

Sparse representation-based synthetic aperture radar imaging

S. Samadi¹ M. Çetin² M.A. Masnadi-Shirazi¹

¹School of Electrical and Computer Engineering, Shiraz University, Zand Street, 71348-51154 Shiraz, Iran

²Faculty of Engineering and Natural Sciences, Sabanci University, Tuzla 34956 Istanbul, Turkey

E-mail: ssamadi@shirazu.ac.ir

Abstract: There is increasing interest in using synthetic aperture radar (SAR) images in automated target recognition and decision-making tasks. The success of such tasks depends on how well the reconstructed SAR images exhibit certain features of the underlying scene. Based on the observation that typical underlying scenes usually exhibit sparsity in terms of such features, this paper presents an image formation method that formulates the SAR imaging problem as a sparse signal representation problem. For problems of complex-valued nature, such as SAR, a key challenge is how to choose the dictionary and the representation scheme for effective sparse representation. Since features of the SAR reflectivity magnitude are usually of interest, the approach is designed to sparsely represent the magnitude of the complex-valued scattered field. This turns the image reconstruction problem into a joint optimisation problem over the representation of magnitude and phase of the underlying field reflectivities. The authors develop the mathematical framework for this method and propose an iterative solution for the corresponding joint optimisation problem. The experimental results demonstrate the superiority of this method over previous approaches in terms of both producing high-quality SAR images and exhibiting robustness to uncertain or limited data.

1 Introduction

Synthetic aperture radar (SAR) is an active microwave sensor that is able to produce high-resolution images of the earth's surface at any time. With day and night capability and operation under all weathers, SAR is one of the most promising remote sensing modalities. SAR makes use of a sensor carried on an airborne or spaceborne platform, transmitting microwave pulses towards an area of interest on the earth, and receives the reflected signal. This signal first undergoes some pre-processing tasks including a demodulation process. The SAR image formation problem is the problem of reconstruction of a spatial reflectivity distribution of the scene from the pre-processed SAR returns. In this paper, we focus on spotlight-mode SAR [1], although the proposed ideas are not limited to a specific mode.

Recently, there has been significant and increasing interest in using SAR images in automated target recognition and decision-making tasks. The success of such tasks depends on how well the reconstructed SAR images exhibit certain features of the underlying scene. The conventional approach for spotlight-mode SAR image formation is the polar format algorithm (PFA) [1], which is based on the premise of clean, full-aperture, full-bandwidth data. This method has no explicit mechanism to counter any imperfection including noise in the data, and also suffers from important shortcomings, such as resolution limitation to system bandwidth, speckle and sidelobe artefacts. These limitations

make it difficult to use conventionally reconstructed SAR images for robust recognition and decision-making tasks.

Based on the observation that typical underlying scenes usually exhibit sparsity in terms of certain features of interest, we develop an image formation method that formulates the SAR imaging problem as a sparse signal representation problem. Sparse signal representation, which has mostly been exploited in real-valued problems, offers many advantages such as super-resolution and feature enhancement for various reconstruction and recognition tasks. Recent work on feature-enhanced SAR image formation based on non-quadratic regularisation [2] has ties to sparse representation (SR). In particular, one interpretation of the technique in [2] involves SR with some fixed and specific dictionaries which makes it useful for scenes containing a limited set of specific feature types. This is only an interpretation as the (fixed) dictionaries involved in the approach are only implicit, and the optimisation problem is formulated over the reflectivities rather than the SR coefficients. The method proposed here extends and generalises the approach in [2] to SR with arbitrary, general dictionaries to handle general scenes containing any type of features. This is achieved by explicitly formulating the SAR image formation problem as a SR problem with arbitrary dictionaries.

A key challenge for exploiting SR in complex-valued problems such as SAR imaging is how to choose the dictionary and the representation scheme. In this paper, we develop a mathematical framework to deal with these

issues. In particular, we apply SR on the magnitude of the complex-valued scattered field, which leads to a joint optimisation problem over the representation of magnitude and phase of the underlying field reflectivities. The framework leads to an iterative algorithm for solving the joint optimisation problem. To demonstrate our approach, we have used some sample dictionaries, including multi-resolution dictionaries, as well as dictionaries adapted to the shapes of the likely objects in the scene. However, this framework has the capability of using any appropriate dictionary for the particular application of interest. The rest of this paper is organised as follows. In Section 2 details of SR-based SAR imaging are presented. In Section 3 experimental results on various SAR images are presented to demonstrate the effectiveness of our approach. Eventually, some concluding remarks are made in Section 4.

2 SR framework for SAR image formation

Sparse signal representation has successfully been used for solving inverse problems in a variety of applications. It has many capabilities for various reconstruction and recognition tasks; however, it has mostly been used in real-valued problems. Owing to the complex-valued and potentially random phase nature of the reflectivities in SAR, our approach is designed to sparsely represent the magnitude of the complex-valued scattered field where we are interested in its features.

2.1 Observation model

We use an observation model for spotlight-mode SAR imaging motivated by the tomographic formulation of SAR [3]. Data are collected using a radar sensor traversing a flight path with an antenna boresight realigned continually to point at a fixed ground patch. The most commonly used transmitting pulse in SAR is a linear FM chirp pulse $s(t) = e^{j(\omega_0 t + \mu t^2)}$ for $|t| \leq T_p/2$ and zero otherwise. Here ω_0 is the carrier frequency and 2μ is the so-called chirp rate. SAR transmits pulses at positions of equal angular increments. The backscattered signal from the scene is mixed with a reference chirp and passed through a low-pass filter. Typically, it is assumed that the distance from radar to the centre of the scene (R_θ) is much greater than the ground patch radius (L). In this case, the demodulated observed signal is given by [1, 2]

$$s_b(\theta, t) = \iint_{x^2+y^2 \leq L^2} f(x, y) \exp\{-j\Omega(t)(x \cos \theta + y \sin \theta)\} dx dy \quad (1)$$

where $\Omega(t) = 2/c(\omega_0 + 2\mu(t - (2R_\theta/c)))$ denotes the radial spatial frequency, $f(x, y)$ is the complex-valued reflectivity field (the unknown image) and c is the speed of light. Note that $s_b(\theta_i, t)$ is a finite slice through the two-dimensional (2D) Fourier transform of the field reflectivity at angle θ_i , and is usually referred to as the phase histories. A discretised version of (1) can be written as [2]

$$\mathbf{S}_b = \mathbf{C}\mathbf{f} \quad (2)$$

where $\mathbf{S}_b = [s_{\theta_1}^t \ s_{\theta_2}^t \ \dots \ s_{\theta_M}^t]^t$ is a vector of sampled phase histories, $\mathbf{C} = [c_{\theta_1}^t \ c_{\theta_2}^t \ \dots \ c_{\theta_M}^t]^t$ is a discretised approximation to the observation kernel in (1) and \mathbf{f} a vector of unknown sampled reflectivity image. Using the projection slice

theorem [4], the observed phase histories can also be written as [2]

$$s_b(\theta, t) = \int_{-L}^L g(\theta, u) \exp\{-j\Omega(t)u\} du \quad (3)$$

where $g(\theta_i, u)$ is the projection of $f(x, y)$ at angle θ_i . Using (1) and (3) a discrete relation between the field \mathbf{f} and the projections \mathbf{g} can be obtained as [2]

$$\mathbf{g} = \mathbf{F}^{-1}\mathbf{S}_b = \mathbf{F}^{-1}\mathbf{C}\mathbf{f} = \mathbf{H}\mathbf{f} \quad (4)$$

where \mathbf{F}^{-1} is a block diagonal matrix with each block performing an inverse discrete Fourier transform (IDFT) on each s_{θ_i} , and \mathbf{H} represents a complex-valued discrete SAR projection operator. The data \mathbf{g} in (4) are called range profiles. In the presence of noise, the range profile observation model becomes

$$\mathbf{y} = \mathbf{H}\mathbf{f} + \mathbf{n} \quad (5)$$

where \mathbf{n} is the additive observation noise. A similar phase history observation model can be obtained based on (2); however, the model in (5) has computational benefits due to the approximately sparse nature of \mathbf{H} [2]. A convolutional linear model can also be defined by replacing \mathbf{y} in (5) with a conventionally reconstructed image, and letting \mathbf{H} be replaced with a matrix with each row containing a spatially shifted version of the corresponding point spread function stacked as a row vector [5]. This form provides some computational (especially memory) advantages as the matrix vector operations can be performed through convolutions.

