# Parameter Selection in Sparsity-driven SAR Imaging

Özge Batu and Müjdat Çetin

**Abstract**

We consider a recently developed sparsity-driven synthetic aperture radar (SAR) imaging approach which can produce superresolution, feature-enhanced images. However, this regularization-based approach requires the selection of a hyper-parameter in order to generate such high-quality images. In this paper we present a number of techniques for automatically selecting the hyper-parameter involved in this problem. In particular, we propose and develop numerical procedures for the use of Stein's unbiased risk estimation, generalized cross-validation, and L-curve techniques for automatic parameter choice. We demonstrate and compare the effectiveness of these procedures through experiments based on both simple synthetic scenes, as well as electromagnetically simulated realistic data. Our results suggest that sparsity-driven SAR imaging coupled with the proposed automatic parameter choice procedures offers significant improvements over conventional SAR imaging.

**Index Terms**

parameter selection, synthetic aperture radar, sparse signal representation, non-quadratic regularization, generalized cross-validation, Stein's unbiased risk estimator, L-curve.

## I. Introduction

Conventional image formation techniques for synthetic aperture radar (SAR) suffer from low resolution, speckle, and sidelobe artifacts. These effects pose challenges for SAR images, in particular when they are used in automatic target detection and recognition tasks. Recently proposed SAR image formation algorithms have been shown to produce high quality images, offering increased resolution and reduced artifacts [1], [2], [3]. We consider the sparsity-driven, non-quadratic regularization-based approach of [1] which aims to produce feature-enhanced SAR images. For a review of this approach as well as other uses of sparsity-based ideas in radar imaging, see [4]. The idea behind this approach is to emphasize appropriate features by regularizing the solution. In fact, regularization methods are well known and widely used for real-valued image restoration and reconstruction problems. However SAR imaging involves some difficulties in application of these methods. As an example, SAR involves complex-valued reflectivities. Considering and addressing such difficulties, extensions of real-valued non-quadratic regularization methods have been developed for SAR imaging [1].

Regularization methods, in general, try to balance the fidelity to data and prior knowledge to obtain a stable solution. This stability is ensured through a scalar parameter which is called the regularization parameter or hyper-parameter. Selection of this parameter is a fundamental problem within a regularization framework. There exist several approaches which are based on statistical considerations such as Stein's unbiased risk estimator (SURE) [5], generalized cross-validation (GCV) [6], [7], Bayesian methods [8], as well as graphical tools such as the L-curve [9]. Most parameter choice methods have been developed

Ö. Batu and M. Çetin are with the Faculty of Engineering and Natural Sciences, Sabancı University, 34956 Istanbul, Turkey (e-mail: ozgebatu@sabanciuniv.edu; mcetin@sabanciuniv.edu).

This work was partially supported by the Scientific and Technological Research Council of Turkey under grant 105E090, and by a Turkish Academy of Sciences Distinguished Young Scientist Award.

in the context of the hyper-parameter choice problem for Tikhonov regularization [10], which is a well-known and widely-used quadratic regularization approach. The quadratic form of the optimization problem in Tikhonov regularization results in a closed-form solution, through a set of linear equations, which simplifies the computation of the regularized solution and the automatic selection of the regularization parameter. As the promise of sparse representation has been discovered in a variety of fields such as optical flow estimation [11], compressed sensing [12] and functional regression [13], regularization constraints which impose sparsity have become more prevalent. It has been shown that a non-quadratic regularizer promotes sparsity in the solution (see e.g., [14]). However, inclusion of such a non-quadratic constraint yields an optimization problem without a closed-form solution. Consequently, iterative procedures need to be used to compute the solution. In this case, the selection of the regularization parameter is more complicated than the quadratic case. For parameter choice in non-quadratic regularization-based techniques, the application of SURE, GCV, and L-curve is limited [15], [16], [8], [17], [18]. Especially for the form of our problem which considers an $\ell_p$-norm penalty with $p \leq 1$ for complex-valued inverse problems, the use and effectiveness of these methods have not been truly explored yet. We propose the use of SURE, GCV, and L-curve in the sparsity-driven SAR image formation framework [1] and develop a number of numerical tools for efficient implementation of the methods considered. We present the effectiveness of the applied methods through our experiments based on both simple synthetic data as well as the Air Force Research Laboratory (AFRL) Backhoe Data Dome [19].

The organization of this paper is in the following manner. Sparsity-driven SAR imaging is formulated in Section II. In Section III, SURE, GCV, and L-curve are adapted to the form of our problem. The optimization tools we propose are discussed in Section IV. Finally, our experimental results are presented in Section V, and the work in this paper is summarized in Section VI.

## II. Sparsity-driven SAR Imaging

We focus on the regularization-based SAR imaging framework proposed in [1]. The framework of [1] relies on the SAR observation process expressed in the following form:

$$g = Hf + w \tag{1}$$

where $H$ represents a complex-valued discrete SAR operator, $w$ stands for additive white Gaussian noise, $g$ and $f$ are data and the reflectivity field, respectively. In SAR imaging, one can obtain the reflectivity field starting from various data domains such as phase history, range profile or conventional image. Here, we first produce a conventional SAR image from SAR returns and then use it as our data $g$. In such a case, the SAR forward model $H$ is a convolutional operator. The framework developed in [1] involves mechanisms for improving sparsity of various features. Here, we consider one form of this approach that imposes sparsity on the reflectivity field. In particular, to emphasize sparsity of the reflectivities, the SAR image reconstruction problem is formulated as the following optimization problem:

$$\hat{f}_\lambda = \text{argmin}_f \left\| g - Hf \right\|_2^2 + \lambda \left\| f \right\|_p^p . \tag{2}$$

Here, $\lambda$ is the regularization parameter, $\left\| f \right\|_p$ denotes the $\ell_p$-norm and is defined as $\left\| f \right\|_p = \left( \sum_i^n |f_i|^p \right)^{1/p}$ where $f_i$ is the $i^{th}$ element of $f$, $n$ is the number of elements in $f$. The first term in (2) is the data fidelity term which incorporates the SAR observation model in (1), and thus information about the observation geometry. The second term which is called the regularizer or side constraint brings in the prior information we would like to impose. When one chooses $p = 2$ in this term, that leads to the well-known Tikhonov regularization method [10]. Unlike the Tikhonov approach, the side constraint in our context is aimed

at indicating a preference for sparsity; hence a choice other than $p = 2$ will be made. It has been known that minimum $\ell_p$-norm reconstruction with $p \leq 1$ provides localized energy concentrations in the resultant image, and thus promotes sparsity. In such images, most elements are forced to be small, on the other hand, a few are allowed to have large values. The outcome of the use of this term is to suppress image artifacts, increase the resolvability of scatterers, and result in a sparse image. Such sparsity constraints have been shown to lead to superresolution [20]. 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 nonzero pixel values. Note that, in general, pixels in the solution do not directly correspond to the scatterers in the reflectivity field because there could be multiple scatterers within one pixel.

