Digital Signal Processing 23 (2013) 1630–1644

Contents lists available at SciVerse ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

# Detection of sparse targets with structurally perturbed echo dictionaries

Mehmet Burak Guldogan [a,*], Orhan Arikan [b]

[a] Department of Electrical and Electronics Engineering, Turgut Ozal University, Ankara TR-06010, Turkey
[b] Department of Electrical and Electronics Engineering, Bilkent University, Ankara TR-06800, Turkey

ARTICLE INFO

*Article history:*
Available online 7 March 2013

*Keywords:*
Compressed sensing (CS)
Orthogonal matching pursuit (OMP)
Cross-ambiguity function (CAF)
Particle swarm optimization (PSO)
Channel identification
Sparse approximation

ABSTRACT

In this paper, a novel algorithm is proposed to achieve robust high resolution detection in sparse multipath channels. Currently used sparse reconstruction techniques are not immediately applicable in multipath channel modeling. Performance of standard compressed sensing formulations based on discretization of the multipath channel parameter space degrade significantly when the actual channel parameters deviate from the assumed discrete set of values. To alleviate this off-grid problem, we make use of the particle swarm optimization (PSO) to perturb each grid point that reside in each multipath component cluster. Orthogonal matching pursuit (OMP) is used to reconstruct sparse multipath components in a greedy fashion. Extensive simulation results quantify the performance gain and robustness obtained by the proposed algorithm against the off-grid problem faced in sparse multipath channels.



## 1. Introduction

In a multipath channel, transmitted signal is reflected off various obstacles and a superposition of multiple delayed, attenuated, frequency-phase shifted copies of the original signal arrive at the receiver. At first, the presence of multipath arrivals seems to degrade the quality of the communication, but a carefully designed communication system can take advantage of the diversity provided by the multipath environment and mitigate fading. Diversity in multipath channels is a result of variation between the direction-of-arrivals (DOA), delays and Doppler shifts of the individual multipath components. In order to use the diversity and mitigate the affect of multipath fading, the channel should be accurately modeled and the channel state information (CSI) should be provided to the receiver. Most of the time, since the CSI is not available to systems, channel should be periodically estimated at the receiver to take advantage of the diversity provided by multipath propagation. Mainly, there exist two approaches: training based and blind channel estimation. In training based methods, information carrying training signals, that are known to the receiver, are emitted to the environment and the CSI is obtained from received and known training signals. Training based channel estimation methods need simple processing and considerably reduce receiver complexity. Therefore, these methods are extensively used in wireless communication systems [1]. On the other hand, blind channel estimation methods obtain the CSI using only the statistics of information carrying signals. They are known to be theoretically efficient but need complex processing [2].

There have been many research efforts in developing training based methods for channel modeling. These efforts basically concentrate on two phases, namely the sensing and reconstruction. In the sensing phase, training signals are designed to probe the channel and in reconstruction phase, receiver output is processed to obtain the CSI. Designing proper training signals and developing efficient reconstruction techniques are highly critical in order to accurately model the channel. The general assumption in most of the important works in wireless communications is that there exists a rich multipath environment and linear reconstruction techniques are known to be optimal in these channels [3,4]. However, recent research show that wireless channels have a sparse structure in time, frequency and space [5–7]. Moreover, it is presented in [8,9], that training based methods using linear reconstruction techniques cannot fully exploit the sparse structure of the channel and causing over utilization of the resources. Recently, by embedding the key concepts from compressed sensing, new training based techniques have been proposed for sparse channels that have better performance than usual least-squares (LS) based approaches to model the sparse wireless channel [8]. In [10,11], authors use a virtual representation of physical multipath channels to model the time–frequency response of sparse multipath channel. In [12,13] the matrix identification problem, where the matrix has a sparse representation in some basis, is discussed. Herman and Strohmer introduced the concept of compressed sensing radar, which provides better time–frequency resolution over classical radar by exploiting the sparse structure [13]. General assumption used in all of these approaches is that the all multipath

* Corresponding author. Fax: +90 (0) 312 5515455.
*E-mail address:* bguldogan@turgutozal.edu.tr (M.B. Guldogan).

1051-2004/$ – see front matter 
http://dx.doi.org/10.1016/j.dsp.2013.03.003

components fall on the grid points, which is practically impossible as the multipath/target parameters are unknown. Hence the true grid, which is possibly irregular, cannot be known beforehand. This so called off-grid problem, results in a mismatch of the dictionary and severely degrades the performance of techniques that exploit sparsity. Furthermore, such methods exhibit an unstable behavior as previously shown in theoretical studies on dictionary errors. In several papers, the problem is pointed out and very simple grid refinement approaches are stated [14,15].

In this paper, we propose a novel algorithm to overcome the off-grid problem based on the sparse approximation theory. Firstly, the receiver output is transformed to delay–Doppler domain by using the CAF for efficient exploitation of the delay–Doppler diversity of the multipath components. In the transform domain, multipath clusters above the noise level are identified. This way, the original channel identification problem is reduced to channel identification problems over the identified multipath clusters in the delay–Doppler domain. Then, we make use of the particle swarm optimization (PSO), to perturb the location of each grid point that reside in each cluster separately and orthogonal matching pursuit (OMP) is used to reconstruct sparse multipath components in a greedy fashion [16,17].

The paper is organized as follows. The parametric channel model is detailed in Section 2. In Section 3, some key concepts of the CS and channel matrix identification are presented. Off-grid problem in sparse approximation is defined in Section 4. Details of the proposed technique are introduced in Section 5. The simulation results are presented in Section 6.

## 2. Sparse multipath channel model

In this section, we present some important key points of the virtual channel model for doubly selective channels ($BW\tau_{\max} \geqslant 1$, $T\nu_{\max} \geqslant 1$) that exploits the relation between the multipath components and the signal space. Here, $BW$ stands for bandwidth, $T$ for duration, $\tau_{\max}$ and $\nu_{\max}$ for delay and Doppler spreads, respectively. Canonical model, also known as virtual channel model, formulizes a lower dimensional approximation of the physical multipath channel by uniformly sampling of the delay–Doppler–spatial domain [18,19]. This alternative modeling exploits the relation between the clustering of multipath components within delay–Doppler–spatial domain and sparsity of degrees of freedom in the multipath channel and prepares the underlying structure to be able to make use of the benefits of the CS theory and sparse approximation tools. Recent multipath channel measurement results show that multipath components are distributed in as clusters within a defined channel spread and impinge onto a receiver in clusters [5,6]. In a scattering environment, clusters of multipath components occur due to the large scale scatters such as buildings and hills. Multipath components within a cluster occur due to small scale scatters of the large scale scatters such as windows of buildings. Moreover, most of the practical multipath channels such as ultra-wideband channels [7], high definition digital television channels [20,21] underwater acoustic channels [22,23] and broadband wireless communication channels [24] exhibit a clustered sparse structure. There exist various efforts in the literature to clarify the underlying theory of clustered sparsity. Therefore, sparse nature of the multipath channels should be exploited in order to accurately estimate the channel parameters [8]. For the sake of simplicity and to be able to introduce the main idea clearly, we provide formulation of the virtual channel model in delay–Doppler domain. Extension to spatial domain is straightforward and can be found in Ref. [19]. Doubly selective multipath channels can be classified as either rich or sparse, depending on the separation between different multipath component clusters. The separations are smaller than $\Delta\tau = 1/BW$ and $\Delta\nu = 1/T$ in delay–Doppler domain for rich multipath component channels. However in sparse multipath component channels, the separations are larger than $\Delta\tau$ and $\Delta\nu$, where each delay–Doppler bin is of size $\Delta\tau \times \Delta\nu$.

The physical multipath communication channel can be modeled as:

$$H(t,f) = \sum_{i=1}^{d} \zeta_i e^{-j2\pi\tau_i f} e^{j2\pi\nu t}, \tag{1}$$

where $d$ is the number of multipath components in the environment, $\zeta_i$, $\tau_i$ and $\nu_i$ are the complex attenuation factor, delay and Doppler shift of the $i$th multipath component, respectively. Although, physical channel model given in (1) is a realistic model, analysis and estimation steps are difficult, due to the presence of large number of parameters, $\zeta_i$, $\tau_i$, $\nu_i$, $i = 1,\ldots,d$. In situations where we have finite signaling duration and channel bandwidth, this multipath model can be approximated by a linear one, known as virtual channel model [19]. By uniformly sampling the physical multipath environment in both delay with $\Delta\tau$ and in Doppler with $\Delta\nu$, a lower dimensional approximation of the multipath channel model can be obtained. The corresponding discrete model is:

$$H(t,f) = \sum_{k=0}^{K-1} \sum_{p=-P}^{P} \mathcal{H}(k,p) e^{j2\pi\frac{p}{T}t} e^{-j2\pi\frac{k}{BW}f}. \tag{2}$$

The virtual channel coefficients can be related to the continuous channel model as:

$$\mathcal{H}(k,p) = \frac{1}{TBW} \int_0^T \int_{-BW/2}^{BW/2} H(t,f) e^{j2\pi\frac{p}{T}t} e^{-j2\pi\frac{k}{BW}f}\, dt\, df. \tag{3}$$

Number of resolvable delay and Doppler cells in each dimension are:

$$K = \left\lceil \frac{\tau_{\max}}{\Delta\tau} \right\rceil + 1 = \lceil BW\tau_{\max} \rceil + 1, \tag{4}$$

$$P = \left\lceil \frac{\nu_{\max}}{2\Delta\nu} \right\rceil + 1 = \lceil T\nu_{\max}/2 \rceil + 1. \tag{5}$$

Hence, in the simplified model, the channel is characterized with virtual channel coefficients $\mathcal{H}(k,p)$, $K$ and $P$ only. Physical and virtual channel models can be related with each other by substituting (1) into (3) as [19]:

$$\mathcal{H}(k,p) = \sum_{i=1}^{d} \zeta_i e^{-j\pi(p-\nu_i T)} \operatorname{sinc}(p-\nu_i T) \operatorname{sinc}(k-\tau_i BW) \tag{6}$$

$$\approx \sum_{i \in S_{\tau,k} \cap S_{\nu,p}} \zeta_i, \tag{7}$$

where $S_{\tau,k} \cap S_{\nu,p}$ is the set of all multipath components whose delays and Doppler's are inside of a delay–Doppler resolution cell of size $\Delta\tau \times \Delta\nu$ and centered on the $k$th virtual delay ($\frac{k}{BW}$) and $p$th virtual Doppler shift ($\frac{p}{T}$). By using the given sampled virtual channel representation, baseband receiver output can be written as:

$$x(t) = \sum_{i=1}^{d} \zeta_i s(t-\tau_i) e^{j2\pi\nu_i t}$$

$$\approx \sum_{k=0}^{K-1} \sum_{p=-P}^{P} \mathcal{H}(k,p) s\left(t - \frac{k}{BW}\right) e^{j2\pi\frac{p}{T}t}, \tag{8}$$

where $s(t)$ is the transmitted signal. Therefore, we can say that the virtual model given above approximately represents the physical discrete doubly selective multipath channel in terms of an $N_h$-dimensional parameter vector containing the virtual channel coefficients $\mathcal{H}(k,p)$. $N_h$ is defined as:

$$N_h = K(2P+1) \tag{9}$$

$$= \big(2\lceil T\nu_{\max}/2\rceil + 1\big)\big(\lceil BW\tau_{\max}\rceil + 1\big) \tag{10}$$

$$\approx \tau_{\max}\nu_{\max}TBW \tag{11}$$

$$\approx \tau_{\max}\nu_{\max}N_b. \tag{12}$$

In the following sections, having summarized some key concepts in sparse approximation, we formulate sensing of sparse doubly selective multipath channels.