2.2 SR scheme

In this subsection we present our SR framework that effectively deals with the complex-valued nature of the SAR signal and can be used for enhanced SAR image formation. Since we are usually interested in features of the magnitude of the SAR signal, our new approach is designed to sparsely represent the magnitude of the complex-valued scattered field. Thus, we consider

$$|\mathbf{f}| = \Phi\alpha \quad (6)$$

where Φ is an appropriate dictionary for our application that can sparsely represent the magnitude of the scattered field (or simply the scene) in terms of the features of interest, and α is the vector of representation coefficients. For any complex-valued vector \mathbf{f} we can write $\mathbf{f} = \mathbf{P}|\mathbf{f}|$, where $\mathbf{P} = \text{diag}\{e^{j\phi_i}\}$ is a diagonal matrix, and ϕ_i represents the unknown phase of image vector element $(\mathbf{f})_i$. Thus, we can rewrite the observation model as

$$\mathbf{y} = \mathbf{H}\mathbf{f} + \mathbf{n} = \mathbf{H}\mathbf{P}|\mathbf{f}| + \mathbf{n} = \mathbf{H}\mathbf{P}\Phi\alpha + \mathbf{n} \quad (7)$$

If we knew \mathbf{P} (or phases of the elements of the unknown image vector), using an atomic decomposition technique such as an extension of basis pursuit denoising [6], we could find an estimate of α , and hence the magnitude image of the unknown scene itself, as follows

$$\hat{\alpha} = \arg \min_{\alpha} \|\mathbf{y} - \mathbf{H}\mathbf{P}\Phi\alpha\|_2^2 + \lambda \|\alpha\|_p^p \quad (8)$$

where $\|\cdot\|_p$ denotes the ℓ_p -norm, and ℓ is a positive real scalar parameter. Note that the ℓ_2 -norm term in (8) is related to the assumption that noise is zero mean white Gaussian (For colored noise, a weighted norm can be used in our data term). Also note that the perfect sparsity condition term would involve an ℓ_0 -norm, which would lead to a combinatorial, hard-to-solve problem. However, it is shown that for fields that admit a sparse enough representation, the ℓ_p -norm with $p \leq 1$, as we use here, also leads to the sparsest of all representations under certain conditions [7–11].

However, the difficulty in solving the optimisation problem in (8) is that we do not know the phase terms of the image vector elements and hence the matrix \mathbf{P} . We propose the following joint optimisation approach to overcome this problem:

- (a) Start with an initial estimate of \mathbf{f} that could be its conventional reconstruction. Using this \mathbf{f} , an initial estimate of the image phase matrix, \mathbf{P} can be obtained.
- (b) Using this estimate of \mathbf{P} , the optimisation problem in (8) can be solved and a new estimate of $\boldsymbol{\alpha}$ can be obtained.
- (c) Using the new estimate of $\boldsymbol{\alpha}$, the new estimate of $|\mathbf{f}|$ can be produced from (6). Now, a new estimate of the phase matrix \mathbf{P} should be found. To do this rewrite the observation model as

$$\mathbf{y} = \mathbf{HP}|\mathbf{f}| + \mathbf{n} = \mathbf{HB}\boldsymbol{\beta} + \mathbf{n} \quad (9)$$

where matrix $\mathbf{B} = \text{diag}\{(|\mathbf{f}|)_i\}$ and $\boldsymbol{\beta}$ is a vector formed by stacking the diagonal elements of matrix \mathbf{P} , hence it contains the unknown phase terms. An estimate of $\boldsymbol{\beta}$ can be obtained through the following estimator

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{HB}\boldsymbol{\beta}\|_2^2 \quad \text{subject to } |(\boldsymbol{\beta})_i| = 1, \quad \forall i \quad (10)$$

where the unconstrained part of (10) is both the maximum likelihood and the minimum variance unbiased estimator of $\boldsymbol{\beta}$ [12]. This is due to the assumption that the observation noise is independent identically distributed complex Gaussian noise, which is the most commonly used statistical model for radar measurement noise [13, 14], and also the linearity of the observation model in (9). However, the prior information on $\boldsymbol{\beta}$ introduces a constraint to the problem, which results in the optimisation problem in (10). For most of the SAR scenes phase of the reflectivity at a certain location could be modelled as random, with a uniform probability density function, and independent of the phase at other locations [15]. Assuming this prior information, it is also interesting to investigate the Bayesian estimation approach for this problem. In particular, the maximum a posteriori (MAP) estimator is considered here, which is obtained as follows

$$\hat{\boldsymbol{\beta}}_{\text{MAP}} = \arg \max_{\boldsymbol{\beta}} p(\mathbf{y}|\boldsymbol{\beta})p(\boldsymbol{\beta}) \quad (11)$$

Since the magnitude of any vector element $(\boldsymbol{\beta})_i$ is a constant independent of its phase, $p(\boldsymbol{\beta})$ can be written as (see

Appendix 1)

$$p(\boldsymbol{\beta}) = \prod_{i=1}^{N^2} \delta(|(\boldsymbol{\beta})_i| - 1)U(\angle(\boldsymbol{\beta})_i) \quad (12)$$

where $U(\cdot)$ is a uniform function over $[-\pi, \pi]$, $\delta(\cdot)$ is a delta function that can be considered as a uniform function over $[-\varepsilon, \varepsilon]$ for a very small value of ε , and N^2 is the number of elements of vector $\boldsymbol{\beta}$ for an image of size $N \times N$. Also, $p(\mathbf{y}|\boldsymbol{\beta})$ is a complex Gaussian pdf. Note that $p(\boldsymbol{\beta})$ is equal to a constant when all the values of $|(\boldsymbol{\beta})_i|$ terms are in the vicinity of 1 and all the phase terms $\angle(\boldsymbol{\beta})_i$ are in the interval of $[-\pi, \pi]$. Otherwise $p(\boldsymbol{\beta})$ is equal to zero. Thus, the maximum of (11) occurs when the first term on the right side is maximum with the constraint that all the magnitude values of $(\boldsymbol{\beta})_i$ terms are in the vicinity of 1 and all their phases are in the interval of $[-\pi, \pi]$. Therefore it can be easily shown that the constrained estimator in (10) is also the MAP estimator of $\boldsymbol{\beta}$, whenever all phases of the solution are considered to be in the interval of $[-\pi, \pi]$.

Using N^2 Lagrange multipliers the constrained problem in (10) can be replaced with the following unconstrained problem

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{HB}\boldsymbol{\beta}\|_2^2 + \sum_{i=1}^{N^2} \lambda'_i (|(\boldsymbol{\beta})_i| - 1) \quad (13)$$

Note that the N^2 constraints $|(\boldsymbol{\beta})_i| = 1, \forall i$ are equivalent to the one constraint $\sum_{i=1}^{N^2} (|(\boldsymbol{\beta})_i| - 1)^2 = 0$, since the summation of some non-negative real terms will be equal to zero if and only if all of them are equal to zero. This reduces the optimisation problem in (13) to the following problem which is much more tractable

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \arg \min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{HB}\boldsymbol{\beta}\|_2^2 + \lambda' \sum_{i=1}^{N^2} (|(\boldsymbol{\beta})_i| - 1)^2 \\ &= \arg \min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{HB}\boldsymbol{\beta}\|_2^2 + \lambda' \|\boldsymbol{\beta}\|_2^2 - 2\lambda' \|\boldsymbol{\beta}\|_1 \end{aligned} \quad (14)$$

Solving this optimisation problem produces a new estimate of the phase vector $\boldsymbol{\beta}$ and hence the matrix $\mathbf{P} = \text{diag}\{(\boldsymbol{\beta})_i\}$.

- (d) Go back to step (b) and repeat this loop until $(\|\hat{\mathbf{f}}^{n+1} - |\hat{\mathbf{f}}^{(n)}\|_2^2) / (\|\hat{\mathbf{f}}^{(n)}\|_2^2) < \delta$, where δ is a small positive real constant, and $|\hat{\mathbf{f}}^{(n)}$ is the estimate of $|\mathbf{f}|$ in step n .

