

Majorization-Minimization-Based Sparse Signal Recovery Method Using Prior Support and Amplitude Information for the Estimation of Time-varying Sparse Channels

Chen Wang¹, Yong Fang^{1*}

¹ Key laboratory of Specialty Fiber Optics and Optical Access Networks, Joint International Research Laboratory of Specialty Fiber Optics and Advanced Communication, Shanghai Institute for Advanced Communication and Data Science, Shanghai University, Shanghai, 200444, China
[e-mail: cwangsh@shu.edu.cn; yfang@staff.shu.edu.cn]

*Corresponding author: Yong Fang

*Received February 8, 2018; revised April 24, 2018; accepted May 18, 2018;
published October 31, 2018*

Abstract

In this paper, we study the sparse signal recovery that uses information of both support and amplitude of the sparse signal. A convergent iterative algorithm for sparse signal recovery is developed using Majorization-Minimization-based Non-convex Optimization (MM-NcO). Furthermore, it is shown that, typically, the sparse signals that are recovered using the proposed iterative algorithm are not globally optimal and the performance of the iterative algorithm depends on the initial point. Therefore, a modified MM-NcO-based iterative algorithm is developed that uses prior information of both support and amplitude of the sparse signal to enhance recovery performance. Finally, the modified MM-NcO-based iterative algorithm is used to estimate the time-varying sparse wireless channels with temporal correlation. The numerical results show that the new algorithm performs better than related algorithms.

Keywords: Channel estimation, non-convex optimization, Majorization-Minimization method, prior information, particle filter, sparse signal recovery

This work was supported by the National Science Foundation (NSF) of China (No. 61271213, No. 61673253) and the Ph.D. Programs Foundation of Ministry of Education of China (No. 20133108110014). We express our thanks to reviewers for constructive comments and hard work.

1. Introduction

The purpose of sparse signal recovery is to reliably recover the signals that are sparse (or approximately sparse) from a number of compressed measurements. Because a wide range of natural and man-made signals, e.g., real time video [1], dynamic Magnetic Resonance Imaging (MRI) signals [2] and broadband wireless channel [3] are sparse (or approximately sparse), there are many possible applications.

Many algorithms have been proposed to solve the sparse signal recovery problem. Among these algorithms, the convex optimization-based methods, such as L_1 minimization [4] and least absolute shrinkage and selection operator (LASSO), which use the alternating direction method of multipliers (ADMM) [5], are able to recover sparse signals. However, a non-convex penalty function, which promotes sparsity more strongly than the L_1 norm, yields more accurate results in many sparse signal recovery problems [6-7]. Hence, many algorithms were developed to solve the Non-convex Optimization (NcO) problem associated with sparse signal recovery. Some representative NcO algorithms for sparse signal recovery include reweighted- L_1 minimization [6], sparse Bayesian learning (SBL) [7], focal underdetermined system solver (FOCUSS) [8], and iteratively reweighted least squares (IRLS) [9-10]. These NcO algorithms were found to be effective at solving problems related to direction-of-arrival (DOA) estimation, neuromagnetic imaging, fetal ECG telemonitoring [11], and computer vision [13].

It seems that there is no connection between the aforementioned NcO algorithms. However, they can be considered special instances of the Majorization-Minimization (MM) algorithm. The main difference between these various NcO is that they use different non-convex sparse penalty functions and different majorization functions to formulate upper bounds. The main advantage of the MM algorithm is that it converts the original hard NcO problem into a simpler problem, which solves a sequence of convex optimization problems iteratively [12-13]. Due to the convexity of the sub-problems, for any arbitrary initial point, the iterative algorithm based on NcO using MM algorithm (MM-NcO) can ensure a convergence to the local optimal solution [11-12] but not a convergence to the global optimal solution, which degenerates performance of the MM-NcO-based iterative algorithm.

To overcome this drawback, we first quantify the error bound of the recovered sparse signal. We show that this error bound greatly depends on the initial point of the iterative algorithm. To the best of our knowledge, this work is the first study that considers the error bound. In addition, we propose a modified MM-NcO-based iterative algorithm for sparse signal recovery. The proposed algorithm exploits the prior information of support and amplitude of the sparse signal to formulate an improved initialization to increase the performance of the MM-NcO-based iterative algorithm for sparse signal recovery.

In practice, prior information of support and amplitude of the sparse signal can be obtained using estimation from the previous time-slot. A sequence of sparse signals, in particular, is usually correlated over time. For instance, consecutive real-time video signals [1], dynamic MRI signals [2], and sparse time-varying wireless channels [20] usually have strong dependencies. In this paper, we exploit the dynamic model for time-varying wireless channels to enable improved initialization and enhance the estimation of time-varying sparse wireless channels with temporal correlation.

The contributions of our paper are as follows. First, we quantify the error bound of the recovered sparse signal and propose a modified MM-NcO-based iterative algorithm that uses prior information of support and amplitude of the sparse signal to enhance recovery performance. Second, we propose a Particle Filtering (PF) aided Modified MM-NcO (PF-Modified MM-NcO) algorithm to estimate the time-varying sparse wireless channels with temporal correlation in an online fashion. Finally, we use the simulation results to evaluate the performance of the new algorithm using different system parameters.

The paper is organized as follows. In Section 2, we analyze the recovery performance of the MM-NcO. In Section 3, we propose a modified MM-NcO algorithm with improved initialization. In Section 4, we propose a PF-Modified MM-NcO algorithm to estimate time-varying sparse channels with temporal correlation. The simulation results are presented in Section 5 and show that the proposed algorithm outperforms the related algorithms. The conclusion is given in Section 6.

Important notations used in this paper include, $\|\mathbf{x}\|_0$, $|\mathbf{x}|$ and $\|\mathbf{x}\|$ to denote the L_0 , L_1 and L_2 norms of a vector \mathbf{x} , respectively. Bold symbols are reserved for vectors and matrices; the i^{th} coordinate of a vector \mathbf{x} is denoted as x_i . The support of \mathbf{x} is the set of indices of non-zero elements of \mathbf{x} , denoted by $\text{supp}(\mathbf{x})$, $\text{supp}(\mathbf{x}) := \{i : x_i \neq 0\}$; \mathbf{A}^\dagger denotes the Moore-Penrose pseudo-inverse of a matrix \mathbf{A} ; $\mathcal{N}(\mu, \sigma^2)$ is the Gaussian distribution with the mean μ and variance σ^2 ; Lastly, \mathbf{I} denotes the identity matrix.

2. Analysis of MM-NcO

In this section, we show that the performance of the MM-based iterative algorithm for sparse signal recovery depends critically on the initial point.

2.1 Sparse Signal Recovery with Non-convex Sparse Penalty

Assuming the compressed measurements $\mathbf{y} \in \mathbb{R}^{M \times 1}$ of an unknown sparse signal matrix $\mathbf{x} \in \mathbb{R}^{M \times 1}$ are given by:

$$\mathbf{y} = \Phi \mathbf{x} + \mathbf{N} \quad (1)$$

where $\mathbf{x} \in \mathbb{R}^{M \times 1}$ is the unknown s -sparse signal ($\|\mathbf{x}\|_0 = s$, and $s \ll N$), $\Phi \in \mathbb{R}^{M \times N}$ is the measure matrix with the dimension of $M \ll N$ (M is far less than N), and $\mathbf{N} \in \mathbb{R}^{M \times 1}$ is the measurement noise. The goal of sparse signal recovery is to recover the sparse signal \mathbf{x} based on \mathbf{y} and Φ . Such a problem can be formulated as:

$$\begin{aligned} \min \quad & \|\mathbf{x}\|_0 \leq s \\ \text{s.t.} \quad & \|\Phi \mathbf{x} - \mathbf{y}\|^2 \leq \varepsilon \end{aligned} \quad (2)$$

where $\|\mathbf{x}\|_0$ denotes the number of nonzero components of \mathbf{x} , and ε is the error tolerance parameter related to noise statistics.