## 3. Sparse approximation and channel estimation

Consider the following model:

$$\mathbf{v} = \boldsymbol{\Upsilon}\boldsymbol{\alpha}, \tag{13}$$

where $\mathbf{v} \in \mathbb{C}^N$ is the discrete signal in time domain which has to be under-sampled, $\boldsymbol{\Upsilon} \in \mathbb{C}^{N\times N}$ is the transform domain matrix and $\boldsymbol{\alpha} \in \mathbb{C}^N$ is the $\mathcal{S}$-sparse ($\|\boldsymbol{\alpha}\|_0 \leqslant \mathcal{S}$) vector with support set $\Lambda_{\mathcal{S}} = \operatorname{supp}(\boldsymbol{\alpha})$. Assume that, instead of directly using $N$ samples of $\mathbf{v}$, take $M$ ($M \ll N$) linear combinations of samples. These linear combinations can be represented with the matrix $\boldsymbol{\Phi} \in \mathbb{C}^{M\times N}$ and new model can be written as:

$$\boldsymbol{\Phi}\mathbf{v} = \boldsymbol{\Phi}\boldsymbol{\Upsilon}\boldsymbol{\alpha},$$

$$\mathbf{x} = \boldsymbol{\Psi}\boldsymbol{\alpha}, \tag{14}$$

where $\mathbf{x} \in \mathbb{C}^M$ can be termed as the observation or measurement vector and $\boldsymbol{\Psi} \in \mathbb{C}^{M\times N}$ is the sensing matrix. Here we are looking for a matrix $\boldsymbol{\Psi}$, which has as few rows as possible and can guarantee recovery of a sparse input.

Our aim is to reliably recover $\boldsymbol{\alpha}$ from knowledge of $\mathbf{x}$ and $\boldsymbol{\Psi}$. However, the dictionary matrix $\boldsymbol{\Psi}$ consists of more columns, called as atoms, than rows. Therefore, in the absence of further prior information, $\boldsymbol{\alpha}$ is unidentifiable from $\mathbf{x}$. This problem can be resolved by regularizing via sparsity constraints. That is, we search for approximate solutions to linear systems in which the unknown vector has few nonzero entries relative to its dimension:

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_0 \quad \text{s.t. } \mathbf{x} = \boldsymbol{\Psi}\boldsymbol{\alpha}. \tag{15}$$

In literature, this formulation is known as sparse approximation [25]. Extensive research is going on the theory and applications of sparse approximation and CS [26–30].

### 3.1. Requirements for the dictionary

In order to reliably recover $\boldsymbol{\alpha}$, we must have a guarantee notifying that different values of $\boldsymbol{\alpha}$ produce different values of $\mathbf{x}$. One way of having such a guarantee is determining all possible $\mathcal{S}$-element sets of atoms called subdictionaries and verifying that the subspaces spanned by these subdictionaries differ from each other. There exists several concepts that formulize the suitability of a dictionary for sparse approximation: the mutual coherence [31], the cumulative coherence [29], the exact recovery coefficient (ERC) [29], the spark [30], the restricted isometry constants (RICs) [32]. Mutual and cumulative coherence measures provide close values and they are easy to calculate but suboptimal when the RICs of $\boldsymbol{\Psi}$ are known. However, for arbitrary dictionary $\boldsymbol{\Psi}$, calculation of other three approaches is not efficient. In this paper, we focus on the mutual coherence concept [27,31].

The mutual coherence $\mu = \mu(\boldsymbol{\Psi})$ is defined as:

$$\mu \triangleq \max_{i\neq j} \big|\boldsymbol{\psi}_i^T \boldsymbol{\psi}_j\big|, \tag{16}$$

where columns $\boldsymbol{\psi}_i$ of dictionary $\boldsymbol{\Psi}$ are atoms of the dictionary. Assuming that each atom $\|\boldsymbol{\psi}_i\|_2 = 1$, then the coherence is bounded by [33]:

$$\sqrt{\frac{N-M}{M(N-1)}} \leqslant \mu(\boldsymbol{\Psi}) \leqslant 1. \tag{17}$$

Two atoms are aligned when $\mu(\boldsymbol{\Psi}) = 1$ and we have the maximal coherence which is the worst case scenario. When $\mu(\boldsymbol{\Psi}) = \sqrt{(N-M)/M(N-1)}$ we have the maximal incoherence which is the best case scenario. In maximal incoherence scenario, the atoms are spread out in $\mathbb{C}^M$.

### 3.2. Orthogonal matching pursuit (OMP)

There exists many different approaches to solve the following problem

$$\mathbf{x} = \boldsymbol{\Psi}\boldsymbol{\alpha} + \mathbf{n}, \tag{18}$$

where $\mathbf{n}$ is the random noise. These approaches can be basically grouped in five classes [34]: convex relaxation [25], greedy pursuit [35], nonconvex optimization [36], brute-force [37] and Bayesian techniques [38]. Convex relaxation based and greedy pursuit techniques are the mostly used techniques in the rapidly growing literature on CS theory. Choosing the proper algorithm for a specific problem is not an easy task. In this paper, we prefer to use OMP, which is the simplest effective greedy algorithm in the literature [17]. The major advantage of OMP is its ease of implementation and low cost of computation compared to relaxation based approaches [26]. Secondly, similar to relaxation based approaches, OMP has also proven performance guarantees, two of them are provided in the subsequent section. Thirdly, it is shown that, in high SNR regimes, OMP shows better performance compared to relaxation based approaches [27]. For this case, support set of $\boldsymbol{\alpha}$ can be recovered accurately and therefore OMP converges to the oracle estimator that is based on the known support set $\Lambda_o$ ($\Lambda_o$ is assumed to have been given by an oracle).

Basically, OMP estimates the support set $\Lambda$ from the measurements $\mathbf{x}$ by iteratively refining the current estimate for the vector $\boldsymbol{\alpha}$ by updating one or several coefficients that yield a considerable improvement in approximating the signal. Having found a support set $\Lambda$, $\boldsymbol{\alpha}$ can be estimated by using LS as:

$$\hat{\boldsymbol{\alpha}} = \boldsymbol{\Psi}_{\Lambda}^{\dagger}\mathbf{x}, \tag{19}$$

and $\mathbf{0}$ elsewhere. Here, $\boldsymbol{\Psi}_{\Lambda}^{\dagger}$ is the Moore–Penrose pseudoinverse of $\boldsymbol{\Psi}_{\Lambda}$. Algorithmic steps of the OMP are given in Table 1. Computationally costly part of the OMP is the identification step [34] which requires $O(M\times N)$ number of multiplications for an unstructured dense matrix. LS technique is used in the reconstruction step. For this purpose, QR factorization of $\boldsymbol{\Psi}_{\Lambda_k}$, which has a cost of $O(Mk)$ in the $k$th iteration, can be used. To stop the algorithm, the following listed criteria can be used:

- stop after a fixed number of iterations, $k = \mathcal{S}$,
- stop when the residual has a small enough magnitude, $\|\mathbf{r}_k\|_2 \leqslant \epsilon$.

An important property of the OMP is that the algorithm never chooses the same atom twice [30]. Therefore, stopping after $\mathcal{S}$ iterations guarantees that $\|\hat{\boldsymbol{\alpha}}\|_0 = \mathcal{S}$. In most of the practical applications similarly in the sparse multipath channel identification, the

**Table 1**
Orthogonal matching pursuit (OMP).

- **Input:** $\mathbf{x} \in \mathbb{C}^M$ and $\boldsymbol{\Psi} \in \mathbb{C}^{M\times N}$
- **Output:** sparse vector $\boldsymbol{\alpha} \in \mathbb{C}^N$

(1) **Initialization:** set $\Lambda_0 = \emptyset$, the residual $\mathbf{r}_0 = \mathbf{x}$ and set counter $k = 1$
(2) **Determination:** find an atom $n_k$ of $\boldsymbol{\Psi}$, which is most strongly correlated with the residual $\mathbf{r}$ as
$n_k = \arg\max_n |\langle \mathbf{r}_{k-1}, \boldsymbol{\psi}_n \rangle|$
$\Lambda_k = \Lambda_{k-1} \cup \{n_k\}$
(3) **Estimation:** using the chosen atoms up to now, find the best coefficients for approximating the signal
$\alpha_k = \arg\min_{\mathbf{b}} \|\mathbf{x} - \boldsymbol{\Psi}_{\Lambda_k}\mathbf{b}\|_2$
(4) **Iteration:** update the residual:
$\mathbf{r}_k = \mathbf{x} - \boldsymbol{\Psi}_{\Lambda_k}\alpha_k$
$k = k + 1$
repeat (2)–(4)
(5) **Output:** return the vector $\boldsymbol{\alpha}$ with components $\alpha(n) = \alpha_k(n)$ for $n \in \Lambda_k$ and $\alpha(n) = 0$ otherwise

focus is on detecting significant components, i.e., indexes of $\boldsymbol{\alpha}$ with large magnitude, instead of recovering whole support set of $\boldsymbol{\alpha}$. It is shown that, with the proper stopping rule, OMP will choose all the most important components of $\boldsymbol{\alpha}$ before possibly selecting incorrect components [39]. Therefore, due to the given reasons and to put forward the novel idea behind the paper, we believe that OMP is the right algorithm for the considered sparse multipath channel identification problem.

### 3.3. Sensing sparse doubly selective multipath channels

In this section, based on the virtual model presented in Section 2, we model the sensing matrix or dictionary matrix. Firstly, we give the discrete time representation of the channel output given in (8) as [9]:

$$x_n = \sum_{k=0}^{K-1}\sum_{p=-P}^{P} \mathcal{H}(k,p) e^{j2\pi \frac{p}{N_b} n} s_{n-k}, \quad n = 0, 1, \ldots, N_b + K - 2 \tag{20}$$

where $N_b = TBW$. Let's define an $\check{N}_b = N_b + K - 1$ length sequence of vectors $\mathbf{s}_n \in \mathbb{C}^K$ as:

$$\mathbf{s}_n = [s_n \quad s_{n-1} \quad \ldots \quad s_{n-K+1}]^T, \quad n = 0, 1, \ldots, \check{N}_b - 1 \tag{21}$$

where $s_\gamma = 0$ for $\gamma \notin (0, 1, 2, \ldots, N_b - 1)$. The $K \times 2(P+1)$ channel matrix $\check{\mathcal{H}}$, each column of which represents the impulse response for a fixed Doppler shift, is defined as

$$\check{\mathcal{H}} = \begin{bmatrix} \mathcal{H}(0,-P) & \ldots & \ldots & \mathcal{H}(0,P) \\ \mathcal{H}(1,-P) & \ldots & \ldots & \mathcal{H}(0,P) \\ \vdots & \ldots & \ldots & \vdots \\ \mathcal{H}(K-1,-P) & \ldots & \ldots & \mathcal{H}(K-1,P) \end{bmatrix}. \tag{22}$$

Lastly, let $\boldsymbol{\varrho} \in \mathbb{C}^{2P+1}$ be an $\check{N}_b$-length sequence of phase vectors with elements $w_{N_b} = e^{j2\pi/N_b}$:

$$\boldsymbol{\varrho}_n = \begin{bmatrix} w_{N_b}^{Pn} & w_{N_b}^{(P-1)n} & \ldots & w_{N_b}^{-(P-1)n} & w_{N_b}^{-Pn} \end{bmatrix}^T \tag{23}$$

where $n = 0, 1, \ldots, \check{N}_b - 1$. Channel output in (20) can be written as follows:

$$x_n = \mathbf{s}_n^T \check{\mathcal{H}} \boldsymbol{\varrho}_n \tag{24}$$

$$= (\boldsymbol{\varrho}_n^T \otimes \mathbf{s}_n^T)\mathfrak{h}, \quad n = 0, 1, \ldots, \check{N}_b - 1 \tag{25}$$

where $\mathfrak{h} = \mathrm{vec}(\check{\mathcal{H}}) \in N_h$ is the channel coefficients vector. In a more compact form, channel system of equations is

$$\mathbf{x} = \boldsymbol{\Psi}\mathfrak{h} \tag{26}$$

where $\boldsymbol{\Psi}$ is the $\check{N}_b \times N_h$ sensing matrix:

$$\boldsymbol{\Psi} = \begin{bmatrix} (\boldsymbol{\varrho}_0 \otimes \mathbf{s}_0)^T \\ (\boldsymbol{\varrho}_1 \otimes \mathbf{s}_1)^T \\ \vdots \\ (\boldsymbol{\varrho}_{\check{N}_b-1} \otimes \mathbf{s}_{\check{N}_b-1})^T \end{bmatrix}. \tag{27}$$

Sensing matrix $\boldsymbol{\Psi}$ can also be expressed as the concatenation of $K$ blocks each of which are $(\check{N}_b \times 2P+1)$-dimensional matrices:

$$\boldsymbol{\Psi} = [\,\boldsymbol{\Psi}_0 \quad \boldsymbol{\Psi}_1 \quad \ldots \quad \boldsymbol{\Psi}_{K-1}\,]. \tag{28}$$

If we have noise, Eq. (26) becomes:

$$\mathbf{x} = \boldsymbol{\Psi}\mathfrak{h} + \mathbf{n}, \tag{29}$$

where $\mathbf{n}$ is a zero-mean white Gaussian noise with variance $\sigma^2$. In the following discussions, $\mathfrak{h}$ is treated as an unknown deterministic vector.