To avoid problems due to nondifferentiability of the objective function when $f_i$ for any $i$ is zero, a smooth approximation to the $\ell_p$-norm is used, and the objective function in (2) takes the following form:

$$\Psi = \|g - Hf\|_2^2 + \lambda \sum_{i=1}^{n} \left(|f_i|^2 + \beta\right)^{p/2} \tag{3}$$

where $\beta$ is a small scalar. As long as $\beta$ is small but positive, the minimizer of the above cost function with $\beta \neq 0$ is close to the minimizer obtained with $\beta = 0$ [21]. However too small $\beta$ values increase the computation time required for the solution of this optimization problem. In our experiments, we choose $\beta$ empirically by considering this tradeoff. In particular, we pick $\beta = 10^{-7}$.

Our goal now is to find the estimate $\hat{f}_\lambda = \text{argmin}_f \Psi$. We note that when $p > 1$, this is a convex optimization problem. We take the gradient of $\Psi$ with respect to $f$:

$$\nabla\Psi = -2H^\dagger g + 2H^\dagger Hf + 2\lambda W_\beta(f) f \tag{4}$$

where $W_\beta(f)$ is a diagonal weight matrix whose $i^{th}$ diagonal element is $\frac{p}{2}\left(|f_i|^2 + \beta\right)^{\frac{p}{2}-1}$, and set the gradient equal to zero. The solution of the optimization problem for any value of $p$ should be a stationary point and should satisfy this equality:

$$\left(H^\dagger H + \lambda W_\beta(\hat{f}_\lambda)\right)\hat{f}_\lambda = H^\dagger g \tag{5}$$

The $i^{th}$ diagonal element of $W_\beta(\hat{f}_\lambda)$ weights the intensity of the $i^{th}$ pixel by a spatially varying penalty. Since the weight matrix depends on $\hat{f}_\lambda$, the equation in (5) is not linear in $\hat{f}_\lambda$, and (5) does not have a closed-form solution. However, one can develop a fixed-point iteration [22], each step of which involves the solution of a linear problem:

$$\left(H^\dagger H + \lambda W_\beta(\hat{f}_\lambda^{(k)})\right)\hat{f}_\lambda^{(k+1)} = H^\dagger g \tag{6}$$

where $\hat{f}_\lambda^{(k)}$ is the solution obtained in the $k^{th}$ iteration. Although equation (6), in principle, leads to a closed form solution for $\hat{f}_\lambda^{(k+1)}$, this would require the inversion of a large matrix. Hence we solve the set of equations in (6) numerically by using the conjugate gradient algorithm. This algorithm has been shown to be a descent algorithm and is likely to converge to a minimum of the cost function [23].

## III. Parameter Selection

The objective function in (3) contains a scalar parameter $\lambda$ which has a role in determining the behavior of the reconstructed field. Small parameter values make the data fidelity term; i.e. first term in (3), dominate the solution, whereas large values of $\lambda$ emphasize the $\ell_p$-norm based penalty term. In order to generate an accurate high-quality image, it is necessary to strike the right balance between these two terms by choosing the value of $\lambda$ approprately. To choose $\lambda$ in a data-driven way, we consider three methods: SURE, GCV, and L-curve.

*1) SURE:* SURE aims to minimize the following predictive risk, i.e. predictive mean-squared error:

$$R_\lambda = \left\| H f_{true} - H\hat{f}_\lambda \right\|_2^2. \tag{7}$$

Here, $\hat{f}_\lambda$ denotes the solution obtained by using $\lambda$ and $f_{true}$ is the true, unknown field. Obviously, the predictive risk cannot be calculated exactly since it depends on $f_{true}$. However, Stein's method achieves an unbiased estimate of the predictive risk [5], [24], by computing the expected value of this risk as:

$$\hat{R}_\lambda = n\sigma^2 + \|e\|_2^2 + 2\sigma^2 \nabla e \tag{8}$$

where $\sigma^2$ is the variance of the Gaussian white noise $w$, $e = H\hat{f}_\lambda - g$ and $\nabla e = \sum \partial e_i / \partial g_i$. Here, $e$ is a measure for the fitness of the estimate $\hat{f}_\lambda$ to the observation $g$, and is usually called the residual. For standard Tikhonov solution, the computation of the gradient in (8) is straightforward since the regularized solution is a linear function of the data. However, when non-quadratic regularization methods are considered, a nonlinear relation arises between the data and the estimate and there does not exist a closed-form solution for the estimate. In this case, it is more convenient to use the chain rule for evaluating $\nabla e$ and calculate the risk estimate in the following form[1] [15]:

$$\hat{R}_\lambda = -n\sigma^2 + \|e\|_2^2 + 2\sigma^2 \text{trace}\left( H \Psi_{\hat{f}\hat{f}}^{-1} \Psi_{\hat{f}g} \right) \tag{9}$$

where $\Psi_{\hat{f}\hat{f}} = \partial^2 \Psi / \partial \hat{f} \partial \hat{f}^\dagger$ is the Hessian, and $\Psi_{\hat{f}g} = \partial^2 \Psi / \partial \hat{f} \partial g^\dagger$. Then, provided that $\sigma^2$ is known or accurately estimated, the problem reduces to finding the parameter $\lambda$ which minimizes (9).

Starting from (9), we develop the SURE function for (3) as:

$$\hat{R}_\lambda = -n\sigma^2 + \left\| H\hat{f}_\lambda - g \right\|_2^2 + 2\sigma^2 \text{trace}\left(T_\lambda\right) \tag{10}$$

where

$$T_\lambda = H\left(2H^\dagger H + \lambda K(\hat{f}_\lambda, \beta)\right)^{-1} 2H^\dagger \tag{11}$$

and $K(\hat{f}_\lambda, \beta)$ is a diagonal matrix whose $i^{th}$ diagonal element is $p\left((p-1)|(\hat{f}_\lambda)_i|^2 + \beta\right)\left(|(\hat{f}_\lambda)_i|^2 + \beta\right)^{\frac{p}{2}-2}$. In summary, the goal is to find $\lambda$ that minimizes $\hat{R}_\lambda$ in (10), and consequently obtain $\hat{f}_\lambda$ which is the image reconstructed with this parameter. See [25] for more details.