The optimization defined by (2) represents an NP-hard problem. The alternative sparsity-promoting functions, such as L_1 norm, can be used to replace L_0 norm to find a sparse solution of \mathbf{x} more efficiently. However, a non-convex penalty function, which promotes sparsity more strongly than L_1 norm, yields more accurate results in many sparse signal recovery problems. Replacing the L_0 norm in (2) with the non-convex penalty sparse function we get:

$$\begin{aligned} \min_{\mathbf{x}} P(\mathbf{x}) &= \sum_{i=1}^N p(x_i) \\ \text{s.t.} \quad &\|\Phi\mathbf{x} - \mathbf{y}\|^2 \leq \varepsilon \end{aligned} \quad (3)$$

where $p(x)$ is a non-convex penalty-function to promote sparsity, and x_i denotes the i^{th} coordinate of the vector \mathbf{x} . In this paper, we consider the use of Type II sparse penalty function, which is originally proposed in SBL [31]. The Type II sparse penalty function, which is defined by (4), is a non-decreasing and strictly concave function of $|x|$. It was shown to be a better sparsity-promoting function than the L_1 norm [31]:

$$p(x) = \lambda^2 (\log(|x| + \sqrt{x^2 + 4\lambda^2}) + \frac{|x|}{\sqrt{x^2 + 4\lambda^2} + |x|}) \quad (4)$$

where $\lambda > 0$ is a positive parameter related to the noise level [34].

Then, the optimization (3) can be further formulated as an unconstrained non-convex optimization problem defined by:

$$\min_{\mathbf{x} \in \mathbf{X}} F(\mathbf{x}) = D(\mathbf{x}) + P(\mathbf{x}) \quad (5)$$

where $D(\mathbf{x}) = \|\Phi\mathbf{x} - \mathbf{y}\|^2$ is a data-fitting term.

2.2 NcO-MM Algorithm

To solve the non-convex problem defined by (5), we use the bounded optimization approach, also known as the majorization-minimization (MM) algorithm. The principle of MM is that it iteratively generates a sequence of a simpler surrogate function, which majorizes the given objective function and computes the next iteration as a minimum of that majorization function. More specifically, starting from an arbitrary initial point \mathbf{x}^0 , the algorithm produces a sequence $\{\mathbf{x}^k\}$ according to the following update rule:

$$\mathbf{x}^{k+1} = \arg \min_{\mathbf{x} \in \mathbf{X}} D(\mathbf{x}) + P^{\mathbf{x}^k}(\mathbf{x}) \quad (6)$$

where \mathbf{x}^k is the point generated by the algorithm at iteration k , and $P^{\mathbf{x}^k}(\mathbf{x})$ is the majorization function of $P(\mathbf{x})$ at \mathbf{x}^k . In the context of MM, the function $P^{\mathbf{x}^k}(\mathbf{x})$ is a convex upper-bound approximation of $P(\mathbf{x})$ that coincides with $P(\mathbf{x})$ at $\mathbf{x} = \mathbf{x}^k$, i.e.:

$$P^{\mathbf{x}^k}(\mathbf{x}) \geq P(\mathbf{x}) \quad (7)$$

$$P^{\mathbf{x}^k}(\mathbf{x}^k) = P(\mathbf{x}^k) \quad (8)$$

Since $D(\mathbf{x})$ and $P^{\mathbf{x}^k}(\mathbf{x})$ are both convex, (6) can be solved using the Fermat rule (Thm. 10.1 in [15]). Using the general MM framework, we can formulate:

$$\underline{F} \leq F(\mathbf{x}^{k+1}) \leq D(\mathbf{x}^{k+1}) + P^{\mathbf{x}^k}(\mathbf{x}^{k+1}) \leq D(\mathbf{x}^k) + P^{\mathbf{x}^k}(\mathbf{x}^k) = F(\mathbf{x}^k) \quad (9)$$

Here we assume that F is bounded from below, i.e., $\inf_{\mathbf{x} \in \mathbf{X}} F(\mathbf{x}) =: \underline{F} > -\infty$.

The MM-based Non-convex Optimization algorithm (MM-NcO) for recovering of sparse signal is given below.

Algorithm 1 MM-NcO

Input: Arbitrary initial point \mathbf{x}^0 , and measure matrix $\Phi \in \mathbb{R}^{M \times N}$

Output: Estimated $\hat{\mathbf{x}}$

Repeat: 1. $\mathbf{x}^{k+1} = \arg \min_{\mathbf{x} \in \mathbf{X}} D(\mathbf{x}) + P^{\mathbf{x}^k}(\mathbf{x})$

2. $k = k + 1$

Until Convergence

In the next subsection, we will present the error bound of the recovered sparse signal

2.3 Error bound of the Recovered Sparse Signal

Before presenting the error bound of the recovered sparse signal, we need to prove that $(\mathbf{x}^k)_{k=0,1,2,\dots}$, which is generated by the MM-NcO algorithm, globally converges to the critical point of F .

Theorem 1 (Global convergence): We assume the sequence $(\mathbf{x}^k)_{k=0,1,2,\dots}$ is generated by the Algorithm 1. If $F(\mathbf{x})$ has a KL property at the accumulation point $\bar{\mathbf{x}} = \lim_{j \rightarrow \infty} \mathbf{x}^{k_j}$, then the sequence $(\mathbf{x}^k)_{k=0,1,2,\dots}$ converges to $\bar{\mathbf{x}} \in \mathbf{X}$ as $k \rightarrow \infty$ and $\bar{\mathbf{x}}$ is a critical point of F . Furthermore, the sequence $(\mathbf{x}^k)_{k=0,1,2,\dots}$ has a finite length:

$$\sum_{k=0}^{\infty} \|\mathbf{x}^k - \mathbf{x}^{k+1}\| < \infty. \quad (10)$$

Proof: See Appendix A.

Theorem 1 indicates that the MM-NcO algorithm converges to the critical point $\bar{\mathbf{x}}$. If Algorithm 1 runs until it converges, the critical point $\bar{\mathbf{x}}$ represents the final recovered sparse signal.

We then use the following definition of a Restricted Isometry Property (RIP).

Definition 1 (RIP): For an integer $t = 1, 2, \dots$, the restricted isometry constant δ_t of the matrix Φ is the smallest number that satisfies (11) for any t -sparse vector \mathbf{x} .

$$(1 - \delta_t) \|\mathbf{x}\|^2 \leq \|\Phi \mathbf{x}\|^2 \leq (1 + \delta_t) \|\mathbf{x}\|^2 \quad (11)$$

Here, we consider that matrix Φ satisfies the RIP of order t with the constant δ_t .

Under the RIP assumption, we can ensure that the critical point $\bar{\mathbf{x}}$ is a reasonable approximation of the sparse solution if $\bar{\mathbf{x}}$ has a very small $\sigma_s(\bar{\mathbf{x}})_2$:

$$\sigma_s(\bar{\mathbf{x}})_2 = \inf_{\|\mathbf{x}^*\|_0 \leq s} \|\bar{\mathbf{x}} - \mathbf{x}^*\| \quad (12)$$

In the previous equation, $\sigma_s(\bar{\mathbf{x}})_2$ represents the error of the best s -term approximation of $\bar{\mathbf{x}}$ in the L_2 -norm [16], while \mathbf{x}^* is the true sparse signal.

Theorem 2 (Recovery Error bound): Assuming that \mathbf{x}^* is the s -sparse vector satisfying $\mathbf{y} = \Phi \mathbf{x}^*$ and that Φ satisfies the RIP of order $2s$ with $\delta_{2s} < 1$, then, the critical point $\bar{\mathbf{x}}$ generated using Algorithm 1 satisfies the following relation:

$$\|\bar{\mathbf{x}} - \mathbf{x}^*\| \leq \sqrt{\frac{F(\mathbf{x}^0)}{1 - \delta_{2s}}} + \left(\frac{1}{\sqrt{1 - \delta_{2s}}} \|\Phi\| + 1 \right) \sigma_s(\bar{\mathbf{x}})_2 \quad (13)$$

Proof: See Appendix B.

Note 1 (how initialization affects performance): From Theorem 2, a \mathbf{x}^0 leading to a smaller $F(\mathbf{x}^0)$ would achieve a better sparse signal recovery performance. This is because δ_{2s} is a constant that depends on Φ and the upper bound of recovery error $\|\bar{\mathbf{x}} - \mathbf{x}^*\|$ is monotonically decreasing as $F(\mathbf{x}^0)$ decreases. Therefore, \mathbf{x}^0 leading to a smaller $F(\mathbf{x}^0)$ produces a smaller upper bound of a recovery error.

3. The Modified MM-NcO Algorithm with Improved Initialization

In this section, we introduce a modified MM-NcO with improved initialization by exploiting prior information of both support and amplitude of the sparse signal.