Linear measurement model in (29) contains $N_h$ unknowns and sensing matrix $\boldsymbol{\Psi}$ is a full rank matrix. Therefore, without a prior sparsity assumption, LS solution of $\mathfrak{h}$ is:

$$\hat{\mathfrak{h}} = \left(\boldsymbol{\Psi}^H\boldsymbol{\Psi}\right)^{-1}\boldsymbol{\Psi}^H\mathbf{x}. \tag{30}$$

This solution is also the maximum likelihood estimate [40]. Clearly, we can write the $\hat{\mathfrak{h}}$ as [9,27]:

$$\hat{\mathfrak{h}} = \mathfrak{h} + \left(\boldsymbol{\Psi}^H\boldsymbol{\Psi}\right)^{-1}\boldsymbol{\Psi}^H\mathbf{n} \tag{31}$$

and the mean squared error of the LS estimator is bounded by the following equation.

$$\mathbb{E}\|\hat{\mathfrak{h}} - \mathfrak{h}\|_2^2 = \mathbb{E}\left\|\left(\boldsymbol{\Psi}^H\boldsymbol{\Psi}\right)^{-1}\boldsymbol{\Psi}^H\mathbf{n}\right\|_2^2 \tag{32}$$

$$= \sigma^2\,\mathrm{tr}\left(\left(\boldsymbol{\Psi}^H\boldsymbol{\Psi}\right)^{-1}\right) \tag{33}$$

$$\geqslant \sigma^2 N_h. \tag{34}$$

If we have a prior sparsity information, oracle estimator solution of $\mathfrak{h}$ is:

$$\hat{\mathfrak{h}} = \left(\boldsymbol{\Psi}_{\Lambda_o}^H\boldsymbol{\Psi}_{\Lambda_o}\right)^{-1}\boldsymbol{\Psi}_{\Lambda_o}^H\mathbf{x}. \tag{35}$$

Similarly, the mean squared error of the oracle estimator is bounded by:

$$\mathbb{E}\|\hat{\mathfrak{h}} - \mathfrak{h}\|_2^2 = \sigma^2\,\mathrm{tr}\left(\left(\boldsymbol{\Psi}_{\Lambda_o}^H\boldsymbol{\Psi}_{\Lambda_o}\right)^{-1}\right) \tag{36}$$

$$\geqslant \sigma^2\mathcal{S}. \tag{37}$$

If we compare the results presented in (34) and (37), conventional LS estimator shows poor performance in the identification of sparse multipath channels. Although constructing an oracle estimator is practically impossible, there exists efficient algorithms, such as OMP, that provide much more better estimates than the conventional LS estimator and have proven performance guarantees

in sparse multipath channels. For example, the following theorem quantifies the MSE upper bound for OMP in terms of mutual coherence in (16):

**Theorem 1.** *(See [27].) Assume that $\mathfrak{h}$ is an unknown deterministic vector with $\|\alpha\| = \mathcal{S}$ and $\mathbf{x} = \boldsymbol{\Psi}\mathfrak{h}$ where $\mathbf{n}$ is a Gaussian random vector with mean 0 and covariance $\sigma^2\mathbf{I}$. Define*

$$|\mathfrak{h}_{\min}| = \min_{i\in\Lambda_o} |\mathfrak{h}_i|, \tag{38}$$

$$|\mathfrak{h}_{\max}| = \max_{i\in\Lambda_o} |\mathfrak{h}_i|. \tag{39}$$

*Assume that*

$$2\sigma\sqrt{2(1+\varepsilon)\log N_h} \leqslant |\mathfrak{h}_{\min}| - (2\mathcal{S}-1)\mu(\boldsymbol{\Psi})|\mathfrak{h}_{\min}|, \tag{40}$$

*for some constant $\varepsilon > 0$. Then with probability exceeding*

$$1 - \frac{1}{N_h^{\varepsilon}\sqrt{\pi(1+\varepsilon)\log N_h}} \tag{41}$$

*the obtained solution $\hat{\mathfrak{h}}$ of the OMP satisfies*

$$\begin{aligned}\|\hat{\mathfrak{h}} - \mathfrak{h}\|_2^2 &\leqslant \frac{2(1+\varepsilon)}{(1-(\mathcal{S}-1)\mu(\boldsymbol{\Psi}))^2}\mathcal{S}\sigma^2\log N_h \\ &\leqslant 8(1+\varepsilon)\mathcal{S}\sigma^2\log N_h. \qquad \square\end{aligned} \tag{42}$$

Although, the solution given by OMP cannot reach the oracle estimator, with very high probability it gives MSE within $8(1+\varepsilon)\log N_h$ multiplied by the MSE bound of oracle estimator given in (37). This result is better than what LS estimator gives. Moreover, we provide another very important theorem which gives the relation between sparsity level and coherency for guaranteed recovery.

**Theorem 2.** *(See [26].) For a general dictionary $\Psi$, every $\mathcal{S}$-sparse signal $\mathfrak{h}$ with*

$$\mathcal{S} < \frac{1}{2}\left(\frac{1}{\mu(\Psi)} + 1\right), \tag{43}$$

*is the unique sparsest representation and is guaranteed to be recovered by OMP when observing*

$$\mathbf{x} = \boldsymbol{\Psi}\mathfrak{h}. \qquad \square \tag{44}$$

As we noticed before, coherency of a dictionary is crucial in representing known data. Atoms in the dictionary should not resemble each other. Namely, the sequence which is used to construct the dictionary should have good incoherency properties and constructed dictionary should have a coherency value close to the lower bound given in (17). In the last part of this section, we present a candidate channel probing sequence called Alltop sequences, which enable us to construct dictionaries with very good incoherence properties [41]. Alltop sequences have been used effectively in several different areas [42,43]. For some prime number $\check{N}_b \geqslant 5$, Alltop sequence, $\mathbf{s}_A = (s_b)_{b=0}^{\check{N}_b-1}$, has the following elements

$$s_b = \frac{1}{\sqrt{\check{N}_b}} e^{2\pi j b^3/\check{N}_b}. \tag{45}$$

Considering that the $\|\mathbf{s}_A\|_2 = 1$ and the dictionary structure in (28), within the same block we have the following property:

$$\left\|\langle\psi_{k,i}, \psi_{k,i'}\rangle\right\| = 0, \quad \text{if } i \neq i', \tag{46}$$

$$\left\|\langle\psi_{k,i}, \psi_{k,i'}\rangle\right\| = 1, \quad \text{if } i = i'. \tag{47}$$

For different blocks, $k \neq k'$, we have the following property:

$$\left\|\langle\psi_{k,i}, \psi_{k',i'}\rangle\right\| = \frac{1}{\sqrt{\check{N}_b}}, \tag{48}$$

for all $i, i' = 0, \ldots, K-1$. In order to emphasize this desirable feature of Alltop sequences, assume that there exists the same number of delay and Doppler bins as $\check{N}_b$ and the resulting dictionary is $\boldsymbol{\Psi} \in \mathbb{C}^{\check{N}_b \times \check{N}_b^2}$. By using (17) we know that the lower coherency bound is $\frac{1}{\sqrt{\check{N}_b+1}}$. Therefore, using Alltop sequences and for large values of $\check{N}_b$, it is clearly seen that this bound can be practically achieved.

## 4. Off-grid problem in sparse signal recovery

General assumption used in all of these sparse multipath/target detection techniques is that all of the multipath components fall on the discrete grid points. Dictionary matrix $\boldsymbol{\Psi}$ is typically constructed based on the assumption of all the possible multipath components are on-grid points. In other words, each atom in the dictionary corresponds to a signal created with a delay–Doppler pair, which fall onto a discrete grid point. However, this situation is practically impossible as the multipath parameters are unknown. In Fig. 1, multipath components that fall on the discrete grid and off-grid points are illustrated. Therefore, the true grid, which is possibly irregular, cannot be known beforehand. This so called off-grid problem, results in a mismatch of the dictionary and severely degrades the performance of techniques that exploit sparsity. If there exists off-grid multipath components, then we won't be able to represent the received signal by using the dictionary $\boldsymbol{\Psi}$ which is created based on the on-grid assumption. Furthermore, such methods exhibit an unstable behavior as previously shown in theoretical studies on dictionary errors. Therefore, atoms of the dictionary $\boldsymbol{\Psi}$ should be properly modified to sparsely represent the receiver output. In several papers, the problem is pointed out and very simple grid refinement approaches are presented [14,15]. In the vicinity of the multipath components, grid is iteratively refined to match with the exact location of the off-grid component. The major drawbacks of these approaches are that this grid refinement is a costly procedure and secondly addition of new atoms to the dictionary adversely affects the recovery guarantees. Therefore, to the best of our knowledge, there exist no viable solution to the off-grid problem in the existing literature up to now.

Negative effects of the off-grid problem can be verified on a four-path scenario. In this scenario we used length-53 Alltop sequence and OMP as a recovery technique. In Fig. 2, all on-grid multipath components are recovered. However, as in Fig. 3, if we perturb the delay–Doppler location of each multipath component in the vicinity of the on-grid point, OMP fails to recover the two paths and makes estimation error in recovery of other two paths. As we pointed out, this result is due to the fact that there exist no atom in the dictionary corresponding to the off-grid multipath components. In the next section, details of the proposed technique to alleviate the off-grid problem by using PSO and OMP will be presented.

## 5. Sparse approximation on cross-ambiguity function surface

We propose a novel technique to overcome the so called off-grid problem in sparse multipath channel modeling. In a multipath environment, as given in (8), the receiver output signal is the superposition of delayed, Doppler shifted and scaled versions of the transmitted signal. In time domain, there exists a considerable overlap between the signals received from different multipaths.

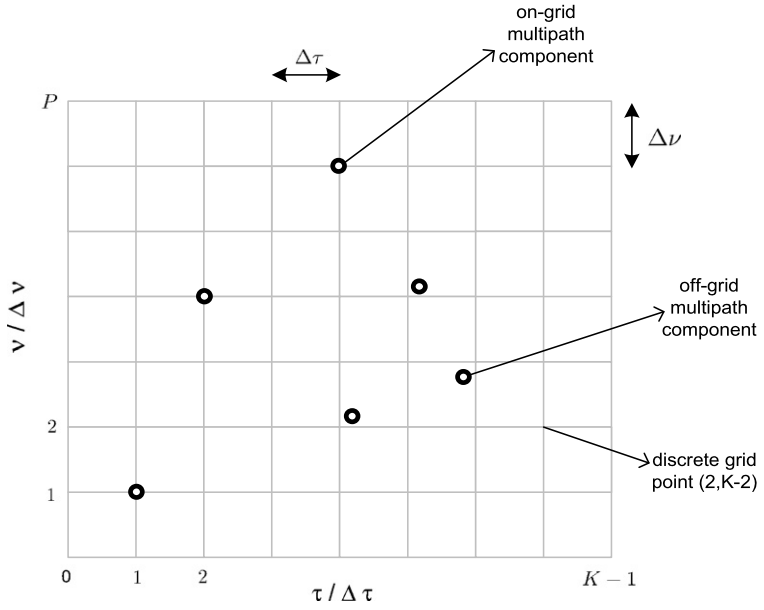


**Fig. 1.** On-grid and off-grid multipath components on delay–Doppler domain.

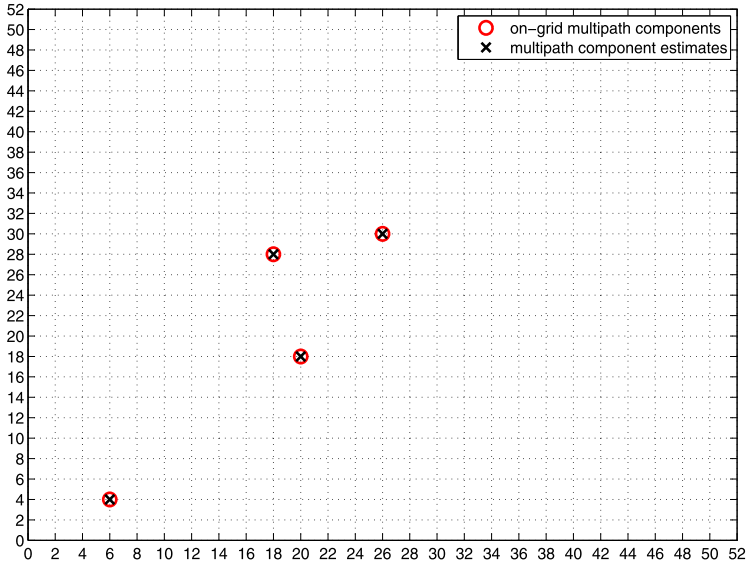


**Fig. 2.** True on-grid and estimated position of each multipath component are illustrated with red circles and black crosses, respectively.

