# A PROJECTED PROXIMAL GRADIENT METHOD FOR EFFICIENT RECOVERY OF SPECTRALLY SPARSE SIGNALS

*Xi Yao and Wei Dai*

Department of Electrical and Electronic Engineering
Imperial College London, London, UK

## ABSTRACT

This paper investigates the recovery of a spectrally sparse signal (SSS) from partially observed entries, with particular emphasis on computational efficiency for large scaled problems. We formulate the SSS recovery as a nonconvex low-rank Hankel matrix recovery problem. A projected proximal gradient method has been developed. It is an iterative process where each iteration involves two steps. The first subspace projection step finds the optimal solution in a low-rank *and* Hankel matrix space for given column and row sub-spaces. The second step optimizes all involved variables including the column and row sub-spaces by using a proximal gradient process. In both steps, sub-problems are formulated so that both the low-rank and the Hankel structures are fully exploited for computational efficiency. The combination of these two steps substantially improves the convergence rate. Numerical simulations demonstrate a significant improvement in efficiency compared with the benchmark algorithms.

***Index Terms*—** Hankel matrix, Low-rank matrix completion, Nonconvex optimization, Spectrally sparse signal

## 1. INTRODUCTION

Spectrally sparse signals (SSS) arise in various applications, including medical imaging [1], fluorescence microsopy [2] and radar imaging [3]. In many practical situations, only parts of the signals can be measured. Hence it is essential to recover SSS from incomplete time-domain samples.

Various recovery algorithms have been developed. One approach is to discretize the frequencies and apply sparse recovery algorithms [4–6]. However, the frequency leakage effect is unavoidable when the ground-truth frequencies are off-the-grid. The gridless approach avoids the issues of frequency discretization. The crux is to cast SSS recovery as a low-rank matrix recovery problem. After solving the low-rank matrix recovery problem, the frequencies of SSS can be extracted using methods such as MUSIC [7], Prony's method [8] or a matrix pencil approach [9].

In the gridless approach, scores of convex techniques have been developed based on the concept of the atomic norm. In [10, 11], the low-rank matrix arising in SSS recovery has a Toeplitz sub-matrix. Later, Hankel matrix based recovery became more popular [12–14] due to its computational efficiency, as the involved matrix as a whole was Hankel structured and supports efficient decomposition. The convex formulation of the low-rank matrix recovery can be solved using semi-definite programming (SDP) [10–14] or alternating direction method of multipliers (ADMM) [15]. In an alternative approach, the Burer-Monteiro heuristic [16, 17] is applied where low-rank matrices are presented by a bilinear outer product so as to avoid explicit matrix decomposition. Performance guarantees of convex optimization techniques have been studied in depth [10–12, 14]. It is noteworthy that the worst-case analysis in [14] shows that convex optimization can fail even when a single element is missing.

Recently, many nonconvex methods have been developed for SSS recovery [18–23]. All these methods assumed prior knowledge of the rank of the involved Hankel matrix and conducted optimization on the space of low-rank Hankel matrices. Cadzow's algorithm [18, 23] performed alternating projections between the set of Hankel matrices and the set of low-rank matrices. This idea was taken further by adding a projection to the direct sum of the column and row spaces [21, 22]. An alternative approach is similar to the Burer-Monteiro heuristic for convex optimization. Writing a low-rank matrix/tensor as an outer-product of low-rank matrices, optimization was performed directly on the component matrices to avoid explicit matrix decomposition [19, 20].

In this paper, we develop a projected proximal gradient (PPG) method for SSS recovery. PPG is an iterative method based on a nonconvex formulation involving the low-rank constraint directly. Each iteration of PPG involves two steps. In the first step, PPG fixes the column and row sub-spaces and performs a projection to a low-rank *and* Hankel sub-space. The second step employs a proximal gradient process to update all the variables including the column and row sub-spaces. In both steps, the corresponding optimization problems are carefully formulated so that the low-rank and Hankel structures are fully exploited to reduce computational complexity. The combination of these two steps substantially improves the convergence rate. Numerical simulations demonstrate the significant efficiency improvement of PPG in solving SSS recovery problems compared to the benchmarks.

