

Sparse Channel Reconstruction With Nonconvex Regularizer via DC Programming for Massive MIMO Systems

Pengxia Wu*, Hui Ma* and Julian Cheng*

*School of Engineering, The University of British Columbia, Kelowna, BC, Canada
Email: pengxia.wu@ubc.ca, hui.ma@alumni.ubc.ca, julian.cheng@ubc.ca

Abstract—Sparse channel estimation for massive multiple-input multiple-output systems has drawn much attention in recent years. The required pilots are substantially reduced when the sparse channel state vectors can be reconstructed from a few numbers of measurements. A popular approach for sparse reconstruction is to solve the least-squares problem with a convex regularizer. However, the convex regularizer is either too loose to force sparsity or lead to biased estimation. In this paper, the sparse channel reconstruction is solved by minimizing the least-squares objective with a nonconvex regularizer, which can exactly express the sparsity constraint and avoid introducing serious bias in the solution. A novel algorithm is proposed for solving the resulting nonconvex optimization via the difference of convex functions programming and the gradient projection descent. Simulation results show that the proposed algorithm is fast and accurate, and it outperforms the existing sparse recovery algorithms in terms of reconstruction errors.

I. INTRODUCTION

Channel estimation for massive multiple-input multiple-output (MIMO) systems has drawn increasing attention in recent years. Due to the large-scale antenna arrays, conventional channel estimation techniques require a large number of time slots as the pilot training overhead. Sparse channel estimation [1]–[5] has the advantage to reduce pilot training overhead substantially by exploiting the channel sparsity in angular domain. Sparse channel estimation requires to reconstruct a sparse channel state vector from a few number of randomly projected measurements. Thus, a sparse channel estimation process also refers to a sparse channel reconstruction. A sparse channel reconstruction is to resolve an underdetermined linear system of equations to achieve a sparse solution closest to the real angular-domain channel vector. A unique sparse solution can be obtained through solving an optimization of least squares constrained with a ℓ_0 -norm term, where the ℓ_0 -norm constraint term is used to enforce sparsity in the solution by limiting the number of non-zero elements. Since ℓ_0 -norm is discrete and nonconvex, the ultimate sparse reconstruction problem is a combinational optimization problem and is NP-hard [6].

Common approaches to solving sparse reconstruction include greedy approach and ℓ_1 -relaxation optimization. Representative greedy algorithms include the orthogonal matching pursuit (OMP) [7], [8], the CoSaMP [9] and the least angle regression (LARS) [10]. These methods work well when the vector is sufficiently sparse, but their performances degrade seriously when the sparsity is reduced. The most popular approach for sparse reconstruction is the ℓ_1 -relaxation optimization, where the nonconvex ℓ_0 -norm constraint is relaxed and approximated by the convex ℓ_1 -norm constraint. Numerous algorithms have been developed for solving the relaxed convex optimization of sparse reconstructions, such as the iterative shrinkage-thresholding algorithm (ISTA) [11], the fast iterative shrinkage-thresholding algorithm (FISTA) [12], the gradient projection sparse recovery algorithm (GPSR) [13] and the sparse reconstruction by separable approximation (SpaRSA) [14]. These ℓ_1 -relaxation based algorithms can guarantee to converge theoretically within a finite number of iterations. However, the ℓ_1 -norm constraint is a loose relaxation of the ℓ_0 -norm constraint, and does not always provide an accurate sparse solution.

In this paper, we propose a novel algorithm for sparse channel reconstruction for massive MIMO systems. More specifically, we introduce the top- $(K, 1)$ norm [15] to represent exactly the ℓ_0 -norm constraint instead of its approximation. Then, we express the problem of sparse reconstructions as the optimization of least squares objective penalised by a nonconvex regularizer term, which is represented using the top- $(K, 1)$ norm. To solve the resulting nonconvex optimization problem of sparse reconstructions, we employ the difference of convex functions (DC) programming [16]. In specific, the DC programming solves nonconvex optimization by decomposing the nonconvex objective into a form of DC and performing iterations through a primal-dual method [16]. At each iteration, the DC algorithm solves a convex subproblem, which is an approximation of the original nonconvex problem. We express the subproblem as a bound-constrained quadratic program (BCQP) with a simple nonnegativity constraint, such that it can be efficiently solved by the gradient projection descent method. The proposed DC gradient projection algorithm is a

double-layer iteration algorithm, which shares the same global convergence property with general DC algorithms [16]. Numerical results show that the proposed DC gradient projection algorithm can accurately reconstruct the sparse channel vectors both in the noiseless and noisy scenario. The proposed DC gradient projection algorithm can achieve solutions that have lower reconstruction errors compared to existing algorithms, including the OMP algorithm and several ℓ_1 -relaxation algorithms.

