCV}$. (e) $\lambda = 10^{2}\lambda_{GCV}$.

change with respect to the parameter choice but this behaviour is common for all methods. Ideally, we would like to be able to observe the scattering centers of the backhoe in a good reconstruction. From this point of view GCV seems to serve the purpose. The under-regularized image in 3.18 (a) is dominated by artifacts and the over-regularized image in Figure 3.18 (e) does not display the the structure of the backhoe correctly because of the unobservable scattering parts.

Chapter 4

Conclusions and Future Work

4.1   Summary and Conclusions

In this thesis, we have explored the application of regularization parameter selection methods in point-enhanced SAR imaging. We have provided extensions of several parameter selection methods to be used in SAR imaging and developed algorithms for parameter selection in point-enhanced imaging of complex-valued SAR reflectivity fields. We have presented experimental results based on real and simulated SAR data and discussed the effectiveness of the applied methods.

In Chapter 3, we reformulated Stein's unbiased risk estimation, generalized cross-validation, and L-curve methods for the point-enhanced SAR imaging framework. These methods were originally proposed for standard Tikhonov type regularization problems. Their use in non-quadratic regularization is still under exploration. We have tested their effectiveness in SAR imaging. SURE and GCV are both aimed at estimating the predictive risk. Although they can successfully estimate the predictive risk, this does not guarantee that the minimizers of the predictive risk and the mean squared error of the solution are the same. From this aspect, we cannot conclude that SURE and GCV always produce satisfying results. However, most of the time their parameter choice leads to reasonable but slightly under-regularized images when compared to the images where the regularization parameter is selected manually. L-curve, on the other hand, tends to select a larger parameter and appears to provide more effective results . This method basically tries to balance the two terms involved in the optimization problem of point-enhanced SAR imaging. Our experiments indicate that L-curve works well in this framework and the L-curve solution is very close to the minimizer of the mean-squared error between the true

and the reconstructed scene.

The implementation of the applied methods, especially SURE and GCV, was challenging. To overcome the memory-intensive computations, we have used several numerical optimization tools. To compute the SURE and GCV costs, we have employed randomized trace estimation and have shown that this method works well for complex-valued problems. To find the minima of the cost function, we have used the golden section search algorithm. This approach has greatly decreased the computational cost. Since the cost functions are quite flat around their minimum, resulting in an interval does not affect the solution noticeably.

Overall, this thesis has addressed an open problem in non-quadratic regularization-based SAR imaging. It is also general enough to cover any complex-valued or ℓ_p-norm regularized image reconstruction problems. The numerical tools can be applied to other types of large-scale problems. This study has provided automatic selection of the regularization parameter, and thus resulting in a new opportunity for advancement in the use of point-enhanced SAR images in ATR systems.

4.2 Future Work

Throughout this thesis, we have considered point-enhanced imaging which is aimed at enhancing point-based features. In some cases, it is also favourable to enhance other features. For example, enhancing region-based features results in a piecewise smooth image. The parameter selection methods considered here can be applied to those image reconstruction problems as well.

As mentioned previously, SURE estimates the predictive risk in regularization-based imaging. We have also mentioned that the predictive risk and the mean-squared error do not need to have the the same minimizer. If SURE can be adapted to estimate the mean-squared error itself, that might lead to better parameter choices.

High-resolution, point-enhanced SAR images are mostly promising to be used in ATR systems. We have not carried out any experiments in an ATR system. Testing the applied methods and obtained SAR images in an ATR system can be very helpful to evaluate their effectiveness in ATR systems.

In this thesis, we consider a spesific cost function which has been proposed for

point-enhanced SAR imaging problem. However, there are other forms of cost functions for imaging problems with different motivations. We note that SURE, GCV and L-curve can also be extended for the selection of the regularization parameter in these problems.

Both SURE and GCV functions depend on the regularized solution and therefore we have to find the reconstructed scene to compute the SURE and GCV costs. In this thesis, for each value of the regularization parameter, we have run the image reconstruction algorithm until convergence, and then have used the reconstructed image to evaluate the SURE and GCV functions. However it might also be possible to update the regularization parameter without running the image reconstruction algorithm until convergence. This way, we would be spending less time in image formation, and hence we could speed up the parameter selection task. Hereafter, we are also planning to consider and try out this approach.

Experiments show that the selection of p is also important as well as the regularization parameter λ. Actually, both p and λ affect the quality of the reconstructed image. Selection of both of these parameters in a data driven fashion can yield effective solutions. We have also considered this problem and tried to explore the relationship between p and λ. We have not reached any final conclusion on this relation but it is one of our ongoing research topics.

In this study, we have not considered any Bayesian approach for parameter selection. There are some pieces of work [10, 38] which consider the image reconstruction problem and the parameter selection problem as part of an overall estimation problem. This approach also requires several approximations and involves computational difficulties. Despite these difficulties we have considered some estimation theoretical methods for parameter selection however we have not observed any advantages over the methods considered and applied in this thesis. Yet, we feel this line of thought is worth further attention.

Bibliography

[1] M. Çetin, "Feature-enhanced synthetic aperture radar imaging," Ph.D. dissertation, Boston University, 2001.

[2] I. G. Gumming and F. H. Wong, *Digital Processing of Synthetic Aperture Radar Data.* Artech House, Boston, 2005.

[3] G. R. Benitz, "High-definition vector imaging," *Lincoln Laboratory Journal*, vol. 10, no. 2, pp. 147–170, 1997.

[4] S. R. Degraaf, "SAR imaging via modern 2-D spectral estimation methods," *IEEE Transactions on Image Processing*, vol. 7, no. 5, pp. 729–761, 1998.