The rest of this paper is structured as follows. Sect. 2 presents our optimization formulation for SSS recovery. The PPG method is developed and detailed in Sect. 3. Numerical results are presented in Sect. 4 to demonstrate PPG's efficiency. Finally, Sect. 5 concludes this paper.

## 2. THE OPTIMIZATION FORMULATION

An SSS of order $r$ is a signal composed of complex sinusoids:

$$\boldsymbol{x}=\sum_{k=1}^{r} b_k \boldsymbol{a}(f_k;n)\in\mathbb{C}^n \tag{1}$$

where $b_k \in \mathbb{C}$ are complex amplitudes, $\boldsymbol{a}(f_k;\, n) = [1,\, e^{i2\pi f_k},\, ...\, ,\, e^{i2\pi(n-1)f_k}]^{\mathsf{T}} \in \mathbb{C}^n$ is often referred to as steering vectors in signal processing, $f_k$ denotes the frequencies, and it is typically assumed $r \ll n$. The problem under consideration is to recover an SSS $\boldsymbol{x}$ from its partial observations $\boldsymbol{s}$ (unobserved entries set to zero), i.e.,

$$\text{find} \quad \boldsymbol{x} \quad \text{s.t.} \quad \mathcal{P}_\Omega \boldsymbol{x}=\boldsymbol{s}, \tag{2}$$

where $\boldsymbol{s}$ is the partially observed signal, $\Omega\subset\{1,...,n\}$ denotes the set of indices of the observed entries, and $\mathcal{P}_\Omega$ is the corresponding projection operator. In this paper, we consider the problem where $r<m=|\Omega|<n$. The sampling rate is denoted by $p=m/n<1$.

**Remark 1.** *When $\boldsymbol{X}$ is a multidimensional SSS, that is, $\boldsymbol{X}=\sum_k b_k \boldsymbol{a}(f_{1,k};n_1)\circ\cdots\circ\boldsymbol{a}(f_{d,k};n_d)\in\mathbb{C}^{n_1\times\cdots\times n_d}$ where the symbol $\circ$ denotes the outer product, a multilevel Hankel matrix is needed. The proposed method can be easily extended for this case. Due to space constraints, we focus on one-dimensional SSS* (1) *in Sections 2 and 3 for technical development but present simulation results including multi-dimensional SSS in Section 4.*

We consider a standard formulation of SSS recovery as a low-rank Hankel matrix recovery problem [12, 17, 19–21]. A Hankel matrix is a matrix with identical elements on ascending skew diagonals. Let $\mathcal{H}$ denote the linear operator mapping a vector $\boldsymbol{x}\in\mathbb{C}^n$ to a Hankel matrix $\mathcal{H}(\boldsymbol{x})\in\mathbb{C}^{n_1\times n_2}$ with $n_1+n_2=n+1$:

$$\mathcal{H}\boldsymbol{x}=\begin{bmatrix} x_1 & x_2 & ... & x_{n_2} \\ x_2 & x_3 & ... & x_{n_2+1} \\ \vdots & & \vdots & \\ x_{n_1} & x_{n_1+1} & ... & x_n \end{bmatrix}\in\mathbb{C}^{n_1\times n_2}. \tag{3}$$

It is typically to choose $n_1\approx n_2\approx n/2$ such that the Hankel matrix is close to a square matrix [12, 20]. The Hankel matrix $\mathcal{H}(\boldsymbol{x})$ admits the Vandermonde decomposition:

$$\mathcal{H}(\boldsymbol{x})=\sum_{k=1}^{r} b_k \boldsymbol{a}(f_k;n_1)\boldsymbol{a}(f_k;n_2)^{\mathsf{T}}, \tag{4}$$