II. SYSTEM MODEL

We consider a downlink massive MIMO system, where the base station (BS) has N antennas and each user equipment (UE) has a single antenna. We let the vector $\mathbf{h}_s \in \mathbb{C}^N$ denote the spatial-domain channel between the BS and a UE, and let the vector $\mathbf{h}_a \in \mathbb{C}^N$ denote channel in virtual angular-domain. Assuming a narrowband blockfading channel, the spatial-domain channel vector \mathbf{h}_s is given by [17]

$$\mathbf{h}_s = \sqrt{\frac{N}{N_p}} \sum_{l=1}^{N_p} \beta^{(l)} \boldsymbol{\alpha}(\phi^{(l)}) \quad (1)$$

where N_p is the number of paths; $l = 1$ is the index for the line-of-sight path; $2 \leq l \leq N_p$ is the index for non-line-of-sight paths; $\beta^{(l)}$ is the complex path gain; $\boldsymbol{\alpha}(\phi^{(l)})$ is the corresponding array steering vector that contains a list of complex spatial sinusoids to represent the relative phase shifts for the incident far-field waveform across the array elements. For the N -element uniform linear array, the array steering vector $\boldsymbol{\alpha}(\phi^{(l)})$ is given by

$$\boldsymbol{\alpha}(\phi^{(l)}) = \frac{1}{\sqrt{N}} [1, e^{-j2\pi\phi^{(l)}}, \dots, e^{-j2\pi\phi^{(l)}(N-1)}]^T \quad (2)$$

where $\phi^{(l)}$ denotes the spatial direction of the l th path, and it is related to the physical angle $\theta^{(l)}$ by $\phi^{(l)} = \frac{d}{\lambda} \sin \theta^{(l)}$ for $-\frac{1}{2} \leq \phi^{(l)} \leq \frac{1}{2}$ and $-\frac{\pi}{2} \leq \theta^{(l)} \leq \frac{\pi}{2}$, where λ is the wavelength, and $d = \frac{\lambda}{2}$ is the antenna spacing.

The spatial-domain channel vector \mathbf{h}_s in (1) can be transformed into the virtual angular-domain by [17]

$$\mathbf{h}_a = \mathbf{U}\mathbf{h}_s \quad (3)$$

where \mathbf{U} denotes the discrete Fourier transform (DFT) matrix having the size $N \times N$, and it can be expressed using a set of orthogonal array steering vectors as

$$\mathbf{U} = [\boldsymbol{\alpha}(\phi_1), \boldsymbol{\alpha}(\phi_2), \dots, \boldsymbol{\alpha}(\phi_N)]^H \quad (4)$$

where $\phi_i = \frac{1}{N}(i - \frac{N+1}{2})$ for $i = 1, 2, \dots, N$ is the spatial direction predefined by the array having half-wavelength spaced antennas. The massive MIMO channels have strong spatially correlations and much-lower degrees of freedom than the number of antennas. Since the BS is usually located at an elevated position far away from the UE, the number of scattering clusters around the BS is limited and each scattering cluster has a small angular spread. Thus, the majority of channel energy $\|\mathbf{h}_a\|_2^2$ is occupied by a limited dimensions in angular domain. Consequently, the massive MIMO channels

have a sparse angular-domain representation, and only a small number of nonzero elements exist in the angular-domain channel vector \mathbf{h}_a .

For the pilot-aided downlink channel estimation schemes, the BS transmits the known pilots \mathbf{P} to the UEs. The received pilot symbols at the UE can be expressed as [18]

$$\mathbf{r} = \mathbf{P}\mathbf{h}_s + \mathbf{w} \quad (5)$$

where $\mathbf{r} \in \mathbb{C}^M$ is the received pilots; $\mathbf{P} \in \mathbb{C}^{M \times N}$ is the downlink pilot matrix transmitted over M time slots; $\mathbf{h}_s \in \mathbb{C}^N$ is the spatial-domain channel vector; \mathbf{w} is the received noise vector and $\mathbf{w} \sim \mathcal{CN}(0, \sigma_n^2 \mathbf{I})$.

By the relationship between the spatial-domain channel \mathbf{h}_s and the angular-domain channel \mathbf{h}_a in (3), the channel vector \mathbf{h}_s in (5) can be replaced by $\mathbf{h}_s = \mathbf{U}^H \mathbf{h}_a$. Thus, the received pilot symbols in (5) can be rewritten as

$$\mathbf{r} = \mathbf{P}\mathbf{U}^H \mathbf{h}_a + \mathbf{w}. \quad (6)$$

By writing $\boldsymbol{\Phi} = \mathbf{P}\mathbf{U}^H$, eq. (6) can be expressed as

$$\mathbf{r} = \boldsymbol{\Phi} \mathbf{h}_a + \mathbf{w} \quad (7)$$

where $\boldsymbol{\Phi} \in \mathbb{C}^{M \times N}$. We aim to estimate the sparse channel vector $\mathbf{h}_a \in \mathbb{C}^N$ from the lower-dimensional measurements $\mathbf{r} \in \mathbb{C}^M$ for $M \ll N$. Since the pilot length M depends on the sparsity level of channel vector \mathbf{h}_a instead of the number of BS antennas, the overheads of pilot transmission and CSI feedback can be largely reduced.

In this paper, we adopt the Gaussian random measurement matrix $\boldsymbol{\Phi} \in \mathbb{R}^{M \times N}$, which has the real-form elements following the standard Gaussian distribution. Thus, the real and imaginary parts of all the complex variables in (7) can be written as

$$\begin{aligned} \Re(\mathbf{r}) &= \boldsymbol{\Phi} \cdot \Re(\mathbf{h}_a) + \Re(\mathbf{w}) \\ \Im(\mathbf{r}) &= \boldsymbol{\Phi} \cdot \Im(\mathbf{h}_a) + \Im(\mathbf{w}) \end{aligned} \quad (8)$$

where $\Re(\cdot)$ and $\Im(\cdot)$ denote the real part and imaginary part of a complex vector. Eq. (8) implies we can equivalently treat a complex channel vector \mathbf{h}_a as a concatenated real-form vector

$$\mathbf{x} = [\Re(\mathbf{h}_a)^T, \Im(\mathbf{h}_a)^T]^T. \quad (9)$$

In the remainder, we uniquely refer \mathbf{x} as the sparse channel vector.