Note that it is possible to view these two optimisation problems as coordinate descent stages of an overall optimisation problem over $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

2.3 Solving the joint optimisation problems

Let us call the cost function of the optimisation problem (8) $J(\boldsymbol{\alpha})$, and use the smooth approximation $\|\boldsymbol{\alpha}\|_p^p \simeq \sum_{i=1}^M (|(\boldsymbol{\alpha})_i|^2 + \varepsilon)^{p/2}$, where ε is a small positive constant, to avoid the non-differentiability problem of the ℓ_p -norm around the origin. The gradient of $J(\boldsymbol{\alpha})$ with respect to $\boldsymbol{\alpha}$ will be

$$\nabla_{\boldsymbol{\alpha}} J(\boldsymbol{\alpha}) = \mathbf{G}(\boldsymbol{\alpha})\boldsymbol{\alpha} - 2(\mathbf{HP}\boldsymbol{\Phi})^H \mathbf{y} \quad (15)$$

where

$$G(\alpha) = 2(HP\Phi)^H(HP\Phi) + \lambda p\Psi(\alpha) \quad (16)$$

and the $\Psi(\cdot)$ function is

$$\Psi(\alpha) = \text{diag}\{1/(|\alpha_i|^2 + \varepsilon)^{1-p/2}\} \quad (17)$$

in which, (α_i) 's are the elements of the vector α . Note that $G(\alpha)$ in (15) is a function of α , therefore this equation generally does not yield a closed-form solution for α and requires numerical optimisation techniques. It has been shown that the standard methods such as Newton's method or quasi-Newton's method with a conventional Hessian update scheme perform poorly for non-quadratic problems of this form [2, 16]. Using the idea in [2], we use here a quasi-Newton's method with a Hessian update scheme that is matched to the structure of our problem. We use $G(\alpha)$ as an approximation to the Hessian and use it in the following quasi-Newton's iterative algorithm

$$\hat{\alpha}^{(n+1)} = \hat{\alpha}^{(n)} - \gamma[G(\hat{\alpha}^{(n)})]^{-1}\nabla_{\alpha}J(\hat{\alpha}^{(n)}) \quad (18)$$

After substituting (15) into (18) and rearranging, we obtain the following iterative algorithm

$$G(\hat{\alpha}^{(n)})\hat{\alpha}^{(n+1)} = (1 - \gamma)G(\hat{\alpha}^{(n)})\hat{\alpha}^{(n)} + 2\gamma(HP\Phi)^H y \quad (19)$$

The algorithm can be started from an initial estimate of α and run until $\|\hat{\alpha}^{(n+1)} - \hat{\alpha}^{(n)}\|_2^2 / \|\hat{\alpha}^{(n)}\|_2^2 < \delta_{\alpha}$, where δ_{α} is a small positive real constant. Note that $G(\hat{\alpha}^{(n)})$ in our problem is approximately sparse, Hermitian, and positive semidefinite, so the set of linear equations in (19) for finding $\hat{\alpha}^{(n+1)}$ can itself be solved efficiently using iterative approaches such as conjugate gradient (CG) [17], which we use in our method.

Although we seek an SR of $|f| = \Phi\alpha$, we do not check if the representation obtained from the above algorithm always yields a positive-valued signal. In some cases, we may obtain negative numbers, which would mean we use some part of the phase (actually a phase shift of π) in the magnitude. This would cause an extra redundancy in our model insofar as a negative magnitude can be compensated by a phase shift of π . Although we have not explored this in detail, this redundancy may sometimes be helping us obtain better sparsity. On the other hand, one could try and limit the approach to guarantee positive numbers in the magnitude representation, however, that would lead to additional computational complexity. In our work, we have not felt the need to limit the solution to be non-negative.

Similarly, we can obtain the following iterative algorithm for solving the optimisation problem in (14)

$$G'\hat{\beta}^{(n+1)} = (1 - \gamma)G'\hat{\beta}^{(n)} + 2\gamma(HB)^H y + 2\gamma\lambda'\varphi(\hat{\beta}^{(n)}) \quad (20)$$

where

$$(\varphi(\hat{\beta}^{(n)}))_i = e^{j(\phi(\hat{\beta}^{(n)}))_i} \quad (21)$$

$$G' = 2(HB)^H(HB) + 2\lambda'I \quad (22)$$

in which I is the identity matrix and $\phi((\hat{\beta}^{(n)}))_i$ is the phase of $(\hat{\beta}^{(n)})_i$. The iteration (20) should also be run until $\|\hat{\beta}^{(n+1)} - \hat{\beta}^{(n)}\|_2^2 / \|\hat{\beta}^{(n)}\|_2^2$ becomes less than a specified

small positive constant. Note that the properties of G' are similar to those of the $G(\hat{\alpha}^{(n)})$ mentioned above and consequently with known $\hat{\beta}^{(n)}$ the set of linear equations in (20) for finding $\hat{\beta}^{(n+1)}$ can also be efficiently solved using the CG algorithm.

2.4 Ties to half-quadratic regularisation and convergence issues

It is shown in [18] that the quasi-Newton-based algorithm of a similar type to the one in (19) has ties to half-quadratic regularisation. 'The main idea in half-quadratic regularisation is to introduce and optimise a new cost functional, which has the same minimum as the original cost functional, but one which can be manipulated with linear algebraic methods. Such a new cost functional is obtained by augmenting the original cost functional with an auxiliary vector' [18, 19].

Let us consider a new cost function $K_{\alpha}(\alpha, b)$, which is quadratic in α (hence the name half-quadratic), and b is an auxiliary vector, such that

$$\inf_b K_{\alpha}(\alpha, b) = J(\alpha) \quad (23)$$

where $J(\alpha)$ is the cost functional in (8). It can be shown that the following augmented cost functional $K_{\alpha}(\alpha, b)$ satisfies (23) (see Appendix 2)

$$K_{\alpha}(\alpha, b) = \|y - HP\Phi\alpha\|_2^2 + \lambda \sum_{i=1}^{N^2} \left[b_i(|\alpha_i|^2 + \varepsilon) + \left(\frac{p}{2b_i}\right)^{p/(2-p)} \left(1 - \frac{p}{2}\right) \right] \quad (24)$$

Note that $K_{\alpha}(\alpha, b)$ is a quadratic function with respect to α , and can be minimised easily in b . According to (23), $K_{\alpha}(\alpha, b)$ and $J(\alpha)$ share the same minima in α , so we can use a block coordinate descent scheme on $K_{\alpha}(\alpha, b)$ to find the $\hat{\alpha}$ that minimises $J(\alpha)$

$$\hat{b}^{(n+1)} = \arg \min_b K_{\alpha}(\hat{\alpha}^{(n)}, b) \quad (25)$$

$$\hat{\alpha}^{(n+1)} = \arg \min_{\alpha} K_{\alpha}(\alpha, \hat{b}^{(n+1)}) \quad (26)$$

Using the results of Appendix 2, we obtain the following iterative algorithm from (25) and (26)

$$\hat{b}_i^{(n+1)} = \frac{p}{2[|\hat{\alpha}^{(n)}|_i^2 + \varepsilon]^{1-p/2}} \quad (27)$$

$$[(HP\Phi)^H(HP\Phi) + \lambda \text{diag}\{\hat{b}_i^{(n+1)}\}]\hat{\alpha}^{(n+1)} = (HP\Phi)^H y \quad (28)$$

Note that the above iterative algorithm is equivalent to the one in (19) with $\gamma = 1$. Based on the above half-quadratic interpretation of the algorithm in (19) we have shown in Appendix 3 that it is convergent in terms of the cost functional.

Now, let us consider the algorithm in (20). Below, we show that it also has a half-quadratic interpretation. The cost function of the optimisation problem of β in (14) can be

written as

$$J(\boldsymbol{\beta}) = \|\mathbf{y} - \mathbf{H}\mathbf{B}\boldsymbol{\beta}\|_2^2 + \lambda' \sum_{i=1}^{N^2} (|(\boldsymbol{\beta})_i| - 1)^2 \quad (29)$$

Let us consider the augmented function $K_\beta(\boldsymbol{\beta}, \mathbf{s})$, which is quadratic in $\boldsymbol{\beta}$ such that

$$\inf_{\mathbf{s}} K_\beta(\boldsymbol{\beta}, \mathbf{s}) = J(\boldsymbol{\beta}) \quad (30)$$

It is shown in Appendix 4 that the following augmented cost function satisfies (30)

$$K_\beta(\boldsymbol{\beta}, \mathbf{s}) = \|\mathbf{y} - \mathbf{H}\mathbf{B}\boldsymbol{\beta}\|_2^2 + \lambda' \sum_{i=1}^{N^2} |(\mathbf{S}\boldsymbol{\beta})_i - 1|^2 \quad (31)$$

where $\mathbf{S} = \text{diag}\{\exp(-j\mathbf{s}_i)\}$, with s_i the i th element of the vector \mathbf{s} . Similar to (25) and (26), a block coordinate descent scheme on $K_\beta(\boldsymbol{\beta}, \mathbf{s})$ can be used to find the optimum $\boldsymbol{\beta}$ that minimises $J(\boldsymbol{\beta})$

$$\hat{\mathbf{s}}^{(n+1)} = \arg \min_{\mathbf{s}} K_\beta(\hat{\boldsymbol{\beta}}^{(n)}, \mathbf{s}) \quad (32)$$

$$\hat{\boldsymbol{\beta}}^{(n+1)} = \arg \min_{\boldsymbol{\beta}} K_\beta(\boldsymbol{\beta}, \hat{\mathbf{s}}^{(n+1)}) \quad (33)$$

According to the results of Appendix 4, we can obtain the following iterative algorithm from (32) and (33)

$$\hat{s}_i^{(n+1)} = \phi[(\hat{\boldsymbol{\beta}}^{(n)})_i] \quad (34)$$

$$\begin{aligned} & [(\mathbf{H}\mathbf{B})^H(\mathbf{H}\mathbf{B}) + \lambda'(\hat{\mathbf{S}}^{(n+1)})^H\hat{\mathbf{S}}^{(n+1)}]\hat{\boldsymbol{\beta}}^{(n+1)} \\ & = (\mathbf{H}\mathbf{B})^H\mathbf{y} + \lambda'(\hat{\mathbf{S}}^{(n+1)})^H\mathbf{1} \end{aligned} \quad (35)$$

where $\mathbf{1}$ is a vector of ones with the same size as $\boldsymbol{\beta}$. It can be easily shown that the above algorithm is equivalent to the one in (20) with $\gamma = 1$. Convergence of the algorithm in terms of cost functional can then be easily established in a similar way to that in Appendix 3.

