

A robust compressive sensing based technique for reconstruction of sparse radar scenes [☆]



Oguzhan Teke ^{a,*}, Ali Cafer Gurbuz ^{b,*}, Orhan Arikan ^a

^a Department of Electrical and Electronics Engineering, Bilkent University, Ankara, Turkey

^b TOBB University of Economics and Technology, Ankara, 06560 Turkey

ARTICLE INFO

Article history:

Available online 31 December 2013

Keywords:

Compressive sensing
Off-grid
Radar
Delay–Doppler
Perturbation

ABSTRACT

Pulse-Doppler radar has been successfully applied to surveillance and tracking of both moving and stationary targets. For efficient processing of radar returns, delay–Doppler plane is discretized and FFT techniques are employed to compute matched filter output on this discrete grid. However, for targets whose delay–Doppler values do not coincide with the computation grid, the detection performance degrades considerably. Especially for detecting strong and closely spaced targets this causes miss detections and false alarms. This phenomena is known as the off-grid problem. Although compressive sensing based techniques provide sparse and high resolution results at sub-Nyquist sampling rates, straightforward application of these techniques is significantly more sensitive to the off-grid problem. Here a novel parameter perturbation based sparse reconstruction technique is proposed for robust delay–Doppler radar processing even under the off-grid case. Although the perturbation idea is general and can be implemented in association with other greedy techniques, presently it is used within an orthogonal matching pursuit (OMP) framework. In the proposed technique, the selected dictionary parameters are perturbed towards directions to decrease the orthogonal residual norm. The obtained results show that accurate and sparse reconstructions can be obtained for off-grid multi target cases. A new performance metric based on Kullback–Leibler Divergence (KLD) is proposed to better characterize the error between actual and reconstructed parameter spaces. Increased performance with lower reconstruction errors are obtained for all the tested performance criteria for the proposed technique compared to conventional OMP and ℓ_1 minimization techniques.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

In many engineering and science applications the objective is to reconstruct an image or a map of the underlying sensed distribution from available set of measurements. Specifically in radar imaging a spatial map of reflectivity is reconstructed from measurements of scattered electric field. State of the art radar systems operate with large bandwidths or high number of channels which generate very large data sets for processing. On the other hand in most of the radar applications the reflectivity scene consists of small number of strong targets. In both cases, significant amount of data is processed mainly to estimate delay and Doppler of relatively few targets. This point raises the applicability of sparse signal processing techniques for radar signal processing.

The emerging field of Compressive Sensing (CS) [1–3] is a recently developed mathematical framework in which the primary

interest is to invert or reconstruct a signal \mathbf{x} from noisy linear measurements \mathbf{y} in the form $\mathbf{y} = \Phi\mathbf{x} + \mathbf{n}$. The focus of CS is to solve this linear problem in the underdetermined case where number of measurements is less than the number of unknowns which is very important in decreasing the required amount of data to tolerable levels in radar applications. For a signal \mathbf{x} of dimension N that has a K -sparse representation in a transform domain Ψ , as $\mathbf{x} = \Psi\mathbf{s}$ and $\|\mathbf{s}\|_0 = K$, CS techniques enable reliable reconstruction of the sparse signal \mathbf{s} , hence \mathbf{x} from $O(K \log N)$ measurements by solving a convex ℓ_1 optimization problem of the following form:

$$\min \|\mathbf{s}\|_1, \quad \text{subject to } \|\mathbf{y} - \Phi\Psi\mathbf{s}\|_2 < \epsilon. \quad (1)$$

CS theory provides strong results which guarantee stable solution of the reconstructed sparse signal for a forward matrix $\mathbf{A} = \Phi\Psi$ if it satisfies the restricted isometry property (RIP) [4–6]. It has also been shown that random measurement matrices Φ with i.i.d. entries guarantees the RIP of \mathbf{A} for known basis [7].

Due to these appealing properties of CS and its important advantages for radar, recently CS has received considerable attention in the radar research community. In one of the earliest papers on CS applied to radar, the possibility of sub-Nyquist sampling and

[☆] This work was supported by TUBITAK grant with project number 109E280 and within FP7 Marie Curie IRG grant with project number PIRG04-GA-2008-239506.

* Corresponding author.

E-mail addresses: teke@ee.bilkent.edu.tr (O. Teke), acgurbuz@etu.edu.tr (A.C. Gurbuz), oarikan@ee.bilkent.edu.tr (O. Arikan).

elimination of match filtering has been discussed [8]. In [9,10], experimental radar imaging results for step frequency and impulse ground penetrating radars have been provided and later extended in [11,12]. To exploit sparsity in the time-frequency domain, high resolution CS-radar has been proposed in [13]. CS based SAR image reconstruction techniques have been proposed in [14]. A CS based MIMO radar has been proposed in [15] for obtaining simultaneous angle and Doppler information. In [16], CS is investigated in distributed radar sensor networks. Further information on the CS based radar applications can be found in [17] and [18].

All the above mentioned sparse reconstruction techniques mainly discretize a continuous parameter space such as range, Doppler or angle and generate a number of grid points where the targets are assumed to be positioned on the nodes of the grid. Under this assumption, the sparsity requirement of CS theory is satisfied and the CS techniques provide satisfactory results. Unfortunately, no matter how fine the grid is, the targets are typically located in off-grid positions. It has been discussed in literature that the off-grid targets creates an important degradation in CS reconstruction performance [19–24]. Off-grid problem is not only observed in CS based radar but many other application areas such as target localization [25], beamforming [26] or shape detection [27], where the sparsity of the signal is in a continuous parameter space and the sparsity basis Ψ is constructed through discretization or gridding of this parameter space.

To reduce the sensitivity of the reconstruction to the off-grid targets, denser grids can be used. However, decreasing grid dimensions causes significant increase in the coherence of the compressive sensing dictionary, beyond a certain limit which causes loss of the RIP [7]. To avoid this problem of increased coherence between dictionary columns, in [28], the dictionary is extended to several orthogonal dictionaries and not in a single dictionary, but in a set of them by using a tree structure, assuming that the given signal is sparse in at least one of them. However, this strategy depend on several set of fixed dictionaries generated through discrete parametrization and the main goal is to select the best set of fixed atoms from all dictionaries rather than focusing on basis mismatch. In the works [21–23] the effect of basis mismatch problem on the reconstruction performance of CS is analyzed and the resultant performance degradation levels and analytical ℓ_2 norm error bounds are given. However these works have not offered a systematic approach for sparse reconstruction under parametric perturbations.

There are several approaches in literature for the basis mismatch problem. In Continuous Basis Pursuit approach [29], recovery of sparse translation invariant signals is performed and perturbations are assumed to be continuously shifted features of the functions on which sparse solution is searched for. A dictionary that includes auxiliary interpolation functions that approximates translates of features via adjustment of their parameters is generated and ℓ_1 based minimization is used on primary coefficients. In [24], an algorithm based on the atomic norm minimization is proposed and the solution is found with a semi-definite programming. In [30], ℓ_1 minimization based algorithms are proposed for linear structured perturbations on the sensing matrix where perturbation vectors are modeled as an unknown constant multiplied by a known vector which specifically defines the direction which is typically unknown in practice. Works based on total least square (TLS) as [31,32] assume that general perturbations appear both on the dictionary and measurements. In [31] for solving TLS problem an optimization over all signal \mathbf{x} , perturbation matrix \mathbf{P} and error vector spaces is performed. To reduce complexity, suboptimal optimization techniques have also been proposed. In [32] a constrained total least squares technique is introduced assuming dictionary mismatches are constrained by errors of grid points and a joint estimate of grid point errors and signal support is found by

general TLS techniques. In [33], non-parametric perturbations in a bounded perturbation space is considered and some reconstruction guarantees are provided.

This paper mainly focuses on reconstruction of sparse parameter scenes and proposes a novel parameter perturbation based adaptive sparse reconstruction technique to provide robust reconstructions in the off-grid case. The proposed technique is an iterative algorithm that works with a selected set of dictionary vectors that can be obtained via one of sparse greedy techniques such as matching pursuit (MP) [34], orthogonal matching pursuit (OMP) [35], iterative hard/soft thresholding (IHT) [36] or the compressive sampling matching pursuit (CoSaMP) [37]. The parameters of the selected dictionary atoms are iteratively adapted within their grids towards directions that decreases the residual norm. The proposed technique presently is used within the general OMP framework hence named as parameter perturbed OMP (PPOMP). As demonstrated in the reconstruction of sparse delay–Doppler radar scenes, the proposed method is successful in recovering the targets with arbitrary positions. Compared to conventional CS reconstruction techniques like OMP or ℓ_1 minimization, proposed PPOMP technique has achieved lower reconstruction errors for a general delay–Doppler scene in all the conducted performance tests. The general idea of proposed parameter perturbation can also be applied to other areas where discrete parameters are selected from continuous parameter spaces such as frequency or angle of arrival estimation problems.