III. DC GRADIENT PROJECTION SPARSE RECONSTRUCTION ALGORITHM

A. Exact Sparsity Constraint Representation

According to (8) and (9), the sparse channel estimation problem can be uniquely expressed using the real-form underdetermined linear system

$$\mathbf{y} = \boldsymbol{\Phi} \mathbf{x} + \mathbf{n} \quad (10)$$

where $\mathbf{y} \in \mathbb{R}^M$ denotes the compressed measurements; $\boldsymbol{\Phi} \in \mathbb{R}^{M \times N}$ is the measurement matrix; $\mathbf{x} \in \mathbb{R}^N$ represents the sparse channel vector; $\mathbf{n} \in \mathbb{R}^M$ is the Gaussian noise vector and $\mathbf{n} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$. Sparse channel reconstruction is

to obtain a solution of $\hat{\mathbf{x}}$ from compressed measurements \mathbf{y} and measurement matrix Φ such that $\hat{\mathbf{x}} \approx \mathbf{x}$. This ℓ_0 -minimization problem is NP-hard and it is defined as

$$\begin{aligned} \min_{\mathbf{x}} \quad & \|\mathbf{x}\|_0 \\ \text{s.t.} \quad & \|\mathbf{y} - \Phi\mathbf{x}\|_2^2 \leq \tau \end{aligned} \quad (11)$$

where τ is a nonnegative real-form parameter. Problem (11) can be rewritten in an equivalent form as [19]

$$\begin{aligned} \min_{\mathbf{x}} \quad & \|\mathbf{y} - \Phi\mathbf{x}\|_2^2 \\ \text{s.t.} \quad & \|\mathbf{x}\|_0 \leq K \end{aligned} \quad (12)$$

where K is the bound of the number of nonzero elements of vector \mathbf{x} , and it is uniquely determined by the parameter τ in problem (11).

To represent exactly the sparsity constraint $\|\mathbf{x}\|_0 \leq K$ in (12), we introduce the top- $(K, 1)$ norm. The top- $(K, 1)$ norm $\|\mathbf{x}\|_{K,1}$ is defined as the sum of the largest K elements of the vector \mathbf{x} in terms of absolute value, namely

$$\|\mathbf{x}\|_{K,1} := |x_{(1)}| + |x_{(2)}| + \cdots + |x_{(K)}| \quad (13)$$

where $|x_{(i)}|$ denotes the element whose absolute value is the i th-largest among the N elements of the vector \mathbf{x} , i.e., $|x_{(1)}| \geq |x_{(2)}| \geq \cdots \geq |x_{(N)}|$. The constraint $\|\mathbf{x}\|_0 \leq K$ is equivalent to the statement that the $(K+1)$ th-largest element of the vector \mathbf{x} is zero, i.e., $\|\mathbf{x}\|_{K+1,1} - \|\mathbf{x}\|_{K,1} = 0$. Thus, we have an equivalent relationship between the following two statements [15]

$$\|\mathbf{x}\|_0 \leq K \Leftrightarrow \|\mathbf{x}\|_1 - \|\mathbf{x}\|_{K,1} = 0. \quad (14)$$

The problem (12) can be rewritten as

$$\begin{aligned} \min_{\mathbf{x}} \quad & \|\mathbf{y} - \Phi\mathbf{x}\|_2^2 \\ \text{s.t.} \quad & \|\mathbf{x}\|_1 - \|\mathbf{x}\|_{K,1} = 0 \end{aligned} \quad (15)$$

where the sparsity constraint $\|\mathbf{x}\|_0 \leq K$ is exactly represented by $\|\mathbf{x}\|_1 - \|\mathbf{x}\|_{K,1} = 0$. Using an appropriate Lagrange multiplier ρ , we rewrite the problem (15) as the following unconstrained optimization problem

$$\min_{\mathbf{x}} \quad \frac{1}{2} \|\mathbf{y} - \Phi\mathbf{x}\|_2^2 + \rho(\|\mathbf{x}\|_1 - \|\mathbf{x}\|_{K,1}) := F(\mathbf{x}) \quad (16)$$

where ρ is the regularization parameter that balance the data consistency and penalty term. Due to the nonnegativity of the penalty term $\rho(\|\mathbf{x}\|_1 - \|\mathbf{x}\|_{K,1}) \geq 0$, it can be ensured that the unconstrained problem (16) is equivalent to the constraint problem (15) when the penalty parameter ρ is taking infinite limit, which can be proved in a similar way with Theorem 17.1 of [20].

To this end, we exactly represent the ℓ_0 -constraint $\|\mathbf{x}\|_0 \leq K$ using the DC constraint $\|\mathbf{x}\|_1 - \|\mathbf{x}\|_{K,1} = 0$ so that the sparse reconstruction problem (12) is expressed in the equivalent form of (15). Then we transform the problem (15) into a unconstrained minimization problem (16). The problem (16) is a nonconvex optimization problem, because the subtracted top- $(K, 1)$ norm $\rho\|\mathbf{x}\|_{K,1}$ in the penalty term results in a nonconvex regularizer $\rho(\|\mathbf{x}\|_1 - \|\mathbf{x}\|_{K,1})$.