Based on the convergence results of the two optimisation problems, Appendix 5 proves the convergence of the overall image reconstruction algorithm presented in Section 2.2 in terms of the cost functional. In general, the algorithm converges to a minimum (may be local or global) and since we use a conventional reconstruction as our initialisation we always come to an improvement over the conventional reconstruction in terms of the cost functional. For a typical 256×256 pixel image that we have used in our experiments, using wavelet dictionary and non-optimised MATLAB code, convergence time of the algorithm is on the order of few minutes on a Pentium IV 3 GHz computer.

2.5 Dictionary selection

Selection of the proper dictionary Φ is an important part of this method. This dictionary should sparsely represent the magnitude of the complex-valued image which contains the features of interest in the scene and so it depends on the application and the type of objects or features of interest in our image.

2.5.1 Overcomplete shape-based dictionaries: If the underlying scene can be represented as a combination of some limited simple shapes such as points, lines and squares of different sizes and so on, then a powerful dictionary can be constructed by gathering all possible positions of these fundamental elements in an overcomplete dictionary. This type of dictionary provides a convenient tool to demonstrate the capabilities of the proposed framework of image formation for synthetic or simple real scenes (e.g. man-made targets in smooth low reflectivity backgrounds). For example, if we are interested in point scatterers along with smooth regions with simple shapes in a SAR image, use of this type of dictionary can lead to excellent results. The use of such an overcomplete dictionary is demonstrated in Section 3 for some synthetic scenes. Although this kind of dictionary has interesting properties, in large-scale problems it may lead to computational problems because of required large number of dictionary atoms. Thus, it is required to provide more efficient dictionaries for practical applications. It is possible to use more general dictionaries that are well known for SR of 2D signals (images).

2.5.2 Wavelet dictionary: Standard multiresolution dictionaries, such as those based on wavelets, are one of the effective options that can be used in our framework. Previous works have established that wavelet transform can sparsely represent natural scene images [20, 21]. The application of wavelet transforms to image compression leads to impressive results over other representations, so depending on the scene this dictionary may have the ability to sparsely represent complicated SAR magnitude images [22]. It is also shown in [23] that the wavelet transform can sparsely represent features of man-made targets with stronger reflectivities than the background.

2.5.3 Other dictionaries: Depending on our application, we may need to use other dictionaries with properties matched to that application. For example, curvelet transform enables directional analysis of an image in different scales [20, 24], so it is well suited for enhancing features such as edges and smooth curves in an image. If we have textures with periodic patterns in our magnitude image then one appropriate dictionary to sparsely represent these patterns could be based on the discrete cosine transform [20].

There exist many other popular dictionaries, which we do not mention here for the sake of brevity. We should just point out that any such dictionary could be used in our framework, if it is appropriate for the particular application of interest.

3 Experimental results

Here we present the results of applying the proposed algorithm in experiments with various synthetic and real SAR scenes and compare them with those of the conventional PFA and the non-quadratic regularisation approach of [2] to demonstrate the achieved improvements in performance. We also present results with different dictionaries to show the effect of dictionary selection and the capability of the proposed framework in using any appropriate dictionary.

3.1 Parameter selection and initialisation of the algorithm

We have used the conventional polar format reconstruction as an initial estimate of \mathbf{f} to be used in both our method and the non-quadratic regularisation method, which provides an initial estimate for both α and β . Based on our discussions on convergence in Section 2.4, for the iterative algorithms in (19) and (20) we have used a fixed step size of $\gamma = 1$, which guarantees convergence in terms of the cost functional. The smoothing parameter, ε , in the definition of the approximate ℓ_p -norm, is set to 10^{-5} , which is small enough not to affect the results given the normalised (to maximum magnitude of 1) input data to our algorithm.

For synthetic image experiments, we have used parameters of a spotlight mode SAR of 10 GHz centre frequency and 0.375 m range and cross-range resolution. For the results that show super-resolution capability, the resolution values are doubled so that the system resolution is a multiple of image pixel size and we can put multiple scatterers in one resolution cell and observe the results. All synthetic scenes consist of 32×32 complex-valued pixels.

The parameters λ , λ' and p reflect the degree of emphasis on the data against the constraints, as well as the nature of our prior information about the scene leading to the type of sparsity constraint used. The value of p could be less than or equal to 1 and smaller values of p usually produce sparser results. Selection of λ and λ' involves a trade-off between relying on data or the prior information. For example, higher values of λ usually produce sparser results. Thus, in general if we have enough information of the scene or we are interested only in special features of the scene, such that a proper dictionary with a good SR of the scene can be chosen, higher values of λ and lower values of p will produce better results. Considering these guidelines, we have used $\lambda = 10$, $p = 0.6$ and $\lambda' = 2$ for the synthetic scene experiments, and $\lambda = 1$, $p = 0.7$ and $\lambda' = 2$ for experiments on the AFRL backhoe data. Our experience as well as the reported experience in [2] show that a set of parameters chosen on a single image of a database can usually be used for other similar images of that database as long as the observation quality does not change. There are also automatic parameter selection methods [25] developed for similar problems that could be an area of research in continuation of this work.

3.2 Quality metrics for evaluation of reconstructed images

To provide a quantitative evaluation for the synthetic scene experiments, we consider metrics that directly use the ground truth image. First idea could be signal-to-noise ratio (SNR) or mean squared error, which are defined as [26]

$$\begin{aligned} \text{MSE} &= \frac{1}{N^2} \|\mathbf{f} - \hat{\mathbf{f}}\|_2^2 \\ \text{SNR}_{(\text{dB})} &= 10 \log_{10} \left(\frac{\sigma_{|\mathbf{f}|}^2}{\text{MSE}} \right) \end{aligned} \quad (36)$$

where $|\mathbf{f}|$ and $|\hat{\mathbf{f}}|$ are true and reconstructed magnitude images, and $\sigma_{|\mathbf{f}|}^2$ is the variance of the true magnitude image. Another idea could be performing an adaptive threshold on the images to separate the target and the background regions and then counting the number of

matched pixels in the binary results [27]. This is a kind of segmentation metric, and could be viewed as a target localisation metric (TLM).

For experiments with AFRL backhoe data we do not have the ground truth, however we have a rich data of different bandwidths. We can produce a reference image using the highest bandwidths and use it for evaluation of lower bandwidth results. To do this we apply the adaptive thresholding on the highest bandwidth (2 GHz) conventional reconstruction, which is the most reasonable and fair one for evaluating our method, to obtain the approximate target region. We then use the 2 GHz conventional image in this region as the ground truth for lower bandwidth reconstructions. Therefore we can compute the above-mentioned metrics for backhoe data too.

We can also use the following quality metrics that are used in the literature to evaluate the quality of reconstructed images of unknown scenes. For these metrics we consider a rectangle surrounding the target as the target region (\mathcal{T}) and out of it as the background region (\mathcal{B}). We can limit the background region to a region with the same area as the target region [28].

1. Target-to-background ratio (TBR) [27, 29]: as a measure of accentuation of the target pixels with respect to the background which is defined as

$$\text{TBR} = 20 \log_{10} \left(\frac{\max_{i \in \mathcal{T}} (|\hat{\mathbf{f}}_i|)}{(1/N_{\mathcal{B}}) \sum_{j \in \mathcal{B}} |\hat{\mathbf{f}}_j|} \right) \quad (37)$$

where $N_{\mathcal{B}}$ denotes the number of pixels in the background region \mathcal{B} .

2. Mainlobe width (MLW) [27]: as a measure of the effective resolution. We obtain an estimate of it by averaging the 3-dB lobe width of the strong scatterers. In each row and column in the target region, we find the nearest point below 3 dB of the maximum value. A better estimate of the 3-dB distance can then be estimated using a linear interpolation between pixels. Finally, the distances obtained from each row and column are averaged to find an overall estimate of the 3-dB lobe width [27].

3. Entropy of the full image (ENT) [30]: as a measure of image sharpness. A sharper image has a smaller entropy value. Let $p(i)$ denote the normalised frequency of occurrence of each gray level, that is, the pixel intensity histogram of the image.

$$\text{ENT} = - \sum_{i=0}^G p(i) \log_2(p(i)) \quad (38)$$

where G is the number of levels in the histogram.