The organization of the paper is as follows. Section 2 outlines the delay and Doppler data model and formulates the sparse reconstruction problem in CS framework. The proposed parameter perturbation technique and the PPOMP algorithm is detailed in Section 3. Simulation results on variety of examples with performance comparisons are given in Section 4. Section 5 covers conclusions, and direction of possible future work.

2. Delay–Doppler radar imaging: data model and formulation

Coherent radar systems transmit a sequence of pulses with known phases and processes the received echoes to perform clutter suppression and detection at each angle of interest. Excellent references on the operation of radar receivers are available in the literature [38,39]. In this paper we consider a classical pulse Doppler radar with a co-located receiver and a transmitter. Although it is not investigated in here, MIMO radar systems can also be considered within CS framework [15,40]. Let radar transmits $s(t)$, a coherent train of N_p pulses:

$$s(t) = \sum_{i=0}^{N_p-1} p(t - iT_{PRI})e^{j2\pi f_c t}, \quad (2)$$

where, $p(t)$ is the individual pulse waveform, T_{PRI} is the uniform pulse repetition interval and f_c is the radar carrier frequency. Assuming K dominant targets with delays of τ_{T_m} and Doppler shifts of ν_{T_m} , $1 \leq m \leq K$, the received signal following the baseband down-conversion can be expressed as:

$$y(t) = \sum_{m=1}^K \alpha_m s(t - \tau_{T_m})e^{j2\pi \nu_{T_m} t} + n(t), \quad (3)$$

where α_m is the complex reflectivity of the individual targets and $n(t)$ is the measurement noise. The above relation between the received signal and target parameters are expressed in terms of the measurable quantities of delay and Doppler. These parameters are related to the range and radial velocity of the m th target as:

$$\tau_{T_m} = \frac{2R_m}{c}, \quad \nu_{T_m} = \frac{2f_c}{c} v_m, \quad (4)$$

where R_m is the range and v_m is the radial velocity of the m th target. At this point, the common practice is to employ matched filtering to individual uniformly spaced samples of pulse returns and use FFT across the delay aligned matched filter outputs. This way the returns are compressed in time and frequency sequentially [39]. In compressive sensing (CS) formulation, a sampled version of the measurement relation given in (3) is adapted to a linear matrix–vector relationship in delay–Doppler domain. For this purpose 2 dimensional delay–Doppler domain which lies in the product space $[\tau_0, \tau_f] \times [v_0, v_f]$ must be discretized where initial and final values of τ_0 and τ_f are determined by the range and v_0 and v_f are determined by the velocity of the potential targets. Discretization generates a finite set of N target points $\mathcal{B} = \{\theta_1, \theta_2, \dots, \theta_N\}$, where each θ_j representing a grid node of (τ_j, v_j) . For each grid node θ_j the data model can be calculated from (3) as:

$$\psi_j = s(\mathbf{t} - \tau_j) \circ e^{j2\pi v_j \mathbf{t}}, \quad (5)$$

where $\mathbf{t} \in \mathbb{R}^{N_t \times 1}$ is the vector holding the time samples and operator “ \circ ” corresponds to Hadamard product. N_t is the number of time samples.

Repeating (5) at each (τ_j, v_j) generates the dictionary Ψ where the j th column of Ψ is ψ_j . The size of the dictionary Ψ becomes $N_t \times N$. If the true target parameters (τ_{T_m}, v_{T_m}) falls exactly on the grid points (τ_j, v_j) then a linear system of equations can be formed as:

$$\mathbf{y}_s = \Psi \mathbf{x} + \mathbf{n}, \quad (6)$$

where \mathbf{y}_s is the sampled measurement vector and \mathbf{x} is a reflectivity vector defining the delay–Doppler space, i.e., if there is a target at θ_j , the value of the j th element of \mathbf{x} should be α_j , otherwise zero. If there are K targets in the scene then the vector \mathbf{x} should be a K sparse vector, that is $\|\mathbf{x}\|_0 = K$. Since actual target positions deviate from the grid centers, (6) is an approximate relationship. Nevertheless, CS framework uses this linear relationship hoping that the noise term \mathbf{n} compensates for any errors due to the discretization of the parameter space, modeling errors and the actual noise of the measurements. In the CS formulation, a very small fraction of the samples obtained at the Nyquist rate carry enough information to represent a sparse signal. Thus a sub-Nyquist sampling can be done and a random subset of M measurements at random times t_s can be measured in CS. In general these new measurements can be represented as $\mathbf{b} = \Phi \mathbf{y}_s$ where Φ is an $M \times N_t$, $M < N_t$ measurement matrix constructed by randomly selecting M rows of an $N_t \times N_t$ identity matrix. The general linear relation is then:

$$\mathbf{b} = \Phi \Psi \mathbf{x} + \mathbf{n} = \mathbf{A} \mathbf{x} + \mathbf{n}. \quad (7)$$

The reflectivity vector \mathbf{x} estimated by the solution to the following constrained ℓ_1 minimization problem,

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{b} - \mathbf{A} \mathbf{x}\|_2 \leq \epsilon. \quad (8)$$

To reduce the computational load, suboptimal greedy algorithms such as MP [34], OMP [35], CoSaMP [37] or IHT [36] are also used in many applications. In the following section, the proposed parameter perturbation technique is introduced within the OMP framework.

3. Parameter perturbation for delay–Doppler reconstruction

Sparse representation of a target scene in delay–Doppler domain requires identification of grid nodes at which the targets are present. This is equivalent to the identification of the support set of the target scene among the columns of \mathbf{A} defined in (7). OMP

method starts with an empty support set and the measured radar signal as the initial residual. Iteratively, the most correlated column of \mathbf{A} with the current residual is added to the support list, increasing the span of the current support at each iteration. Then, projection of the measurements onto the current support is computed to obtain an estimate at that iteration. This procedure is repeated until the residual norm is less than a given tolerance level of ϵ or a predetermined sparsity level is reached. Note that at the k th iteration of the OMP algorithm, the measured signal can be represented in a k -sparse manner as a linear combination of the k support vectors as:

$$\mathbf{b} = \mathbf{b}_\perp + \sum_{i=1}^k \alpha_i \mathbf{a}(\theta_i), \quad (9)$$

where \mathbf{b}_\perp is the orthogonal residual of \mathbf{b} within the span of the k chosen support vectors $\mathbf{a}(\theta_i)$, $i = 1, \dots, k$ and $\mathbf{a}(\theta_i)$ is a column of dictionary \mathbf{A} with grid parameters θ_i . Hence at each iteration of OMP, the vectors in the support set, their coefficients α_i , and the orthogonal residual, \mathbf{b}_\perp , are obtained.

In general, a target with parameters (τ_{T_i}, v_{T_i}) may not be located at a grid node but is positioned within a grid area with an unknown perturbation from the closest grid node as:

$$\tau_{T_i} = \tau_i + \delta\tau_i \quad \text{and} \quad v_{T_i} = v_i + \delta v_i, \quad (10)$$

where (τ_i, v_i) are the nearest grid node parameters, $|\delta\tau_i| < \Delta_\tau/2$ and $|\delta v_i| < \Delta_v/2$ with Δ_τ and Δ_v defining the grid dimensions in delay and Doppler, respectively. If there were no noise in (7), the measurement vector \mathbf{b} would be in the span of $\mathbf{a}(\tau_{T_i}, v_{T_i})$. Our goal is to perturb the selected grid parameters, hence the corresponding column vectors in \mathbf{A} , so that a better fit to the measurements can be accomplished. This goal can be formulated as the following optimization problem:

$$\arg \min_{\alpha_i, \delta\tau_i, \delta v_i} \left\| \mathbf{b} - \sum_{i=1}^k \alpha_i \mathbf{a}(\tau_i + \delta\tau_i, v_i + \delta v_i) \right\|_2$$

s.t. $|\delta\tau_i| < \Delta_\tau/2, \quad |\delta v_i| < \Delta_v/2. \quad (11)$

Solution of the problem in (11) provides the perturbation parameters $\delta\theta_i = (\delta\tau_i, \delta v_i)$ and the representation coefficients α_i for the selected set of k column vectors.

Assume that there exist a solver for the problem, namely $\mathbb{S}(\cdot)$, which takes the measurement vector \mathbf{b} and the initial grid points, then returns the solution of the problem in (11). In an abstract sense, this solver can be written as:

$$(\boldsymbol{\alpha}, [\delta\theta_1 \dots \delta\theta_k]) = \mathbb{S}(\mathbf{b}, [\theta_1 \dots \theta_k]). \quad (12)$$

Note that solver $\mathbb{S}(\cdot)$ is not dependent on the OMP technique itself. Therefore, it is possible to integrate $\mathbb{S}(\cdot)$ into any algorithm that provides a suitable estimation of the grid points. In this study OMP is preferred due to its simplicity. When such a solver is utilized within the OMP iterations, an “ideal” parameter perturbed OMP (I-PPOMP) procedure, which is provided in Table 1, can be implemented.