Therefore, it is desirable to have a preprocessing that enables localization of the multipath signal components and reduction of the significant overlapping of components in the time domain. Since typical communication signals are phase or frequency modulated, with large time–bandwidth products, as in radar detection their CAFs are highly localized in the delay–Doppler domain. Therefore, the transformation of the signal outputs to the CAF domain localizes different multipath signals in clusters to their respective delay and Doppler cell. Although there exist several different representation, symmetrical version of the CAF between the transmitted signal and the received signal can be written as [44,45]:

$$\chi_{x(t),s(t)}(\tau,\nu)=\int_{-\infty}^{\infty} x\left(t+\frac{\tau}{2}\right)s^{*}\left(t-\frac{\tau}{2}\right)e^{-j2\pi\nu t}\,dt. \tag{49}$$

To illustrate the localization on delay–Doppler surface, consider a synthetic multipath channel with 6 distinct paths. As shown in Fig. 4(a), the individual multipath signals overlap significantly in time at the output of the receiver. However, as shown in Fig. 4(b), the CAF given in (49) between the received signal and the transmitted signal localizes the contribution of different multipath components in delay–Doppler domain. Matched filtering is the optimum solution for detection, in terms of signal-to-noise ratio (SNR). Performance of the receiver that makes use of a matched filter that matched to the transmitted signal may significantly degrade, when the Doppler shift is not known. The CAF characterizes the output of a matched filter when the input signal is delayed and Doppler shifted. CAF calculation is the optimal solution for detection, in the case of one multipath component. If there exist two or more multipath components separated enough in delay–Doppler domain, again, CAF surface offers a very useful detection surface by using properly chosen waveforms for the application of interest [45–48].

Proposed algorithm starts with the detection of multipath clusters on ambiguity surface exceeding a predefined detection threshold. For that purpose, peak point of the ambiguity surface,

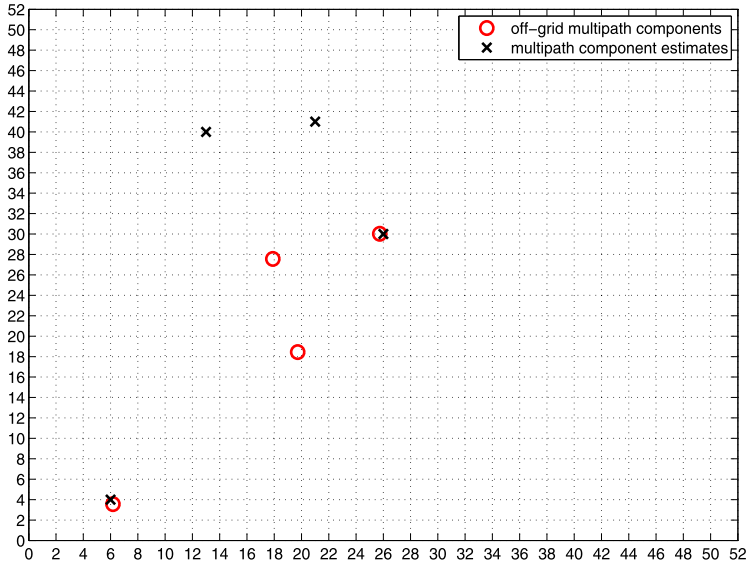


**Fig. 3.** True off-grid and estimated position of each multipath component is illustrated with red circles and black crosses, respectively.

$|\chi(\tau,\nu)|$, is determined and compared with the noise level. Similar strategies are commonly employed in radar target detection [49]. Since it is relatively insensitive to possible outliers, noise level on the ambiguity surface is quantified with the median operator. Specifically, ratio between maximum and median values of $|\chi(\tau,\nu)|$ is computed as:

$$\frac{\max(|\chi_{x(t),s(t)}(\tau,\nu)|)}{\text{median}(|\chi_{x(t),s(t)}(\tau,\nu)|)} \geqslant \xi \tag{50}$$

and compared with a properly chosen threshold value, $\xi$. If the calculated ratio is higher than the determined threshold value then that peak point is considered as the center location of the strongest multipath cluster [50,51]. This detection phase enable us to determine corresponding grid points that will be perturbed to be able to detect multipath components that reside on possible off-grid locations. Having estimated the parameters of each multipath component in the cluster, effect of the cluster is eliminated from receiver output to recurse on the residual for detection of the remaining multipath clusters.

As stated before, we make use of PSO as a global optimization tool in this work [16]. Although, there are exist many evolutionary techniques like differential evolution [52], genetic algorithm [53], artificial bee colony algorithm [54], shuffled frog leaping [55], MCMC-based methods and their variations, PSO is by far the most well-known evolutionary algorithm in the literature. Main reasons for its popularity are simplicity, ease of modification with which it can be adapted to various practical applications, ability to be hybridized with other methods and faster in convergence. A very good reference paper categorizing a large number of publications on applications of PSO can be found in [56].

Block diagram of the proposed technique is presented in Figs. 5 and 6. $C$ clusters of multipath components present on delay–Doppler domain and the number of multipath components in cluster $c$ is $d_c$ for $1 \leqslant c \leqslant C$. For example, as shown in Fig. 4(b), 6 multipath components are localized in $C = 2$ clusters and each cluster consists of 3 multipath components. Having identified the location of each cluster, individual PSO searches are conducted to perturb the assumed discrete set of delay and Doppler values that are in the support of each cluster, separately. Then, with this optimized dictionary matrix, OMP is used as a sparse reconstruction method to estimate delay–Doppler parameters of multipaths. Following PSO searches and multipath reconstruction in each cluster,

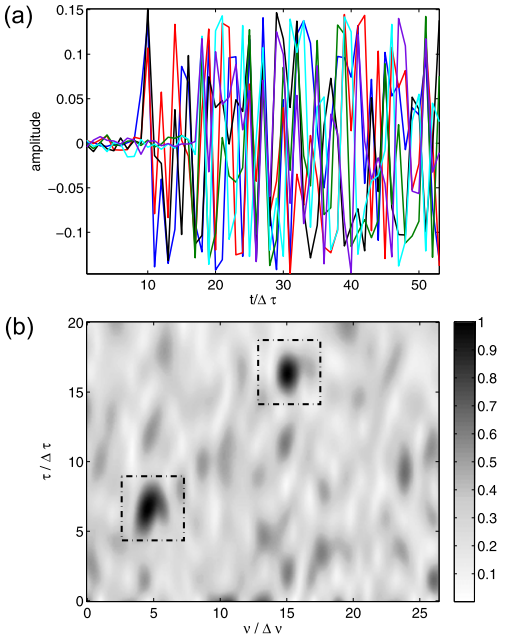


**Fig. 4.** 6 Alltop sequences (a) in time domain, and (b) in delay–Doppler domain localized in 2 clusters each of which has 3 paths.

effects of the estimated multipath components are eliminated for a better estimation in the remaining clusters. Since, optimization in each cluster has to be performed multiple times, PSO iterations in each cluster need not to be pursued until convergence is established. Therefore, by cycling over the identified set of clusters, the proposed technique iteratively provides estimates for each path in each cluster. Before getting into to the details of the proposed technique, it is better to clarify the differences between the PSO–OMP and the technique presented in [47]. Firstly, main goals of this work are to analyze the level of performance degradation when off-grid problem is not considered and propose a computationally efficient sparse channel estimation method using sparse approximation tools and compressed sensing theory. Secondly, PSO is used to optimize the dictionary atoms to better represent the received signal. PSO optimization is conducted in a small dimensional search space. Search dimension is equal to the number of delay–Doppler grid points of the cluster that is under consider-

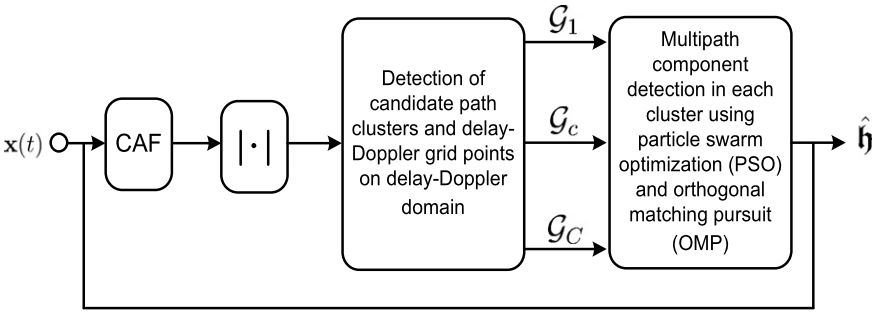


**Fig. 5.** Signal flow diagram of the algorithm.

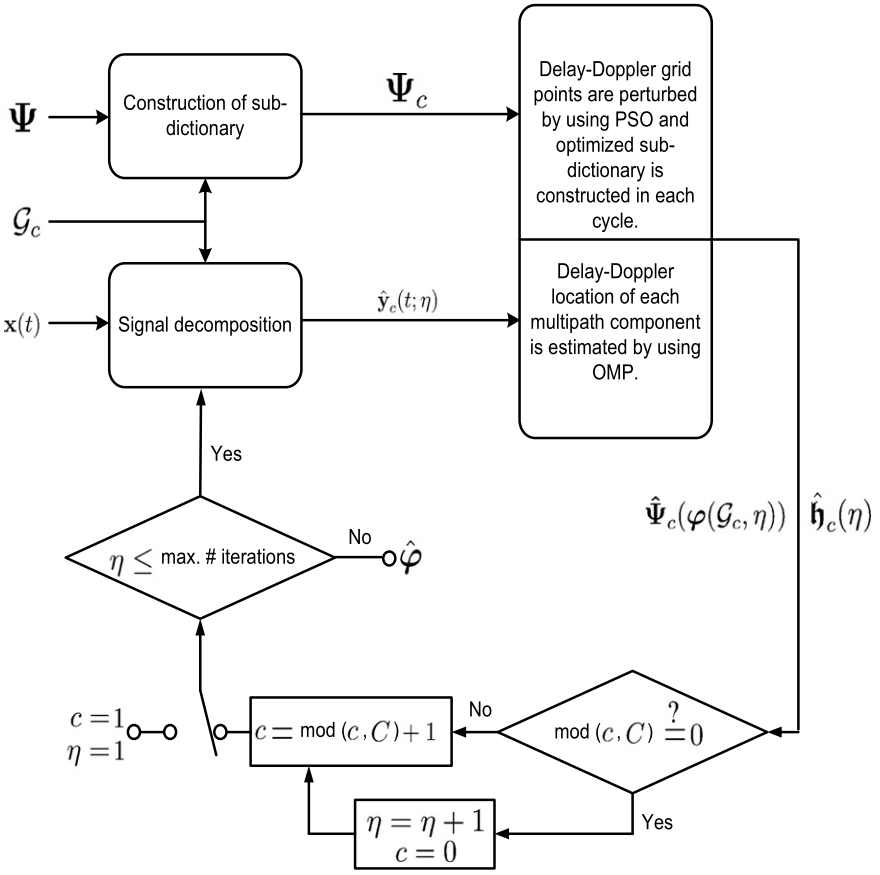


**Fig. 6.** Signal flow sub-block diagram of the parameter estimation in each cluster using PSO and OMP block in Fig. 5.

ation. Therefore, optimization search dimension is constant irrespective of the number of multipath components. However, in [56], a search is conducted to estimate delay, Doppler, azimuth and elevation in a four times number of multipath components dimensional space. Namely, search dimension depends on number of multipath components under the considered cluster. Lastly, in each PSO cycle, a linear system of equation based on the modified dictionary is solved using OMP in a computationally efficient way. On the other hand, in [56], a cost function formulated in the ambiguity function domain is evaluated in each PSO cycle and fitness of each particle is determined. The only similar point between this manuscript and [56] is in the detection mechanism of the clusters on the cross-ambiguity function domain. However, as stated previously, peak detection procedure is a strategy which is commonly employed in the literature. In the following, formulation of the proposed technique for each cluster is presented.