*2) GCV:* The method of generalized cross-validation [6], [7] provides an estimate for $\lambda$ which approximately minimizes the expected value of the predictive risk, without requiring knowledge of $\sigma$. Let us define the so-called influence matrix $A_\lambda$ as:

$$H\hat{f}_\lambda = A_\lambda g. \tag{12}$$

Then the GCV estimate of $\lambda$ is the minimizer of (see [6]):

$$V_\lambda = \frac{\frac{1}{n}\|e_\lambda\|_2^2}{\left[\frac{1}{n}\text{trace}(I - A_\lambda)\right]^2}. \tag{13}$$

The GCV method was originally designed for problems in which $A_\lambda$ is independent of $g$. If $A_\lambda$ depends on $g$, then $A_\lambda$ can be approximated by $H\partial\hat{f}_\lambda/\partial g$, where $\partial\hat{f}_\lambda/\partial g$ is the Jacobian of $\hat{f}_\lambda$ with respect to $g$ [26]. (Note that $A_\lambda = H\partial\hat{f}_\lambda/\partial g$ if $\hat{f}_\lambda$ is linearly dependent on $g$.) Thus, we set $A_\lambda = H\partial\hat{f}_\lambda/\partial g$. We also note $H\partial\hat{f}_\lambda/\partial g = H\Psi_{\hat{f}\hat{f}}^{-1}\Psi_{\hat{f}g} = T_\lambda$ where $T_\lambda$ is given in (11), and obtain the GCV function as:

$$V_\lambda = \frac{\frac{1}{n}\|e_\lambda\|_2^2}{\left[\frac{1}{n}\text{trace}(I - T_\lambda)\right]^2}. \tag{14}$$

[1]For the sake of notational simplicity, we replace $\hat{f}_\lambda$ with $\hat{f}$ in subscripts.

*3) L-curve:* L-curve was first defined in the Tikhonov context as a parametric $\log - \log$ plot of the norm $||\hat{f}_\lambda||_2$, versus the corresponding residual norm $||H\hat{f}_\lambda - g||_2$, with the regularization parameter $\lambda$ as the parameter [9]. Then, it was extended to different regularization methods [27], [18]. In many applications, L-curve appears as an L-shaped curve as shown in Figure 1. The corner of the L-shaped curve is considered as the region containing good parameter choices that balance the regularization errors and perturbation errors in $\hat{f}_\lambda$. The L-curve criterion for choosing the regularization parameter is based on this feature. Although this intuition is natural and quite simple, computing the corner of the L-curve may not be straightforward. Several ideas have been proposed to determine the corner including the point of maximum curvature, the point closest to a reference location, such as the origin [27], and the point of tangency with a line of slope $-1$ [18]. Below we adapt L-curve to the parameter selection problem in sparsity-driven SAR imaging, and develop our own procedure for finding the corner.

## IV. Optimization Tools

### A. Computation of SURE and GCV: Randomized Trace Estimation

For large scale problems, $T_\lambda$ in (11) cannot be easily constructed due to the memory limitations of computers. In such cases, it is more convenient to find an estimate of $trace\,(T_\lambda)$, which is what SURE and GCV need, without explicitly constructing $T_\lambda$. The method in [28] calculates an estimate of the trace of the influence matrix in regularization of linear equations and enables the use of the SURE and GCV methods in large-scale problems. The method can be applied through the following algorithm:

1) generate $k$ independent realizations $q_i$ of a white noise vector with zero mean and unit variance, where $i \in \{1, ..., k\}$,
2) compute $t_i\,(\lambda) = q_i^\dagger T_\lambda q_i$,
3) take the sample mean $\bar{t}\,(\lambda) = \sum_{1=1}^{k} t_i\,(\lambda)\,/k$ to be the trace estimate.

This algorithm appears to have an explicit dependence on the matrix $T_\lambda$. However, here we do not construct $T_\lambda$ explicitly. All the matrix-vector products involved 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. It is well known that a conventional SAR image, which is the data in our case, can be described by the convolution of the original reflectivity field with a sinc function [29]. Hence the operator $H$ as well as $H^\dagger$ in our problem are convolutional. Note that $T_\lambda$ itself is also a convolutional operator. Therefore the computation required in step 2 above is also carried out through convolution operations. Finally note that the computation of $T_\lambda$ in (11) requires the inversion of a large matrix. Rather than performing that inversion explicitly, we perform numerical computation through the conjugate gradient algorithm by posing this calculation as the problem of solving a set of linear equations.

The accuracy of the trace estimate obtained through the 3-step procedure above depends on the variability of the $t_i\,(\lambda)$'s, and this variability can be quantified in terms of the variance of $t_i\,(\lambda)$. It has been shown that this variance is minimized by taking the white noise $q_i$ to be a random vector whose components are independent and take values +1 and -1, with equal probability [30].

### B. Minimization of SURE and GCV: Golden Section Search

SURE and GCV are aimed at finding the value of $\lambda$ which minimizes the expressions given in (10) and (13), respectively. Note that the differentiation of these expressions is not straightforward and these minimization problems do not have closed-form solutions. One might consider a brute force search but we should also take into account that evaluating points on SURE and GCV curves is computationally demanding and one would prefer to compute as few points as possible. Based on our experience with these methods, SURE and GCV curves have a unimodal structure in most cases although, to our knowledge,

there is no guarantee that this will always be the case. Even if they exhibit unimodal behavior in their overall structure, due to numerical approximations involved in their computation, the curves might not be strictly unimodal due to oscillations in some parts, especially around small lambda values. Fortunately, these small oscillations do not cause significant problems unless the optimization method requires derivatives and small step size in its search. Therefore we employ golden section search which is a derivative-free optimization method for unimodal functions [31]. Unlike a gradient based approach, which seeks the minimum with local movements, the golden section search approach has a more global perspective and first aims to locate the general basin of attraction. It then focuses on smaller regions in subsequent steps. In this sense, it is less likely that it will be trapped in local minima with large values of the function to be optimized, far away from the actual minimum. Of course, there is no guarantee that it will perfectly locate the global minimum. We find the minima of SURE and GCV functions through the following algorithm:

1) determine an initial interval $I = [\lambda_{min}, \lambda_{max}]$ (we start with a quite large initial interval, e.g., $I = \left[10^{-8}, 10^{2}\right]$)
2) determine two test values $\lambda_1, \lambda_2 \in I$ according to the golden ratio $\alpha = 0.618$
   (since $\lambda$ covers a large range of values, we choose golden section in the logarithmic scale), i.e.,
   $\log \lambda_1 = \log \lambda_{min} + (1 - \alpha)\,(\log \lambda_{max} - \log \lambda_{min})$ and
   $\log \lambda_2 = \log \lambda_{min} + \alpha\,(\log \lambda_{max} - \log \lambda_{min})$
3) compute $\hat{R}_{\lambda_1}$ and $\hat{R}_{\lambda_2}$ ($V_{\lambda_1}$ and $V_{\lambda_2}$)
4) determine a new interval $\tilde{I}$ through golden section search, i.e.,
   if $\hat{R}_{\lambda_2} - \hat{R}_{\lambda_1} > \delta$ ($V_{\lambda_2} - V_{\lambda_1} > \delta$) (where $\delta$ is a small positive constant) then
   $\tilde{I} = [\lambda_{min}, \lambda_2]$
   else
   $\tilde{I} = [\lambda_1, \lambda_{max}]$
5) set $I = \tilde{I}$ and repeat starting from step 2 until the interval is sufficiently small (e.g., $\log \lambda_{max} - \log \lambda_{min} = 10^{-2}$)