3.1 Improved Initialization using Prior Support and Amplitude Information

Often, sparse patterns of the signals are correlated across time. This property implies that the sparse signal $\mathbf{x}(t)$ at the time slot t is correlated to all previous sparse signals, namely $\mathbf{x}(t-1), \mathbf{x}(t-2), \dots, \mathbf{x}(0)$. Here, we consider the case where $\mathbf{x}(t)$ only depends on $\mathbf{x}(t-1)$.

Definition 2 (Prior Information of sparse signals with temporal correlation): Let $\mathbf{x}_p(t)$ be the signal obtained by exploiting the prior information on $\mathbf{x}(t)$, for a time-varying scenario, $\mathbf{x}_p(t)$ can be characterized by the function of $\mathbf{x}(t-1)$, which implies that $\mathbf{x}_p(t) = G\{\mathbf{x}(t-1)\}$, where G is the function that transfers $\mathbf{x}(t-1)$ to $\mathbf{x}_p(t)$.

An example for the above definition is that $\mathbf{x}(t)$ is time invariant, hence $\mathbf{x}_p(t) = \mathbf{x}(t-1)$. Moreover, as discussed later in the context of time-varying wireless channels, G is a first-order Gauss-Markov random process, which leads to the following relationship between $\mathbf{x}_p(t)$ and $\mathbf{x}(t-1)$:

$$\mathbf{x}_p(t) = \beta \mathbf{x}(t-1) + \mathbf{u}(t) \quad (14)$$

Details of (14) are presented in Section 4. However, in practice only an estimate of $\mathbf{x}(t-1)$, i.e., $\hat{\mathbf{x}}(t-1)$ is available. Thus, in practice we modify Definition 2 to yield:

$$\mathbf{x}_p(t) = G\{\hat{\mathbf{x}}(t-1)\} \quad (15)$$

Proposition 1: Because $\hat{\mathbf{x}}(t-1)$ is estimated using a sparse recovery algorithm, such as ADMM or conventional MM-based NcO, $\hat{\mathbf{x}}(t-1)$ would have a smaller distance to the true sparse signal $\mathbf{x}^*(t-1)$ at time slot $t-1$ than an arbitrary sparse signal $\mathbf{x}(t-1)$.

Proof: See Appendix C.

Proposition 2: A sparse \mathbf{x}^0 with a smaller $\|\mathbf{x}^0 - \mathbf{x}^*\|$ would lead to a smaller $F(\mathbf{x}^0)$

Proof: See Appendix C.

Proposition 3: A method to determine a improved initialization point: $\mathbf{x}_p(t)$ generated by (15) represents a improved initialization point than the arbitrary initialization point \mathbf{x}^0 .

Proof: See Appendix C.

3.2 Modified MM-NcO

Proposition 3 implies that we can find an improved initialization \mathbf{x}_p by exploiting the prior information of temporal correlated sparse signals. Using the improved initialization \mathbf{x}_p , the recovery performance of a conventional MM-NcO can be improved. Hence, we introduce the modified MM-NcO algorithm. In contrast to the conventional MM-NcO algorithm, which uses the arbitrary initial point \mathbf{x}^0 , the proposed modified MM-NcO algorithm uses an improved initialization \mathbf{x}_p by exploiting the prior information of support and amplitude of temporal correlated sparse signals to improve recovery.

Moreover, the prior information of support and amplitude of the sparse signal incurs inevitable errors. To eliminate performance degradation caused by these errors, we propose

adding a decreasing smoothening parameter ε^k , where $\varepsilon^0 > \varepsilon^1 > \dots > \varepsilon^k \geq \varepsilon_{\min}$. Inspired by the selection method of ε^0 in [6], we let $\varepsilon^0 \in (x_{p_{\max}}/10, 10x_{p_{\min}})$, where $x_{p_{\max}}$ is the maximal element of $|\mathbf{x}_p|$, and $x_{p_{\min}}$ is the minimal element of $|\mathbf{x}_p|$. This choice of ε^k can preserve the dominant prior information of the sparse signal and smooth out the undesirable points. As a result, the initialization of the modified MM-NcO is $(\mathbf{x}_p + \varepsilon^0)$.

Here, we seek the convex upper bound approximation $P^{\mathbf{x}^k}(\mathbf{x})$ in (6) using the Local Quadratic Approximation (LQA) [17]. In (6), $p(x)$ is concave and therefore below its tangent [6]. Thus, linearizing $p(x)$ at $x = x^k$ yields the following inequality: $p(x) \leq p(x^k) + p'(x^k)(x - x^k)$, where $p'(x^k)$ is the gradient of $p(x)$ at $x = x^k$. Using the inequality $x \leq \frac{x^2}{2x^k} + \frac{x^k}{2}$. We obtain the convex upper bound approximation as $p^{x^k}(x, x^k) = \frac{p'(x^k)}{2x^k}x^2 + p(x^k) - \frac{x^k}{2}p'(x^k)$. Then, $P^{\mathbf{x}^k}(\mathbf{x})$ can be obtained by:

$$P^{\mathbf{x}^k}(\mathbf{x}) = \sum_{i=1}^N p(x_i, x^k) = \frac{1}{2} \mathbf{x}^T \mathbf{W}^k \mathbf{x} + c \quad (16)$$

$$\text{where } \mathbf{W}^k = \begin{bmatrix} p'(x_1^k)/x_1^k & & \\ & \ddots & \\ & & p'(x_N^k)/x_N^k \end{bmatrix} = \text{diag}(p'(\mathbf{x}^k)/\mathbf{x}^k), \text{ and } c \text{ is independent of } \mathbf{x}.$$

Using the initialization $(\mathbf{x}_p + \varepsilon^0)$ and the convex upper bound approximation $P^{\mathbf{x}^k}(\mathbf{x})$, the pseudo code for a modified MM-NcO algorithm (Modified MM-NcO) can be presented as **Algorithm 2**:

Algorithm 2 Modified MM-NcO

Input: 1) Improved initialization point \mathbf{x}_p obtained by exploiting the prior support and amplitude information of temporal correlated sparse signals

2) Measure matrix $\Phi \in \mathbb{R}^{M \times N}$

Output: Estimated $\hat{\mathbf{x}}$

1. Initialization: Let $\varepsilon^0 \in (x_{p_{\max}}/10, 10x_{p_{\min}})$, $\mathbf{x}_{p\varepsilon}^0 = \mathbf{x}_p + \varepsilon^0$, $\varepsilon_{\min} = 10^{-8}$, and $k = 0$

Repeat:

2. $\mathbf{x}_{p\varepsilon}^k = \mathbf{x}_p + \varepsilon^k$

3. $\mathbf{x}_p^{k+1} = (\Phi^T \Phi + \mathbf{W}_{p\varepsilon}^k)^{-1} \Phi^T \mathbf{y}$, where $\mathbf{W}_{p\varepsilon}^k = \text{diag}(p'(\mathbf{x}_{p\varepsilon}^k)/\mathbf{x}_{p\varepsilon}^k)$

4. $\varepsilon^{k+1} = \varepsilon^k / 10$, If $\varepsilon^{k+1} \leq \varepsilon_{\min}$, then $\varepsilon^{k+1} = \varepsilon_{\min}$

5. $k = k + 1$

Until convergence

6. Let $\hat{\mathbf{x}} \leftarrow \mathbf{x}_p^{k+1}$, and **return** $\hat{\mathbf{x}}$

- (1) *Improved initialization point formulation (Step 1)*: we let $\varepsilon^0 \in (x_{p\max}^0 / 10, 10x_{p\min}^0)$ and $\mathbf{x}_{p\varepsilon}^k = \mathbf{x}_p^k + \varepsilon^k$ to formulate an improved initialization $\mathbf{x}_{p\varepsilon}^0$.
- (2) *Sparse signal estimation (Step 2- 6)*: In *Step 2*, the estimation from the last iteration \mathbf{x}_p^k is used to formulate the upper bound for the current iteration. Minimizing $\|\Phi\mathbf{x} - \mathbf{y}\|^2 + \frac{1}{2}\mathbf{x}^T \mathbf{W}_{p\varepsilon}^k \mathbf{x}$ with respect to \mathbf{x} and setting it to 0, we obtain $\mathbf{x}_p^{k+1} = (\Phi^T \Phi + \mathbf{W}_{p\varepsilon}^k)^{-1} \Phi^T \mathbf{y}$, which represents the estimation of current iteration – see *Step 3*. *Step 4* decreases ε^k , and checks if ε^k reaches its minimal value. The main difference between our modified MM-NcO algorithm and the conventional MM-NcO algorithm is that the new algorithm uses prior information of both support and amplitude of temporal correlated sparse signals. In addition, it uses ε^k to preserve the dominant prior information of the sparse signal and smooths out irrelevant data points.