Associated with the $c$th cluster, the following fitness function is optimized:

$$f_c\big(\boldsymbol{\Psi}_c\big(\boldsymbol{\varphi}(\mathcal{G}_c),\eta\big),\mathfrak{h}_\eta\big)=\big\|\hat{\mathbf{y}}_c(t,\eta)-\boldsymbol{\Psi}_c\big(\boldsymbol{\varphi}(\mathcal{G}_c),\eta\big)\mathfrak{h}_c(\eta)\big\|_2^2, \quad (51)$$

where $\eta$ represents the iteration index of the algorithm, $\boldsymbol{\varphi}\in\mathbb{R}^{N^2}$ is the vector containing all possible discrete delay–Doppler values:

$$\boldsymbol{\varphi}=[\varphi_{11},\ldots,\varphi_{1N},\varphi_{21},\ldots,\varphi_{NN}], \quad (52)$$

$\varphi_{1N}=[\tau_1,\nu_N]$, $\mathcal{G}_c$ is the set containing index of grid points (each grid point corresponds to a delay–Doppler value pair) inside the $c$th cluster, $\hat{\mathbf{y}}_c(t,\eta)$ is the estimated output signal and $\boldsymbol{\Psi}_c$ is the sub-dictionary created using the columns of dictionary matrix $\boldsymbol{\Psi}$ that are in the set $\Lambda_c$ which contains column index of vectors of $\boldsymbol{\Psi}$, that are in support of cluster $c$:

$$\boldsymbol{\Psi}_c=(\boldsymbol{\psi}_i\colon i\in\Lambda_c). \quad (53)$$

In Fig. 7, two clusters of on-grid points are shown. For each cluster, during PSO cycles, location of these grid points are changed to update the corresponding atoms that belong to the cluster. With these definitions, $\boldsymbol{\varphi}(\mathcal{G}_c)$ holds the delay–Doppler pairs that will

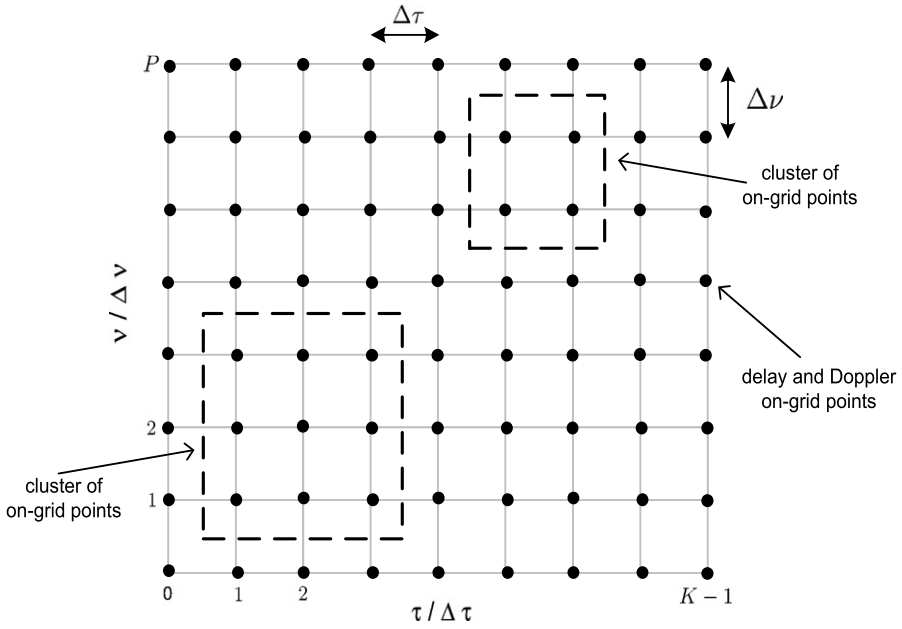


**Fig. 7.** Equally spaced $P+1 \times K$ discrete on-grid points on delay–Doppler domain. 2 clusters of on-grid points are selected.

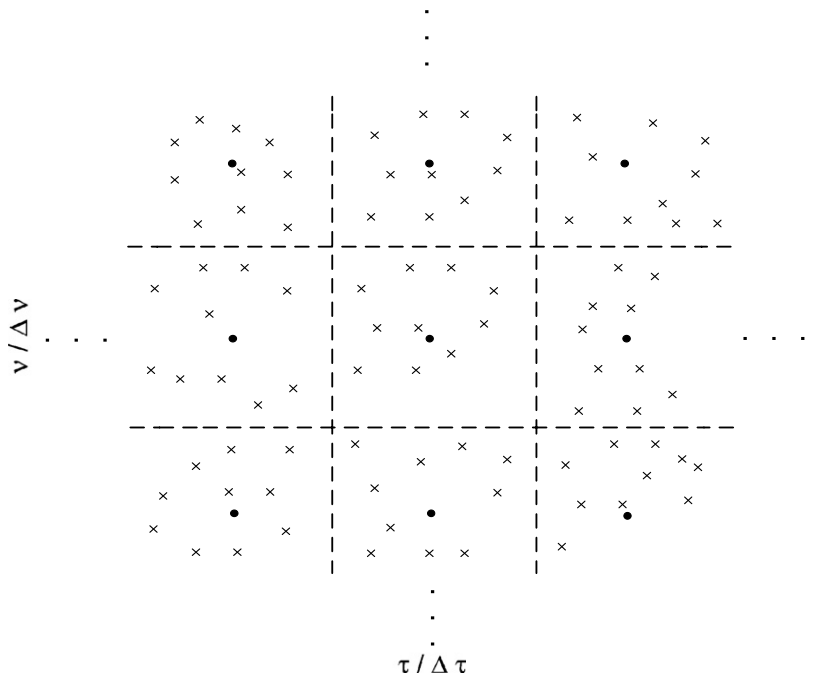


**Fig. 8.** On-grid points that reside in a cluster are zoomed. Boundaries around each on-grid point is marked with dash lines. Crosses represent particles. 10 particles are used. Corresponding vector (here, 9-dimensional) elements of each particle is notated with crosses. That is why many crosses are seen in the figure, although there are 10 particles.

be perturbed and $\boldsymbol{\Psi}_c(\boldsymbol{\varphi}(\mathcal{G}_c), \eta)$ holds the vectors that are created with these perturbed delay–Doppler pairs during the PSO cycles at $\eta$th iteration. Proposed algorithm works in an expectation-maximization (EM) like manner [57]. EM includes two steps called as expectation and maximization steps. In the expectation step: unobservable (complete) data is estimated. In the maximization step: an optimization is conducted to find the unknown parameters of the expected complete data. Estimated output signal, so called the complete data $\hat{\mathbf{y}}_\eta$, can be formed as:

$$\hat{\mathbf{y}}_c(t;\eta) = \mathbf{x}(t) - \sum_{\gamma=1,\gamma\neq c}^{C} \hat{\boldsymbol{\Psi}}_\gamma\big(\boldsymbol{\varphi}(\mathcal{G}_\gamma,\eta)\big)\hat{\mathbf{h}}_c(\eta). \tag{54}$$

For the first iteration, $\eta = 1$, for the first cluster, $\hat{\mathbf{y}}_c(t;\eta)$ is initialized as $\hat{\mathbf{y}}_c(t;\eta) = \mathbf{x}(t)$. In the maximization step, the channel parameter estimates and proper sub-dictionary to represent off-grid multipaths for the $c$th cluster at $\eta$th iteration are obtained by maximizing the following optimization problem:

$$\hat{\boldsymbol{\Psi}}_c\big(\boldsymbol{\varphi}(\mathcal{G}_c,\eta)\big)\hat{\mathbf{h}}_c(\eta) = \arg\max_{\boldsymbol{\Psi}_c,\mathbf{h}_c} \frac{1}{f_c(\boldsymbol{\Psi}_c(\boldsymbol{\varphi}(\mathcal{G}_c),\eta),\mathbf{h}_c(\eta))}. \tag{55}$$

Channel parameters and sub-dictionary corresponding to the $c$th cluster are estimated using swarm of particles in a $|\mathcal{G}_c|$-dimensional search space. As shown in Fig. 8, at the beginning of the PSO cycles, particle locations (each of which is a solution candidate) are randomly initialized as:

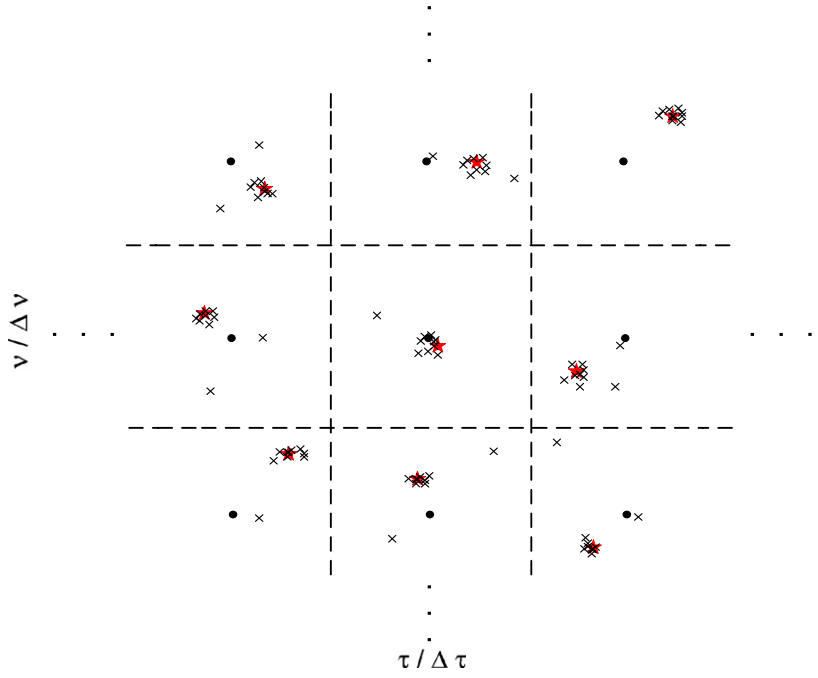


**Fig. 9.** One snapshot coordinates of particles $\mathbf{z}$ ($\times$) and **globalBest**($\mathbf{p}_g, \star$) distributed on the delay–Doppler domain. Particles swarm to the **globalBest** position.