### *C. Finding the L-corner*

Evaluating points on the L-curve involves less computational cost since it does not require the computation of the matrix trace involved in SURE and GCV. Still, it is desirable to compute as few points as possible. Besides, defining the corner of the L-curve is an important issue. Our approach involves the definition and solution of an optimization problem for this task. In particular, we determine the L-corner through the following algorithm:

1) let $I = [I_1, I_2]$ denote the search interval for $\lambda$ where $I_1$ and $I_2$ are the lower and the upper limits of $I$, respectively.
2) set initial limits of the interval such that $I_1 = \lambda^0_{min}$ and $I_2 = \lambda^0_{max}$
3) consider the $\lambda$ values $\lambda^{k-1}_{min}$, $\lambda^{k}_{min} = \lambda^{k-1}_{min} + \Delta\lambda$, $\lambda^{l-1}_{max}$ and $\lambda^{l}_{max} = \lambda^{l-1}_{max} - \Delta\lambda$ where $k$ and $l$ are iteration numbers, and $\Delta\lambda$ is a predefined step size
4) compute slopes $m^{k-1}_{min}$, $m^{k}_{min}$, $m^{l-1}_{max}$ and $m^{l}_{max}$ of the tangent lines at the points on the L-curve corresponding to $\lambda^{k-1}_{min}$, $\lambda^{k}_{min}$, $\lambda^{l-1}_{max}$ and $\lambda^{l}_{max}$, respectively
   (note that the derivatives are computed numerically)
5) if $m^{k-1}_{min} > m^{k}_{min}$ then
   $I_1 = \lambda^{k}_{min}$
   increment $k$

else

$\lambda_{min} = \lambda_{min}^{k-1}$

and similarly,

if $m_{max}^{l-1} < m_{max}^{l}$ then

$I_2 = \lambda_{max}^{l}$

increment $l$

else

$\lambda_{max} = \lambda_{max}^{l-1}$

continue shrinking the interval by iterating through step 3 to 5.

(note that the conditions for $k$ and $l$ in step 5 are performed independently in parallel as illustrated in Figure 1)

6) set a reference point $(x_0, y_0)$ which is the intersection point of the tangent lines at $\lambda_{min}$ and $\lambda_{max}$
7) determine two test values $\lambda_1, \lambda_2 \in I$ according to the golden ratio
8) compute the residual norm $r_{\lambda_i} = ||g - H\hat{f}_{\lambda_i}||_2^2$ and the solution norm $\rho_{\lambda_i} = ||\hat{f}_{\lambda_i}||_p^p$ for $i = 1, 2$
9) compute the distance from $(r_{\lambda_i}, \rho_{\lambda_i})$ to the reference as $d_i = (\log r_{\lambda_i} - \log x_0)^2 + (\log \rho_{\lambda_i} - \log y_0)^2$
10) determine a new interval $\tilde{I}$ through golden section search, i.e.,

if $d_1 > d_2$ then

$\tilde{I} = [\lambda_1, \lambda_{max}]$

else

$\tilde{I} = [\lambda_{min}, \lambda_2]$

11) set $I = \tilde{I}$ and repeat starting from step 7 until the interval is sufficiently small

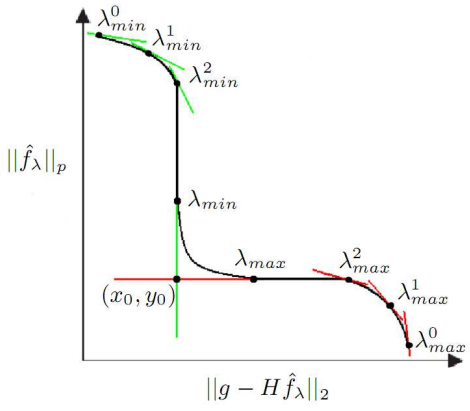


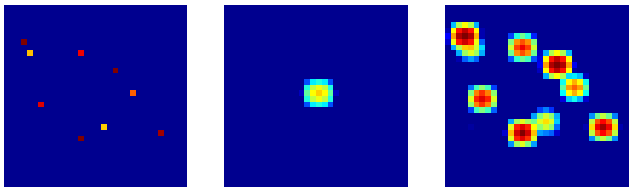

Fig. 1. The generic form of the L-curve and path for L-corner search.

## V. Experimental Results

We demonstrate the effectiveness of the parameter choice algorithms we have described on both synthetic and electromagnetically simulated realistic data. We present sparsity-driven SAR images with selected parameters and compare these results to different parameter choices and conventional reconstructions.

### A. Synthetic Scene

We first present experimental results on a synthetic example. As we know the ground truth, we can use this example to provide a quantitative performance analysis in terms of the reconstructed scenes. The results we present on a single scene and a single noise realization here are typical over multiple scenarios not presented here. The complex-valued synthetic scene consists of 9 randomly chosen point scatterers as shown in Figure 2(a). Throughout our work, we display the magnitude (in dB) of the complex-valued reflectivities. Figures 2(b) and (c) show the point spread function (PSF) of the SAR imaging system and the conventional SAR image of the synthetic scene, respectively. Here the collected SAR data involve bandlimited (through a rectangular window) Fourier samples, and as a result, the PSF is a 2-D sinc function. The vertical spread of the PSF corresponds to the resolution in the range direction in which the radar pulses are transmitted. The horizontal spread of the PSF corresponds to the cross-range resolution. Thus, the conventional image is a filtered or smoothed version of the true scene. We perform experiments for different noise levels, adding complex Gaussian noise to the simulated SAR data. We take the SNR to be the power ratio of the noise-free data to noise in dB.