4. Target to background entropy difference (TBED) [28]: as a measure of extractability of the target from its background. It is defined as the absolute difference between the entropy of target region and the entropy of background region, normalised to the entropy of the full image. This measure, when low, indicates that the target and background region have very similar entropy levels, making the target more difficult to extract from its local background. Clark and Velten [28] have examined nine different image quality measures and found that this measure is significantly related to the performance of an automatic target recognition algorithm.

The above metrics have been exploited to evaluate the results of the experiments in the following subsections.

3.3 Synthetic scene reconstruction

Experiment 1: First we show an important capability of this method that is super-resolution, which means that it can reconstruct image details under bandwidth limitations. To demonstrate this property, we apply our method on a synthetic scene composed of eight point scatterers with unit reflectivity magnitude and random (uniform) uncorrelated phase. The magnitude field for this scene is shown in Fig. 1a. We set our system parameters so that the system resolution is twice the image pixel size, so we seek super-resolution reconstructions. Fig. 1b shows conventional spotlight mode SAR reconstruction using the PFA [1] that cannot resolve closely spaced scatterers and suffers from high sidelobes. Fig. 1c shows the result of non-quadratic regularisation reconstruction method [2] by using regularisers for both point and region feature enhancement. Note that this method also has super-resolution capability when we just use the point enhancement regulariser. However, as we usually do not have enough prior knowledge about the scene, here we consider the scene may contain both point and region-based features. As it is seen, this technique fails to reconstruct the scene accurately.

Fig. 1d shows the result of the proposed SR method in which we observe that the reconstructed image is very close to the true image. To be comparable with the result of non-quadratic regularisation method, here we have exploited an overcomplete shape-based (SB) dictionary that consists of points as well as squares of various sizes at every possible location in the scene. Note that such a dictionary can be used to sparsely represent many scenes containing point-like targets as well as smooth regions.

Experiment 2: To demonstrate the capabilities of the proposed framework and contrast it with existing methods, in this experiment we consider a more general synthetic scene composed of point scatterers as well as a smooth distributed region as shown in Fig. 2a. The image consists of 32×32 complex-valued pixels, and we show only the magnitude

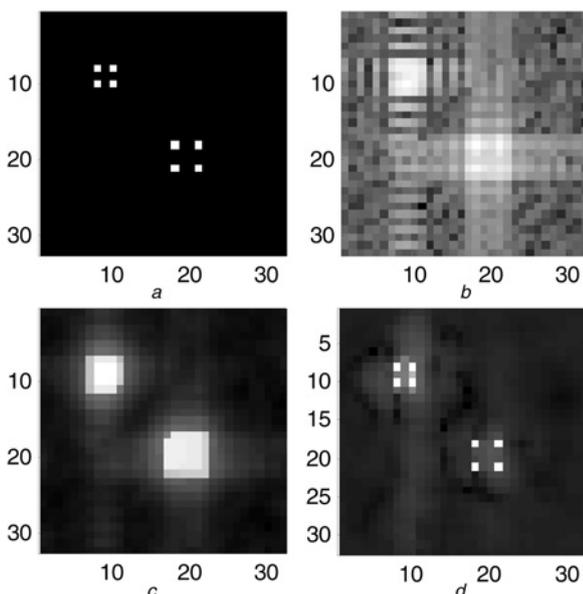


Fig. 1 Super-resolving a scene with isolated point scatterers

- a True scene
- b Conventional reconstruction
- c Point-region-enhanced non-quadratic regularisation
- d SR-based reconstruction

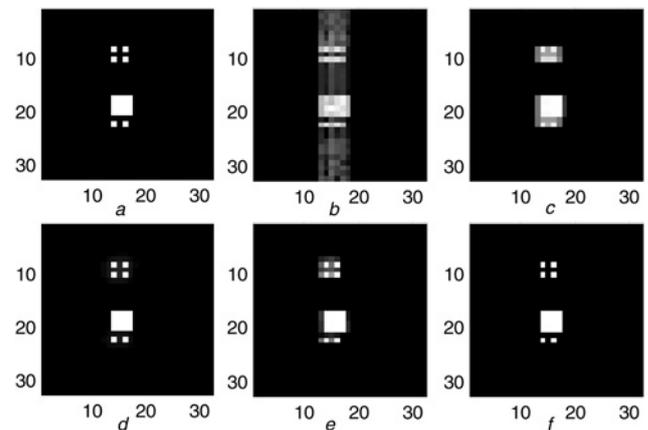


Fig. 2 Synthetic scene reconstruction

- a Synthetic scene
- b Conventional reconstruction
- c Point-region-enhanced non-quadratic regularisation
- d–f SR based reconstructions
- d SB dictionary
- e Wavelet dictionary
- f Spike-wavelet overcomplete dictionary

field in Fig. 2a. Phase is randomly distributed with a uniform density function in $[-\pi, \pi]$.

The conventional image reconstruction based on PFA is shown in Fig. 2b, which is a poor result. Fig. 2c shows the non-quadratic regularisation method with regularisers for both point and region enhancement. Note that this method cannot enhance both types of features simultaneously in the reconstructed image due to applying two contrary regularisers on the whole image. Whether the points or the regions are better preserved is a trade-off that could be adjusted through the regularisation parameters.

Fig. 2d through Fig. 2f show the reconstructed images with the proposed SR method using dictionaries described in Subsections 2.5.1 and 2.5.2. The SB dictionary here, which consists of points as well as squares of various sizes at every possible location of the scene, has a good SR for the magnitude of this synthetic image, so the reconstructed image with this dictionary in Fig. 2d is very close to the perfect reconstruction. Note that for better comparison of all results, we have not shown the values below 50 dB of the maximum value of the image.

The multi-resolution wavelet dictionary is a much more general dictionary that can be used for unknown complicated scenes. Here we have used the Haar wavelet. Its result, which is shown in Fig. 2e, shows an interesting and relatively good agreement with the true scene. As this figure shows, it seems that using wavelet dictionary alone in the proposed framework is not so powerful to reconstruct the point scatterers. To overcome this problem, we propose to use an overcomplete dictionary composed of point dictionary (i.e. spikes) and the wavelet. The result obtained by using this overcomplete dictionary is shown in Fig. 2f, where it clearly represents both the smooth region and point targets. Consider that while the SB dictionary appears to be very good in terms of reconstruction quality, computationally it is the most demanding one as it is based on a highly redundant overcomplete dictionary. Evaluation results using quality metrics defined in Section 3.2 are shown in Table 1, which demonstrate the superiority of the proposed SR-based methods.

Experiment 3: To be more realistic, in this experiment we use a synthetic image, constructed from the MIT Lincoln Laboratory Advanced Detection Technology Sensor

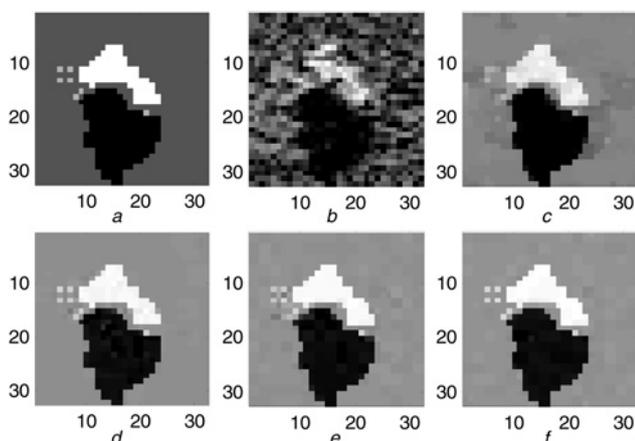
Table 1 Evaluation results of experiments 2 and 3

		Conventional	Non-quadratic	SR-SB	SR-wavelet	SR-spkwav
Experiment 2	SNR, dB	11.50	16.22	27.76	22.07	27.95
	TLM, %	85.25	92.57	98.14	96.87	99.70
Experiment 3	SNR, dB	13.87	23.01	29.72	28.31	30.03
	TLM, %	78.22	98.82	99.80	99.31	100

(ADTS) data set [31] by segmentation techniques, as well as addition of some point scatterers (with random phase), as shown in Fig. 3a. The conventional and non-quadratic regularisation reconstructions are shown in Figs. 3b and c, respectively. Fig. 3d through Fig. 3f show the reconstructed images with the SR method with different dictionaries. We have used the same SB dictionary described in the previous experiment. Because of non-zero background and arbitrary distributed regions, the representation of this image is not as sparse as the one in the previous synthetic scene; however, the resultant reconstructed image still is in good agreement with the true image. We can see in Fig. 3f that for this more realistic scene, the overcomplete spike-wavelet dictionary has excellent result. We have used the Haar wavelet for this experiment. Evaluation results using previously defined quality metrics are depicted in Table 1, which show the improvements achieved using the proposed SR-based methods.