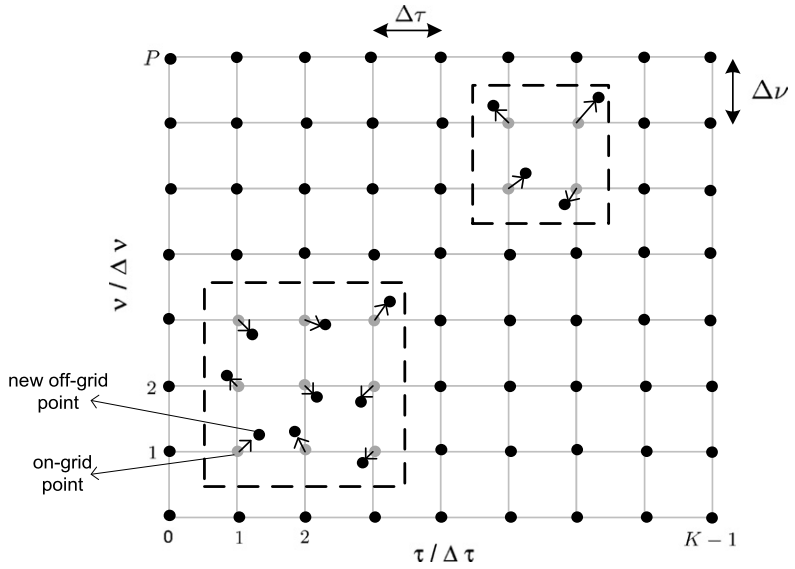


**Fig. 10.** Position update of each grid point, that reside in a cluster, to the estimated new off-grid position.

$$\mathbf{z}_l = \boldsymbol{\varphi}(\mathcal{G}_c) + \mathcal{U}(-\kappa/2, \kappa/2), \quad l = 1, \ldots, \#\text{particles} \tag{56}$$

and updated according to PSO equations as follows

$$\boldsymbol{\upsilon}_l = \iota\big(\boldsymbol{\upsilon}_l + c_1\epsilon_1(\mathbf{p}_l - \mathbf{z}_l) + c_2\epsilon_2(\mathbf{p}_g - \mathbf{z}_l)\big), \tag{57}$$

$$\mathbf{z}_l = \mathbf{z}_l + \boldsymbol{\upsilon}_l. \tag{58}$$

Here, $\mathcal{U}$ represents uniform random distribution, $\kappa$ is the spacing between discrete grid points, $\boldsymbol{\upsilon}_l$ is the velocity vector, $\mathbf{p}_l$ is the personal best position vector, $\mathbf{p}_g$ is the global best position vector, $c_1$ is the cognitive factor that adjusts how much a particle is influenced by the historical best position of his own, $c_2$ is the social factor that adjusts how much a particle is influenced by the historical best of the swarm, $\epsilon_1$ and $\epsilon_2$ are two uniformly distributed random numbers and $\iota$ is the constriction factor, that balances global and local searches and defined as:

$$\iota = \frac{2}{|2 - \varsigma - \sqrt{\varsigma^2 - 4\varsigma}|}, \tag{59}$$

where $\varsigma = c_1 + c_2$. Recommended values for these constants are $c_1 = c_2 = 2.05$ and $\iota = 0.72984$ [58].

Location, $\mathbf{z}_l \in \mathbb{R}^{|\mathcal{G}_c|}$, of each particle in the $|\mathcal{G}_c|$-dimensional search space is a candidate off-grid location solution. Fig. 9 illustrates the search of particles around each discrete grid point and the convergence of particles to a possible solution. In each PSO cycle, the following linear system of equation:

$$\hat{\mathbf{y}}_c(t;\eta) \approx \hat{\boldsymbol{\Psi}}_c\big(\boldsymbol{\varphi}(\mathcal{G}_c, \eta)\big)\mathfrak{h}_c(\eta), \tag{60}$$

is solved using the OMP in a greedy fashion and very efficiently by minimizing

$$\big\|\hat{\mathbf{y}}_c(t;\eta) - \hat{\boldsymbol{\Psi}}_c\big(\boldsymbol{\varphi}(\mathcal{G}_c, \eta)\big)\mathfrak{h}_c(\eta)\big\|_2^2, \tag{61}$$

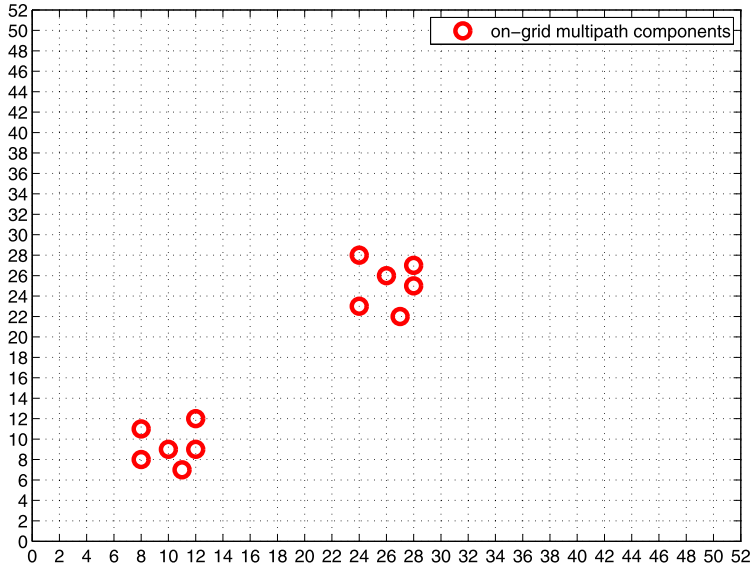


**Fig. 11.** Location of on-grid multipath components in two separate clusters on delay–Doppler domain.

in order to compare the performance of each particle. Eq. (55) is evaluated using the location values of each particle and the location that gives the best fitness chosen as the **globalBest**. In Fig. 10, it is shown that, in the end of PSO cycles, initial on-grid points are updated based on the minimization results and new off-grid points are estimated to better support existing off-grid multipath components. Having estimated the parameters of each multipath component in the $c$th cluster, effects of these multipath components are eliminated as in (54) from the receiver output for a better estimation in remaining clusters. Iterations, $\eta$, continue until convergence is established or a preset number of iterations is reached.

## 6. Simulation results

In this section, we present numerical results to clarify the performance gains obtained by exploiting the clustered structure and handling the off-grid problem. In all simulations, length-53 ($\check{N}_b = 53$) Alltop sequences are used as a probing signal and OMP is used as a sparse recovery technique. 500 Monte Carlo simulations are conducted for each scenario. Number of multipath components, in other words sparsity level $\mathcal{S}$ is changed in between 2 and 12. It is assumed that there exist two multipath clusters exists on the delay–Doppler domain. Locations of multipath components in two separate clusters on delay–Doppler domain are shown in Fig. 11. In each Monte Carlo realization, cluster locations are preserved but multipath component locations are randomly changed.

Firstly, we observe the effect of off-grid problem in terms of recovery percentage for different sparsity $\mathcal{S} = 2, \ldots, 12$ and perturbation levels $\kappa = 0, 0.2, 0.3, 0.4, 0.5$. On-grid delay–Doppler location of each multipath components is perturbed as follows:

$$\tau_i = \tau_i + \mathcal{U}(-\kappa, \kappa)/\Delta\tau, \tag{62}$$

$$\nu_i = \nu_i + \mathcal{U}(-\kappa, \kappa)/\Delta\nu. \tag{63}$$

Results obtained by OMP are shown in Fig. 12. Note that, for $\mathcal{S} \leqslant 4$ we have a recovery rate of 100% when all multipath components are on-grid as suggested by Theorem 2. However, when we perturb the delay–Doppler location of each multipath randomly within a limit $\kappa$, performance degrades severely. Since we cover all probable delay–Doppler parameter pairs, most realistic scenario is when $\kappa = 0.5$. Even for sparsity level smaller than 4, $\mathcal{S} \leqslant 4$, we have an approximately 30% decrease in recovery percentage.

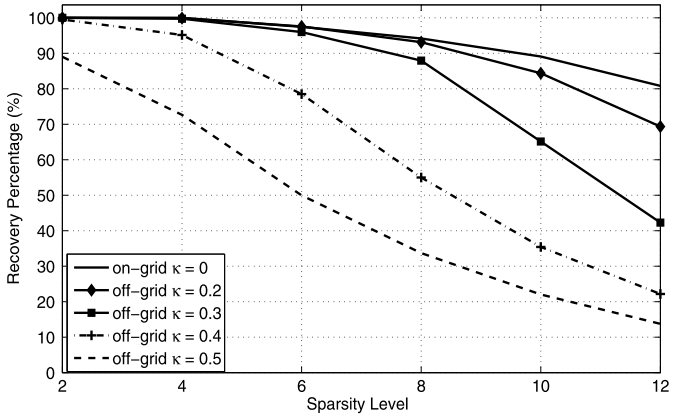


**Fig. 12.** Recovery percentage the OMP technique for various sparsity and perturbation levels.

In the following experiments, we provide results obtained by using proposed technique for various different settings. In the first experiment, we look for how much we can improve recovery ability with minimum resources. In other words, using minimum number of particles, PSO cycles and EM iterations. Choose number of particles as 2, number of PSO cycles as $10, 30, 50$, and number of EM iterations $\eta = 1$ and $\kappa = 0.5$. All simulations are conducted on an HP Desktop with Intel Core-2 2.13 GHz processor. Parameter estimation time of standard OMP technique is recorded as 0.012 s and PSO–OMP with #particles $= 2$, $\eta = 1$, PSO cycles $=$ 10 is recorded as 0.3 s for this specified scenario. In Fig. 13 performance of PSO–OMP with #particles $= 2$ and $\eta = 1$ is compared with standard OMP and results are presented in terms of recovery percentage (%), rMSE and rMSE of detected multipath components. Perturbation limit $\kappa$ is set to 0.5. $10, 30$ and 50 PSO iterations are conducted. Even for 10 PSO iterations, PSO–OMP outperform OMP and solves the off-grid problem. For example, it seen that, for sparsity level 10, PSO–OMP with 10 PSO iterations increase the recovery 20%. Moreover it is obvious that performance is increased with higher number of PSO iterations due to the increased chance of converging the global solution. In the second set of results that are presented in Fig. 14, we provide the performance improvements in recovery percentage, rMSE and rMSE of detected multipaths when we have 2 EM iterations instead of

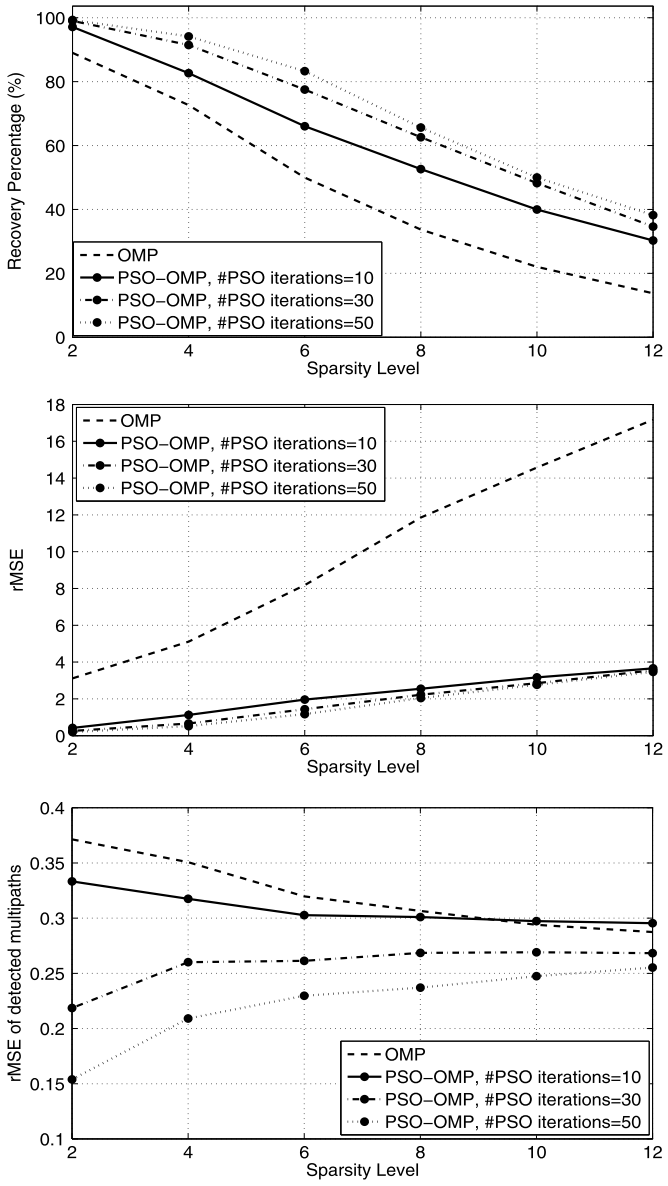


**Fig. 13.** Recovery percentage, rMSE and rMSE of detected multipath components of OMP and PSO–OMP (number of EM iterations is 1 and number of particles is 2), for various sparsity levels and number of PSO iterations, respectively. Perturbation limit, $\kappa = 0.5$.