It is clear that $\text{rank}(\mathcal{H}(\boldsymbol{x}))\le r$ [9]. Hence, the SSS recovery problem can be formulated as a constrained nonconvex problem:[1]

$$\min_{\boldsymbol{H},\boldsymbol{x}} \delta(\text{rank}(\boldsymbol{H})\le r)+\frac{1}{2}\|\boldsymbol{s}-\mathcal{P}_\Omega\boldsymbol{x}\|^2 \quad \text{s.t. } \boldsymbol{H}=\mathcal{H}\boldsymbol{x}, \tag{5}$$

where $\delta(\cdot)$ is the characteristic function taking values zero and infinity. Due to the difficulty to solve a constrained nonconvex optimization problem, the following relaxation is used in this paper:

$$\min_{\boldsymbol{H},\boldsymbol{x}} \delta(\text{rank}(\boldsymbol{H})\le r)+\frac{1}{2}\|\boldsymbol{s}-\mathcal{P}_\Omega\boldsymbol{x}\|^2+\frac{\beta}{2}\|\boldsymbol{H}-\mathcal{H}\boldsymbol{x}\|_F^2+\frac{\alpha}{2}\|\boldsymbol{x}\|^2, \tag{6}$$

where $\beta>0$ is the Hankel enforcement parameter and its value should be large to enforce a low-rank Hankel solution; $\alpha$ is a small positive scalar for regularization.

## 3. A PROJECTED PROXIMAL GRADIENT METHOD

A projected proximal gradient (PPG) method is proposed in this section. The PPG is an iterative process composed of two steps: the first subspace projection optimizes all variables except some low-rank column and row subspaces, and the second is to update the column and row subspaces together with all other variables via a proximal gradient (PG). Each step is carefully designed to use low-rank and Hankel structures for computational efficiency.

### 3.1. The Subspace Projection Step

The subspace projection step is to solve (6) with fixed low-rank column and row subspaces of $\boldsymbol{H}$. Specifically, a rank-$r$ matrix $\boldsymbol{H}$ admits a decomposition $\boldsymbol{H}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathsf{H}}$ where $\boldsymbol{U}\in\mathbb{C}^{n_1\times r}$ and $\boldsymbol{V}\in\mathbb{C}^{n_2\times r}$ contain orthonormal columns. By fixing $\boldsymbol{U}$ and $\boldsymbol{V}$, the optimization problem (6) becomes

$$\begin{aligned}&\min_{\boldsymbol{H}\text{ with fixed }\boldsymbol{U}\&\boldsymbol{V},\boldsymbol{x}} \delta(\text{rank}(\boldsymbol{H})\le r)+\frac{1}{2}\|\boldsymbol{s}-\mathcal{P}_\Omega\boldsymbol{x}\|^2 \\ &+\frac{\beta}{2}\|\boldsymbol{H}-\mathcal{H}\boldsymbol{x}\|_F^2+\frac{\alpha}{2}\|\boldsymbol{x}\|^2 \\ &=\min_{\boldsymbol{\Sigma},\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{s}-\mathcal{P}_\Omega\boldsymbol{x}\|^2+\frac{\beta}{2}\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathsf{H}}-\mathcal{H}\boldsymbol{x}\|_F^2+\frac{\alpha}{2}\|\boldsymbol{x}\|^2,\end{aligned} \tag{7}$$

which is a least-squares problem of variables $\boldsymbol{\Sigma}^2$ and $\boldsymbol{x}$, and admits a closed-form solution. As the dimension of the variable in (7) is $n+r^2$, the complexity of solving (7) is in general $O((n+r^2)^3)$. Our preliminary simulations show that this complexity can be too much for large $n$.

Our novelty in this step is to recast (7) into (8), which brings down the complexity from $O((n+r^2)^3)$ to $O(r^3 n\log n+r^4 n)$. Specifically, reformulate (7) as

$$\begin{aligned}\min_{\boldsymbol{\Sigma}} h(\boldsymbol{\Sigma})=\min_{\boldsymbol{\Sigma}}\min_{\boldsymbol{x}} &\frac{1}{2}\|\boldsymbol{s}-\mathcal{P}_\Omega\boldsymbol{x}\|^2+\frac{\beta}{2}\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathsf{H}}-\mathcal{H}\boldsymbol{x}\|_F^2 \\ &+\frac{\alpha}{2}\|\boldsymbol{x}\|^2.\end{aligned} \tag{8}$$

Here, the objective function $h(\boldsymbol{\Sigma})$ involves only one variable $\boldsymbol{\Sigma}$ but the evaluation of $h(\boldsymbol{\Sigma})$ involves another optimization of $\boldsymbol{x}$. The optimal solution to (8) is given below.