(a) (b) (c)

Fig. 2. The plot of the magnitude of the (a) $32 \times 32$ synthetic scene, (b) PSF and (c) conventional SAR image.

| SNR | $\left\|f - \hat{f}_\lambda\right\|_2^2$ | $\left\|Hf - H\hat{f}_\lambda\right\|_2^2$ | SURE- GCV | L-curve |
|---|---|---|---|---|
| 30 dB | 0.024 | 0.028 | 0.028 | 0.004 |
| 20 dB | 0.080 | 0.081 | 0.083 | 0.026 |
| 10 dB | 0.302 | 0.271 | 0.342 | 0.104 |

TABLE I

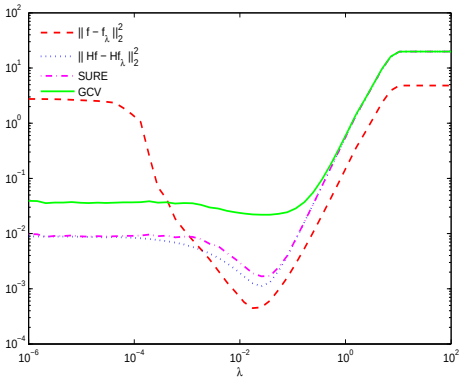
VALUES OF PARAMETER $\lambda$ MINIMIZING VARIOUS COSTS FOR THE SYNTHETIC SCENE IN FIGURE 2 WHEN P=1.

| SNR | $\left\|f - \hat{f}_\lambda\right\|_2^2$ | $\left\|Hf - H\hat{f}_\lambda\right\|_2^2$ | SURE- GCV | L-curve |
|---|---|---|---|---|
| 30 dB | 1.108 | 0.020 | 0.018 | 0.054 |
| 20 dB | 1.720 | 0.162 | 0.173 | 0.125 |
| 10 dB | 2.864 | 0.828 | 0.826 | 0.854 |

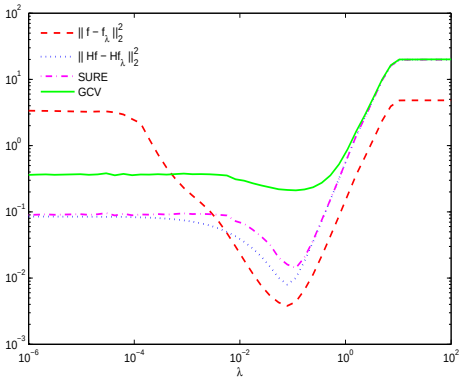
TABLE II

VALUES OF PARAMETER $\lambda$ MINIMIZING VARIOUS COSTS FOR THE SYNTHETIC SCENE IN FIGURE 6, WHEN P=1.

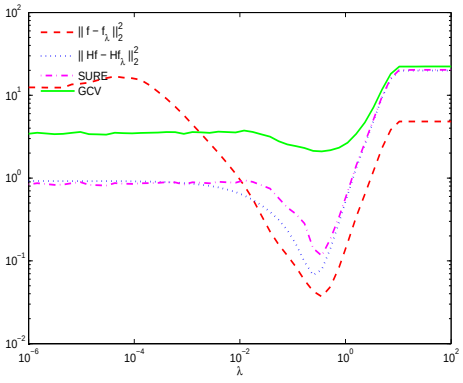
Here, we consider sparsity-driven reconstructions with $p = 1$. Figure 3 shows the estimation error $\left\|f_{true} - \hat{f}_\lambda\right\|_2^2$, predictive risk $\left\|Hf_{true} - H\hat{f}_\lambda\right\|_2^2$, as well as SURE and GCV curves for three different SNR values. SURE has a similar structure with




(a)




(b)




(c)

Fig. 3. The estimation error, predictive risk, SURE and GCV cost functions for the synthetic image with (a) 30 dB, (b) 20 dB and (c) 10 dB SNR.

the predictive risk whereas GCV has a different structure. Note that GCV is very flat around its minimum and this sometimes makes it difficult to locate the minimum. Yet, the minima of SURE and GCV are close to the minimum of the predictive risk. As we have mentioned before, SURE and GCV estimate the predictive risk in (7), but the quality of the reconstruction is, naturally, better measured by the estimation error. In this example, the minima of the estimation error and the predictive risk are very close, and as a result SURE and GCV provide good parameter choices in the sense of minimizing the estimation error. Figure 4

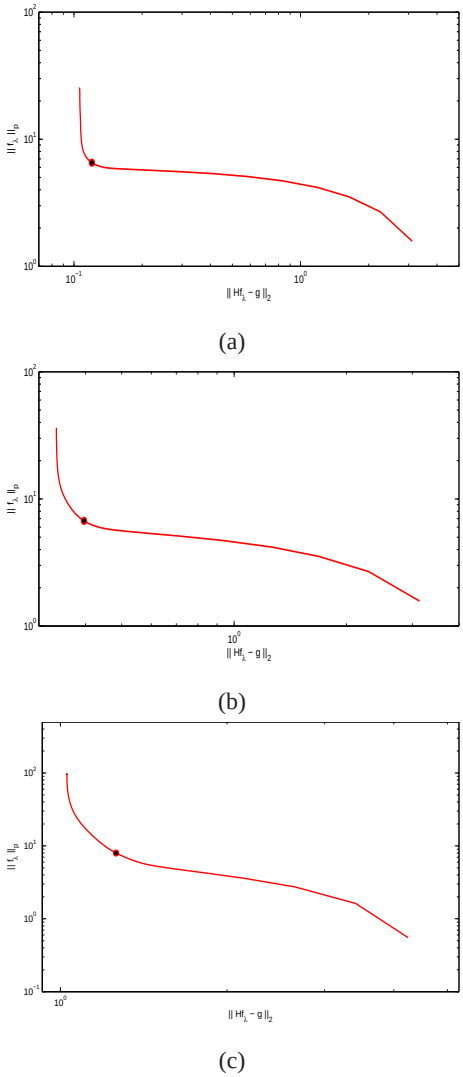
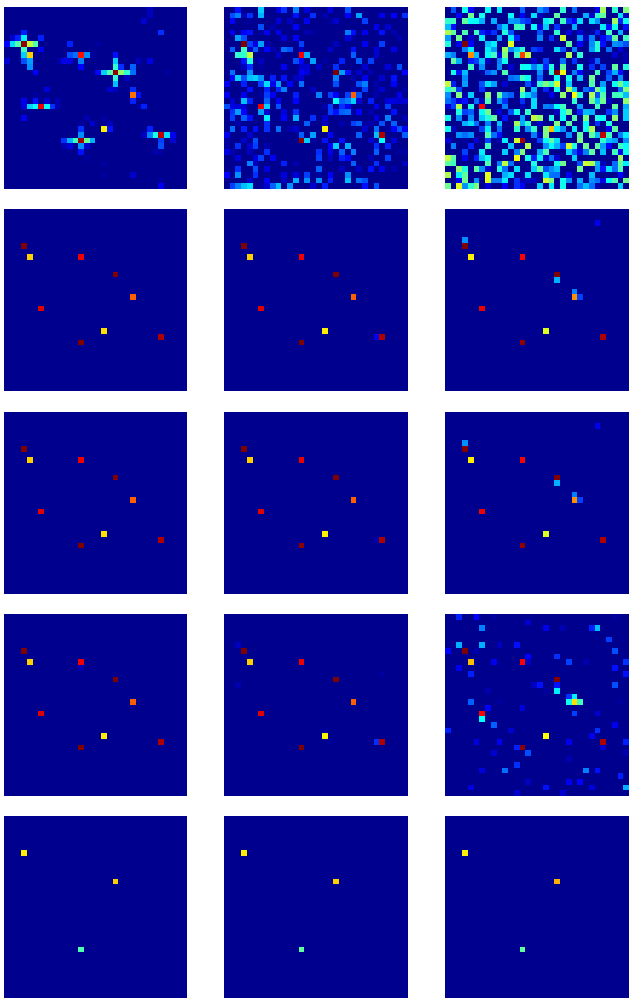