3.4 AFRL backhoe data

Experiment 4: We now present our experimental results based on the AFRL backhoe data [32], which is a wideband, full polarisation, complex-valued backscattered data from a backhoe vehicle (shown in Fig. 4) in free space. In our experiment, we use VV polarisation data centred at 10 GHz with three available bandwidths of 500 MHz, 1 GHz and 2 GHz, with an azimuthal span of 110° (centred at 45°). Fig. 5a shows the conventional composite reconstructions of these data as described in [33] for bandwidths of 500 MHz, 1 GHz and 2 GHz. We obtain the composite images through combination of 19 subapertures, in each of

**Fig. 3** Synthetic ADTS scene reconstruction

- a Synthetic scene
- b Conventional reconstruction
- c Point-region-enhanced non-quadratic regularisation
- d–f SR based reconstructions
- d SB dictionary
- e Wavelet dictionary
- f Spike-wavelet overcomplete dictionary

which we use the spotlight-mode SAR formulation. Also, a point-enhanced composite reconstruction method using non-quadratic regularisation has been recently presented in [5] for backhoe data. Fig. 5b shows the reconstructed images based on this technique. Fig. 5c shows the results based on our proposed method using the wavelet dictionary. We should expect to obtain the sparsest representation of the important/interested features (according to the selected dictionary) in the reconstructed image. For better comparison, in all results of Fig. 5, the values below 55 dB of the image maximum value are not shown. As it is expected, for this complicated scene using the wavelet dictionary, the SR method produces very good results with very little artefacts. Also based on our knowledge of the shape and structure of the scene, there are very little false targets/features out of the expected zone of the backhoe vehicle. We have used the Daubechies 2 (db2) wavelet for this experiment. Quantitative evaluation results depicted in Tables 2 and 3 demonstrate the improved quality of SR-based reconstructions in terms of all quality metrics except the MLW, for which the non-quadratic regularisation method is the best with the proposed SR-based method following close behind. Note that the point-enhanced non-quadratic method is optimised for enhancing spike-like targets and therefore it is expected to have better MLW.

Experiment 5: Finally, in the last experiment we show the improved robustness of the SR method in data limitation scenarios. One important data limitation scenario that may occur in many SAR applications is frequency band omission that may be encountered in several situations such as jamming and data dropouts or in VHF/UHF frequency band systems such as foliage penetration radar in which it is likely that we will not be able to use an uninterrupted frequency band [5]. In this experiment for each of the three available bandwidths of backhoe data, we consider the case of frequency-band omissions where 20% of the spectral data within that bandwidth are available. We have selected the available band randomly, with a preference for contiguous bands (expressed through a parameter used in random band generation). The corresponding results for this

**Fig. 4** Backhoe vehicle model

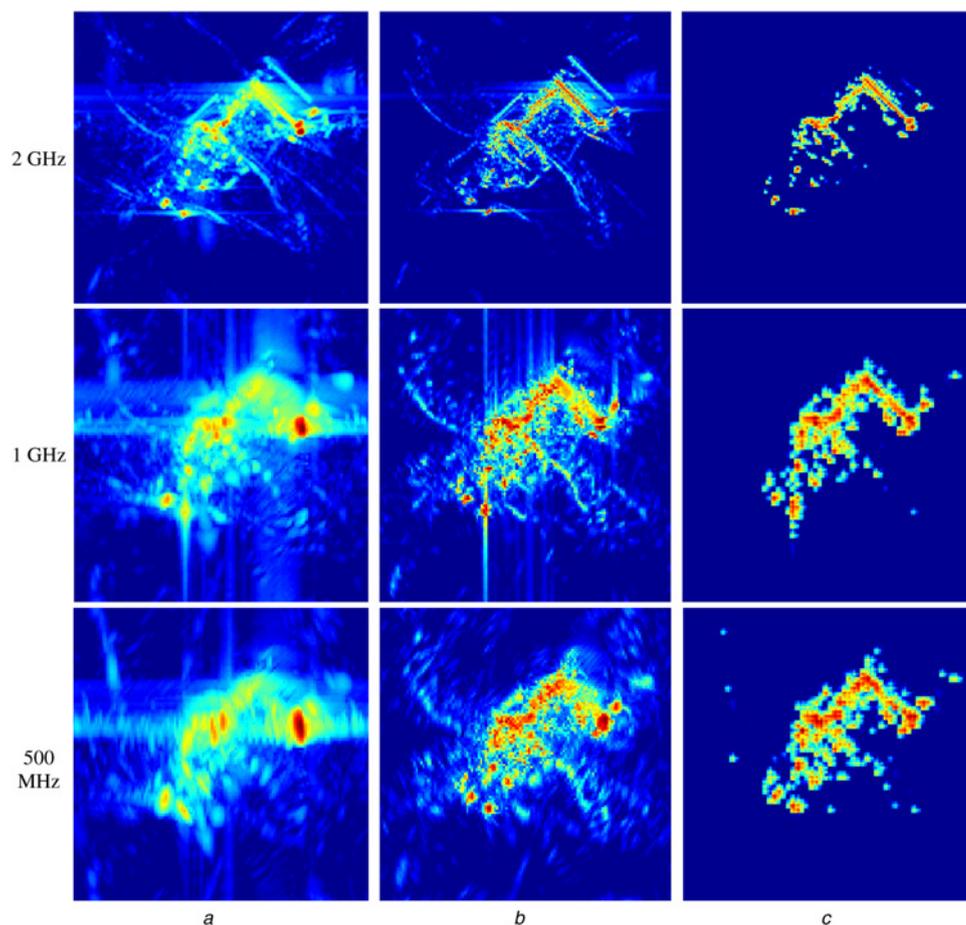


Fig. 5 Results with the AFRL Backhoe data of 2 GHz, 1 GHz and 500 MHz bandwidth

- a Conventional composite reconstruction
- b Point-enhanced (non-quadratic regularisation) composite reconstruction
- c SR-based (with wavelet dictionary) composite reconstruction

Table 2 Evaluation results of experiments 4 and 5 (true scene-dependent metrics)

		1 GHz			500 MHz		
		Conventional	Non-quadratic	SR-wav	Conventional	Non-quadratic	SR-wav
Experiment 4	SNR, dB	28.01	28.65	31.33	28.19	28.55	30.87
	TLM, %	80.71	85.11	91.02	80.88	85.89	91.01
Experiment 5	SNR, dB	26.27	27.39	30.34	26.03	26.50	29.87
	TLM, %	77.76	86.36	89.24	77.60	84.90	89.85

Table 3 Evaluation results of experiments 4 and 5 (unknown scene metrics)

		2 GHz			1 GHz			500 MHz		
		Conventional	Non-quadratic	SR-wav	Conventional	Non-quadratic	SR-wav	Conventional	Non-quadratic	SR-wav
Experiment 4	TBR, dB	50.01	55.31	71.99	43.07	47.42	60.47	42.47	47.86	58.42
	MLW, m	0.061	0.016	0.019	0.141	0.034	0.040	0.293	0.074	0.082
	ENT	1.485	1.003	0.475	2.324	1.789	1.026	2.138	1.865	1.140
	TBED	1.633	2.546	3.890	0.930	1.735	3.362	0.973	1.779	3.230
Experiment 5	TBR, dB	36.48	45.88	60.04	29.49	40.88	55.87	27.66	33.36	51.55
	MLW, m	0.099	0.028	0.020	0.185	0.040	0.040	0.456	0.091	0.084
	ENT	3.250	2.118	0.804	4.048	2.690	1.271	4.403	3.657	1.323
	TBED	0.576	1.396	3.445	0.337	1.244	2.979	0.287	0.814	2.877