[1] We choose the data fidelity term $\|\boldsymbol{s}-\mathcal{P}_\Omega\boldsymbol{x}\|^2$ rather than $\|(\mathcal{H}^*\mathcal{H})^{\frac{1}{2}}(\boldsymbol{s}-\mathcal{P}_\Omega(\boldsymbol{x}))\|_F^2$ in [20, Eq.(5)]. The latter may lead to biased solutions by repeatedly counting some entries [22].

[2] Note that $\boldsymbol{\Sigma}$ is not assumed to be diagonal

**Proposition 1.** *Let* $\mathcal{P}_{\boldsymbol{U},\boldsymbol{V}}(\boldsymbol{\Sigma})=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathsf{H}}$, $\mathcal{P}^*_{\boldsymbol{U},\boldsymbol{V}}(\boldsymbol{H})=\boldsymbol{U}^{\mathsf{H}}\boldsymbol{H}\boldsymbol{V}$, $\mathcal{H}^*$ *be the adjoint operator of* $\mathcal{H}$. *Then the optimal solution to* (8) *and* (7) *is given by :*

$$\boldsymbol{x}^\star=\underbrace{(\alpha\mathcal{I}+\mathcal{P}^*_\Omega\mathcal{P}_\Omega+\beta\mathcal{H}^*\mathcal{H})^{-1}}_{\mathcal{L}}(\mathcal{P}^*_\Omega\boldsymbol{s}+\beta\mathcal{H}^*\mathcal{P}_{\boldsymbol{U},\boldsymbol{V}}\boldsymbol{\Sigma}^\star), \quad (9)$$

$$\boldsymbol{\Sigma}^\star=\left(\mathcal{I}-\beta\mathcal{P}^*_{\boldsymbol{U},\boldsymbol{V}}\mathcal{H}\mathcal{L}\mathcal{H}^*\mathcal{P}_{\boldsymbol{U},\boldsymbol{V}}\right)^{-1}\left(\mathcal{P}^*_{\boldsymbol{U},\boldsymbol{V}}\mathcal{H}\mathcal{L}\mathcal{P}^*_\Omega\boldsymbol{s}\right). \quad (10)$$

*Proof Sketch.* For any given $\boldsymbol{\Sigma}$, the corresponding optimal $\boldsymbol{x}$ is obtained by solving the inner least squares problem in (8). Substitute it back to (8) and set the gradient of $h(\boldsymbol{\Sigma})$ to zero. Eq. (10) can be reached after some elementary algebra. □

Computation of $\boldsymbol{\Sigma}^\star$ is efficient. Note that $\mathcal{H}^*\mathcal{H}$ is a linear operator that scales each entry of an $n$-dimensional vector by $w_a$, where $w_a$ is the number of elements in the $a$th skew-diagonal of a $n_1\times n_2$ matrix. Additionally, $\mathcal{P}^*_\Omega\mathcal{P}_\Omega=\mathcal{P}_\Omega$, which can be represented by diagonal matrices with 1 on the observed entries and 0 for the others. Therefore, the operator $\mathcal{L}$ exists and can be represented by diagonal matrices with positive entries. Similarly, because of the positive constant $\alpha$, $\left(\mathcal{I}-\beta\mathcal{P}^*_{\boldsymbol{U},\boldsymbol{V}}\mathcal{H}\mathcal{L}\mathcal{H}^*\mathcal{P}_{\boldsymbol{U},\boldsymbol{V}}\right)$ is positive semi-definite. Using the low-rank and the Hankel structure, the complexity of the evaluation of $\mathcal{P}^*_{\boldsymbol{U},\boldsymbol{V}}\mathcal{H}$ and $\mathcal{H}^*\mathcal{P}_{\boldsymbol{U},\boldsymbol{V}}$ is $O(rn\log n+r^2n)$ via the Fast Fourier Transform [24]. Note that the dimension of $\boldsymbol{\Sigma}$ is $r^2$. The overall complexity of subspace projection is $O(r^3n\log n+r^4n)$ via conjugate gradient (CG) method [25].