4. Application in time-varying sparse channels

In this section, we use the proposed sparse signal recovery algorithm (modified MM-NcO) to solve a sparse time-varying channels estimation problem. In many wireless communication systems, channel estimation is crucial, especially when wireless channels change over time [21]. Recently, it was found that, in several scenarios, wireless channels show intrinsic sparsity, i.e., only a few channel-gains are dominant [3, 22]. Further studies revealed that, besides sparsity, practical wireless channels show temporal correlation [24, 28, 29]. The common method to leverage both temporal correlation and sparsity is to use batch processing methods [23-26, 33].

These batch processing approaches [23-26, 33], however, have two drawbacks. For example, they require several measurements over a period of time and they have to process these measurements simultaneously to obtain channel estimation results. This procedure excludes the batch processing approaches from an online estimation of time-varying sparse wireless channels. Another drawback of Structured Compressive Sensing (SCS) based batch processing approaches (e.g. [23], [24], [26], [33]) is that they use the L_2 norm to select the supports of sparse signals. This implies that they use only the prior information of support of the sparse signal, which decreases SCS robustness with regard to noise.

To solve the disadvantages of batch processing methods, we use the proposed Modified MM-NcO algorithm to estimate the sparse time-varying channels with temporal correlation, as a result, we introduce a Particle Filtering (PF) aided Modified MM-NcO algorithm. Unlike previous studies that aimed to eliminate inter-relay interference (IRI) in cooperative OFDM channels [35], our work here aims to improve estimation performance for time-varying sparse wireless channels.

4.1 Problem Formulation

We now consider the estimation of a time-varying sparse channel $\mathbf{h}(t) \in \mathbb{R}^{N \times 1}$ from measurements $\mathbf{y}(t) \in \mathbb{R}^{N \times 1}$, $t = 0, 1, 2, \dots$. Here, t depends on a sample period of the wireless communication system. For instance, if we consider the LTE-Advanced system working at the carrier frequency of 2 GHz with a signal bandwidth f_s of 10 MHz [24], then, time slot is equal

to the system sample period $T_s = 1/f_s = 0.1\mu s$. The measurement vector $\mathbf{y}(t)$ at time slot t can be obtained using a linear measurement:

$$\mathbf{y}(t) = \Phi \mathbf{h}(t) + \mathbf{N} \quad (17)$$

where $\Phi \in \mathbb{R}^{M \times N}$ is the matrix of training pilots (M is far less than N , $M \ll N$) [27], Φ is time-invariant and drawn from a zero-mean Gaussian distribution, $\mathbf{N} \in \mathbb{R}^{M \times 1}$ is the measurement noise, and $\mathbf{N} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$.

Recent studies [28-29] show that a time-varying sparse channel is temporarily correlated with the channel state from previous time slot. This correlation of time-varying sparse channels can be modeled using the first-order Gauss-Markov random process:

$$\mathbf{h}(t) = \beta \mathbf{h}(t-1) + \mathbf{u}(t) \quad (18)$$

where $\beta \in (0,1)$ is based on the priori knowledge of the propagation environment, and $\mathbf{u}(t) \sim \mathcal{N}(0, \sigma_d^2 \mathbf{I})$ is the driving noise [25, 28-29].

Equation (18) indicates that, although the amplitude of $\mathbf{h}(t)$ continuously changes, $\text{supp}(\mathbf{h}(t+1)), \dots, \text{supp}(\mathbf{h}(t+I))$ are invariant over I time slots. Based on this common sparsity, the SCS algorithm waits for I measurements and estimates $\mathbf{h}(t+1), \dots, \mathbf{h}(t+I)$ simultaneously to enhance channel estimation performance. However, the SCS algorithm still has the two main drawbacks as previously mentioned.

4.2 Proposed Algorithm

To solve previously mentioned drawbacks of SCS, we propose a Particle Filtering (PF) aided modified MM-NcO algorithm (PF-Modified MM-NcO). We first use a conventional sparse signal recovery method, such as YALL1 [18], to obtain the estimated $\hat{\mathbf{h}}(0)$. Then, by utilizing the temporal correlation model defined by (18), we use the Particle Filter [30, 32] to obtain $\mathbf{h}_p(t)$. Finally, we let $\mathbf{x}_p \leftarrow \mathbf{h}_p(t)$, and run Algorithm 2 (Modified MM-NcO) to obtain an accurate estimate of $\hat{\mathbf{h}}(t)$. The proposed algorithm is depicted in Fig. 1, and its corresponding pseudo code is given in Algorithm 3.

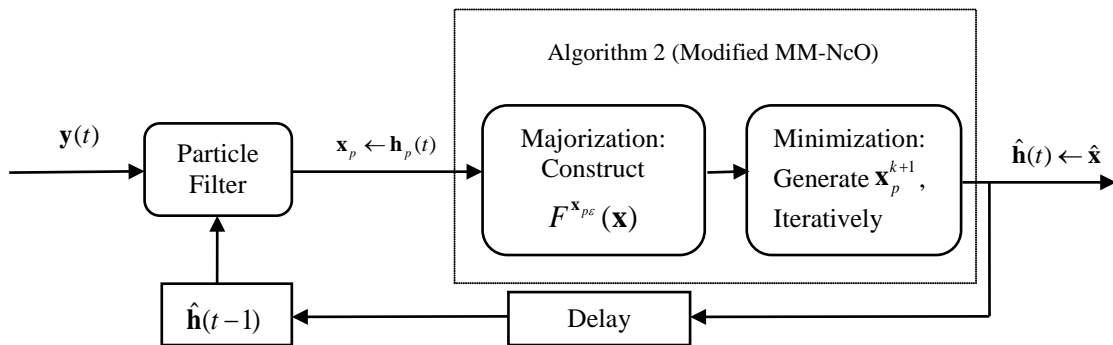


Fig. 1. Graphical model of the proposed Algorithm 3

Algorithm 3 PF-Modified MM-NcO

Input: Received signals: $\mathbf{y}(t)$, matrix of training pilots: Φ

Output: the estimated channels $\hat{\mathbf{h}}(t)$

1. Initialization: $t \leftarrow 0$, $\hat{\mathbf{h}}(0) = \arg \min_{\mathbf{h}} \|\Phi \mathbf{h} - \mathbf{y}(0)\|^2 + \|\mathbf{h}\|$

Repeat:

2. **For** $t = 1, 2, \dots$

3. **For** $j = 1 : N_p$

4. Prediction: $\mathbf{h}^j(t) = \mathcal{N}(\beta \hat{\mathbf{h}}(t-1), \sigma_d^2 \mathbf{I})$

5. Particle weight update: $\mathbf{w}^j(t) = p(\mathbf{y}(t) | \mathbf{h}^j(t))$, where $p(\mathbf{y}(t) | \mathbf{h}^j(t))$ is the likelihood function of $\mathbf{y}(t)$

End For

6. Particle weight normalization: $\mathbf{w}^j(t) = \mathbf{w}^j(t) / \sum_{j=1}^{N_p} \mathbf{w}^j(t)$

7. Estimation: $\mathbf{h}_p(t) = \sum_j^{N_p} \mathbf{w}^j(t) \mathbf{h}^j(t)$

8. $\mathbf{x}_p \leftarrow \mathbf{h}_p(t)$

9. **Initiate** Algorithm 2 using \mathbf{x}_p , and **run** Algorithm 2 to obtain $\hat{\mathbf{x}}$.

10. Let $\hat{\mathbf{h}}(t) \leftarrow \hat{\mathbf{x}}$, and **return** $\hat{\mathbf{h}}(t)$

(1) Particle Filtering (*Step 3 to Step 7*): Given the number of particles N_p , *Step 4* predicts the j^{th} particle $\mathbf{h}^j(t)$ using the temporal correlation model (18). *Step 5* updates the associated weight $\mathbf{w}^j(t)$ of the j^{th} particle using $\mathbf{w}^j(t) = p(\mathbf{y}(t) | \mathbf{h}^j(t))$, where $p(\mathbf{y}(t) | \mathbf{h}^j(t))$ is the likelihood function of $\mathbf{y}(t)$. The weights are normalized in *Step 6*. The estimation $\mathbf{h}_p(t)$ by Particle Filtering is finally obtained in *Step 7*.