B. DC Programming Algorithm Framework

Our goal now is to solve the nonconvex unconstrained optimization problem (16). We employ the DC programming and decompose the objective function $F(\mathbf{x})$ in problem (16) as the difference of the two convex functions of $f(\mathbf{x})$ and $g(\mathbf{x})$

$$\min_{\mathbf{x}} \quad \underbrace{\frac{1}{2} \|\mathbf{y} - \Phi\mathbf{x}\|_2^2 + \rho\|\mathbf{x}\|_1}_{f(\mathbf{x})} - \underbrace{\rho\|\mathbf{x}\|_{K,1}}_{g(\mathbf{x})}. \quad (17)$$

At the t th-iteration, we solve the following convex subproblem

$$\min_{\mathbf{x}} \quad f(\mathbf{x}) - \mathbf{x}^T \partial g(\mathbf{x}^{t-1}). \quad (18)$$

The second convex function $g(\mathbf{x})$ in (17) is linearized by $\mathbf{x}^T \partial g(\mathbf{x}^{t-1})$ in (18), where $\partial g(\mathbf{x}^{t-1})$ is the subgradient of $g(\mathbf{x})$ at the $(t-1)$ th update \mathbf{x}^{t-1} , that is

$$\partial g(\mathbf{x}^{t-1}) = \rho \partial \|\mathbf{x}^{t-1}\|_{K,1} \quad (19)$$

where $\partial \|\mathbf{x}^{t-1}\|_{K,1}$ denotes the subgradient of $\|\mathbf{x}^{t-1}\|_{K,1}$. The subgradient $\partial \|\mathbf{x}\|_{K,1}$ of $\|\mathbf{x}\|_{K,1}$ is defined as [15]

$$\partial \|\mathbf{x}\|_{K,1} := \operatorname{argmax}_{\mathbf{w}} \left\{ \sum_{i=1}^N x_i w_i \mid \sum_{i=1}^N |w_i| = K, w_i \in [-1, 1] \right\}. \quad (20)$$

A subgradient $\mathbf{w}_x^{t-1} \in \partial \|\mathbf{x}^{t-1}\|_{K,1}$ can be simply obtained by assigning the sign of the first K largest elements of $|\mathbf{x}^{t-1}|$ to the corresponding elements of \mathbf{w}_x^{t-1} , i.e., $(\mathbf{w}_x^{t-1})_{(i)} = \operatorname{sign}(x_{(i)}^{t-1})$, where the subscript (i) represents the i th element of a vector, and setting the other elements of \mathbf{w}_x^{t-1} to be zeros.

The DC algorithm framework for sparse reconstruction is outlined as follows:

1. **Start:** Given a starting point \mathbf{x}^0 , and a small threshold parameter $\epsilon > 0$.
2. **Repeat:** For $t = 1, 2, \dots$
 Select a subgradient $\mathbf{w}_x^{t-1} \in \partial \|\mathbf{x}^{t-1}\|_{K,1}$;
 Solve the convex subproblem (18), i.e., $\min_{\mathbf{x}} f(\mathbf{x}) - \rho \mathbf{x}^T \mathbf{w}_x^{t-1}$, and obtain \mathbf{x}^t ;
 Increment t ;
3. **End:** Until terminate condition satisfies.

C. DC Gradient Projection Algorithm for Sparse Reconstruction

Following the aforementioned DC algorithm framework to solve the problem (16), at the t th iteration we need to solve a nonsmooth convex subproblem (18), which is

$$\min_{\mathbf{x}} \quad \frac{1}{2} \|\mathbf{y} - \Phi\mathbf{x}\|_2^2 + \rho\|\mathbf{x}\|_1 - \rho \mathbf{x}^T \mathbf{w}_x^{t-1} \quad (21)$$

where $\mathbf{w}_x^{t-1} = \partial \|\mathbf{x}^{t-1}\|_{K,1}$. We turn it into a constraint quadratic problem and solve it using the projected gradient descent method. We split the positive and negative part of \mathbf{x} , and represent \mathbf{x} as the difference of its positive part \mathbf{u} and its negative part \mathbf{v} , that is

$$\mathbf{x} = \mathbf{u} - \mathbf{v}, \mathbf{u} \geq \mathbf{0}, \mathbf{v} \geq \mathbf{0} \quad (22)$$

where $\mathbf{u} = (\mathbf{x})_+$, $\mathbf{v} = (-\mathbf{x})_+$, where $(\cdot)_+$ represents a nonnegative-clipper operation that retains the nonnegative elements and sets negative elements be zeros. More precisely, $(\mathbf{x})_+$ represents for each element x in vector \mathbf{x} we take $(x)_+ = \max\{0, x\}$; $(-\mathbf{x})_+$ represents for each element $-x$ in vector $-\mathbf{x}$ we take $(-x)_+ = \max\{0, -x\}$. Noticing that $\|\mathbf{x}\|_1 = \mathbf{1}^T \mathbf{u} + \mathbf{1}^T \mathbf{v}$, the subproblem (21) can be written as a bound-constrained quadratic program (BCQP)