### 3.2. The Proximal Gradient (PG) Step

In the second step of PPG iterations, the PG with line search is adopted to update both $\boldsymbol{H}$ and $\boldsymbol{x}$, and hence the column and row sub-spaces of $\boldsymbol{H}$. One novelty in this step is to reformulate the objective function to (13), which supports fast convergence.

To proceed, the standard PG method is introduced in Algorithm 1. PG solves unconstrained optimization problems of the form

$$\min_{\boldsymbol{x}} F(\boldsymbol{x}):=\min_{\boldsymbol{x}} f(\boldsymbol{x})+g(\boldsymbol{x}) \quad (11)$$

where $f(\cdot)$ is Lipschitz differentiable[3] and $g(\cdot)$ is proximable, i.e.,

$$\text{prox}_{\gamma g}(\boldsymbol{v})=\arg\min_{\boldsymbol{H}}\left(g(\boldsymbol{x})+\frac{1}{2\gamma}\|\boldsymbol{x}-\boldsymbol{v}\|^2\right) \quad (12)$$

is easy to solve. Note that the step-size $0<\gamma\le 1/L_f$ is small when the Lipschitz constant $L_f$ is large.

**Algorithm 1** The standard PG method [15]

Set $\boldsymbol{x}_0$ and $0<\gamma\le\frac{1}{L_f}$ ($L_f$: the Lipschitz constant of $\nabla f$).
**for** $k=0,1...$ **do**
  Update $\boldsymbol{x}_{k+1}\leftarrow\text{prox}_{\gamma g}(\boldsymbol{x}_k-\gamma\nabla f(\boldsymbol{x}_k))$.

A naive application of the PG method to solve (6) would define $f(\boldsymbol{H},\boldsymbol{x})=\frac{1}{2}\|\boldsymbol{s}-\mathcal{P}_\Omega\boldsymbol{x}\|^2+\frac{\beta}{2}\|\boldsymbol{H}-\mathcal{H}\boldsymbol{x}\|_F^2+\frac{\alpha}{2}\|\boldsymbol{x}\|^2$. The Lipschitz constant of $\nabla f$ is $1+\beta\min(n_1,n_2)+\alpha=O(\beta n)$ where $n_1$ and $n_2$ come from the Hankel mapping (3).

We take a different approach where the objective function involves only a single variable $\boldsymbol{H}$:

$$F(\boldsymbol{H}):=f(\boldsymbol{H})+g(\boldsymbol{H}), \text{ where} \quad (13)$$

$$f(\boldsymbol{H}):=\min_{\boldsymbol{x}}\frac{1}{2}\|\boldsymbol{s}-\mathcal{P}_\Omega\boldsymbol{x}\|^2+\frac{\beta}{2}\|\boldsymbol{H}-\mathcal{H}\boldsymbol{x}\|_F^2+\frac{\alpha}{2}\|\boldsymbol{x}\|^2, \quad (14)$$

$$g(\boldsymbol{H}):=\delta(\text{rank}(\boldsymbol{H})\le r). \quad (15)$$

The minimizer of the inner optimization in (14) is given by

$$\boldsymbol{x}^\star=\mathcal{L}(\mathcal{P}^*_\Omega\boldsymbol{s}+\beta\mathcal{H}^*\boldsymbol{H}), \text{ where } \mathcal{L}=(\alpha\mathcal{I}+\mathcal{P}^*_\Omega\mathcal{P}_\Omega+\beta\mathcal{H}^*\mathcal{H})^{-1}. \quad (16)$$

An important feature of the formulation (13) is that the Lipschitz constant of $\nabla f(\boldsymbol{H})$ is upper bounded by $\beta$ which is dimension independent, as shown in Proposition 2 below. It allows the corresponding proximal gradient method to take a larger step-size and converge faster for large scale problems.

**Proposition 2.** *Denote the Lipschitz constant of* $\nabla f(\boldsymbol{H})$ *by* $L_f$. *It holds that* $L_f\le\beta$.