(2) Modified MM-NcO (*Step 8 to Step 9*): In *step 8* and *step 9*, we initiate Algorithm 2 by using $\mathbf{x}_p = \mathbf{h}_p(t)$ and run Algorithm 2 to obtain $\hat{\mathbf{x}}$, respectively. In *Step 10*, we use $\hat{\mathbf{x}}$ as a final estimation of $\hat{\mathbf{h}}(t)$ at time slot t .

In contrast to the SCS, the proposed PF-Modified MM-NcO has several desirable features. First, the PF-Modified MM-NcO algorithm estimates time-varying sparse channels in an online fashion (the proposed algorithm has access only to the present and previous time slots when estimating the channel of the present time slot). Second, unlike the SCS that uses L_2 norm to select supports of the sparse signals, the PF-Modified MM-NcO is based on a non-convex sparse promoting function, which is more robust with respect to noise. Third, the PF-Modified MM-NcO algorithm simultaneously exploits the prior information of support and amplitude of the sparse signal.

Note 2 (the relationship between the proposed Algorithm 2 and 3): Algorithm 2 (Modified MM-NcO) represents a general method to improve the recovery performance of the conventional MM-NcO by exploiting the prior information of both support and amplitude of the sparse signal with temporal correlation. However, the method to obtain the prior information depends on the specific application. Algorithm 3 uses PF to obtain the prior information by utilizing the temporal correlation model of time-varying sparse wireless channels. In other words, Algorithm 3 is a specific application of Algorithm 2 to estimate the time-varying sparse wireless channels with temporal correlation.

4.3 Computational Complexity

The computational complexity of the proposed PF-Modified MM-NcO relates to two parts: (1) Particle Filtering, and (2) modified MM-NcO. The computational complexity of the Particle Filtering part is mainly determined by the complexity of *Step 5* in Algorithm 3. For each time slot of Algorithm 3, the computational complexity of *Step 5* is $\mathcal{O}(MNN_p)$. The computational complexity of the modified MM-NcO part is determined by the complexity of *Step 3* in Algorithm 2. For each iteration of Algorithm 2, the computational complexity of *Step 3* is $\mathcal{O}(N^3 + MN)$. Then, the total complexity of the proposed algorithm at each time slot is: $\mathcal{O}(MNN_p + (N^3 + MN)k)$. Due to the property of compressed measurements, the value for M is usually small. Moreover, we found that the number of iterations needed for Algorithm 2 to converge is small ($k < 10$ at some time slots). This means that the value for k is also small because Algorithm 2 exploits the prior information of support and amplitude of sparse signal to formulate an improved initialization. These two properties are responsible for the acceptable complexity of the proposed Algorithm 3.

5. Simulation Results and Analysis

In this section, we consider the estimation of time-varying sparse channels to determine the effectiveness of the proposed algorithm. Specifically, we compare the performance of the proposed PF-Modified MM-NcO with the performance of the following algorithms:

Mixed-L1L2: Deploys the ADMM-based mixed L_1 / L_2 minimization algorithm [18] to estimate the time-varying sparse channels. The μ in YALL1 is set to 0.01.

SCS-6: Structured Compressed Sensing [23] with 6 received measurements, i.e., $I=6$.

MM-NcO-ran: Deploys the MM-NcO with random initialization, i.e., $\mathbf{x}^0 = \mathcal{N}(0, \mathbf{I})$.

Genie-aided LS: Serves as a performance upper-bound scenario, in which the channel support $\text{supp}(\mathbf{h}(t))$ is assumed to be known, and we directly use the least-square method to recover the channel coefficients on $\text{supp}(\mathbf{h}(t))$.

We consider the time-varying sparse signal $\mathbf{h}(t) \in \mathbb{R}^{N \times 1}$ with a channel length of $N = 180$. The number of non-zero taps of $\mathbf{h}(t)$ is K , and is set to $K = 12$. The unit of t is explained in subsection 4.1. The matrix of the training pilots $\Phi \in \mathbb{R}^{M \times N}$ was generated as a zero mean random Gaussian matrix with columns normalized to the unit L_2 norm, and the length of training pilots is M . Furthermore, the received measurements were contaminated by an additive Gaussian noise \mathbf{N} , and $\mathbf{N} = \mathcal{N}(0, \sigma^2 \mathbf{I})$. For the channel correlation model (25), the temporal correlation parameter β is chosen as $0 < \beta < 1$, and the driving noise parameter is

$\sigma_d = \sqrt{1 - \beta^2}$. The number of particles used in the proposed algorithm is N_p , λ is chosen to be $\lambda = 3\sigma$. The estimation performance of the algorithms was measured based on the normalized mean-square error (NMSE), i.e., $\text{NMSE} = \frac{\|\mathbf{h}(t) - \hat{\mathbf{h}}(t)\|}{\|\mathbf{h}(t)\|}$, and the time-averaged normalized mean-squared error (A-NMSE), i.e., $\text{A-NMSE} = \frac{1}{T} \sum_{t=1}^T \frac{\|\mathbf{h}(t) - \hat{\mathbf{h}}(t)\|}{\|\mathbf{h}(t)\|}$. In the simulations, the time length is set to $T = 60$.

Fig. 2 shows the NMSE performance for the estimated time-varying sparse channels at each time slot. The parameters are set to $M = 45$, $\sigma = 0.01$, $\beta = 0.7$, and $N_p = 5000$. Using this figure, we find that the MM-NcO-ran fluctuates greatly and performs poorest due to random initialization. The SCS outperforms the traditional Mixed-L1L2 by exploiting the common sparsity of the time-varying channels. As expected, the proposed PF-Modified MM-NcO performs better than other competing algorithms, and it is stable. This is because the proposed PF-Modified MM-NcO uses the prior support and amplitude information generated by the PF technique to initialize the MM-NcO. Specifically, the performance gain of the PF-Modified MM-NcO compared with Mixed L1L2 and MM-NcO-ran demonstrates the advantage of using prior information of support and amplitude of the sparse signal. The performance gain of the PF-Modified MM-NcO compared with SCS-6 suggests the following benefits: (1) Use of a non-convex sparse promoting function, (2) exploitation of prior information of both amplitude and support of the sparse signal.

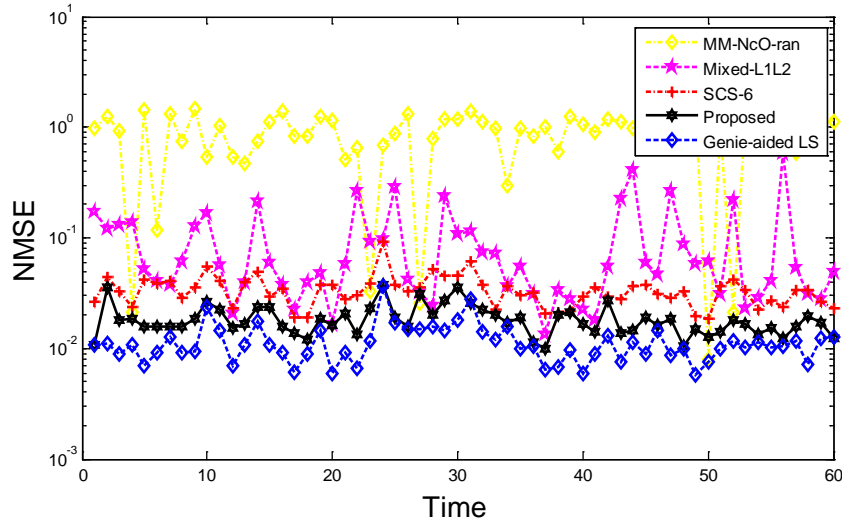


Fig. 2. The NMSE of the estimated sparse channel at each time slot

In **Fig. 3**, we compare the A-NMSE for the estimated time-varying sparse channels as a function of the noise level σ . The parameters are $M = 45$, $\beta = 0.7$, and $N_p = 5000$. Using this figure, we can see that the A-NMSE performance of the Mixed-L1L2, SCS-6, the proposed PF-Modified MM-NcO, and the genie-aided LS improves monotonically as the noise level σ decreases. On the other hand, the A-NMSE performance of the MM-NcO-ran fluctuates and performs worst. Moreover, the performance loss between the proposed PF-Modified MM-NcO and the genie-aided LS decreases as the noise level σ decreases.