Fig. 4. L-curve and corresponding L-corner for the synthetic image with (a) 30 dB, (b) 20 dB and (c) 10 dB SNR.

shows the L-curve and corresponding L-corner. Note that for lower noise levels, the corner of the L-curve is sharper, and thus it is easier to locate the L-corner. Parameter choices of SURE, GCV and L-curve are given and compared to the parameter values minimizing the estimation error and the predictive risk in Table I. When compared to the estimation error and the predictive risk, SURE and GCV lead to very good parameter choices whereas L-curve chooses a smaller parameter. In general, the L-curve choice of $\lambda$ is $3-10$ times smaller than those of SURE and GCV, and thus leads to less sparse images. The observation that L-curve usually leads to underregularization has been made by others as well (see, e.g., [32]).

We now show the reconstruction results based on these parameters in Figure 5. We observe that SAR images reconstructed using the SURE, GCV, and L-curve parameters are very similar and hard to distinguish visually. The reconstructed SAR image is noisy when $\lambda$ is too small, and some of the scatterers are not found when $\lambda$ is too large. This confirms the need for a parameter choice method. In addition, the similarity of the scenes reconstructed through our parameter choice algorithms to the scene reconstructed by the optimal parameter $\lambda_{optimum}$ (minimizing the estimation error), as well as to the underlying true

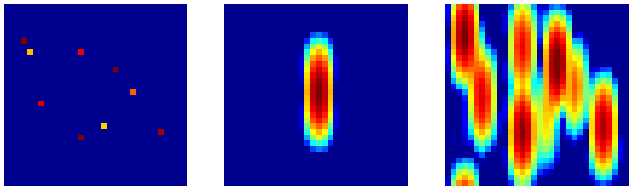


(a) (b) (c)

Fig. 5. SAR reconstructions for (a) 30 dB, (b) 20 dB and (c) 10 dB. The following parameters are used from top to bottom: $10^{-2}\lambda_{optimum}$, $\lambda_{optimum}$, $\lambda_{SURE}$ (SURE and GCV choices of $\lambda$ are almost the same and their reconstructions are visually indistinguishable therefore we show one reconstruction for both), $\lambda_{L-curve}$, and $10^{2}\lambda_{optimum}$ .

scene demonstrates the success of our approach.

In certain cases, we have observed significant differences between the minima of the estimation error and the predictive risk. An example of this occurs when we consider a different SAR imaging scenario where the resolution is very poor resulting in a PSF as in Figure 6. In particular, in the 30 dB SNR case of the scenario in Figure 6, the minimum of the predictive risk is attained at a significantly smaller value of $\lambda$ as compared to the estimation error, as shown in Figure 7. As SURE and GCV are based on the predictive risk, they also choose small parameter values and may lead to underregularization in this case. In fact, as shown in Table II, SURE and GCV produce smaller parameters than L-curve for 30 dB and 10 dB SNR in this particular scenario. The main difference between the scenario in Figure 2 and the one in Figure 6 is that the resolution is much lower in the latter case. When we have high resolution (such that the pixel spacing and the resolution are close), the operator $H$ is close to unitary. When that is the case, the predictive risk and the estimation error have a similar structure. On the other hand, in



(a) (b) (c)

Fig. 6. The plot of the magnitude of the (a) $32 \times 32$ synthetic scene, (b) PSF and (c) conventional SAR image.

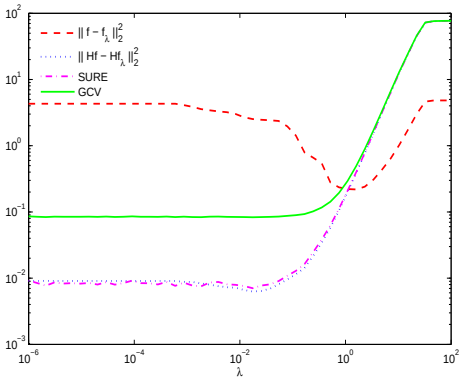



Fig. 7. The estimation error, predictive risk, SURE and GCV cost functions for the synthetic image in Figure 6 with 30 dB SNR.

scenarios such as the one in Figure 6, where $H$ is further away from being unitary, we are not guaranteed that the two curves will have similar structures. This explains our empirical observations in this experiment in terms of the differences between the minima of the predictive risk and the estimation error.

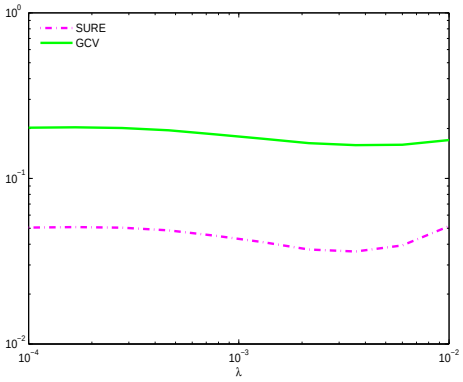
### *B. Realistic Data*

Fig. 8. Backhoe model used in Xpatch scattering predictions. The view to the right corresponds approximately to the view in the images in our experiments.

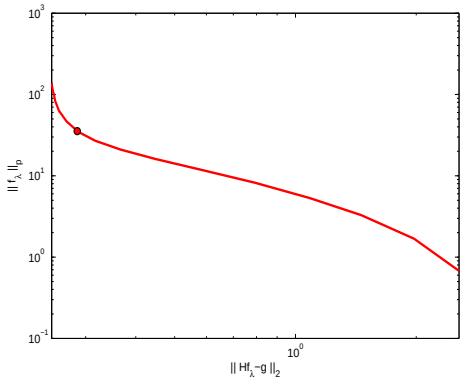
| SNR | SURE-GCV | L-curve |
|---|---|---|
| 30 dB | $1.9 \times 10^{-3}$ | $2.0 \times 10^{-4}$ |
| 20 dB | $3.6 \times 10^{-3}$ | $1.5 \times 10^{-3}$ |
| 10 dB | $3.0 \times 10^{-2}$ | $7.8 \times 10^{-3}$ |

TABLE III

VALUES OF PARAMETER $\lambda$ MINIMIZING VARIOUS COSTS FOR THE BACKHOE SCENE WHEN P=1.