Since the optimization problem defined in (11) is non-convex, it may not be possible to obtain an ideal solver as specified in (12). Hence we propose to use a gradient descent optimization of the cost function in (11). Therefore starting from the grid nodes, selected parameters will be gradually perturbed until a convergence criteria is met. To simplify the iterations further α_i 's and $\delta\theta_i = (\delta\tau_i, \delta v_i)$'s will be sequentially updated in the following way:

First initialize $\theta_{i,1} = \theta_i = (\tau_i, v_i)$, $i = 1, \dots, k$, to grid centers and obtain an initial representation coefficient vector $\boldsymbol{\alpha}_1$ as:

Table 1

Ideal Parameter Perturbed-OMP (I-PPOMP) Algorithm.

Inputs: $(\mathbf{A}, \mathbf{b}, \epsilon)$
Initialization: $\mathbf{b}_{\perp,0} = \mathbf{b}$, $\mathbf{T}_0 = \{\}$, $e = \ \mathbf{b}_{\perp,0}\ _2$, $k = 1$
Keep iterating until $e < \epsilon$
$j^* = \arg \max_{1 \leq j \leq N} \mathbf{a}(\theta_j)^H \mathbf{b}_{\perp,k-1} $
$\mathbf{T}_k = \mathbf{T}_{k-1} \cup \{\theta_{j^*}\}$
$(\boldsymbol{\alpha}, [\delta\theta_1 \dots \delta\theta_k]) = \mathbb{S}(\mathbf{b}, \mathbf{T}_k)$
$\mathbf{b}_{\perp,k} = \mathbf{b} - \sum_{i=1}^k \alpha_i \mathbf{a}(\theta_i + \delta\theta_i)$
$e = \ \mathbf{b}_{\perp,k}\ _2$
$k = k + 1$
Output: $(\boldsymbol{\alpha}, [\delta\theta_1 \dots \delta\theta_k], \mathbf{T}_k)$

$$\boldsymbol{\alpha}_1 = \arg \min_{\boldsymbol{\alpha}} \left\| \mathbf{b} - \sum_{i=1}^k \alpha_i \mathbf{a}(\theta_{i,1}) \right\|_2. \quad (13)$$

Starting from $l = 1$, until convergence, perform updates:

$$\boldsymbol{\theta}_{i,l+1} = \boldsymbol{\theta}_{i,l} + \delta\boldsymbol{\theta}_{i,l},$$

where l represents the perturbation index, i represents the target index and

$$[\delta\theta_{1,l} \dots \delta\theta_{k,l}] = \arg \min_{\substack{\delta\tau_i: |\delta\tau_i| \leq \Delta\tau/2 \\ \delta\nu_i: |\delta\nu_i| \leq \Delta\nu/2}} \left\| \mathbf{b} - \sum_{i=1}^k \alpha_i \mathbf{a}(\tau_{i,l} + \delta\tau_i, \nu_{i,l} + \delta\nu_i) \right\|_2, \quad (14a)$$

$$\boldsymbol{\alpha}_l = \arg \min_{\boldsymbol{\alpha}} \left\| \mathbf{b} - \sum_{i=1}^k \alpha_i \mathbf{a}(\boldsymbol{\theta}_{i,l}) \right\|_2. \quad (14b)$$

The problem defined in (14b) is a standard least squares formulation, however obtaining solution to the constrained nonlinear optimization problem in (14a) is not practical for radar applications. Linearization of the cost function in (14a) around $\boldsymbol{\theta}_{i,l} = (\tau_{i,l}, \nu_{i,l})$ significantly reduces the complexity of the optimization. For this purpose, $\mathbf{a}(\tau_{i,l} + \delta\tau_i, \nu_{i,l} + \delta\nu_i)$ can be approximated by using the first order Taylor series as:

$$\mathbf{a}(\tau_{i,l} + \delta\tau_i, \nu_{i,l} + \delta\nu_i) \approx \mathbf{a}(\tau_{i,l}, \nu_{i,l}) + \frac{\partial \mathbf{a}}{\partial \tau_{i,l}} \delta\tau_i + \frac{\partial \mathbf{a}}{\partial \nu_{i,l}} \delta\nu_i. \quad (15)$$

By using (15), and ignoring the constraints on the perturbations, problem in (14a) can be re-written as:

$$[\delta\theta_{1,l} \dots \delta\theta_{k,l}] = \arg \min_{\mathbf{u}} \|\mathbf{r}_l - \mathbf{B}_l \mathbf{u}\|_2, \quad (16)$$

where $\mathbf{r}_l = \mathbf{b} - \sum_{i=1}^k \alpha_i \mathbf{a}(\boldsymbol{\theta}_{i,l})$ is the orthogonal residual from the least squares in (14b), $\mathbf{B}_l \in \mathbb{C}^{M \times 2k}$ is the matrix holding the weighted partial derivatives at the linearization point and is defined as:

$$\mathbf{B}_l = \begin{bmatrix} \Delta\tau \alpha_{1,l} \frac{\partial \mathbf{a}}{\partial \tau_{1,l}}, \dots, \Delta\tau \alpha_{k,l} \frac{\partial \mathbf{a}}{\partial \tau_{k,l}}, \\ \Delta\nu \alpha_{1,l} \frac{\partial \mathbf{a}}{\partial \nu_{1,l}}, \dots, \Delta\nu \alpha_{k,l} \frac{\partial \mathbf{a}}{\partial \nu_{k,l}} \end{bmatrix}, \quad (17)$$

and $\mathbf{u} = [\delta\tau_1, \dots, \delta\tau_k, \delta\nu_1, \dots, \delta\nu_k]^T \in \mathbb{R}^{2k \times 1}$ is the dummy vector variable containing updates in the l th iteration on the corresponding parameters. Each partial derivative in \mathbf{B}_l is scaled by its corresponding grid size so that corresponding updates become unitless. Notice that \mathbf{B}_l is different in each iteration since the linearization points $\boldsymbol{\theta}_{i,l}$ are updated. A new linearization is made at each updated parameter point.

Due to errors in linearization, direct solution of (16) will not produce the desired parameter perturbations. Instead we adapt a

Table 2Proposed Solver $\widehat{\mathbb{S}}(\cdot)$.

Inputs: $(\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_k\}, \mathbf{b}, \boldsymbol{\mu})$
Initialize: $l = 0$, $\boldsymbol{\theta}_{i,0} = \boldsymbol{\theta}_i$ for $1 \leq i \leq k$
Until stopping condition met,
$\mathbf{A}_l = [\mathbf{a}(\boldsymbol{\theta}_{1,l}) \mathbf{a}(\boldsymbol{\theta}_{2,l}) \dots \mathbf{a}(\boldsymbol{\theta}_{k,l})]$,
$\boldsymbol{\alpha}_l = \mathbf{A}_l^\dagger \mathbf{b}$,
$\mathbf{r}_l = \mathbf{b} - \mathbf{A}_l \boldsymbol{\alpha}_l$,
$\mathbf{B}_l = [\Delta\tau \alpha_{1,l} \frac{\partial \mathbf{a}}{\partial \tau_{1,l}}, \dots, \Delta\tau \alpha_{k,l} \frac{\partial \mathbf{a}}{\partial \tau_{k,l}}, \Delta\nu \alpha_{1,l} \frac{\partial \mathbf{a}}{\partial \nu_{1,l}}, \dots, \Delta\nu \alpha_{k,l} \frac{\partial \mathbf{a}}{\partial \nu_{k,l}}]$,
$\mathbf{g}_l = \text{Re}\{\mathbf{B}_l^H \mathbf{r}_l\}$,
For all i , $1 \leq i \leq k$,
$\tau_{i,l+1} = \tau_{i,l} + \Delta\tau \mu_{i,l} \mathbf{g}_{i,l}$,
$\nu_{i,l+1} = \nu_{i,l} + \Delta\nu \mu_{i+k,l} \mathbf{g}_{i+k,l}$,
Check if $\boldsymbol{\theta}_{i,l+1} = (\tau_{i,l+1}, \nu_{i,l+1})$ is within grid
$\delta\boldsymbol{\theta}_i = \boldsymbol{\theta}_{i,l+1} - \boldsymbol{\theta}_{i,0}$,
$l = l + 1$,
Output: $(\boldsymbol{\alpha}_l, \{\delta\boldsymbol{\theta}_1, \delta\boldsymbol{\theta}_2, \dots, \delta\boldsymbol{\theta}_k\})$

gradient descent type algorithm to solve (16) and take a small step in the direction that decreases the norm the most, i.e., direction of negative gradient. Then the new parameter point will be used in the next iteration and so on until the convergence. Let $J(\mathbf{u}) = \|\mathbf{r}_l - \mathbf{B}_l \mathbf{u}\|_2^2$ and negative of the gradient of J will be $-\nabla_{\mathbf{u}} J(\mathbf{u}) = 2\mathbf{B}_l^H (\mathbf{r}_l - \mathbf{B}_l \mathbf{u})$. Since we have intention of taking a small step from the linearization point, we need the gradient of $J(\mathbf{u})$ at $\mathbf{u} = \mathbf{0}$. Therefore, $-\nabla_{\mathbf{u}} J(\mathbf{u})|_{\mathbf{u}=\mathbf{0}} = 2\mathbf{B}_l^H \mathbf{r}_l$. Remember that both \mathbf{B}_l and \mathbf{r}_l are complex valued whereas perturbations need to be real. When solution is forced to be real, step direction is found to be as $\text{Re}\{-\nabla_{\mathbf{u}} J(\mathbf{u})|_{\mathbf{u}=\mathbf{0}}\} = \text{Re}\{2\mathbf{B}_l^H \mathbf{r}_l\}$. With this important modification, alternating gradient descent solution of the main problem in (11) can be written as:

$$\boldsymbol{\alpha}_l = [\mathbf{a}(\boldsymbol{\theta}_{1,l}) \mathbf{a}(\boldsymbol{\theta}_{2,l}) \dots \mathbf{a}(\boldsymbol{\theta}_{k,l})]^\dagger \mathbf{b}, \quad (18a)$$

$$\boldsymbol{\theta}_{i,l+1} = \boldsymbol{\theta}_{i,l} + \mu_{i,l} \text{Re}\{\mathbf{B}_l^H \mathbf{r}_l\}, \quad (18b)$$

where $\mu_{i,l}$ is the step size. To keep the updated points within the grid, the algorithm will also check that the total perturbations will not exceed the grid size. Eq. (18) defines the main update iterations of the proposed gradient based perturbation solver (GS)- $\widehat{\mathbb{S}}(\cdot)$ for (11) which is summarized in Table 2. Notice that, when $\mathbb{S}(\cdot)$ in Table 1 is replaced with the $\widehat{\mathbb{S}}(\cdot)$, proposed PPOMP algorithm is obtained.

For the gradient based parameter perturbation solver in Table 2, one could make several selections for the stopping criteria and the step size μ . It is possible to monitor the residual, \mathbf{r}_l , during the iterations, and terminate the solver in the l th iteration if the residual norm $\|\mathbf{r}_l\|_2$, or amount of decrease $\|\mathbf{r}_l\|_2 - \|\mathbf{r}_{l-1}\|_2$, or rate of decrease $\|\mathbf{r}_l\|_2 / \|\mathbf{r}_{l-1}\|_2$ is below a certain threshold. It is also possible to observe the parameters $\boldsymbol{\theta}_{i,l}$ and terminate the iterations when $|\boldsymbol{\theta}_{i,l} - \boldsymbol{\theta}_{i,l-1}|$ is below a certain threshold. Also iterations can be terminated, when norm of the gradient $\|\mathbf{B}_l^H \mathbf{r}_l\|_2$ is smaller than a given threshold value or when a maximum iteration count is reached. Several metrics can also be used in conjunction with the stopping criteria. In the presented results, the iterations were terminated when the rate of decrease of the residual is less than a certain threshold.

For the selection of step size μ there are several possibilities. It is possible to use a fixed step size μ , that is $\mu_{i,l} = \mu$. If μ is small enough, after sufficient number of iterations convergence can be achieved. A more efficient approach is to use step sizes with constant rate of decrease, that is $\mu_{i,l} = \gamma \mu_{i,l-1}$ where γ is fixed and $0 < \gamma < 1$. If the gradient of a function is Lipschitz continuous with a constant L , gradient descent steps converges to a local optima by using constant step size that satisfies $\mu < 2/L$ [41,42]. In addition, line search techniques can also be used to select locally optimal step sizes that guarantees convergence with at least linear convergence rates [41]. As shown in Appendix A, normalized form of the

nonlinear objective function in (14a) is Lipschitz continuous with $L = 10\pi^2$, therefore gradient descent is guaranteed to converge to a local minima.

In the presented results, step size is selected as $\mu_{i,l} = z\gamma_{i,l}$, where z is a pre-selected value of $z = 0.01$. For the i th point, $\gamma_{i,l}$ is the ratio of the norm of the gradient in the l th iteration to the maximum observed norm of the gradient through the perturbation iterations. As a result, $\gamma_{i,l} \leq 1$ and decreases as the norm of the gradient decreases. With this selection, smaller steps are taken as the gradient decreases when approaching a local minima. Notice that $\mu_{i,l} \leq 0.01 < 2/L \approx 0.02$, thus our selection of the step size is guaranteed to converge to a local minima.

The additional computational complexity of PPOMP algorithm compared to OMP is the calculation of the gradient directions, and this requires a matrix vector multiplication which can be performed significantly faster than solving constrained nonlinear problem in (14a). For the pulse-Doppler radar application, the required gradient computations simplify further as:

$$\begin{aligned} \frac{\partial \mathbf{a}}{\partial \tau} &= e^{j2\pi \nu t} \circ \frac{d}{d\tau} s(\mathbf{t} - \tau) = -e^{j2\pi \nu t} \circ \left. \frac{ds(t)}{dt} \right|_{t=\mathbf{t}-\tau}, \\ \frac{\partial \mathbf{a}}{\partial \nu} &= j2\pi \mathbf{t} \circ \mathbf{a}(t). \end{aligned} \quad (19)$$

Note that for pre-computed and stored $ds(t)/dt$ values, calculation of these partial derivatives require only component-wise multiplication of vectors that has M multiplications each. Hence \mathbf{B}_l can be computed efficiently and the total computational complexity of PPOMP will be in the same order as OMP algorithm due to mainly solution of least squares in both techniques.

4. Simulation results

In this section, performance of the proposed PPOMP technique is analyzed for sparse reconstruction of delay and Doppler radar scenes in the case of targets that can be arbitrarily located within the grid cells. In the simulations, a classical single receiver-single transmitter pulsed-Doppler radar transmitting a linear chirp signal $p(t)$ with bandwidth of $B = 1.5$ MHz and pulse width of $T_p = 20 \mu\text{s}$ is considered. In the coherent processing, a pulse train of $N_p = 8$ pulses are used with $T_{PRI} = 50 \mu\text{s}$. The delay and Doppler space is chosen as the maximum unambiguous ranges of $[T_p, T_{PRI} - T_p]$ in delay and $[-1/(2T_{PRI}), 1/(2T_{PRI})]$ in Doppler. To create the forward linear model the space is discretized to grids with Rayleigh resolution spacing in both parameter axis which is $\Delta_\nu = 1/(N_p T_{PRI})$ in Doppler and $\Delta_\tau = 1/(2B)$ in delay. For the simulated case this discretization creates a total of $N = 279$ grid nodes. Sparse target scene is modeled as $K = 9$ point reflectors that are generated with delay and Doppler parameters randomly selected from the defined continuous delay-Doppler space where none of them exactly coincides with the chosen grid nodes. The complex reflectivity of the parameters are selected randomly with magnitudes selected from a normal distribution of $N(5, 1)$ and phases selected uniformly from $[0, 2\pi]$. For $M = 2N/3 = 186$ randomly spaced time samples in $[0, N_p T_{PRI}]$, the received signal is computed using (3). If the samples are taken at the Nyquist rate, total number of samples is $(N_p T_{PRI})(2B) = 1200$. Therefore M corresponds to only 15% of the Nyquist rate samples. Measurement noise corresponding to an SNR of 27.3 dB is added to the computed time samples. Here SNR is defined as $20 \log_{10}(\|\mathbf{b}_0\|_2 / \|\sigma \mathbf{n}\|_2)$ where $\sigma \mathbf{n}$ is the noise component in the measurements.

The actual target reflectivity and its reconstruction by the proposed PPOMP technique are shown in Fig. 1(a) and (b), respectively. It can be seen that even for off-grid targets, PPOMP could provide accurate reconstruction of the sparse target scene. Note that PPOMP result is obtained in the absence of prior information about the target sparsity level. OMP technique using the same