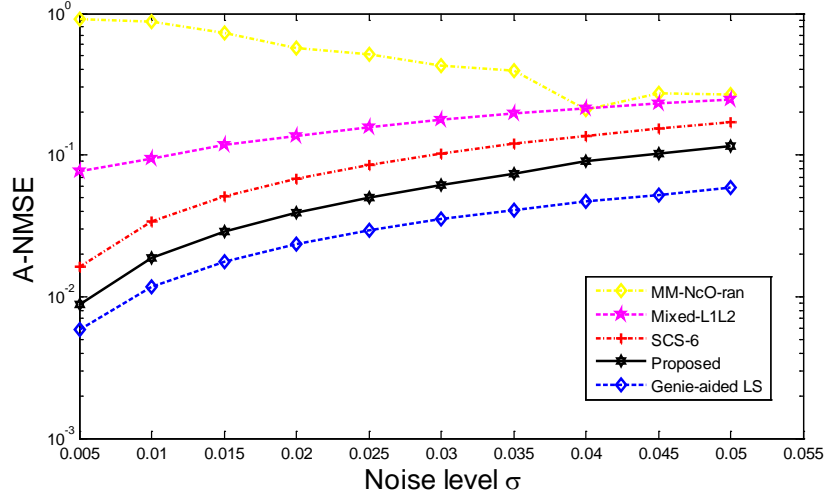


Fig. 3. The A-NMSE of estimated sparse channels versus noise level σ

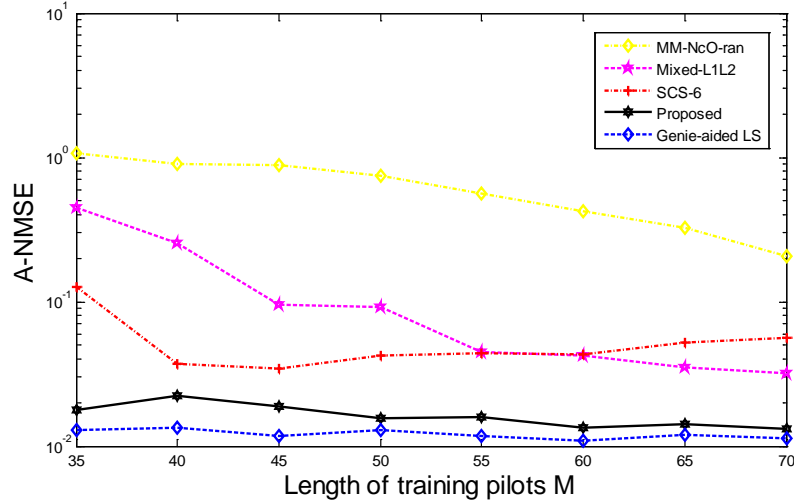


Fig. 4. The A-NMSE of estimated sparse channels versus length of training pilots M

The impact of the length of the training pilots M is shown in **Fig. 4**, the parameters are set to $\sigma = 0.01$, $\beta = 0.7$, and $N_p = 5000$. We can see that the A-NMSE performance of the Mixed-L1L2 and MM-NcO-ran improves monotonically as M increases. This is because more training pilots can bring more channel-state information and enable a more accurate channel estimation. We also observe that the performance of SCS-6, the proposed PF-Modified MM-NcO, and the genie-aided LS are not sensitive to the length of the training pilots. This is because these three algorithms exploit priori information about the sparse signal. Note that the performance of the SCS-6 is worse than the Mixed-L1L2 when M becomes large. This is because, no matter how many pilots are used by SCS-6, SCS-6 would always select similar supports of the sparse channels. By contrast, the proposed algorithm approaches the genie-aided LS as M increases. This shows that the proposed algorithm is robust with respect to different lengths of the training pilots and can achieve better channel estimation performance with substantially fewer training pilots.

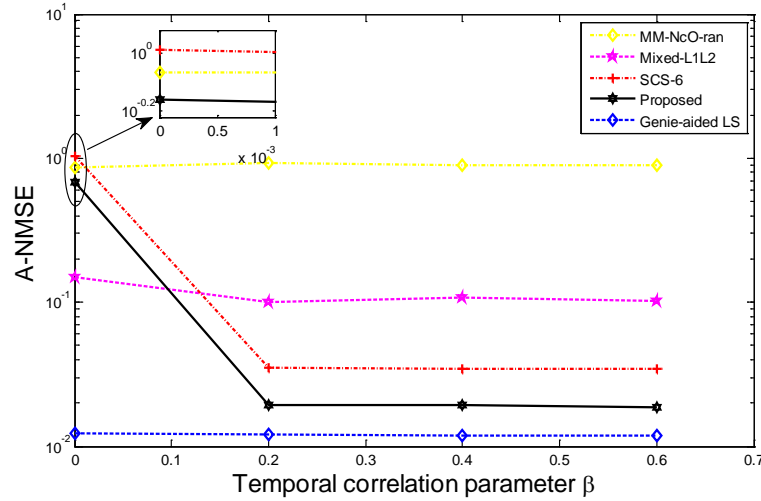


Fig. 5. The A-NMSE of estimated sparse channels versus the temporal correlation parameter β

In **Fig. 5**, we compare the A-NMSE of estimated time-varying sparse channels as a function of the temporal correlation parameter β . The parameters are set to $\sigma = 0.01$, $M = 45$, and $N_p = 5000$. The time-varying sparse channels with $\beta = 0$ mean that $\text{supp}(\mathbf{h}(t))$ is different between consecutive time slots. Using this figure, we find that the performance of SCS-6 and the proposed PF-Modified MM-NcO is worse than the Mixed-L1L2 when $\beta = 0$. This is because SCS-6 still tries to estimate the sparse channels with common sparsity, and the proposed algorithm cannot generate effective particles to initiate MM-NcO. However, the SCS-6 and the proposed PF-Modified MM-NcO show a substantially better performance than the Mixed-L1L2 when $\beta > 0$. The performance of the investigated algorithms shows no obvious relation with the temporal correlation when $\beta > 0$. Particularly, when $\beta = 0$, the proposed algorithm performs better than both the MM-NcO-ran and SCS-6 because the proposed algorithm still tries to obtain improved initialization by running the PF.

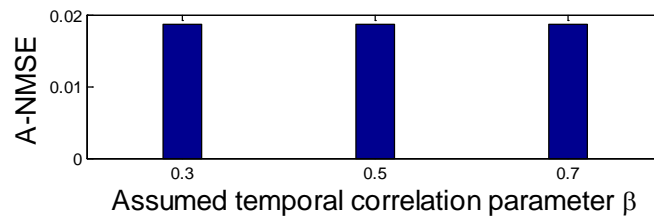


Fig. 6. The A-NMSE of estimated sparse channels versus assumed β

In **Fig. 6**, we investigate the impact of a mismatch between the assumed β and the true β on the estimation performances. The true β is $\beta = 0.5$, while the assumed β is 0.3 and 0.7, respectively. The parameters are $\sigma = 0.01$, $M = 45$, and $N_p = 5000$. Clearly, the mismatch of β causes almost no performance degradation. Therefore, it can be stated that our proposed algorithm is very robust with respect to mismatching β .

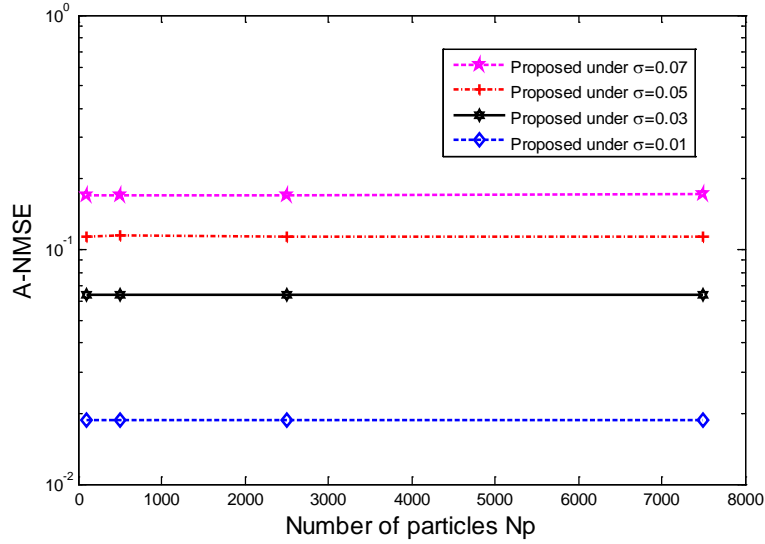


Fig. 7. The A-NMSE of estimated sparse channels versus the number of particles N_p

In **Fig. 7**, we can see the impact of the number of particles N_p for different noise levels σ . The parameters are $M = 45$, and $\beta = 0.7$. **Fig. 7** shows that the proposed algorithm is very robust with respect to different numbers of particles.