$$\begin{aligned} \min_{\mathbf{u}, \mathbf{v}} \quad & \frac{1}{2} \|\mathbf{y} - \Phi(\mathbf{u} - \mathbf{v})\|_2^2 + \rho \mathbf{1}^T \mathbf{u} + \rho \mathbf{1}^T \mathbf{v} - \rho \mathbf{u}^T \mathbf{w}_u^{t-1} \\ & - \rho \mathbf{v}^T \mathbf{w}_v^{t-1} \\ \text{s.t.} \quad & \mathbf{u} \geq \mathbf{0}, \mathbf{v} \geq \mathbf{0} \end{aligned} \quad (23)$$

where \mathbf{w}_u^{t-1} and \mathbf{w}_v^{t-1} represent the positive and negative part of \mathbf{w}_x^{t-1} , i.e., $\mathbf{w}_u^{t-1} = (\mathbf{w}_x^{t-1})_+$, $\mathbf{w}_v^{t-1} = (-\mathbf{w}_x^{t-1})_+$. Let \mathbf{z} denote the concatenation of \mathbf{u} and \mathbf{v} , i.e., $\mathbf{z} = [\mathbf{u}^T, \mathbf{v}^T]^T$, we rewrite (23) into a compact form

$$\begin{aligned} \min_{\mathbf{z}} \quad & \frac{1}{2} \mathbf{z}^T \mathbf{B} \mathbf{z} + \mathbf{c}^T \mathbf{z} := G(\mathbf{z}), \\ \text{s.t.} \quad & \mathbf{z} \geq \mathbf{0} \end{aligned} \quad (24)$$

where

$$\begin{aligned} \mathbf{z} &= \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \Phi^T \Phi & -\Phi^T \Phi \\ -\Phi^T \Phi & \Phi^T \Phi \end{bmatrix}, \\ \mathbf{c} &= \begin{bmatrix} \Phi^T \mathbf{y} \\ -\Phi^T \mathbf{y} \end{bmatrix} + \rho \mathbf{1}^T - \rho \mathbf{w}_z^{t-1} \end{aligned}$$

where $\mathbf{1}^T$ represents the all-ones column vector in the same size with \mathbf{z} ; $\mathbf{s}_z^{t-1} = [(\mathbf{s}_u^{t-1})^T, (\mathbf{s}_v^{t-1})^T]^T \in \partial \|\mathbf{z}^{t-1}\|_{K,1}$ is a subgradient of $\|\mathbf{z}^{t-1}\|_{K,1}$, and \mathbf{w}_z^{t-1} can have either zero-valued or one-valued elements for $\mathbf{z}^{t-1} \geq \mathbf{0}$.

Now, we can apply the gradient projection descent to solve the subproblem (24). The k th-step update is expressed as

$$\begin{aligned} \mathbf{z}^{(k+\frac{1}{2})} &= \text{Proj} \left(\mathbf{z}^{(k)} - \alpha^k \nabla G(\mathbf{z}^{(k)}) \right), \\ \mathbf{z}^{(k+1)} &= \mathbf{z}^{(k)} + \beta^k (\mathbf{z}^{(k+\frac{1}{2})} - \mathbf{z}^{(k)}) \end{aligned} \quad (25)$$

where $\alpha^k > 0$ is the step size which can be determined by the Barzilai-Borwein (BB) approach, and $\beta^k \in (0, 1]$ is another step size can be found by line search [13]; $\text{Proj}(\cdot)$ represents the operation of orthogonal projection that projects the vector to the nonnegative orthant; $\nabla G(\mathbf{z}^{(k)})$ represents the gradient of $G(\mathbf{z})$ in terms of $\mathbf{z}^{(k)}$ which is calculated as

$$\begin{aligned} \nabla G(\mathbf{z}^{(k)}) &= \begin{bmatrix} \Phi^T \Phi (\mathbf{u}^{(k)} - \mathbf{v}^{(k)}) \\ -\Phi^T \Phi (\mathbf{u}^{(k)} - \mathbf{v}^{(k)}) \end{bmatrix} + \begin{bmatrix} -\Phi^T \mathbf{y} \\ \Phi^T \mathbf{y} \end{bmatrix} \\ & - \rho \mathbf{w}_z^{t-1} + \rho \mathbf{1}^T. \end{aligned} \quad (26)$$

In summary, the DC algorithm can be simplified as iteratively performing the following two steps until convergence:

$$\begin{aligned} \text{(a)} \quad & \mathbf{w}_z^{t-1} \in \partial \|\mathbf{z}^{t-1}\|_{K,1} \\ \text{(b)} \quad & \mathbf{z}^t = \underset{\mathbf{z} \geq \mathbf{0}}{\text{argmin}} \left\{ \frac{1}{2} \mathbf{z}^T \mathbf{B} \mathbf{z} + \mathbf{c}^T \mathbf{z} \right\} \end{aligned} \quad (27)$$

where \mathbf{B} and \mathbf{c} are defined in (24). The subproblem (b) in (27) is solved by applying the gradient projection descent (25). We summarize this DC gradient projection algorithm for sparse reconstruction (DC-GPSR) in Algorithm 1.