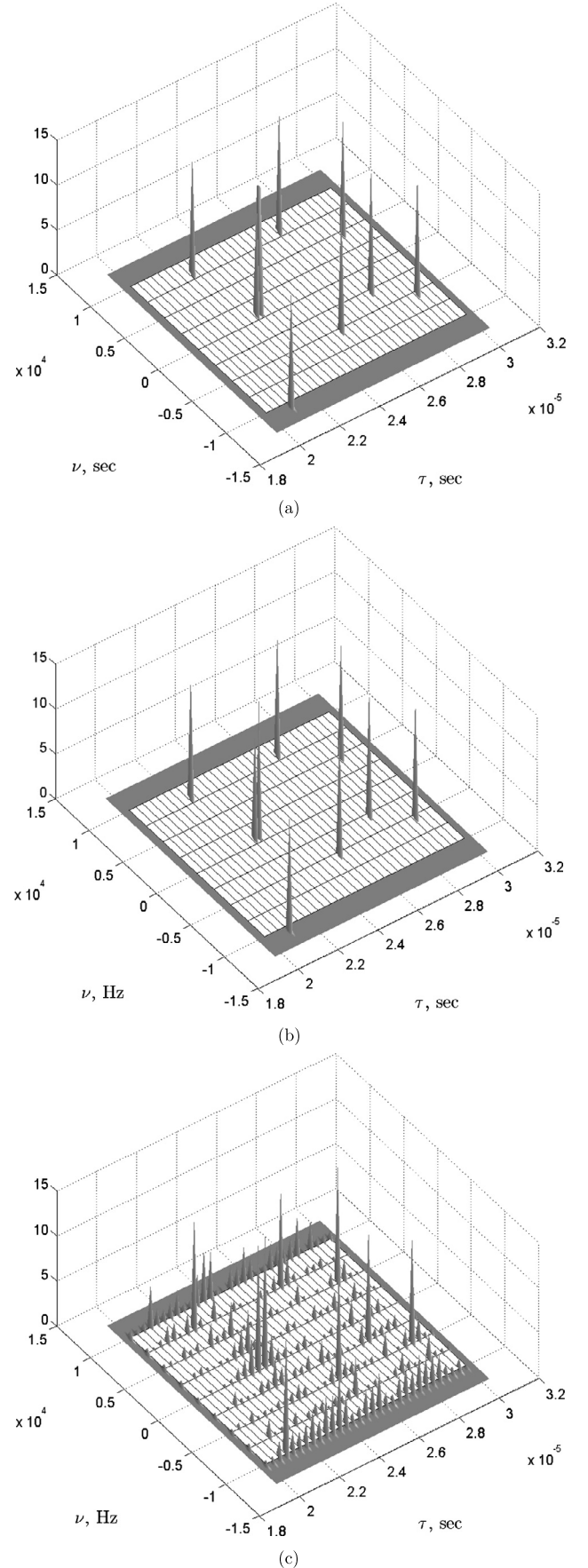


Fig. 1. (a) True delay-Doppler space reflectivity with $K = 9$ off the grid targets, (b) PPOMP reconstruction result, (c) OMP reconstruction result.

measurements and the same termination criteria with PPOMP generated the result shown in Fig. 1(c). Due to presence of off grid targets, OMP generates large number of significant peaks resulting in excessively many false target detections even at high level of detection threshold.

Fig. 2 shows the same simulation result as a 2D image with underlying grids and their centers. It can be seen that PPOMP could find all the target parameters very close to their actual values. Fig. 3(a) shows the gradient based steps taken for one of the targets starting from the grid center. It can be seen that with decreasing step sizes the algorithm converges to the actual target parameters. Similarly, Fig. 3(b) shows gradient steps taken for two closely spaced targets. Note that the separation of these two targets is closer than a grid size corresponding to the classical Rayleigh resolution limit both in delay and Doppler axis. While a matched filter won't be able to resolve these two targets, the proposed PPOMP technique could detect and identify their actual parameters accurately. This shows the high resolution capability of the proposed PPOMP technique, which is an attribute of other sparse signal reconstruction techniques as well [13,43]. Here this phenomenon is also observed for off-grid targets.

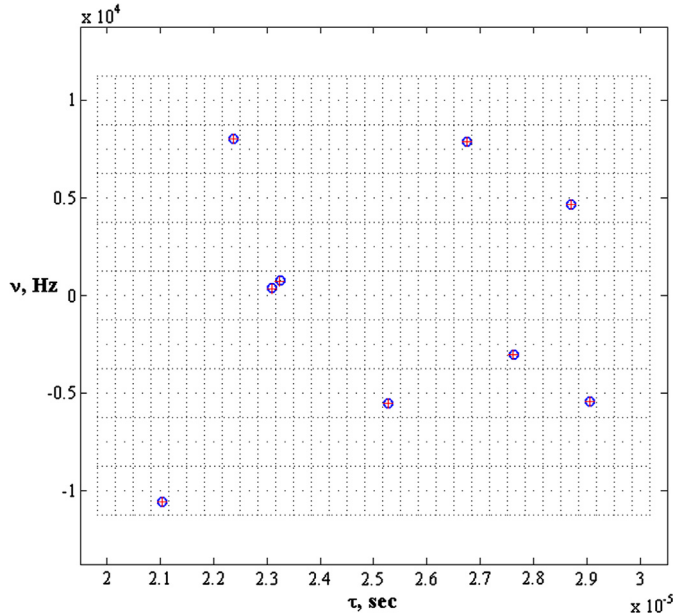


Fig. 2. Actual and reconstructed target positions in the delay-Doppler domain. Circles (o) correspond to the actual target parameters where plus signs (+) correspond to the reconstructed target parameters by the proposed PPOMP technique.

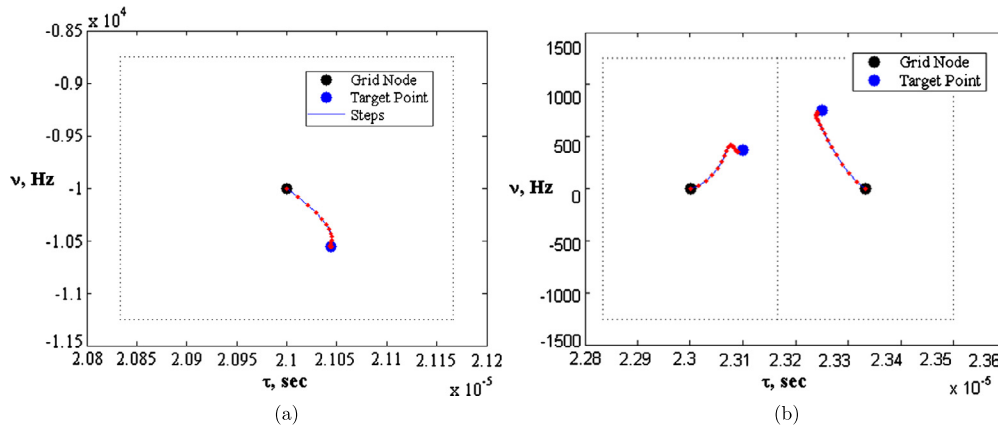


Fig. 3. Gradient based steps taken within the PPOMP algorithm at (a) one of the target grids, with (b) two targets grids where the two target parameters are closer than a grid size in both τ and ν . Grid node corresponds to a discretized point as in (5) and target point corresponds to the actual off-grid target point.

In this part of the simulations, the average performance and robustness of the proposed technique is investigated as a function of sparsity, number of measurements and SNR levels. One of the important problems of standard CS based reconstruction techniques is that in the presence of off-grid targets, they tend to generate non-sparse reconstructions. In such a case, the reconstruction error should be carefully defined. One approach would be to match the closest points in the correct scene and the reconstructed one, then compute the parameter error between them. However, when sparsity levels do not match, this error criterion is not appropriate. Hence it is a necessity to find a suitable metric in order to compare the parameter estimation performance of the sparse reconstruction techniques. Here, we propose to use Kullback–Leibler Divergence (KLD) between the actual and reconstructed target scenes, which is defined as follows:

$$D(p \parallel q) \triangleq \int_{-\infty}^{\infty} p(\mathbf{x}) \ln \left(\frac{p(\mathbf{x})}{q(\mathbf{x})} \right) d\mathbf{x}, \quad (20)$$

where $p(\mathbf{x})$ and $q(\mathbf{x})$ are probability density functions of the corresponding scenes. Even though a given target scene has no probabilistic behavior, we can consider it as a 2-dimensional Gaussian Mixture Model (GMM), where each mixture element has the following covariance matrix:

$$\mathbf{C} = \begin{pmatrix} \Delta_{\tau}^2 & 0 \\ 0 & \Delta_{\nu}^2 \end{pmatrix}, \quad (21)$$

where Δ_{τ} and Δ_{ν} are the resolutions of delay-Doppler grid. Hence, if a scene has K targets with parameters τ_{T_i} and ν_{T_i} ; reflection coefficients with α_i for $i = 1, \dots, K$, then we define its corresponding GMM as:

$$p(\mathbf{x}) = \sum_{i=1}^K \frac{\alpha'_i}{2\pi |\mathbf{C}|^{1/2}} \exp \left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_i)^T \mathbf{C}^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) \right), \quad (22)$$

where α'_i are the normalized coefficients such that $\alpha'_i = |\alpha_i| / \sum_j |\alpha_j|$ and $\boldsymbol{\mu}_i$ are the corresponding delay-Doppler parameters such that $\boldsymbol{\mu}_i = [\tau_{T_i} \nu_{T_i}]^T$. Using the definition as in (22), $p(\mathbf{x})$ becomes a valid pdf, hence KLD defined in (20) can be used.