6. Conclusion

In this paper, we show that the performance of the convergent MM-NcO-based iterative algorithm for sparse signal recovery depends critically on the initial point. Therefore, we developed a modified MM-NcO algorithm that uses prior information of both support and amplitude of the sparse signal. We used the modified MM-NcO algorithm to estimate the time-varying sparse wireless channels with temporal correlation. We then introduced a PF-Modified MM-NcO algorithm. The obtained results show that the PF-Modified MM-NcO algorithm is effective to estimate time-varying sparse channels with temporal correlation. The PF-Modified MM-NcO algorithm proposed in this study estimates time-varying sparse channels in an online fashion, which makes it possible to bypass a common limitation of batch processing algorithms, such as SCS. Moreover, the proposed algorithm performs better than the SCS. The simulation results indicate that our proposed algorithm can effectively estimate time-varying sparse wireless channels with temporal correlation using only a small number of training pilots.

Appendix A

Proof of Theorem 1

Proof: Based on Theorem 2 in [13], if we can show that: (1) the proposed function $F(\mathbf{x})$ has the Kurdyka-Łojasiewicz (KL) property, (2) $\nabla P(\mathbf{x})$ and $\nabla P^{\mathbf{x}^k}$ are Lipschitz continuous, then our Algorithm 1 converges to a critical point of (6) from every initial point. It is shown in [13-14] that the real analytic and sub-analytic functions satisfy the KL property. $F(\mathbf{x})$ here is a sub-analytic function, thus, $F(\mathbf{x})$ has the KL property. We now proceed to prove (2): The gradient of $p(x)$ is $p'(x) = \frac{1}{2}(\sqrt{x^2 + 4\lambda^2} - x)$, when $x > 0$, note that both

$f(x) = \sqrt{x^2 + 4\lambda^2}$ and $f(x) = x$ are Lipschitz continuous with the Lipschitz constant $K = 1$. Then, $p'(x)$ is Lipschitz continuous with the Lipschitz constant $L_i \geq 0$. Therefore, the gradient of $P(\mathbf{x}) = \sum_i p(x_i)$, i.e., $\nabla P(\mathbf{x}) = \sum_i p'(x_i)$ is Lipschitz continuous with the Lipschitz constant $L_p \geq \sum_i L_i \geq 0$. When $x < 0$, it is clear that we can draw the same conclusion. Next, we prove that $\nabla P^{\mathbf{x}^k}$ are Lipschitz continuous: the gradient of $P^{\mathbf{x}^k}$ is $\nabla P^{\mathbf{x}^k} = \mathbf{W}^k \mathbf{x}$, $\mathbf{W}^k = \text{diag}(\frac{2\lambda^2}{|x^k|(\sqrt{(x^k)^2 + 4\lambda^2} + |x^k|)})$. Note that x^k is bounded. If we let the

maximal value of \mathbf{W}^k be W_{\max}^k , the gradients of $P^{\mathbf{x}^k}$, i.e., $\nabla P^{\mathbf{x}^k}$ is locally Lipschitz continuous on a bounded set B for all $\mathbf{x} \in \mathbf{X}$ with the common Lipschitz constant $\tilde{L}_p \geq W_{\max}^k > 0$.

Appendix B

Proof of Theorem 2

Proof: We let S be an index set of nonzero entries of \mathbf{x}^* and \bar{S} be an index set of s largest entries of the absolute value of $\bar{\mathbf{x}}$. Since, $\|\mathbf{x}^*\|_0 \leq s$, we obtain:

$$\begin{aligned} \|\bar{\mathbf{x}} - \mathbf{x}^*\| &\leq \|(\bar{\mathbf{x}} - \mathbf{x}^*)_{S \cup S^*}\| + \|(\bar{\mathbf{x}})_{(S \cup S^*)^c}\| \\ &\leq \frac{1}{\sqrt{1 - \delta_{2s}}} \|\Phi(\bar{\mathbf{x}} - \mathbf{x}^*)_{S \cup S^*}\| + \|(\bar{\mathbf{x}})_{(S \cup S^*)^c}\| \\ &\leq \frac{1}{\sqrt{1 - \delta_{2s}}} \|\Phi \bar{\mathbf{x}} - \mathbf{y}\| + (\frac{1}{\sqrt{1 - \delta_{2s}}} \|\Phi\| + 1) \|(\bar{\mathbf{x}})_{(S \cup S^*)^c}\| \end{aligned} \quad (19)$$

In addition, since $\|\Phi\bar{\mathbf{x}} - \mathbf{y}\| \leq \sqrt{F(\bar{\mathbf{x}})} \leq \sqrt{F(\mathbf{x}^0)}$, we finally have:

$$\|\bar{\mathbf{x}} - \mathbf{x}^*\| \leq \sqrt{\frac{F(\mathbf{x}^0)}{1 - \delta_{2s}}} + \left(\frac{1}{\sqrt{1 - \delta_{2s}}} \|\Phi\| + 1\right) \sigma_s(\bar{\mathbf{x}})_2 \quad (20)$$

Appendix C

Proof of Proposition 1-3

Proof of proposition 1: If the true signal vector is $\mathbf{x}^*(t-1) = [0, 0, 1, 0, 0]^T$, since $\hat{\mathbf{x}}(t-1)$ is estimated by a sparse recovery algorithm, $\|\hat{\mathbf{x}}(t-1) - \mathbf{x}^*(t-1)\|$ is often smaller than 0.1 [18]. However, assuming an arbitrary sparse signal is $\mathbf{x}(t-1) = [1, 0, 0, 0, 1]^T$, then, $\|\mathbf{x}(t-1) - \mathbf{x}^*(t-1)\| \approx 1.7$ which is bigger than $\|\hat{\mathbf{x}}(t-1) - \mathbf{x}^*(t-1)\|$.

Proof of proposition 2: $F(\mathbf{x})$ is the sum of the data-fitting term and sparse penalty term. So a sparse \mathbf{x}^0 with a smaller $\|\mathbf{x}^0 - \mathbf{x}^*\|$ would lead to a smaller $F(\mathbf{x}^0)$.

Proof of proposition 3: $\hat{\mathbf{x}}(t-1)$ has a smaller distance to the true sparse signal $\mathbf{x}^*(t-1)$ than an arbitrary sparse signal $\mathbf{x}(t-1)$. If \mathbf{G} can be accurately determined, then $\mathbf{x}_p(t)$ has a smaller distance to the true sparse signal $\mathbf{x}^*(t)$. Based on proposition 2, we use $\mathbf{x}_p(t)$ as the initialization point \mathbf{x}^0 leads to a smaller $F(\mathbf{x}^0)$. According to Note 1, improved sparse signal recovery performance can be achieved by using $\mathbf{x}_p(t)$ as the initialization point \mathbf{x}^0 .

References

- [1] M. Cossalter, G. Valenzise, M. Tagliasacchi and S. Tubaro, "Joint Compressive Video Coding and Analysis," *IEEE Trans. Multimedia*, vol. 12, no. 3, pp. 168-183, April, 2010. [Article \(CrossRef Link\)](#).
- [2] N. Vaswani and W. Lu, "Modified-CS: Modifying Compressive Sensing for Problems With Partially Known Support," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4595-4607, September, 2010. [Article \(CrossRef Link\)](#).
- [3] W. U. Bajwa, J. Haupt, A. M. Sayeed and R. Nowak, "Compressed Channel Sensing: A New Approach to Estimating Sparse Multipath Channels," in *Proc. of IEEE*, vol. 98, no. 6, pp. 1058-1076, June, 2010. [Article \(CrossRef Link\)](#).
- [4] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4203-4215, December, 2005. [Article \(CrossRef Link\)](#).
- [5] W. Shi, Q. Ling, K. Yuan, G. Wu and W. Yin, "On the Linear Convergence of the ADMM in Decentralized Consensus Optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 7, pp. 1750-1761, April, 2014. [Article \(CrossRef Link\)](#).
- [6] E. J. Candès, M. B. Wakin and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5-6, pp. 877-905, December, 2008. [Article \(CrossRef Link\)](#).