Algorithm 1 DC gradient projection descent algorithm for sparse reconstruction (DC-GPSR)

Input: measurements \mathbf{y} , measurement matrix Φ and a small number ϵ

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

Initialization: $\mathbf{u}^0, \mathbf{v}^0, \mathbf{z}^0 \leftarrow [(\mathbf{u}^0)^T, (\mathbf{v}^0)^T]^T$

- 1: **for** $t = 1, 2, \dots$ **do**
 - 2: Compute the subgradient $\mathbf{w}_z^{t-1} \in \partial \|\mathbf{z}^{t-1}\|_{K,1}$
 - 3: **for** $k = 1, 2, \dots$ **do**
 - 4: Compute gradient $\nabla G(\mathbf{z}^{(k)})$ by (26)
 - 5: Perform gradient projection descent (25) for obtaining $\mathbf{z}^{(k+1)}$
 - 6: Check convergence, set $\mathbf{z}^* \leftarrow \mathbf{z}^{(k+1)}$ and proceed to step 7 if convergence is satisfied; otherwise return to step 3.
 - 7: **end for**
 - 8: $\mathbf{z}^t \leftarrow \mathbf{z}^*$
 - 9: Check the terminate condition $\|\mathbf{z}^t - \mathbf{z}^{t-1}\|_2 \leq \epsilon$, and return to step 1 if not satisfied; otherwise terminate with approximate solution $\mathbf{z}^t = [(\mathbf{u}^t)^T, (\mathbf{v}^t)^T]^T$, and obtain the reconstruction $\hat{\mathbf{x}} = \mathbf{u}^t - \mathbf{v}^t$.
 - 10: **end for**
-

IV. NUMERICAL RESULT

In this section, the performance of the proposed DC-GPSR algorithm is numerically evaluated and compared with existing recovery algorithms. In our simulation, the number of BS antennas is set as $N = 256$; the sparsity level of the angular-domain channel is set as 16; the dimension of measurements is set as $M = 128$. The real and imaginary part of the channel vector is concatenated together being a sample of sparse channel vector. Thus, we are reconstructing a sparse vector $\hat{\mathbf{x}} \in \mathbb{R}^{512}$ with a number of nonzero elements $K = 32$ from compressed measurements $\mathbf{y} \in \mathbb{R}^{128}$. A random Gaussian matrix drawn from standard Gaussian distribution is adopted as the measurement matrix Φ ; The terminate condition for our proposed algorithm is set as $\|\hat{\mathbf{x}}^t - \hat{\mathbf{x}}^{t-1}\|_2 \leq 10^{-30}$.

A. Performance in a Noiseless Scenario

We perform the proposed DC-GPSR algorithm in noiseless scenario to reconstruct a channel vector. The optimal sparse channel vector is denoted by \mathbf{x}_{opt} ; and the reconstructed vector is denoted by $\hat{\mathbf{x}}$. The CPU time of running DC-GPSR algorithm is 0.086 seconds on a desktop computer equipped with 3.2 GHz Intel Core i7-8700 CPU. The objective evolutions versus iterations are shown in Fig. 1 and Fig. 2. The reconstruction errors versus iterations is shown in Fig. 3. The optimal and reconstructed vectors are plotted in Fig. 4. We plot the results of conventional GPSR algorithm [13] in Figs. 1–4 as comparisons.

Figure 1 shows the objective values versus iterations, where the red dots indicate iterations of the DC-GPSR algorithm, and the solid line shows the objective evolutions of gradient projection for solving the subproblem at each iteration. The dashed line shows the objective of conventional GPSR algorithm for

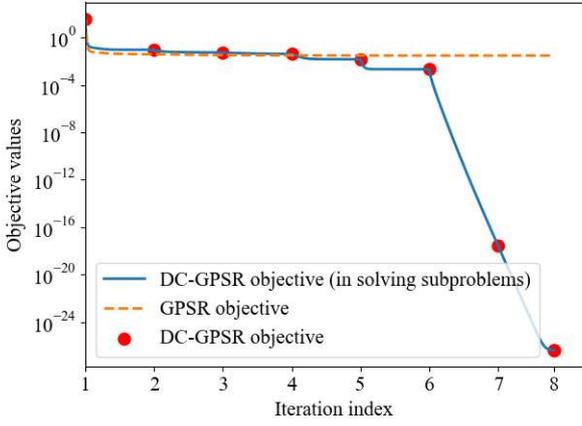


Figure 1: Objective values versus iterations; DC-GPSR objective is $\frac{1}{2}\|\mathbf{y} - \Phi\hat{\mathbf{x}}\|_2^2 + \rho(\|\hat{\mathbf{x}}\|_1 - \|\hat{\mathbf{x}}\|_{K,1})$, and GPSR objective is $\frac{1}{2}\|\mathbf{y} - \Phi\hat{\mathbf{x}}\|_2^2 + \rho\|\hat{\mathbf{x}}\|_1$

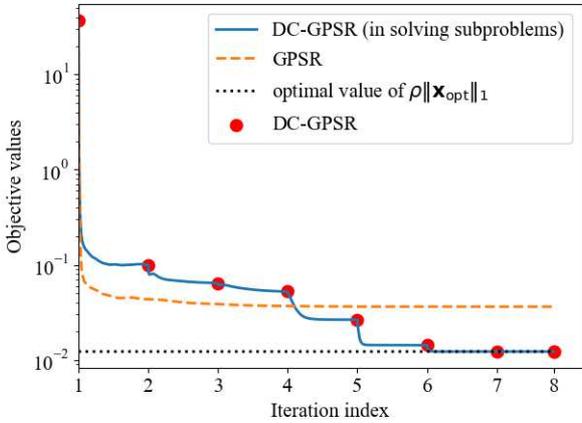


Figure 2: Objective values ($\frac{1}{2}\|\mathbf{y} - \Phi\hat{\mathbf{x}}\|_2^2 + \rho\|\hat{\mathbf{x}}\|_1$) versus iterations for DC-GPSR and GPSR algorithm

comparison. We can see the proposed DC-GPSR algorithm quickly decreases its objective after the sixth outer-step, and terminates at the eighth step, while the objective of GPSR is stuck at a relatively large value.