For a single Gaussian, a closed form of the KLD is available in terms of defining parameters. However, for GMM, there is no closed form solution of the integral in (20). An efficient approximation can be obtained by using Monte Carlo techniques since KLD defined in (20) can also be considered as an integral to compute the expectation of $\ln \left(\frac{p(\mathbf{x})}{q(\mathbf{x})} \right)$ under the distribution of $p(\mathbf{x})$. Therefore, it can be written as:

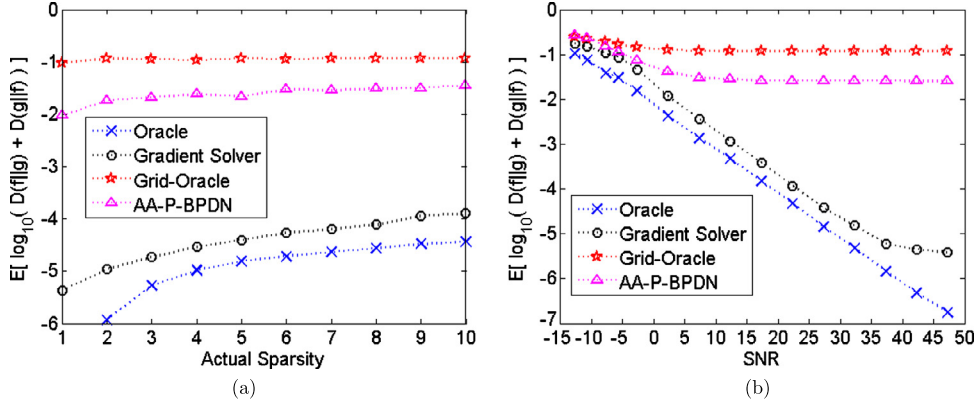


Fig. 4. Mean of the KLD metric for tested techniques in comparison with the oracle result at different (a) sparsity levels, (b) SNR levels.

$$D(p \parallel q) \triangleq \int_{-\infty}^{\infty} p(\mathbf{x}) \ln \left(\frac{p(\mathbf{x})}{q(\mathbf{x})} \right) d\mathbf{x} = E_{p(\mathbf{x})} \left[\ln \left(\frac{p(\mathbf{x})}{q(\mathbf{x})} \right) \right]. \quad (23)$$

Sample mean can be used to approximate the actual value of the expectation relying on the law of large numbers as:

$$D(p \parallel q) \simeq \frac{1}{Z} \sum_{j=1}^Z \ln \left(\frac{p(\mathbf{x}_j)}{q(\mathbf{x}_j)} \right), \quad (24)$$

where each \mathbf{x}_j is drawn independently and identically from $p(\mathbf{x})$ as defined in (22). In the following simulations, each KLD is approximated by using $Z = 10^6$ samples.

First, we would like to define an oracle estimator that would be the lower bound for the reconstruction performance of the tested algorithms. We assume that oracle estimator exactly knows the off-grid target parameters and hence the oracle estimation of the coefficients is given as a Least-Squares (LS) solution as:

$$\hat{\mathbf{x}} = [\mathbf{a}(\tau_{T_1}, \nu_{T_1}) \mathbf{a}(\tau_{T_2}, \nu_{T_2}) \dots \mathbf{a}(\tau_{T_K}, \nu_{T_K})]^\dagger \mathbf{b}. \quad (25)$$

We also define a “grid-oracle” solution, which is the LS solution not on the actual parameter points, but on the grid points closest to the actual off-grid ones. This provides the lower bound for the unperturbed techniques.

To illustrate the performance of the proposed GS algorithm defined in Table 2, reconstructions are compared with the oracle solution, grid-oracle solution and AA-P-BPDN algorithm proposed in [30] while actual sparsity is changing from 1 to 10 at a fixed SNR of 27.3 dB and measurement number of $M = 186$. In this simulation all techniques are given the grid points closest to the actual ones, basically to measure how well the proposed perturbation technique and alternatives can estimate the actual parameters given the correct grid points compared to oracle performances. For a fair comparison, AA-P-BPDN is modified and ℓ_1 reconstruction in the iterations is replaced with the LS solution on the given grid location. Iterations in AA-P-BPDN is terminated when normalized norm of the difference between the solutions in two consecutive iterations are smaller than 10^{-2} . It is expected that both algorithms will be better than the grid-oracle due to their perturbation mechanism, yet oracle solution will still be the lower bound.

For the error metric, KLD as defined in (24) is used. Since KLD is not symmetric, we consider the difference between two radar scenes, namely p and q , as $d(p, q) = D(p \parallel q) + D(q \parallel p)$. For the average reconstruction performance, simulations were repeated 300 times at each sparsity level with independent delay-Doppler domain target scenes. The average of base-10 logarithm of the $d(p, q)$ distances is provided in Fig. 4(a). It can be observed that as the sparsity level increases, reconstruction performance decreases. However, performance of the proposed gradient solver

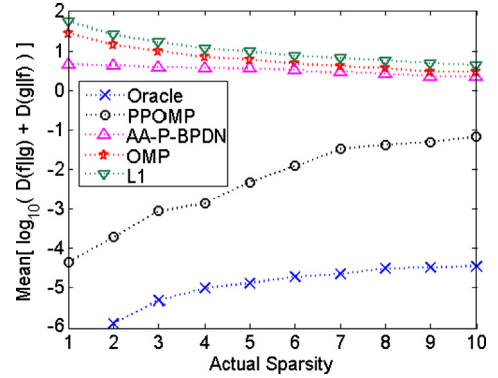


Fig. 5. Mean of the KLD metric for tested techniques in comparison with the oracle result at varying sparsity levels.

follows the oracle performance closely with similar performance gap for sparsity ranging from 3 to 10. AA-P-BPDN, on the other hand, has a better performance than the grid-oracle solution but significantly inferior than the proposed gradient solver and the oracle solution. This is partly because AA-P-BPDN linearizes the basis functions only on the grid points and the accuracy of its approximation decreases for larger perturbations. However, GS updates its approximation at each iteration with a gradient descent update which converges to a local minima.

To investigate the effect of noise level, different SNR levels in the range of -15 dB to 50 dB are tested for a fixed sparsity level of $K = 5$ and number of samples $M = 186$. Fig. 4(b) shows the average KLD in the logarithmic scale for the tested algorithms. It can be observed that oracle KLD decreases linearly with increasing SNR. Proposed GS closely follows oracle performance with a small performance gap until an SNR up to 35 dB and levels of after that due to termination of the iterations. Although AA-P-BPDN has a better performance than grid oracle, it performs worse than proposed GS for all SNR levels. Results presented in Fig. 4 show that if the closest grid locations are given to the proposed perturbation procedure, a close performance to the oracle solution can be obtained. Now we would like to test the total PPOMP algorithm when the perturbation is done on the grid locations selected by OMP.

In the following simulations, performance of the proposed PPOMP algorithm is compared to AA-P-BPDN, standard OMP, ℓ_1 reconstruction and the oracle solution for varying sparsity, measurement number and SNR levels. At each test case 50 independent random trails are performed. Noise fit level for all suitable techniques is set to be $\epsilon = 1.3 \|\sigma \mathbf{n}\|_2 / \|\mathbf{b}\|_2$.

Fig. 5 shows the mean of logarithm base-10 of KLD metric for varying sparsity levels. PPOMP is closer to the oracle performance,

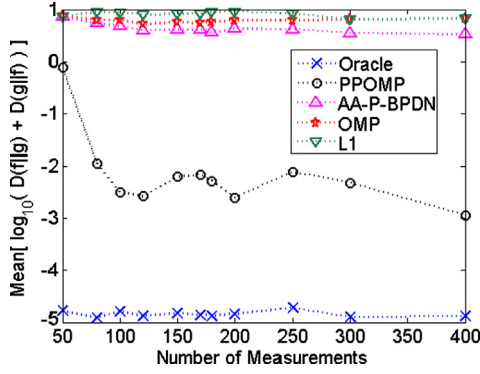


Fig. 6. Mean of the KLD metric for tested techniques in comparison with the oracle result at varying number of measurements.

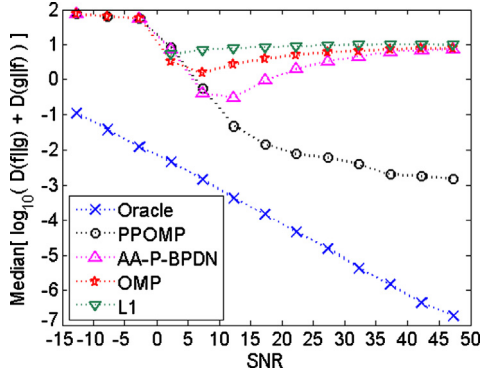


Fig. 7. Mean of the KLD metric for tested techniques in comparison with the oracle result at varying SNR levels.

whereas other techniques are significantly inferior compared to the proposed PPOMP technique. When compared to the small gap between the gradient solver and the oracle performance in Fig. 4, gap between PPOMP and oracle solution in Fig. 5 is larger. This is because OMP iterations within the PPOMP technique are not always able to provide correct grid points. This performance gap is more apparent in the less sparse region since OMP iterations are more prone to fail in that range. The same case is also valid for the AA-P-BPDN; AA-P-BPDN is inferior to its oracle counter-part. It is important to notice that since PPOMP and AA-P-BPDN have perturbation mechanism, both are superior to the classical unperturbed techniques.

Fig. 6 shows a similar comparison for a range of number of measurements. It can be seen that after a minimum required number of measurements which seems to be around 100 for this case, performance of PPOMP do not increase with the increase in number of measurements. Even if there is performance gap between the PPOMP and the oracle solution, PPOMP is significantly superior to the other compared techniques.