(a)




(b)

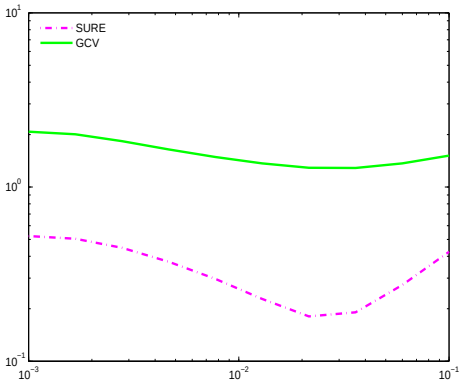
Fig. 9. (a) SURE and GCV curves, (b) L-curve and corresponding L-corner for the backhoe scene when $p = 1$ and SNR=20 dB.

| SNR | SURE-GCV | L-curve |
|---|---|---|
| 30 dB | $8.6 \times 10^{-5}$ | $7.6 \times 10^{-5}$ |
| 20 dB | $4.6 \times 10^{-4}$ | $3.0 \times 10^{-4}$ |
| 10 dB | $2.8 \times 10^{-3}$ | $1.7 \times 10^{-3}$ |

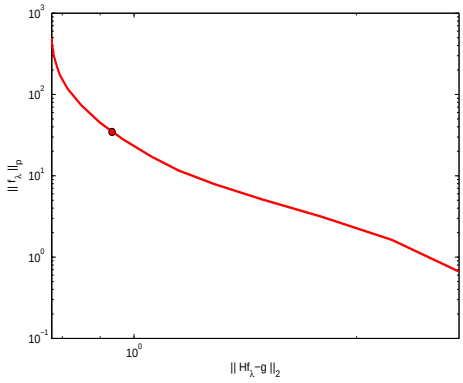
TABLE IV

VALUES OF PARAMETER $\lambda$ MINIMIZING VARIOUS COSTS FOR THE BACKHOE SCENE WHEN P=0.7.

We now present 2-D 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 [19]. The backhoe model is shown in Figure 8. The backscatter data are available over a full upper $2\pi$ steradian viewing hemisphere. In our experiments, we use VV polarization data, centered at 10 GHz, and with an azimuthal span of $110^\circ$ and a peak elevation angle of $30^\circ$ (at azimuth center). Note that this is a wide-angle imaging scenario. Sparsity-driven SAR imaging was extended and applied to SAR data collections that span a wide angular aperture [33]. Here, we consider the sparsity-driven composite imaging technique of [33] and show experimental results based on this framework. In this framework, the whole angular aperture is divided into subapertures and a separate image is formed for each subaperture. For composite imaging, we use 19 overlapping subapertures, with azimuth centers at $0^\circ, 5^\circ, \ldots, 90^\circ$, and each with an azimuthal width of




(a)




(b)

Fig. 10. (a) SURE and GCV curves, (b) L-curve and corresponding L-corner for the backhoe scene when $p = 1$ and SNR=10 dB.

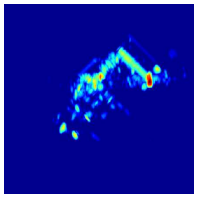
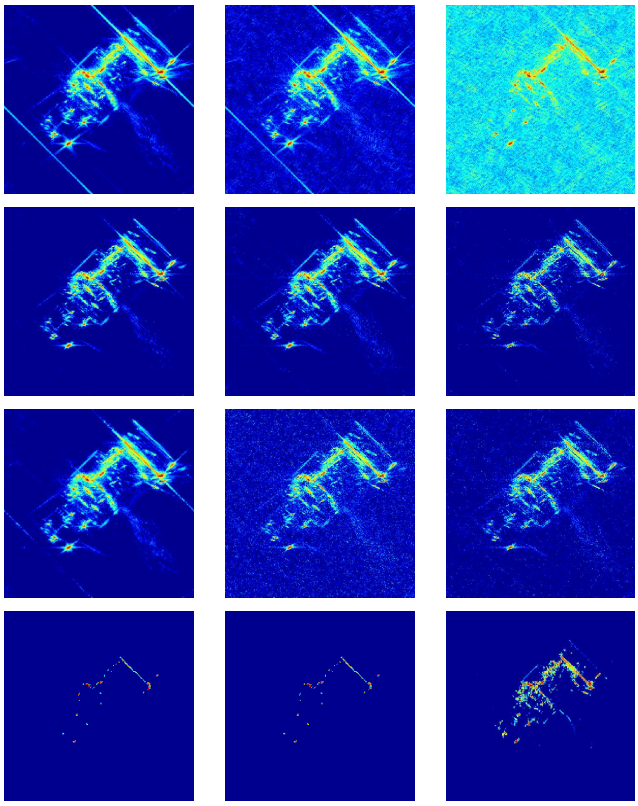


Fig. 11. Conventional SAR image

$20^\circ$. Then, the maximum pixel magnitude among all subapertures is retained for each pixel location in the composite image. In our experiments, we have observed that the $\lambda$ choice in each subaperture image is very similar. Hence we have chosen a $\lambda$ value in one subaperture image and then used that $\lambda$ for the reconstruction of all subaperture images. In a different scenario, one may need to choose it separately for each subaperture.

Figures 9 and 10 show SURE, GCV curves, and the L-curve for 20 dB and 10 dB SNR, respectively, when $p = 1$. Tables III and IV display selected $\lambda$ values for three different noise levels when $p = 1$ and $p = 0.7$, respectively. It can be observed that, larger parameters are selected for data with lower SNR. This behavior makes sense from a Bayesian estimation-theoretic view on image reconstruction: noisier data result in relatively more emphasis on the prior than the data through the use of a larger regularization parameter. As in the synthetic example in Figure 2, L-curve chooses smaller parameters than SURE and GCV. To



(a) (b) (c)