[5] A. N. Tikhonov, "Solution of incorrectly formulated problems and the regularization method," *Soviet Mathematics - Doklady*, vol. 4, no. 6, pp. 1035–1038, 1963.

[6] V. A. Morozov, "On the solution of functional equations by the method of regularization," *Soviet Mathematics Doklady*, vol. 7, pp. 414–417, 1966.

[7] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *The Annals of Statistics*, vol. 9, no. 6, pp. 1135–1151, 1981.

[8] G. Wahba, "Practical approximate solution of linear operator equations when the data are noisy," *SIAM Journal on Numerical Analysis*, vol. 14, no. 4, pp. 651–667, 1977.

[9] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215–223, 1979.

[10] M. Ting, R. Raich, and A. O. Hero, "Sparse image reconstruction using sparse priors," in *Image Processing, 2006 IEEE International Conference on*, 2006, pp. 1261–1264.

[11] P. C. Hanson, "Analysis of discrete ill-posed problems by means of the l-curve," *SIAM Review*, vol. 34, no. 4, pp. 561–580, 1992.

[12] L. Ng and V. Solo, "Optical flow estimation using adaptive wavelet zeroing," in *Proceedings of the IEEE Intl. Conf. on Image Processing,*, vol. 3, 1999, pp. 722–726.

[13] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 4036–4048, 2006.

[14] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[15] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society*, vol. 58, no. 1, pp. 267–288, 1996.

[16] V. Solo, "A sure-fired way to choose smoothing parameters in ill-conditioned inverse problems," in *Image Processing, 1996. Proceedings., International Conference on*, vol. 3, 1996, pp. 89–92.

[17] ——, "Selection of regularisation parameters for total variation denoising," in *Acoustics, Speech, and Signal Processing, 1999. ICASSP '99. Proceedings., 1999 IEEE International Conference on*, vol. 3, 1999, pp. 1653–1655.

[18] W. G. Carrara, R. S. Goodman, and R. M. Majewski, *Spotlight Synthetic Aperture Radar: Signal Processing Algorithms.* Artech House, Boston, 1995.

[19] J. Hadamard, *Lectures on Cauchy's Problem in Linear Partial Differential Equations.* Yale University Press, New Haven, CT, 1923.

[20] D. L. Phillips, "A technique for the numerical solution of certain integral equations of the first kind," *Journal of the Association for Computing Machinery*, vol. 9, no. 1, pp. 84–97, 1962.

[21] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," in *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, 2003, pp. 2197–2202.

[22] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1-4, pp. 259–268, 1992.

[23] W. C. Karl, *Handbook of Image and Video Processing.* Academic Press, 2000, ch. Regularization in image restoration and reconstruction, pp. 141–160.

[24] C. R. Vogel and M. E. Oman, "Fast, robust total variation-based reconstruction of noisy, blurred images," *IEEE Transactions on Image Processing*, vol. 7, no. 6, pp. 813–824, 1998.

[25] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Transactions on Image Processing*, vol. 6, no. 2, pp. 298–310, 1997.

[26] C. L. Mallows, "Some comments on cp," *Technometrics*, vol. 15, no. 4, pp. 661–675, 1973.

[27] S. Geisser, "The predictive sample reuse method with applications," *Journal of the American Statistical Association*, vol. 70, no. 350, pp. 320–328, 1975.

[28] D. M. Allen, "The relationship between variable selection and data agumentation and a method for prediction," *Technometrics*, vol. 16, no. 1, pp. 125–127, 1974.

[29] R. J. Santos and A. R. DePierro, "Generalized cross-validation applied to the conjugate gradients method as a stopping rule," Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, Tech. Rep., 1995.

[30] P. C. Hanson and D. P. O'Leary, "The use of the l-curve in the regularization of discrete ill-posed problems," *SIAM Journal on Scientific Computation*, vol. 12, no. 6, pp. 1487–1503, 1993.

[31] S. Oraintara, W. C. Karl, D. A. Castanon, and T. Q. Nguyen, "A method for choosing the regularization parameter in generalizedtikhonov regularized linear inverse problems," in *Image Processing, 2000. Proceedings. 2000 International Conference on*, vol. 1, 2000, pp. 93–96.

[32] D. A. Girard, "A fast monte-carlo cross-validation procedure for large least squares problems with noisy data," *Numerische Mathematik*, vol. 56, no. 1, pp. 1–23, 1989.

[33] M. F. Hutchinson, "A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines," *Numerische Mathematik*, vol. 19, no. 2, pp. 433–450, 1990.

[34] J. Kiefer, "Sequential minimax search for a maximum," in *Proceedings of the American Mathematical Society*, vol. 4, no. 3, 1953, pp. 502–506.

[35] "Backhoe data dome and visual-d challange problem," Air Force Research Laboratory, Model Based Vision Laboratory, Sensor Data Management System MSTAR, (https://www.sdms.afrl.af.mil/datasets/mstar/), 1997.

[36] "Backhoe data dome and visual-d challange problem," Air Force Research Laboratory Sensor Data Management System, (https://www.sdms.afrl.af.mil/main.php), 2004.

[37] R. L. Moses, L. C. Potter, and M. Çetin, "Wide angle sar imaging," in *Proceedings of SPIE*, E. G. Zelnio and F. D. Garber, Eds., vol. 5427. Algorithms for Synthetic Aperture Radar Imagery XI, 2004, pp. 164–175.

[38] Y. Lin and D. D. Lee, "Bayesian l1-norm sparse learning," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, vol. 5, 2006, pp. 605–608.