Fig. 7 shows the mean of logarithm base-10 of KLD for a range of SNR levels. While PPOMP performs closer to oracle for high SNR, its performance degrades and becomes similar to OMP or ℓ_1 for lower SNR regime. Even though the proposed GS is able to follow the oracle performance for varying SNR as in Fig. 4(b), PPOMP performs worse in low SNR regime since OMP is not able to provide the correct grid points to the gradient solver in that regime. As a result, we observe a degraded performance in the overall PPOMP algorithm. At lower SNR, impressions due to off-grid error are washed out by noise, hence results of all the investigated techniques are at about the same level of performance. For the investigated application whose results are shown in Fig. 7, the benefit of using the proposed technique becomes noticeable beyond SNR of 10 dB, which is commonly encountered in practice.

Also notice that in the analysis, AA-P-BPDN performs better than both OMP and ℓ_1 minimization due to its perturbation scheme. However, the performance gap between AA-P-BPDN and PPOMP is significant for high SNR.

In addition to the higher performance of PPOMP, it is also less computationally complex compared to AA-P-BPDN. Since in each iteration of AA-P-BPDN, one ℓ_1 optimization and one constraint least-squares problems are solved, the computational complexity is significantly higher than the proposed PPOMP technique. The reported simulations performed on a workstation with Intel E5450 processor using CVX toolbox [41]. While AA-P-BPDN takes approximately 360 seconds in average for a single reconstruction, PPOMP converges in approximately 37 seconds in average for the same test problem of $K = 5$, $M = 186$, SNR = 27.3 dB.

5. Conclusions

In this paper a novel compressive sensing technique is proposed to alleviate the issues related with the reconstruction of the radar targets whose positions do not coincide with the assumed delay-Doppler grid. The proposed PPOMP technique adapts the signal dictionary to the actual measurements by performing perturbations of the parameters governing the signal dictionary. To quantify the performance, Kullback-Leibler Divergence is proposed as the error metric for off-grid target reconstruction performance. Compared to the tested techniques, proposed method provides significantly higher performance for a wide range of sparsity and SNR levels. Furthermore, due to its significantly lower complexity of implementation, PPOMP technique is more feasible in radar applications than the convex optimization based techniques.

Appendix A. Lipschitz continuity of the objective function

After the linearization of (14a), normalized cost function becomes $J(\mathbf{u}) = \|\mathbf{r}_l - \mathbf{B}_l \mathbf{u}\|_2^2 / \|\mathbf{b}\|_2^2$ and its gradient is $\nabla J(\mathbf{u}) = -(2/\|\mathbf{b}\|_2^2) \mathbf{B}_l^H (\mathbf{r}_l - \mathbf{B}_l \mathbf{u})$. Therefore,

$$\begin{aligned} \frac{\|\nabla J(\mathbf{u}_1) - \nabla J(\mathbf{u}_2)\|_2}{\|\mathbf{u}_1 - \mathbf{u}_2\|_2} &\leq \frac{2}{\|\mathbf{b}\|_2^2} \|\mathbf{B}_l^H \mathbf{B}_l\|_2 \\ &\leq \frac{2}{\|\mathbf{b}\|_2^2} \|\mathbf{B}_l^H \mathbf{B}_l\|_* = \frac{2}{\|\mathbf{b}\|_2^2} \text{tr}(\mathbf{B}_l^H \mathbf{B}_l), \end{aligned} \quad (\text{A.1})$$

where $\|\cdot\|_2$, $\|\cdot\|_*$ and $\text{tr}(\cdot)$ represents the spectral norm, nuclear norm and the trace of the argument respectively. Furthermore, we can expand the trace as:

$$\begin{aligned} \text{tr}(\mathbf{B}_l^H \mathbf{B}_l) &= \sum_{i=1}^k \left\| \Delta_\tau \alpha_{i,l} \frac{\partial \mathbf{a}}{\partial \tau_{i,l}} \right\|_2^2 + \sum_{i=1}^k \left\| \Delta_\nu \alpha_{i,l} \frac{\partial \mathbf{a}}{\partial \nu_{i,l}} \right\|_2^2 \\ &= \sum_{i=1}^k |\alpha_{i,l}|^2 \left(\Delta_\tau^2 \left\| \frac{\partial \mathbf{a}}{\partial \tau_{i,l}} \right\|_2^2 + \Delta_\nu^2 \left\| \frac{\partial \mathbf{a}}{\partial \nu_{i,l}} \right\|_2^2 \right). \end{aligned} \quad (\text{A.2})$$

Depending on the definition of the basis function in (5), the i th index of $\mathbf{a}(\tau, \nu)$ is $a_i(t_i; \tau, \nu) = s(t_i - \tau) e^{j2\pi \nu t_i}$ where t_i is the i th time sample and $s(t) = e^{jf(t)}$ is the radar waveform. Hence, $a_i(t_i; \tau, \nu) = e^{j(2\pi \nu t_i + f(t_i - \tau))}$ and norm of the partial derivative of $a_i(t_i; \tau, \nu)$ with respect to τ can be written as:

$$\begin{aligned} \left| \frac{\partial a_i(t_i; \tau, \nu)}{\partial \tau} \right| &= \left| -jf'(t_i - \tau) e^{j(2\pi \nu t_i + f(t_i - \tau))} \right| \\ &= |f'(t_i - \tau)|. \end{aligned} \quad (\text{A.3})$$

For linear chirp signals, $f(t) = \pi \beta (t - T_p/2)^2$, and $f'(t) = 2\pi \beta (t - T_p/2)$. Since pulse duration is $0 \leq t \leq T_p$, we have $|f'(t)| \leq \pi \beta T_p$. Therefore, norm of the partial derivative can be bounded as:

$$\left| \frac{\partial a(t_i; \tau, \nu)}{\partial \tau} \right| = |f'(t_i - \tau)| \leq \pi \beta T_p. \quad (\text{A.4})$$

Using (A.4), $\|\partial \mathbf{a} / \partial \tau\|_2^2$ can be bounded as:

$$\left\| \frac{\partial \mathbf{a}}{\partial \tau} \right\|_2^2 = \sum_{i=1}^M \left| \frac{\partial a(t_i; \tau, \nu)}{\partial \tau} \right|^2 \leq M \pi^2 \beta^2 T_p^2 = M \pi^2 \Delta_\tau^{-2}, \quad (\text{A.5})$$

where the last part follows with the selection of $\beta = 2B/T_p$, and $\Delta_\tau = 1/(2B)$ is the Rayleigh resolution spacing. Similarly, norm of the partial derivative of $a(t_i; \tau, \nu)$ with respect to ν can be written as:

$$\left| \frac{\partial a(t_i; \tau, \nu)}{\partial \nu} \right| = |-j2\pi t_i e^{j(2\pi \nu t_i + f(t_i - \tau))}| = |2\pi t_i|. \quad (\text{A.6})$$

Since time samples are taken from $[0, N_p T_{PRI}]$ range, we have $|2\pi t_i| \leq 2\pi N_p T_{PRI}$. Norm of the partial derivative can be bounded as:

$$\left| \frac{\partial a(t_i; \tau, \nu)}{\partial \nu} \right| = |2\pi t_i| \leq 2\pi N_p T_{PRI}. \quad (\text{A.7})$$

Using (A.7), $\|\partial \mathbf{a} / \partial \nu\|_2^2$ can be bounded as:

$$\left\| \frac{\partial \mathbf{a}}{\partial \nu} \right\|_2^2 = \sum_{i=1}^M \left| \frac{\partial a(t_i; \tau, \nu)}{\partial \nu} \right|^2 \leq M 4\pi^2 N_p^2 T_{PRI}^2 = M 4\pi^2 \Delta_\nu^{-2}, \quad (\text{A.8})$$

where $\Delta_\nu = 1/(N_p T_{PRI})$ is the Rayleigh resolution spacing. Combining (A.2), (A.5) and (A.8), $\text{tr}(\mathbf{B}_i^H \mathbf{B}_i)$ can be upper bounded as:

$$\text{tr}(\mathbf{B}_i^H \mathbf{B}_i) \leq \sum_{i=1}^k |\alpha_{1,i}|^2 (M\pi^2 + 4M\pi^2) = 5\pi^2 M \|\boldsymbol{\alpha}_i\|_2^2. \quad (\text{A.9})$$

Notice that $\boldsymbol{\alpha}_i$ is the coefficients of the projection of \mathbf{b} onto the estimated parameters. Since $\|\mathbf{a}(\tau, \nu)\|_2^2 = M$, we have $M \|\boldsymbol{\alpha}_i\|_2^2 \approx \|\mathbf{b}\|_2^2$. Using (A.9), (A.1) can be upper bounded as:

$$\frac{\|\nabla J(\mathbf{u}_1) - \nabla J(\mathbf{u}_2)\|_2}{\|\mathbf{u}_1 - \mathbf{u}_2\|_2} \leq 10\pi^2 = L. \quad (\text{A.10})$$