We investigate the values of least square objective with the ℓ_1 -norm penalty, i.e., $\|\mathbf{y} - \Phi\hat{\mathbf{x}}\|_2^2 + \rho\|\hat{\mathbf{x}}\|_1$, and show the values versus iterations in Fig. 2. The solid line marked by the red dots indicates the result of DC-GPSR, and the dashed line indicates the result of GPSR. We also draw the optimal values of the penalty term $\rho\|\mathbf{x}_{\text{opt}}\|_1$. We can see the DC-GPSR algorithm arrives and stays at the value of $\rho\|\mathbf{x}_{\text{opt}}\|_1$ after the sixth iteration, which is also the optimal value that the objective $\|\mathbf{y} - \Phi\hat{\mathbf{x}}\|_2^2 + \rho\|\hat{\mathbf{x}}\|_1$ can achieve when $\|\mathbf{y} - \Phi\hat{\mathbf{x}}\|_2^2 = 0$ and $\rho\|\hat{\mathbf{x}}\|_1 = \rho\|\mathbf{x}_{\text{opt}}\|_1$. However, the GPSR cannot achieve this optimal value with a gap. It is meaningful to observe this is the gap of minimal objective values between GPSR and DC-GPSR, because it help us understand the approximation error from relaxing the ℓ_0 -norm by the ℓ_1 -norm.

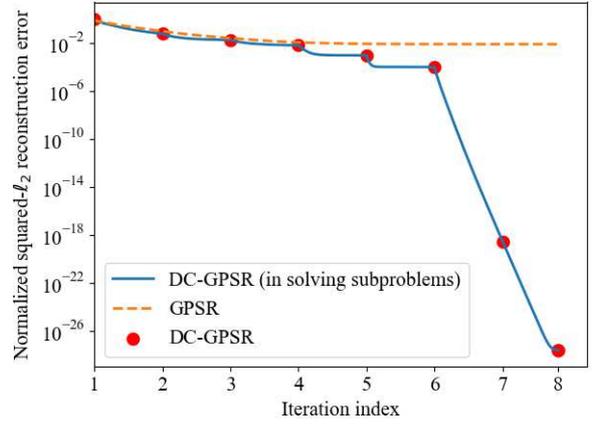


Figure 3: Normalized squared- ℓ_2 -error of reconstructions versus iterations for DC-GPSR and GPSR algorithm

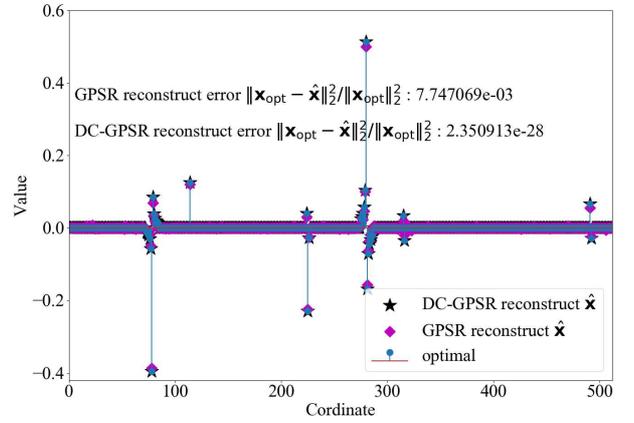


Figure 4: Illustration of the optimal sample and the final reconstructions by DC-GPSR and GPSR algorithm

Figure 3 shows the normalized squared- ℓ_2 error ($\|\mathbf{x}_{\text{opt}} - \hat{\mathbf{x}}\|_2^2 / \|\mathbf{x}_{\text{opt}}\|_2^2$) versus iterations for reconstructions by DC-GPSR and GPSR algorithm. Fig. 4 illustrates the optimal sample and the finally reconstructed vectors by DC-GPSR and GPSR algorithm. We can see the proposed DC-GPSR algorithm achieves an accurate reconstruction with a small error on the order of 10^{-28} , which is far more accurate than the reconstruction by the GPSR algorithm with error on the order of 10^{-3} .