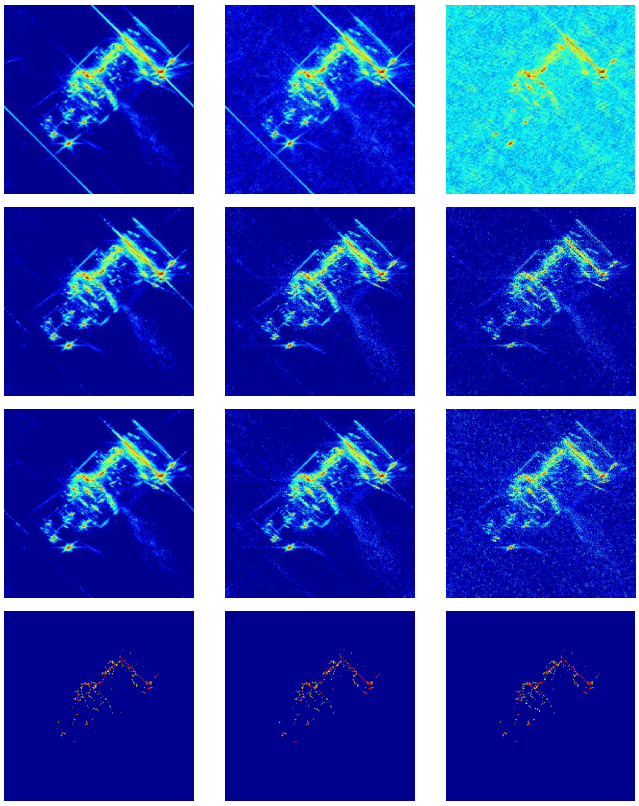
Fig. 12. Sparsity-driven images for $p = 1$ and (a) 30 dB, (b) 20 dB and (c) 10 dB. The following parameters are used from top to bottom: Too small $\lambda$, $\lambda_{SURE}$ (SURE and GCV choices of $\lambda$ are almost the same and their reconstructions are visually indistinguishable therefore we show one reconstruction for both), $\lambda_{L-curve}$, and too large $\lambda$.

provide a visual comparison, we first present the conventional SAR image in Figure 11. Sparsity-driven SAR reconstructions with parameters selected by SURE, GCV, and L-curve for $p = 1$ and $p = 0.7$ are shown in Figures 12 and 13, respectively. We cannot carry out any quantitative performance analysis for this experiment since ground truth data are not directly available to us. While our quantitative results on the synthetic scene experiments provide a useful characterization of the proposed methods, further quantitative analysis on more complicated scenes, such as the backhoe, would be of interest in future work.

## VI. Conclusion

We have considered the problem of automatic regularization parameter selection in sparsity-driven SAR imaging. We have provided extensions of several parameter selection methods to be used in SAR imaging and developed numerical algorithms for automatic parameter selection in sparsity-driven imaging of complex-valued SAR reflectivity fields.

SURE and GCV are both aimed at estimating the predictive risk and in many cases the minimizers of the predictive risk and the mean squared error of the solution are close. Under these conditions, we can conclude that SURE and GCV usually produce satisfying results. Thus, their parameter choice leads to reasonable images when compared to the images where the regularization parameter is selected manually. L-curve, on the other hand, tends to select slightly smaller parameters than those chosen by SURE and GCV with less computational cost. Sparsity-driven backhoe images appear to be somewhat underregularized, but still



(a) (b) (c)

Fig. 13. Sparsity-driven images for $p = 0.7$ and (a) 30 dB, (b) 20 dB and (c) 10 dB. The following parameters are used from top to bottom: Too small $\lambda$, $\lambda_{SURE}$ (SURE and GCV choices of $\lambda$ are almost the same and their reconstructions are visually indistinguishable therefore we show one reconstruction for both), $\lambda_{L-curve}$, and too large $\lambda$.

provide a reasonable tradeoff between artifact suppression and feature preservation for all three methods.

Overall, this work has addressed an open problem in sparsity-driven SAR imaging. It is also general enough to be applied to any complex-valued $\ell_p$-norm regularized image reconstruction problem. The numerical tools we have developed can be used in other types of large-scale problems. This study has provided mechanisms for automatic selection of the regularization parameter, thus resulting in a new opportunity for advancement in the use of sparsity-driven SAR images in automatic target recognition systems.

## REFERENCES


[1] M. Çetin and W. C. Karl, "Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization," *IEEE Transactions on Image Processing*, vol. 10, no. 4, pp. 623–631, 2001.

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

[3] 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.

[4] L. Potter, E. Ertin, J. T. Parker, and M. Çetin, "Sparsity and compressed sensing in radar imaging," in *Proceedings of the IEEE,*, vol. 98, no. 6, 2010, pp. 1006–1020.

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

[6] 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.

[7] 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.

[8] M. Ting, R. Raich, and A. O. Hero, "Sparse image reconstruction using sparse priors," in *Proceedings of the IEEE Intl. Conf. on Image Processing,*, 2006, pp. 1261–1264.

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

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

[11] 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.

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

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

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

[15] 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.

[16] ——, "Selection of regularisation parameters for total variation denoising," in *Proceedings of the IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing,*, vol. 3, 1999, pp. 1653–1655.

[17] S. Ramani, T. Blu, and M. Unser, "Blind optimization of algorithm parameters for signal denoising by Monte-Carlo SURE," in *Proceedings of the IEEE Intl. Conf. on Acoustic, Speech and Signal Processing,*, 2008, pp. 905–908.

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

[19] "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.

[20] D. L. Donoho, "Superresolution via sparsity constraints," *SIAM J. Math. Anal.*, vol. 23, no. 5, pp. 1309–1331, 1992.

[21] C. Vogel and M. Oman, "Iterative methods for total variation denoising," *SIAM Journal of Scientific Computing*, vol. 17, no. 1, pp. 227–238, 1996.

[22] ——, "Fast, robust total variation-based reconstruction of noisy, blurred images," *IEEE Transactions on Image Processing*, vol. 7, no. 6, pp. 813–824, 1998.

[23] M. Çetin, C. Karl, and A. S. Willsky, "A feature-preserving regularization method for complex-valued inverse problems with application to coherent imaging," *Optical Engineering*, vol. 54, no. 1, p. 017003, 2006.

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

[25] O. Batu, "Parameter selection in non-quadratic regularization-based SAR imaging," Master's thesis, Sabanci University, 2008, available through http://students.sabanciuniv.edu/ozgebatu/.

[26] 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.

[27] P. C. Hansen and D. P. O'Leary, "The use of the L-curve in the regularization of discrete ill-posed problems," *SIAM Journal of Scientific Computing*, vol. 14, no. 6, pp. 1487–1503, 1993.

[28] 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.

[29] C. V. Jakowatz Jr., D. E. Wahl, P. H. Eichel, D. C. Ghiglia, and P. A. Thompson, *Spotlight-mode Synthetic Aperture Radar: a Signal Processing Approach*. Kluwer Academic Publishers, Norwell, MA, 1996.

[30] 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.

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

[32] C. R. Vogel, *Computational methods for inverse problems*. SIAM, 2002.

[33] 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.