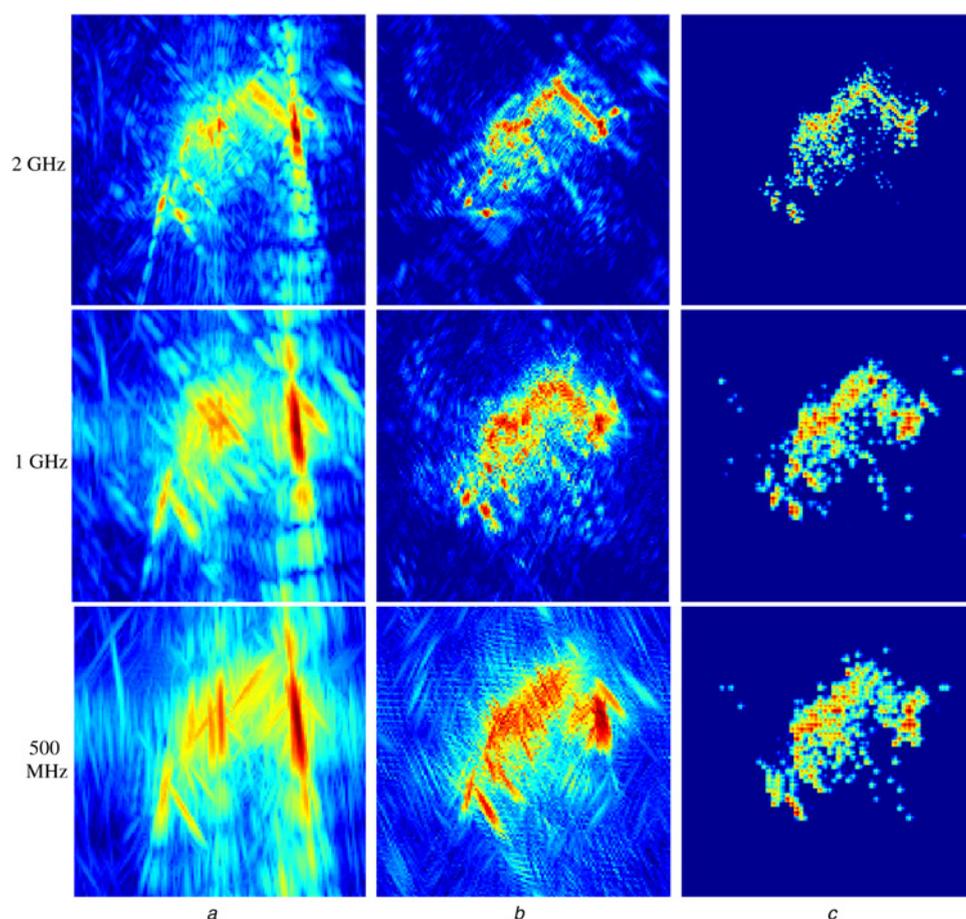


Fig. 6 Results with the AFRL Backhoe data of 2 GHz, 1 GHz, and 500 MHz bandwidth with frequency band omissions (20% of the full band data available)

- a* Conventional composite reconstruction
b Point-enhanced (non-quadratic regularisation) composite reconstruction
c SR-based (with wavelet dictionary) composite reconstruction

case are shown in Fig. 6. In these results the top 50 dB part of all images are shown for better comparison. Robustness of the proposed SR framework to bandwidth limitations as well as frequency band data limitations is clearly revealed in these results. In Tables 2 and 3 computed quality metrics are depicted, which show the improvements provided by the proposed SR-based reconstructions in terms of all quality metrics. It is interesting that here MLW of SR-based are better than that of non-quadratic method, which shows its excellent robustness to data limitation scenarios.

4 Conclusions

In this paper we have proposed a new approach for SAR image formation based on sparse signal representation. Owing to the complex-valued nature of the SAR reflectivities, we have designed our approach to sparsely represent the magnitude of the complex-valued scattered field in terms of the features of interest. We have formulated the mathematical framework for this method and proposed an iterative algorithm to solve the corresponding joint optimisation problem over the representation of magnitude and phase of the underlying field reflectivities. In various experiments, we have demonstrated the performance of this approach that produces high-quality SAR images with enhanced features and very little artefacts, which is ideal for automatic recognition tasks.

Selection of the dictionary depends on the application. For SAR images of natural scenes wavelet dictionary seems to be a good choice and for images of man-made targets with known-shape parts, SB dictionaries could result in better reconstructions than the standard dictionaries at the expense of higher computational load. Also as we have demonstrated, this approach exhibits interesting features including super-resolution, and robustness to bandwidth limitations as well as to uncertain and limited data. Thus, it could be a good choice in such scenarios. In addition to these characteristics, the proposed approach has the potential to provide enhanced image quality in non-conventional data collection scenarios, for example, those involving sparse apertures.

5 Acknowledgment

This work was partially supported by an Iranian Ministry of Science, Research and Technology Scholarship, and by the Scientific and Technological Research Council of Turkey under Grant 10SE090, and by a Turkish Academy of Sciences Distinguished Young Scientist Award.

6 References

- 1 Carrara, W.G., Goodman, R.S., Majewski, R.M.: 'Spotlight synthetic aperture radar: signal processing algorithms' (Artech House, Boston, MA, 1995)

2 Çetin, M., Karl, W.C.: 'Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization', *IEEE Trans. Image Process.*, 2001, **10**, (4), pp. 623–631

3 Munson, Jr D.C., O'Brien, J.D., Jenkins, W.K.: 'A tomographic formulation of spotlight-mode synthetic aperture radar', *Proc. IEEE*, 1983, **71**, pp. 917–925

4 Kak, A.C., Slaney, M.: 'Principles of computerized tomographic imaging' (IEEE Press, New York, 1988)

5 Çetin, M., Moses, R.L.: 'SAR imaging from partial-aperture data with frequency-band omissions'. SPIE Defense and Security Symposium, Algorithms for Synthetic Aperture Radar Imagery XII, Orlando, FL, USA, March 2005, pp. 32–43

6 Chen, S.S., Donoho, D.L., Saunders, M.A.: 'Atomic decomposition by basis pursuit', *SIAM J. Sci. Comput.*, 1998, **20**, pp. 33–61

7 Donoho, D.L., Huo, X.: 'Uncertainty principles and ideal atomic decomposition', *IEEE Trans. Inf. Theory*, 2001, **47**, (7), pp. 2845–2862

8 Donoho, D.L., Elad, M.: 'Maximal sparsity representation via l_1 minimization', *Proc. Natl. Acad. Sci.*, 2003, **100**, pp. 2197–2202

9 Elad, M., Bruckstein, A.: 'A generalized uncertainty principle and sparse representation in pairs of bases', *IEEE Trans. Inf. Theory*, 2002, **48**, (9), pp. 2558–2567

10 Chartrand, R.: 'Exact reconstruction of sparse signals via nonconvex minimization', *IEEE Signal Process. Lett.*, 2007, **14**, (10), pp. 707–710

11 Malioutov, D.M., Cetin, M., Willsky, A.S.: 'Optimal sparse representations in general overcomplete bases'. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP), May 2004, pp. 793–796

12 Kay, S.M.: 'Fundamentals of statistical signal processing: estimation theory' (Prentice-Hall, Englewood Cliffs, NJ, 1993)

13 Borden, B.: 'Maximum entropy regularization in inverse synthetic aperture radar imagery', *IEEE Trans. Signal Process.*, 1992, **40**, (4), pp. 969–973

14 Mensa, D.L.: 'High resolution radar imaging' (Artech House, Dedham, MA, 1981)

15 Munson, Jr D.C., Sanz, J.L.C.: 'Image reconstruction from frequency-offset Fourier data', *Proc. IEEE*, 1984, **72**, pp. 661–669

16 Vogel, C.R., Oman, M.E.: 'Iterative methods for total variation denoising', *SIAM J. Sci. Comput.*, 1996, **17**, (1), pp. 227–238

17 Golub, G.H., Van Loan, C.F.: 'Matrix computations' (The Johns Hopkins University Press, Baltimore, MD, 1996)

18 Çetin, M., Karl, W.C., Willsky, A.S.: 'Feature-preserving regularization method for complex-valued inverse problems with application to coherent imaging', *Opt. Eng.*, 2006, **45**, (1), p. 017003

19 Geman, D., Yang, C.: 'Nonlinear image recovery with half-quadratic regularization', *IEEE Trans. Image Process.*, 1995, **4**, (7), pp. 932–946

20 Starck, J.L., Elad, M., Donoho, D.L.: 'Image decomposition via the combination of sparse representations and a variational approach', *IEEE Trans. Image Process.*, 2005, **14**, (10), pp. 1570–1582

21 Donoho, D.L., Johnstone, I.: 'Ideal spatial adaptation via wavelet shrinkage', *Biometrika*, 1994, **81**, pp. 425–455

22 Zeng, Z., Cumming, I.G.: 'SAR image data compression using a tree-structured wavelet transform', *IEEE Trans. Geosci. Remote Sens.*, 2001, **39**, (3), pp. 546–552

23 Rilling, G., Davies, M., Mulgrew, B.: 'Compressed sensing based compression of SAR raw data'. Signal Processing with Adaptive Sparse Structured Representations Workshop, Saint-Malo, France, April 2009

24 Starck, J.-L., Candès, E., Donoho, D.L.: 'The curvelet transform for image denoising', *IEEE Trans. Image Process.*, 2002, **11**, (6), pp. 131–141

25 Batu, O., Çetin, M.: 'Hyper-parameter selection in non-quadratic regularization-based radar image formation'. SPIE Defense and Security Symp., Algorithms for Synthetic Aperture Radar Imagery XV, Orlando, FL, USA, March 2008

26 Argenti, F., Alparone, L.: 'Speckle removal from SAR images in the undecimated wavelet domain', *IEEE Trans. Geosci. Remote Sens.*, 2002, **40**, (11), pp. 2363–2374

27 Çetin, M., Karl, W.C., Castañón, D.A.: 'Feature enhancement and ATR performance using non-quadratic optimization-based SAR imaging', *IEEE Trans. Aerospace Electr. Syst.*, 2003, **39**, (4), pp. 1375–1395