- [7] D. P. Wipf, B. D. Rao and S. Nagarajan, "Latent Variable Bayesian Models for Promoting Sparsity," *IEEE Trans. Inform. Theory*, vol. 57, no. 9, pp. 6236-6255, September, 2011. [Article \(CrossRef Link\)](#).
- [8] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600-616, March, 1997. [Article \(CrossRef Link\)](#).
- [9] C. Lu, Z. Lin and S. Yan, "Smoothed Low Rank and Sparse Matrix Recovery by Iteratively Reweighted Least Squares Minimization," *IEEE Trans. Image Process.*, vol. 24, no. 2, pp. 646-654, Feb. 2015. [Article \(CrossRef Link\)](#).
- [10] C. J. Miosso, R. von Borries, M. Argaez, L. Velazquez, C. Quintero and C. M. Potes, "Compressive Sensing Reconstruction With Prior Information by Iteratively Reweighted Least-Squares," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2424-2431, June, 2009. [Article \(CrossRef Link\)](#).
- [11] Z. Zhang, T. P. Jung, S. Makeig and B. D. Rao, "Compressed Sensing for Energy-Efficient Wireless Telemonitoring of Noninvasive Fetal ECG Via Block Sparse Bayesian Learning," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 2, pp.300-309, February, 2013. [Article \(CrossRef Link\)](#).
- [12] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *Am. Stat.*, vol. 58, no. 1, pp. 30-37, 2004. [Article \(CrossRef Link\)](#).
- [13] P. Ochs, A. Dosovitskiy, T. Brox, and T. Pock, "On Iteratively Reweighted Algorithms for Nonsmooth Nonconvex Optimization in Computer Vision," *SIAM J. Imaging Sci.*, vol. 8, no. 1, pp. 331-372, February, 2015. [Article \(CrossRef Link\)](#).
- [14] H. Attouch, J. Bolte and B. F. Svaiter, "Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods," *Math. Program.*, vol. 137, no. 1, pp. 91-129, February, 2013. [Article \(CrossRef Link\)](#).
- [15] R. T. Rockafellar and R. J. B. Wets, *Variational analysis*, 1st Edition, Springer-Verlag, Berlin, 1998. [Article \(CrossRef Link\)](#).
- [16] M. J. Lai, Y. Y. Xu, and W. Yin, "Improved iteratively reweighted least squares for unconstrained smoothed ℓ_q minimization," *SIAM J. Numer. Anal.*, vol. 51, no. 2, pp. 927-957, March, 2013. [Article \(CrossRef Link\)](#).
- [17] L. Qin, Z. Lin, Y. She and C. Zhang, "A comparison of typical ℓ_p minimization algorithms," *Neurocomputing*, vol. 119, pp. 413-424, November, 2013. [Article \(CrossRef Link\)](#).
- [18] J. f. Yang and Y. Zhang, "Alternating Direction Algorithms for ℓ_1 -Problems in Compressive Sensing," *SIAM J. Sci. Comput.*, vol. 33, no. 1, pp. 250-278, February, 2011. [Article \(CrossRef Link\)](#).
- [19] S. Boyd and L. Vandenberghe, *Convex optimization*, 1st Edition, Cambridge University Press, New York, 2004. [Article \(CrossRef Link\)](#).
- [20] Z. Gao, L. Dai, Z. Wang and S. Chen, "Spatially Common Sparsity Based Adaptive Channel Estimation and Feedback for FDD Massive MIMO," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6169-6183, December, 2015. [Article \(CrossRef Link\)](#).
- [21] F. Pena-Campos, R. Carrasco-Alvarez, O. Longoria-Gandara and R. Parra-Michel, "Estimation of Fast Time-Varying Channels in OFDM Systems Using Two-Dimensional Prolate," *IEEE Trans. Wirel. Commun.*, vol. 12, no. 2, pp. 898-907, February, 2013. [Article \(CrossRef Link\)](#).
- [22] A. Zhang, S. Yang, J. Li, C. Li and Z. Liu, "Sparsity Adaptive Expectation Maximization Algorithm for Estimating Channels in MIMO Cooperation systems," *Ksii Trans. Internet Inf.*, vol. 10, no. 8, p3498-3511, August, 2016. [Article \(CrossRef Link\)](#).
- [23] M. F. Duarte and Y. C. Eldar, "Structured Compressed Sensing: From Theory to Applications," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4053-4085, September, 2011. [Article \(CrossRef Link\)](#).
- [24] Z. Gao, L. Dai, W. Dai, B. Shim and Z. Wang, "Structured Compressive Sensing-Based Spatio-Temporal Joint Channel Estimation for FDD Massive MIMO," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 601-617, February, 2016. [Article \(CrossRef Link\)](#).

- [25] Z. Zhang and B. D. Rao, "Sparse Signal Recovery With Temporally Correlated Source Vectors Using Sparse Bayesian Learning," *IEEE J. Sel. Top. Sign. Proces.*, vol. 5, no. 5, pp. 912-926, September, 2011. [Article \(CrossRef Link\)](#).
- [26] Z. Gao, L. Dai, C. Qi, C. Yuen and Z. Wang, "Near-Optimal Signal Detector Based on Structured Compressive Sensing for Massive SM-MIMO," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1860-1865, February, 2017. [Article \(CrossRef Link\)](#).
- [27] S. Beygi, U. Mitra and E. G. Ström, "Nested Sparse Approximation: Structured Estimation of V2V Channels Using Geometry-Based Stochastic Channel Model," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4940-4955, September, 2015. [Article \(CrossRef Link\)](#).
- [28] R. Prasad, C. R. Murthy and B. D. Rao, "Joint Approximately Sparse Channel Estimation and Data Detection in OFDM Systems Using Sparse Bayesian Learning," *IEEE Trans. Signal Process.*, vol. 62, no. 14, pp. 3591-3603, July, 2014. [Article \(CrossRef Link\)](#).
- [29] J. W. Choi and B. Shim, "Statistical Recovery of Simultaneously Sparse Time-Varying Signals From Multiple Measurement Vectors," *IEEE Trans. Signal Process.*, vol. 63, no. 22, pp. 6136-6148, November, 2015. [Article \(CrossRef Link\)](#).
- [30] F. Gustafsson et al., "Particle filters for positioning, navigation, and tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 425-437, February, 2002. [Article \(CrossRef Link\)](#).
- [31] D. P. Wipf, B. D. Rao and S. Nagarajan, "Latent Variable Bayesian Models for Promoting Sparsity," *IEEE Trans. Inform. Theory*, vol. 57, no. 9, pp. 6236-6255, September, 2011. [Article \(CrossRef Link\)](#).
- [32] M. S. Arulampalam, S. Maskell, N. Gordon and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174-188, February, 2002. [Article \(CrossRef Link\)](#).
- [33] Q. Qin, L. Gui, B. Gong, X. Ren and W. Chen, "Structured Distributed Compressive Channel Estimation Over Doubly Selective Channels," *IEEE Trans. Broadcast.*, vol. 62, no. 3, pp. 521-531, September, 2016. [Article \(CrossRef Link\)](#).
- [34] D. P. Wipf and B. D. Rao, "An Empirical Bayesian Strategy for Solving the Simultaneous Sparse Approximation Problem," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3704-3716, July, 2007. [Article \(CrossRef Link\)](#).
- [35] Z. Zhang, W. Zhang and C. Tellambura, "Cooperative OFDM Channel Estimation in the Presence of Frequency Offsets," *IEEE Trans. Veh. Technol.*, vol. 58, no. 7, pp. 3447-3459, September, 2009. [Article \(CrossRef Link\)](#).



Chen Wang received the B.S. degree from Tongji University in 2010, the M.S. degree from Hong Kong University of Science and Technology in 2012. He is currently working toward the Ph.D. degree with the School of Communication and Information Engineering, Shanghai University, Shanghai, China. His research interest is Compressive Sensing for Wireless Communication Systems.



Yong Fang received the Ph.D. degree from Electronic Engineering, City University of Hong Kong, Hong Kong, in 1999. Since 2002, he has been a Professor in School of Communication and Information Engineering, Shanghai University, Shanghai. His research interests include Communication Signal Processing, Blind Signal Processing, and Adaptive Information System.