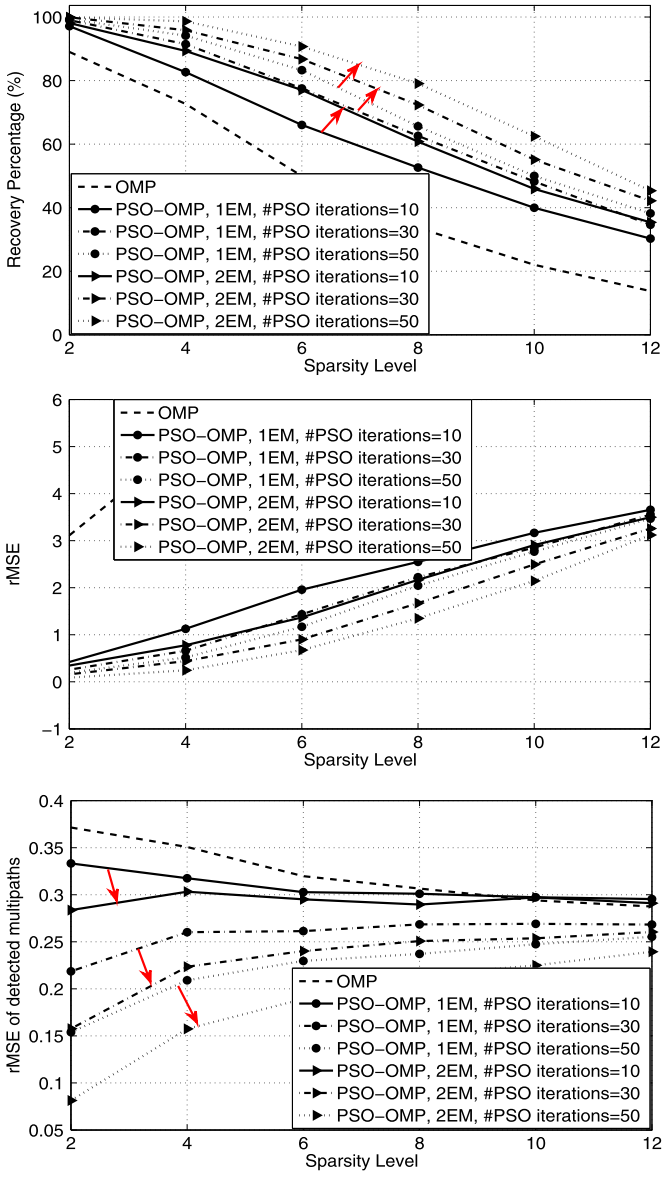


**Fig. 14.** Recovery percentage, rMSE and rMSE of detected multipath components of OMP, PSO–OMP (number of EM iterations is 1 and number of particles is 2), and PSO–OMP (number of EM iterations is 2 and number of particles is 2), for various sparsity levels and number of PSO iterations, respectively. $\kappa = 0.5$.

1 EM iteration. As expected, since we better isolate the effect of multipath clusters to each other with addition of second EM iteration, we obtained better results. For sparsity level 6, recovery percentage is increased approximately 15%, with addition of 1 EM iteration.

In the third set of results we tested the performance of the algorithm by using more resources. Namely, number of particles is increased to 4 and number of PSO iterations are increased to 80. It is seen in Fig. 15 that performance is increased. For sparsity level 6, recovery percentage is increased approximately 10%.

We also analyzed the error progress curves of EM and PSO iterations. In Fig. 16, the rMSE between measured and estimated receiver outputs obtained with PSO–OMP, are shown for 20 EM iterations values and for different sparsity levels. Number of PSO iterations and number of particles are chosen as 30 and 2, respectively. The rMSE is monotonically decreases and saturates around after 20 iterations. Sharp decreases occur in between 1th–4th iterations. However, as proofed in previously conducted experiments, even for 1 EM iteration, off-grid problem is successfully handled. In Fig. 17, normalized error versus number of PSO iterations curve obtained with PSO–OMP is shown. Number of EM iterations and number of particles are chosen as 2 and 2, respectively. Similar to the curve in Fig. 16, the rMSE is monotonically decreases and saturates around after 100 iterations. Meaning that, particles converged to a point and movement of particles does not change the estimation error anymore. Note from the results shown in Fig. 13 that only 10 PSO iterations are enough to get good results. Finally, similar to the results shown in Fig. 12, for various perturbation limit values performance of the PSO–OMP is tested (see Fig. 18). Number of EM iterations, number of particles and number of PSO iterations are chosen as 1, 2 and 10, respectively. As expected, when we lower the perturbation limit we get better results. This is due to the fact that, since particles firstly search for the global optimum in the very vicinity of the on-grid point, they found the correct off-grid point rapidly and inrush onto the point.

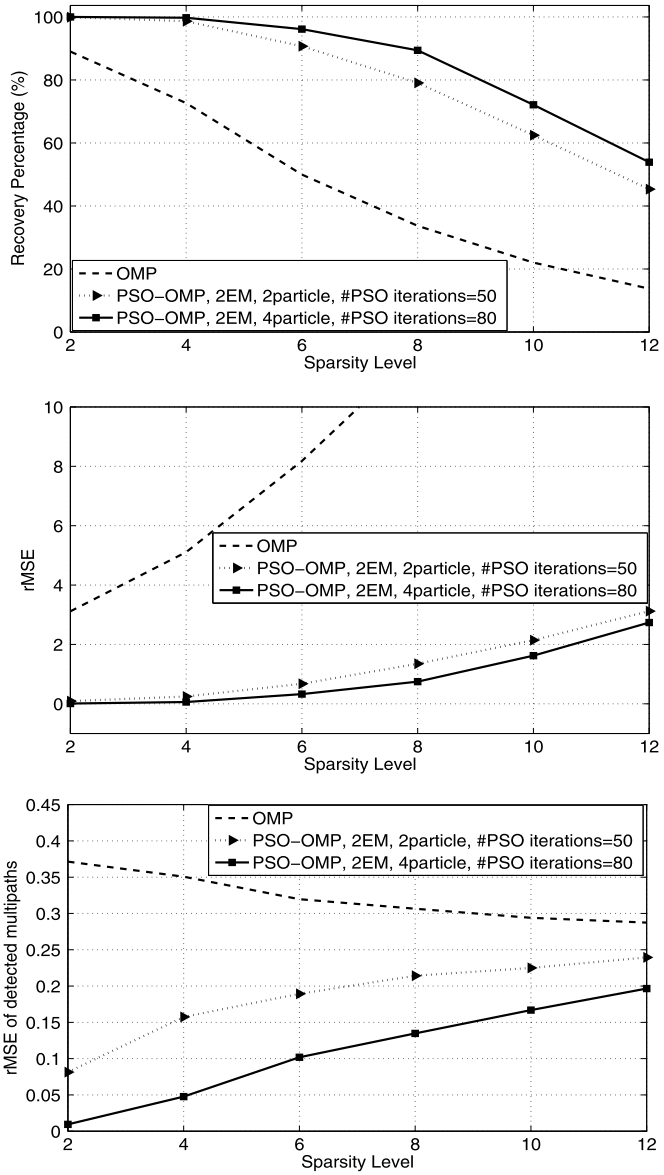


**Fig. 15.** Recovery percentage, rMSE and rMSE of detected multipath components of OMP, PSO–OMP (number of EM iterations is 2 and number of particles is 2), and PSO–OMP (number of EM iterations is 2 and number of particles is 4), for various sparsity levels and number of PSO iterations, respectively. $\kappa = 0.5$.

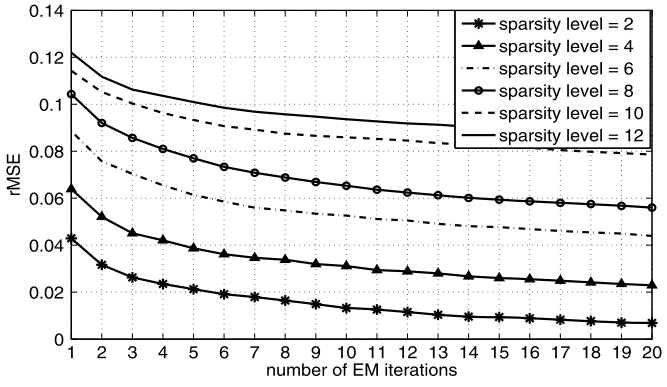


**Fig. 16.** rMSE values for various EM iteration values and for various sparsity levels obtained with PSO–OMP. Number of particles is 2 and number of PSO iterations is 30.

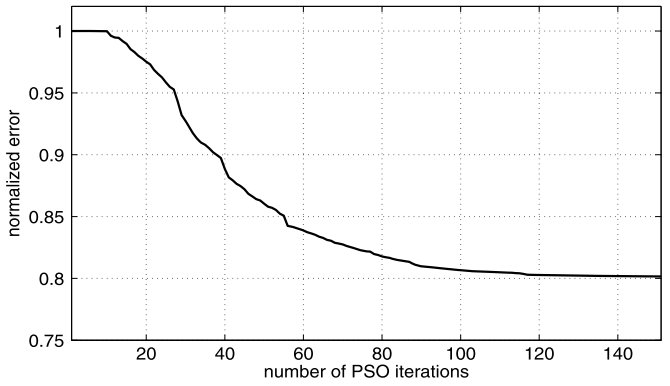


**Fig. 17.** Normalized error during PSO iterations obtained with PSO–OMP. Number of particles is 2 and number of EM iterations is 2.

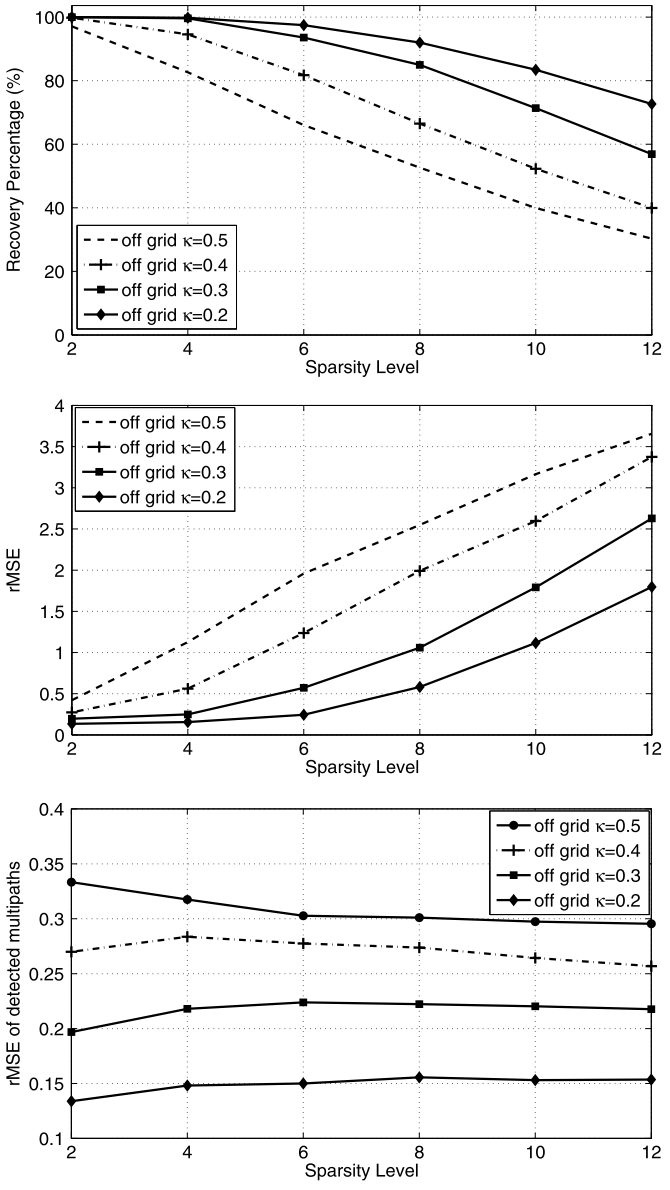


**Fig. 18.** Recovery percentage, rMSE and rMSE of detected multipath components of PSO–OMP (number of EM iterations is 1 and number of particles is 2, number of PSO iterations is 10) for various sparsity levels and $\kappa$ values. $\kappa = 0.5$.