References

- [1] D. Donoho, Compressed sensing, *IEEE Trans. Inform. Theory* 52 (4) (2006) 1289–1306.
- [2] E. Candes, J. Romberg, T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, *IEEE Trans. Inform. Theory* 52 (2006) 489–509.
- [3] R. Baraniuk, Compressive sensing, *IEEE Signal Process. Mag.* 24 (4) (July 2007) 118–121.
- [4] E. Candes, J. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, *Comm. Pure Appl. Math.* 59 (8) (2006) 1207–1223.
- [5] J. Haupt, R. Nowak, Signal reconstruction from noisy random projections, *IEEE Trans. Inform. Theory* 52 (9) (2006) 4036–4048.
- [6] D. Donoho, M. Elad, V. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise, *IEEE Trans. Inform. Theory* 52 (1) (2006) 6–18.
- [7] R. Baraniuk, M. Davenport, R. DeVore, M. Wakin, A simple proof of the restricted isometry property for random matrices, *Constr. Approx.* 28 (3) (2008) 253–263.
- [8] R. Baraniuk, P. Steeghs, Compressive radar imaging, in: *IEEE Radar Conf.*, 2007, pp. 128–133.
- [9] A.C. Gurbuz, J.H. McClellan, W.R. Scott Jr., Compressive sensing for GPR imaging, in: *Asilomar Conf.*, 2007, pp. 2223–2227.
- [10] S. Shah, Y. Yu, A. Petropulu, Step-frequency radar with compressive sampling (SFR-CS), in: 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), 2010, pp. 1686–1689.
- [11] A.C. Gurbuz, J.H. McClellan, W.R. Scott Jr., Compressive sensing for subsurface imaging using ground penetrating radars, *Signal Process.* 89 (10) (2009) 1959–1972.
- [12] A.C. Gurbuz, J.H. McClellan, W.R. Scott Jr., A compressive sensing data acquisition and imaging method for stepped frequency GPRs, *IEEE Trans. Signal Process.* 57 (7) (2009) 2640–2650.
- [13] M. Herman, T. Strohmer, High-resolution radar via compressed sensing, *IEEE Trans. Signal Process.* 57 (6) (2009) 2275–2284.
- [14] S. Ugur, O. Arikan, SAR image reconstruction and autofocus by compressed sensing, *Digit. Signal Process.* 22 (6) (2012) 923–932.
- [15] Y. Yu, A.P. Petropulu, H.V. Poor, MIMO radar using compressive sampling, *IEEE J. Sel. Top. Signal Process.* 4 (1) (2010) 146–163.
- [16] L. Xu, Q. Liang, X. Cheng, D. Chen, Compressive sensing in distributed radar sensor networks using pulse compression waveforms, *EURASIP J. Wirel. Commun. Netw.* 36 (2013).
- [17] J.H. Ender, On compressive sensing applied to radar, *Signal Process.* 90 (5) (2010) 1402–1414.
- [18] L. Potter, E. Ertin, J. Parker, M. Cetin, Sparsity and compressed sensing in radar imaging, *Proc. IEEE* 98 (6) (2010) 1006–1020.
- [19] Y. Chi, L. Scharf, A. Pezeshki, R. Calderbank, The sensitivity to basis mismatch of compressed sensing in spectrum analysis and beamforming, in: *Sixth Workshop on Defense Applications of Signal Processing (DASP)*, Lihue, HI, October, 2009.
- [20] M.A. Tuncer, A.C. Gurbuz, Analysis of unknown velocity and target off the grid problems in compressive sensing based subsurface imaging, in: *ICASSP*, Prague, Czech Republic, 2011, pp. 2880–2883.
- [21] M. Herman, T. Strohmer, General deviants: An analysis of perturbations in compressed sensing, *IEEE J. Sel. Top. Signal Process.* 4 (2) (2010) 342–349.
- [22] Y. Chi, L. Scharf, A. Pezeshki, R. Calderbank, Sensitivity of basis mismatch to compressed sensing, *IEEE Trans. Signal Process.* 59 (2011) 2182–2195.
- [23] D.H. Chae, P. Sadeghi, R. Kennedy, Effects of basis-mismatch in compressive sampling of continuous sinusoidal signals, in: *International Conference on Future Computer and Communication (ICFCC)*, 2010, pp. 739–743.
- [24] G. Tang, B.N. Bhaskar, P. Shah, B. Recht, Compressed sensing off the grid, [arXiv:abs/1207.6053](https://arxiv.org/abs/1207.6053).
- [25] V. Cevher, M. Duarte, R. Baraniuk, Distributed target localization via spatial sparsity, in: *European Signal Processing Conf. (EUSIPCO)*, August, 2008, pp. 134–142.
- [26] A. Gurbuz, V. Cevher, J. McClellan, Bearing estimation via spatial sparsity using compressive sensing, *IEEE Trans. Aerosp. Electron. Syst.* 48 (2) (2012) 1358–1369.
- [27] N. Aggarwal, W.C. Karl, Line detection in images through Regularized Hough Transform, *IEEE Trans. Image Process.* 15 (2006) 582–590.
- [28] G. Peyre, Best basis compressed sensing, *IEEE Trans. Signal Process.* 58 (5) (May 2010) 2613–2622.
- [29] C. Ekanadham, D. Tranchina, E. Simoncelli, Recovery of sparse translation-invariant signals with continuous basis pursuit, *IEEE Trans. Signal Process.* 59 (10) (2011) 4735–4744.
- [30] Z. Yang, C. Zhang, L. Xie, Robustly stable signal recovery in compressed sensing with structured matrix perturbation, *IEEE Trans. Signal Process.* 60 (9) (2012) 4658–4671.
- [31] H. Zhu, G. Leus, G. Giannakis, Sparsity-cognizant total least-squares for perturbed compressive sampling, *IEEE Trans. Signal Process.* 59 (5) (2011) 2002–2016.
- [32] T. Huang, Y. Liu, H. Meng, X. Wang, Adaptive matching pursuit with constrained total least squares, *EURASIP J. Adv. Signal Process.* 76 (2012).
- [33] O. Teke, A. Gurbuz, O. Arikan, Perturbed orthogonal matching pursuit, *IEEE Trans. Signal Process.* 61 (24) (2013) 6220–6231.
- [34] S. Mallat, Z. Zhang, Matching pursuits with time–frequency dictionaries, *IEEE Trans. Signal Process.* 41 (1993) 3397–3415.
- [35] J. Tropp, A. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, *IEEE Trans. Inform. Theory* 53 (12) (Dec. 2007) 4655–4666.
- [36] T. Blumensath, M.E. Davies, Iterative hard thresholding for compressed sensing, *Appl. Comput. Harmon. Anal.* 27 (3) (2009) 265–274.
- [37] D. Needell, J.A. Tropp, CoSaMP: Iterative signal recovery from incomplete and inaccurate samples, *Appl. Comput. Harmon. Anal.* 26 (3) (2009) 301–321.
- [38] M. Skolnik, *Introduction to Radar Systems*, McGraw Hill, 1980.
- [39] M. Richards, *Fundamentals of Radar Signal Processing*, McGraw Hill, 2005.
- [40] J. Li, P. Stoica, *MIMO Radar Signal Processing*, Wiley, 2008.
- [41] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [42] D.P. Bertsekas, *Nonlinear Programming*, Athena Scientific, 1999.
- [43] D. Donoho, Superresolution via sparsity constraints, *SIAM J. Math. Anal.* 23 (5) (1992) 1309–1331.



Oguzhan Teke was born in 1990, in Tokat, Turkey. In 2012, he received his B.Sc. degree in Electrical and Electronics Engineering from Bilkent University, Ankara, Turkey. He is currently working toward the M.S. degree under the supervision of Professor Orhan Arikan in the Department of Electrical and Electronics Engineering, Bilkent University. His research interests are in applied linear algebra and numerical analysis, inverse problems, and compressed sensing.



Ali Cafer Gurbuz received the B.S. degree from Bilkent University, Ankara, Turkey, in 2003, in Electrical and Electronics Engineering, and the M.S. and Ph.D. degrees from the Georgia Institute of Technology, Atlanta, in 2005 and 2008, both in electrical and computer engineering. From 2003 to 2008, he participated in multimodal landmine detection system research as a graduate research assistant and, from 2008 to 2009, as postdoctoral fellow, all with Georgia Tech.

He is currently an Associate Professor with TOBB University of Economics and Technology, Ankara, Turkey, with the Department of Electrical and Electronics Engineering. His research interests include compressive sensing applications, ground penetrating radar, array signal processing, remote sensing, and imaging.



Orhan Arikan was born in 1964, in Manisa, Turkey. In 1986, he received his B.Sc. degree in Electrical and Electronics Engineering from the Middle East Technical University, Ankara, Turkey. He received both his M.S. and Ph.D. degrees in Electrical and Computer Engineering from the University of Illinois, Urbana-Champaign, in 1988 and 1990, respectively. Following his graduate studies, he was employed as a Research Scientist at Schlumberger-Doll Research Center, Ridgefield, CT. In 1993, he joined the Electrical and Electronics Engineering Department of Bilkent University, Ankara, Turkey. Since 2011, he is serving as the department chairman. His current research interests include statistical signal processing, time–frequency analysis and remote sensing. Dr. Arikan has served as Chairman of IEEE Signal Processing Society Turkey Chapter and President of IEEE Turkey Section.