28 Clark, L.G., Velten, W.J.: 'Image characterization for automatic target recognition algorithm evaluations', *Opt. Eng.*, 1991, **30**, (2), pp. 147–153

29 Benitz, G.R.: 'High-definition vector imaging', *Lincoln Lab. J.*, 1997, **10**, (2), pp. 147–170

30 Wang, J., Liu, X.: 'SAR minimum-entropy autofocus using an adaptive-order polynomial model', *IEEE Trans. Geosci. Remote Sens.*, 2006, **3**, (4), pp. 512–516

31 Air Force Research Laboratory, Model Based Vision Laboratory, Sensor Data Management System ADTS: <http://www.mbvlab.wpafb.af.mil/public/sdms/datasets/adts/>

32 Backhoe Data Sample & Visual-D Challenge Problem, Air Force Research Laboratory, Sensor Data Management System: <https://www.sdms.af.mil/main.htm>

33 Moses, R.L., Potter, L.C., Cetin, M.: 'Wide angle SAR imaging'. Proc. SPIE, Algorithms for Synthetic Aperture Radar Imagery XI, Orlando, FL, USA, April 2004, pp. 164–175

34 Papoulis, A.: 'Probability, random variables, and stochastic processes' (McGraw-Hill, 1991, 3rd edn.)

7 Appendix 1

7.1 Derivation of the PDF of β

Consider the complex random variable $(\beta)_i = u_i + j v_i = r_i e^{j\phi_i}$. Here, we know that $r_i = 1$ independent of ϕ_i , and ϕ_i possesses a uniform pdf over $|\phi_i| < \pi$ which we denote it by $U(\phi_i)$. The pdf of the complex-valued random variable $(\beta)_i$ can be defined as [34]

$$p((\beta)_i) = p(u_i, v_i) = \frac{p(r_i, \phi_i)}{|J(r_i, \phi_i)|} \quad (39)$$

where the Jacobian $J(r_i, \phi_i) = r_i$. Since r_i and ϕ_i are independent

$$p(r_i, \phi_i) = p(r_i)p(\phi_i) = \delta(r_i - 1)U(\phi_i) \quad (40)$$

Thus, $p((\beta)_i)$ can be written as

$$\begin{aligned} p((\beta)_i) &= p(u_i, v_i) = \frac{p(r_i, \phi_i)}{|J(r_i, \phi_i)|} = \frac{1}{r_i} \delta(r_i - 1)U(\phi_i) \\ &= \delta(r_i - 1)U(\phi_i) = \delta(|(\beta)_i| - 1)U(\angle(\beta)_i) \end{aligned} \quad (41)$$

Since $|(\beta)_i| = 1$, $\forall i$ independent of each other and the phase terms, and also since all ϕ_i random variables are assumed to be independent, (12) follows.

8 Appendix 2

Let us differentiate $K_\alpha(\alpha, \mathbf{b})$ with respect to \mathbf{b}_i and set it to zero. We obtain the following relationship for the optimum (minimising) \mathbf{b}

$$\mathbf{b}_i = \frac{P}{2[|(\alpha)_i|^2 + \varepsilon]^{1-p/2}} \quad (42)$$

Substituting (42) into (24) we obtain

$$\inf_{\mathbf{b}} K_\alpha(\alpha, \mathbf{b}) = \|\mathbf{y} - \mathbf{HP}\Phi\alpha\|_2^2 + \lambda \sum_{i=1}^{N^2} [|(\alpha)_i|^2 + \varepsilon]^{p/2} = J(\alpha) \quad (43)$$

9 Appendix 3

Let us define the sequence $K_n = K_\alpha(\hat{\alpha}^{(n)}, \hat{\mathbf{b}}^{(n+1)})$ and show that it is convergent. From (25) and (26) we can deduce that

$$K_\alpha(\hat{\alpha}^{(n)}, \hat{\mathbf{b}}^{(n+1)}) \leq K_\alpha(\hat{\alpha}^{(n)}, \hat{\mathbf{b}}^{(n)}) \quad \forall n \quad (44)$$

$$K_\alpha(\hat{\alpha}^{(n+1)}, \hat{\mathbf{b}}^{(n+1)}) \leq K_\alpha(\hat{\alpha}^{(n)}, \hat{\mathbf{b}}^{(n+1)}) \quad \forall n \quad (45)$$

The difference $K_n - K_{n-1}$ can be written as

$$K_n - K_{n-1} = [K_\alpha(\hat{\alpha}^{(n)}, \hat{\mathbf{b}}^{(n+1)}) - K_\alpha(\hat{\alpha}^{(n)}, \hat{\mathbf{b}}^{(n)})] + [K_\alpha(\hat{\alpha}^{(n)}, \hat{\mathbf{b}}^{(n)}) - K_\alpha(\hat{\alpha}^{(n-1)}, \hat{\mathbf{b}}^{(n)})] \quad (46)$$

Using (44) and (43) we obtain

$$K_n - K_{n-1} \leq 0 \quad \forall n \quad (47)$$

which means that the sequence K_n is decreasing. Since it is bounded below and decreasing, the sequence converges. Hence the algorithm is convergent in terms of the cost functional.

10 Appendix 4

We want to find \mathbf{s} that minimises $K_\beta(\boldsymbol{\beta}, \mathbf{s})$ of (31). The portion of $K_\beta(\boldsymbol{\beta}, \mathbf{s})$ that depends on \mathbf{s} is its second term, as follows

$$\sum_{i=1}^{N^2} |(\mathbf{S}\boldsymbol{\beta})_i - 1|^2 \quad (48)$$

$$\begin{aligned} |(\mathbf{S}\boldsymbol{\beta})_i - 1|^2 &= |\exp(-js_i)(\boldsymbol{\beta})_i - 1|^2 \\ &= |(\boldsymbol{\beta})_i|^2 + 1 - 2\Re[(\boldsymbol{\beta})_i \exp(j\{\phi[(\boldsymbol{\beta})_i] - s_i\})] \end{aligned} \quad (49)$$

where $\phi[(\boldsymbol{\beta})_i]$ denotes the phase of the complex number $(\boldsymbol{\beta})_i$. The sum in (48) takes its minimum value when the term inside the bracket in (49) has a zero imaginary part for all i . Therefore the \mathbf{s} that minimises the sum in (48) and hence (31) satisfies

$$s_i = \phi[(\boldsymbol{\beta})_i] \quad \forall i \quad (50)$$

With this \mathbf{s} , we have $\mathbf{S}\boldsymbol{\beta} = |\boldsymbol{\beta}|$, and hence

$$\inf_{\mathbf{s}} K_\beta(\boldsymbol{\beta}, \mathbf{s}) = \|\mathbf{y} - \mathbf{H}\mathbf{B}\boldsymbol{\beta}\|_2^2 + \lambda' \sum_{i=1}^{N^2} (|(\boldsymbol{\beta})_i| - 1)^2 = J(\boldsymbol{\beta}) \quad (51)$$

11 Appendix 5

Considering that $\mathbf{B} = \text{diag}\{\Phi\boldsymbol{\alpha}\}$ and $\mathbf{P} = \text{diag}\{\boldsymbol{\beta}\}$, the cost functions of both (8) and (14) are actually functions of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Note that the first part (ℓ_2 -norm data dependent term) of both cost functions are equal for given $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, so we denote the data term by $J_1(\boldsymbol{\alpha}, \boldsymbol{\beta})$. Based on that, let us define a total cost function as below

$$J_t(\boldsymbol{\alpha}, \boldsymbol{\beta}) = J_1(\boldsymbol{\alpha}, \boldsymbol{\beta}) + \lambda \|\boldsymbol{\alpha}\|_p^p + \lambda' \sum_{i=1}^{N^2} (|(\boldsymbol{\beta})_i| - 1)^2 \quad (52)$$

Note that finding the minimising $\boldsymbol{\alpha}$ for $J_t(\boldsymbol{\alpha}, \boldsymbol{\beta})$ reduces to the optimisation problem in (8), and finding the minimising $\boldsymbol{\beta}$ for $J_t(\boldsymbol{\alpha}, \boldsymbol{\beta})$ reduces to the optimisation problem in (14). Thus, the overall image reconstruction algorithm of Section 2.2 is actually a block coordinate descent approach as follows

$$\hat{\boldsymbol{\alpha}}^{(l+1)} = \arg \min_{\boldsymbol{\alpha}} J_t(\boldsymbol{\alpha}, \hat{\boldsymbol{\beta}}^{(l)}) \quad (53)$$

$$\hat{\boldsymbol{\beta}}^{(l+1)} = \arg \min_{\boldsymbol{\beta}} J_t(\hat{\boldsymbol{\alpha}}^{(l+1)}, \boldsymbol{\beta}) \quad (54)$$

where l denotes the iteration number of the overall algorithm. Using (53) and (54), convergence of the overall algorithm in terms of cost functional can be easily established in a similar way to the one in Appendix 3.