*Proof Sketch.* Plug (16) into (14). It holds that

$$\nabla f(\boldsymbol{H})=-\beta\mathcal{H}\mathcal{L}\mathcal{P}^*_\Omega\boldsymbol{s}-\beta^2\mathcal{H}\mathcal{L}\mathcal{H}^*\boldsymbol{H}+\beta\boldsymbol{H}. \quad (17)$$

The Hessian matrix of $f(\boldsymbol{H})$ is then $\beta(\mathcal{I}-\beta\mathcal{H}\mathcal{L}\mathcal{H}^*)$. As

$$0\le\|\beta\mathcal{H}\mathcal{L}\mathcal{H}^*\|=\|\beta\mathcal{H}(\alpha\mathcal{I}+\mathcal{P}^*_\Omega\mathcal{P}_\Omega+\beta\mathcal{H}^*\mathcal{H})^{-1}\mathcal{H}^*\|<1, \quad (18)$$

we can conclude that $L_f\le\beta$.[4] □

Our proximal gradient step is detailed in Algorithm 2. It uses $1/\beta-\eta$ as the minimum value (initial value) for the step-size and employs a back-tracking process to allow an even larger step-size. The small positive constant $\eta$ is added to ensure sufficient decrease for nonconvex proximal mapping.

**Algorithm 2** Line search for step size

Set constants $c\in(0,1)$, $\alpha=\frac{1}{1000}$,$\eta=\frac{1}{1000\beta}$, and $\gamma\leftarrow\frac{1}{\beta}-\eta$.
$l_k\leftarrow\arg\min_i\{F(\boldsymbol{H}_k^i):i=0,1,\cdots\}$ ;
where, $\boldsymbol{H}_k^i\leftarrow\text{prox}_{\frac{\gamma}{c^i}g}\left(\boldsymbol{H}_k^0-\frac{\gamma}{c^i}\nabla f(\boldsymbol{H}_k^0)\right)$ ;
**Output:** $\boldsymbol{H}_k\leftarrow\boldsymbol{H}_k^{l_k}$ ;

Our PG step also enjoys efficient computations. In our PG step, the key computations involve the evaluations of $\nabla f(\boldsymbol{H})$ in (17), and the proximal operator

$$\text{prox}_{\frac{\gamma}{c^i}g}\left(\boldsymbol{H}_k-\frac{\gamma}{c^i}\nabla f(\boldsymbol{H}_k)\right)=\mathcal{T}_r\left(\boldsymbol{H}_k-\frac{\gamma}{c^i}\nabla f(\boldsymbol{H}_k)\right), \quad (19)$$

where $\mathcal{T}_r(\cdot)$ is the truncated SVD operator [26]. By applying the Lanczos method for truncated SVD [25], it can be shown the computation complexity of the PG step is $O(r^2n+rn\log n)$. The detailed computation procedure and complexity analysis will be presented in the journal version of this paper but are omitted here due to space constraints.

[3] the gradient $\nabla f$ is Lipschitz continuous and the corresponding Lipschitz constant is denoted by $L_f$

[4] The operator norm is used in (18). It is the maximum eigenvalue of the corresponding matrix representation.

### 3.3. PPG: Convergence

The overall PPG algorithm is summarized in Algorithm 3. As analyzed above, the complexity of each iteration is $O(r^3 n \log n + r^4 n)$.

**Algorithm 3** PPG for spectrally sparse signal reconstruction

**Input:** $\boldsymbol{s}$, $\beta$. Set $\gamma \leftarrow \frac{1}{\beta} - \eta$, $\boldsymbol{H}_0 \leftarrow \mathcal{T}_r \mathcal{H} \boldsymbol{s}$

**while** not converged **do**

Compute $\boldsymbol{H}_{k,\frac{1}{2}} \leftarrow \boldsymbol{U}_k \boldsymbol{\Sigma}^{\star}{}_k \boldsymbol{V}_k^{\mathsf{H}}$ by subspace projection $\boldsymbol{H}_k$ ;

Compute $\boldsymbol{H}_{k+1}^0 \leftarrow \mathcal{T}_r\Big(\boldsymbol{H}_{k,\frac{1}{2}} - \gamma \nabla f(\boldsymbol{H}_{k,\frac{1}{2}})\Big)$ ;

Line search for $\boldsymbol{H}_{k+1}^0$ in Algorithm 2 to get $\boldsymbol{H}_{k+1}$ ;