B. Performance in a Noisy Scenario

We perform sparse reconstructions by DC-GPSR in the noisy scenarios for 1,000 channel vector samples, and compare the normalized mean square error (NMSE) with several existing reconstruction algorithms. The NMSE is defined as

$$\text{NMSE} = \frac{1}{m} \sum_{i=0}^{m-1} \frac{\|\mathbf{x}_i - \hat{\mathbf{x}}_i\|_2^2}{\|\mathbf{x}_i\|_2^2}$$

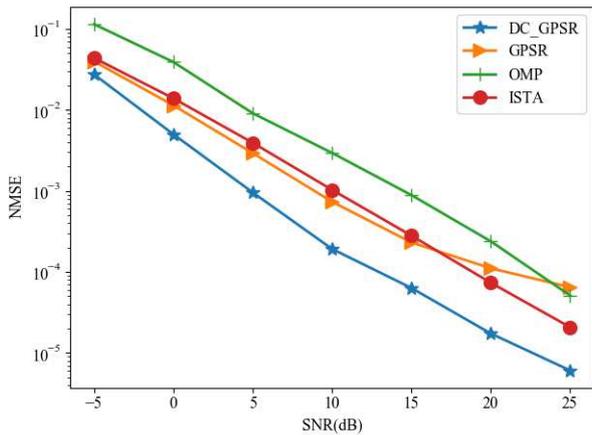


Figure 5: Reconstruction NMSE in noisy scenarios by different sparse reconstruction algorithms

where $m = 1000$ is the number of samples. For the i th channel vector sample \mathbf{x}_i , we add noise $\mathbf{n}_i \in \mathbb{R}^{128}$ on the compressed measurements \mathbf{y}_i to form a noise-corrupted compressed measurements \mathbf{y}'_i , then we reconstruct the sparse vector $\hat{\mathbf{x}}_i$ from \mathbf{y}'_i and measurement matrix Φ . The signal-to-noise ratio (SNR) is defined as

$$\text{SNR} = \frac{\|\mathbf{x}_i\|_2^2}{\mathbb{E}(\|\mathbf{n}_i\|_2^2)} = \frac{\|\mathbf{x}_i\|_2^2}{M \cdot \sigma^2}$$

where the subscript $i = 1, 2, \dots, 1000$ is the sample index; $\mathbb{E}(\cdot)$ is the expectation operator; M is the dimension of compressed measurements; σ is the standard deviation of the noise that follows zero-mean Gaussian distribution.

The benchmark algorithms include ISTA [11], GPSR [13] and OMP [8]. The results of reconstruction NMSE are shown in Fig. 5. We can see the proposed DC-GPSR has a lower reconstruction NMSE compared with the existing sparse reconstruction algorithms. Although DC-GPSR performance degrades due to noise, the DC-GPSR algorithm exhibits considerable robustness in the noisy scenario. For example, when the SNR is 25 dB, the reconstruction NMSE of DC-GPSR algorithm is 6.07×10^{-6} , which is sufficiently accurate for most of the applications.

V. CONCLUSION

A novel sparse reconstruction algorithm DC-GPSR was proposed for the sparse channel estimation in massive MIMO systems. The sparse recover problem was formulated as a least squares problem with a nonconvex regularizer. The nonconvex regularizer is an exact representation for the ℓ_0 -norm constraint, which leads to more accurate sparse solution than the convex ℓ_1 -norm regularizer. In the proposed DC-GPSR algorithm, the nonconvex optimization problem is decomposed and approximated by a list of convex subproblems; the convex subproblems are expressed as BCQP problems and solved by the gradient projection descent method. The proposed DC-GPSR algorithm shares the global convergence property

with the general DC algorithm. Numerical results showed the DC-GPSR algorithm is fast, accurate and robust, and it outperforms several existing sparse recovery algorithms.

REFERENCES

- [1] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, June 2010.
- [2] X. Rao and V. K. N. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261–3271, June 2014.
- [3] 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, Dec. 2015.
- [4] 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, Feb. 2016.
- [5] X. Gao, L. Dai, S. Han, C. I, and X. Wang, "Reliable beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6010–6021, Sep. 2017.
- [6] E. Amaldi and V. Kann, "On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems," *Theoretical Computer Science*, vol. 209, pp. 237–260, 1998.
- [7] Y. C. Pati, R. Rezaeiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Conf. Rec. 27th Asilomar Conf. Signals, Syst. Comput.*, vol. 1, 1993, pp. 40–44.
- [8] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inform. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
- [9] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2008.
- [10] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *The Annals of Statistics*, vol. 32, no. 2, pp. 407–499, 2004.
- [11] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *Journal of Fourier Analysis and Applications*, vol. 14, pp. 629–654, 2008.
- [12] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.
- [13] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 586–597, Dec 2007.
- [14] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2479–2493, July 2009.
- [15] J. Gotoh, A. Takeda, and K. Tono, "DC formulations and algorithms for sparse optimization problems," *Mathematical Programming*, vol. 169, no. 1, pp. 141–176, 2018.
- [16] P. D. Tao and L. T. H. An, "Convex analysis approach to DC programming: theory, algorithms and applications," *Acta Mathematica Vietnamica*, vol. 22, no. 1, pp. 289–355, 1997.
- [17] D. Tse and V. Prasad, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [18] J. W. Choi, B. Shim, Y. Ding, B. Rao, and D. I. Kim, "Compressed sensing for wireless communications: Useful tips and tricks," *IEEE Commun. Surveys Tut.*, vol. 19, no. 3, pp. 1527–1550, Third Quart. 2017.
- [19] I. Rish and G. Y. Grabarnik, *Sparse Modeling: Theory, Algorithms, and Applications*. Boca Raton, Florida: Chapman & Hall/CRC Press, 2015.
- [20] S. J. W. Jorge Nocedal, *Numerical Optimization*, 2nd ed. Berlin: Springer, 2006.