## 7. Conclusions

Based on sparse approximation tools and compressed sensing theory, a new approach for identification of sparse multipath channels is presented. A general assumption used in all of the sparse multipath channel estimation techniques is that the all multipath components fall on the grid points, which is practically impossible as the target parameters are unknown. Performance of standard compressed sensing formulations based on discretization of the multipath channel parameter space degrade significantly when the actual channel parameters deviate from the assumed discrete set of values. To solve this so called "off-grid", we proposed a novel algorithm that can also be used in applications other than the multipath channel identification. The proposed algorithm, firstly makes use of the cross-ambiguity function calculation and transform the receiver output to the delay–Doppler domain for efficient exploitation of the delay–Doppler diversity of the multipath signals. Then by detecting the candidate multipath clusters, the original channel identification problem is reduced to channel identification problems over the identified clusters in the delay–Doppler domain. After that, on-grid points that reside in each cluster are perturbed by using PSO and multipath components are recovered by using OMP in a greedy fashion. Superior performance of the proposed algorithm verified on various test scenarios.

## References


[1] L. Tong, B.M. Sadler, M. Dong, Pilot-assisted wireless transmissions, IEEE Signal Process. Mag. 21 (6) (2004) 12–25.
[2] L. Tong, S. Perreau, Multichannel blind identification: From subspace to maximum likelihood methods, Proc. IEEE 86 (10) (1998) 1951–1968.
[3] X. Ma, G.B. Giannakis, S. Ohno, Optimal training for block transmissions over doubly selective wireless fading channels, IEEE Trans. Signal Process. 51 (5) (2003) 1351–1366.
[4] I. Barhumia, G. Leus, M. Moonen, Optimal training design for MIMO OFDM systems in mobile wireless channels, IEEE Trans. Signal Process. 51 (6) (2003) 1615–1624.
[5] N. Czink, X. Yin, H. Ozcelik, M. Herdin, E. Bonek, B.H. Fleury, Cluster characteristics in a MIMO indoor propagation environment, IEEE Trans. Wireless Commun. 6 (4) (2007) 1465–1475.
[6] A. Saleh, R. Valenzuela, A statistical model for indoor multipath propagation, IEEE J. Sel. Areas Commun. (1987) 128–137.
[7] A. Molisch, Ultrawideband propagation channels – theory, measurement and modeling, IEEE Trans. Veh. Technol. (2005) 1528–1545.
[8] S. Cotter, B. Rao, Sparse channel estimation via matching pursuit with applications to equalization, IEEE Trans. Commun. 50 (3) (2002) 374–377.
[9] W. Bajwa, A. Sayeed, R. Nowak, Learning sparse doubly-selective channels, in: Proc. 46th Ann. Allerton Conf., 2008.
[10] A. Sayeed, B. Aazhang, Joint multipath–Doppler diversity in mobile wireless communications, IEEE Trans. Commun. (1999) 123–132.
[11] W. Bajwa, J. Haupt, A. Sayeed, R. Nowak, Compressed channel sensing: A new approach to estimating sparse multipath channels, Proc. IEEE 98 (6) (2010) 1058–1076.
[12] G.E. Pfander, H. Rauhut, J. Tanner, Identification of matrices having a sparse representation, IEEE Trans. Signal Process. 56 (11) (2008) 257–275.
[13] M. Herman, T. Strohmer, High-resolution radar via compressed sensing, IEEE Trans. Signal Process. 57 (6) (2009) 2275–2285.
[14] D. Malioutov, M. Cetin, A.S. Willsky, A sparse signal reconstruction perspective for source localization with sensor arrays, IEEE Trans. Signal Process. 53 (8) (2005) 3010–3023.
[15] Y. Yu, A.P. Petropulu, H.V. Poor, MIMO radar using compressive sampling, IEEE J. Sel. Topics Signal Process. 4 (1) (2010) 146–163.
[16] J. Kennedy, R. Eberhart, Particle swarm optimization, in: IEEE Int. Conf. Neural Networks, 1995.
[17] Y.C. Pati, R. Rezaiifar, P.S. Krishnaprasad, Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition, in: Proc. 27th Ann. Asilomar Conf. Signals, Systems and Computers, 1993.
[18] A. Sayeed, A virtual representation for time- and frequency-selective correlated MIMO channels, in: IEEE Int. Conf. Acoust. Speech Signal Processing (ICASSP), 2003.
[19] W. Bajwa, A. Sayeed, R. Nowak, Sparse multipath channels: Modeling and estimation, in: IEEE Workshop on Digital Signal Processing, 2008.
[20] W.F. Schreiber, Advanced television systems for terrestrial broadcasting: Some problems and some proposed solutions, Proc. IEEE 83 (1995) 958–981.
[21] I.J. Fevrier, S.B. Gelfand, M.P. Fitz, Reduced complexity decision feedback equalization for multipath channels with large delay spreads, IEEE Trans. Commun. 47 (1999) 927–937.
[22] M. Kocic, Sparse equalization for real-time digital underwater acoustic communications, in: Proc. OCEAN, 1995.
[23] W. Li, J. Preisig, Estimation of rapidly time-varying sparse channels, IEEE J. Ocean. Eng. (2007) 927–939.
[24] S. Ariyavisitakul, N.R. Sollenberger, L.J. Greenstein, Tap selectable decision-feedback equalization, IEEE Trans. Commun. 45 (1997) 1497–1500.
[25] S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, SIAM Rev. 43 (1) (2001) 129–159.
[26] J.A. Tropp, Greed is good: Algorithmic results for sparse approximation, IEEE Trans. Inf. Theory 50 (10) (2004) 2231–2242.
[27] Z.B. Haim, Y.C. Eldar, M. Elad, Coherence-based performance guarantees for estimating a sparse vector under random noise, IEEE Trans. Signal Process. 58 (10) (2010) 5030–5043.
[28] E.J. Candes, The restricted isometry property and its implications for compressed sensing, C. R. Math. 346 (9) (2008) 589–592.
[29] J.A. Tropp, Just relax: Convex programming methods for identifying sparse signals in noise, IEEE Trans. Inf. Theory 52 (3) (2006) 1030–1051.
[30] D.L. Donoho, M. Elad, V.N. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise, IEEE Trans. Inf. Theory 52 (1) (2006) 6–18.
[31] D. Donoho, X. Huo, Uncertainty principles and ideal atomic decomposition, IEEE Trans. Inf. Theory 47 (7) (2001) 2845–2862.
[32] E.J. Candes, T. Tao, The Dantzig selector: Statistical estimation when $p$ is much larger than $n$, Ann. Statist. 35 (6) (2007) 2313–2351.
[33] L.R. Welch, Lower bounds on the maximum cross-correlation of signals, IEEE Trans. Inf. Theory 20 (3) (1974) 397–399.
[34] J.A. Tropp, S.J. Wright, Computational methods for sparse solutions of linear inverse problems, CALTECH ACM Tech. Report 1, 2009, pp. 1–8.
[35] S. Mallat, Z. Zhang, Matching pursuits with time–frequency dictionaries, IEEE Trans. Signal Process. 41 (12) (1993) 3397–3415.
[36] R. Chartrand, Exact reconstruction of sparse signals via nonconvex minimization, IEEE Signal Process. Lett. 14 (10) (2007) 707–710.
[37] A.J. Miller, Subset Selection in Regression, Chapman and Hall, London, 2002.
[38] D. Wipf, B. Rao, Sparse bayesian learning for basis selection, IEEE Trans. Signal Process. 52 (8) (2004) 2153–2164.
[39] T. Cai, L. Wang, Orthogonal matching pursuit for sparse signal recovery with noise, IEEE Trans. Inf. Theory 57 (7) (2011) 4680–4688.
[40] S. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory, Prentice Hall, 1993.
[41] W.O. Alltop, Complex sequences with low periodic correlations, IEEE Trans. Inf. Theory 26 (3) (1980) 350–354.
[42] T. Strohmer, R.H. Heath Jr., Grassmannian frames with applications to coding and communications, Appl. Comput. Harmon. Anal. 14 (3) (2003) 257–275.
[43] H.C.R.M. Planat, S. Perrine, A survey of finite algebraic geometrical structures underlying mutually unbiased quantum measurements, Found. Phys. 36 (11) (2006) 1662–1680.
[44] P.M. Woodward, Probability and Information Theory with Application to Radar, Pergamon, London, 1953.
[45] N. Levanon, E. Mozeson, Radar Signals, Wiley–IEEE Press, 2004.
[46] H.L.V. Trees, Detection, Estimation and Modulation Theory, Part III: Radar–Sonar Signal Processing and Gaussian Signals in Noise, Wiley, 2001.
[47] M.B. Guldogan, O. Arikan, Multipath channel identification by using global optimization in ambiguity function domain, Signal Process. 91 (11) (2011) 2647–2660.
[48] M.B. Guldogan, O. Arikan, Cross-ambiguity function domain multipath channel parameter estimation, Digital Signal Process. 22 (2) (2012) 275–287.
[49] M.I. Skolnik, Introduction to Radar Systems, McGraw–Hill, 2001.
[50] G.T. Zhou, M. Ikram, Unsupervised detection and parameter estimation of multi-component sinusoidal signals in noise, in: Proc. 27th Ann. Asilomar Conf. Signals, Systems and Computers, 2000.
[51] H. Hindberg, A. Hanssen, S. Olhede, Thresholding the ambiguity function, in: IEEE Int. Conf. Acoust. Speech Signal Processing (ICASSP), 2008.
[52] K.V. Price, R. Storn, J.A. Lampinen, Differential Evolution: A Practical Approach to Global Optimization, Springer, 2005.
[53] D.E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning, Addison–Wesley Longman, 1989.
[54] D. Karaboga, B. Basturk, On the performance of artificial bee colony (abc) algorithm, Appl. Soft Comput. 8 (2007) 687–697.
[55] K.L.M.M. Eusuff, Optimization of water distribution network design using the shuffled frog leaping algorithm, J. Water Resour. Plan. Manage. 129 (3) (2003) 210–225.
[56] R. Poli, Analysis of the publications on the applications of particle swarm optimisation, J. Artif. Evol. Appl. 2008 (2008) 685175.
[57] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. R. Stat. Soc. 39 (1) (1977) 1–38.
[58] D. Bratton, J. Kennedy, Defining a standard for particle swarm optimization, in: IEEE Int. Symp. Swarm Intelligence (SIS), 2007.

**Mehmet Burak Guldogan** received the B.S., M.S. and Ph.D. degrees all in Electrical and Electronics Engineering from Bilkent University, Turkey, in 2003, 2006 and 2010, respectively. Between 2010 and 2012, he was a postdoctoral fellow in the Automatic Control Group at Linköping University, Sweden. He joined Turgut Ozal University in 2012, where he is presently an Assistant Professor at the Department of Electrical and Electronics Engineering, Ankara, Turkey. His current research interests are in statistical signal processing, time–frequency analysis, estimation theory and target tracking.

**Orhan Arikan** received the B.Sc. degree in Electrical and Electronics Engineering from the Middle East Technical University in 1986 and the 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 worked for three years as a Research Scientist at the Schlumberger–Doll Research, Ridgefield, CT. He joined Bilkent University in 1993, where since 2006 he is a Professor of Electrical Engineering. His current research interests are in statistical signal processing, time–frequency analysis and array signal processing.