**Output:** $\{x^{\dagger} \leftarrow \mathcal{L}(\mathcal{P}_{\Omega}^{*}\boldsymbol{s} + \beta \mathcal{H}^{*} \boldsymbol{H}^{\dagger})\}$

The PPG method is guaranteed to converge. It is straightforward to see that the objective function (6) decreases after each step of the iterations. Note that the step-size of our proximal gradient step is at least the constant $1/\beta - \eta$. Moreover, Proposition 2 illustrated $\beta \geq L_f$, which supports a sufficient decrease as shown in [27]. Therefore, the convergence of the developed PPG method is guaranteed.

## 4. SIMULATIONS

In this section, numerical simulations are performed to demonstrate the fast convergence rate and the overall computational efficiency (in wall-clock time) of the PPG method. In all the tests, the ground-truth SSS and their partial observations are generated as follows. The frequencies $f_k$ of SSS are randomly generated from the uniform distribution on $[0,1)$. The complex amplitudes $b_k = |b_k| e^{j\phi_k}$ where the modulus $|b_k| = 1 + 10^{0.5c}$ and $c$ is randomly sampled from the uniform distribution on $[0,1]$, and the phase $\phi_k$ is randomly generated from the uniform distribution on $[0,2\pi)$. For a given sample size $m = pn$, the index set of observed entries $\Omega$ is randomly taken from the uniform distribution on $\binom{[n]}{m}$. In the tests, our PPG method is compared against the benchmark nonconvex algorithms including FIHT [21], PGD [20], and the naive PG algorithm [15] (without the subspace projection step), and two variations of the proposed PPG method, i.e., LPG (PPG without the subspace projection step) and PGG (PPG without the back-tracking line search).

Figure 1 compares the convergence rates of algorithms. The recovery root mean square error (RMSE) is depicted as a function of the number of iterations. The results demonstrate that the proposed PPG converges the fastest. Furthermore, by comparing the proposed PPG against PG, LPG, and PGG, the benefits of various steps in the PPG become explicit.

Table 1 compares both the convergence rate and the wall-clock time of the algorithms in some challenging scenarios: the sampling ratio $p$ is low, the Hankel enforcement parameter $\beta > 0$ is medium or large, and the SSS $\boldsymbol{x}$ is a multi-dimensional array which leads to a large multilevel Hankel block matrix [24]. The results show that the convergence rate of our PPG remains the fastest. Although the complexity of each iteration goes up for the purpose of fast convergence, the computational efficiency (wall-clock time) of the overall algorithm is still the best compared to the benchmark algorithms in the literature. It is interesting to observe that in the case FIHT diverges.

**Fig. 1**. Convergence rate comparison: $n = 400$, $(n_1, n_2) = (200,201)$, $p = 0.3$, and $r = 15$.

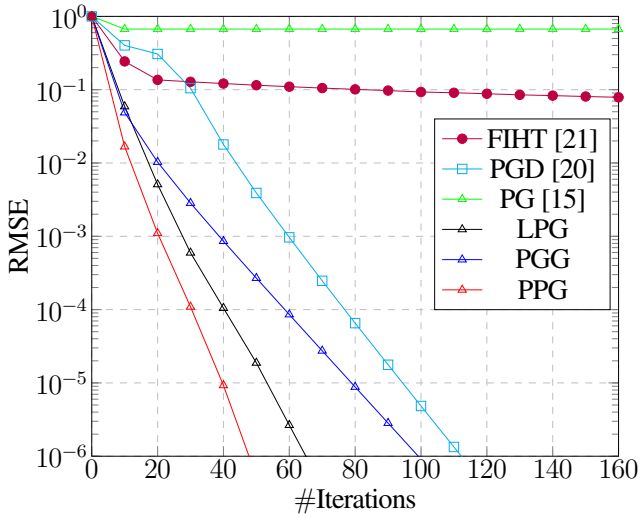


**Table 1**. Convergence rate and wall-clock time comparisons: the SSS $\boldsymbol{X} \in \mathbb{C}^{16\times16\times16}$ is a multi-dimensional array, the resulting $\boldsymbol{H} \in \mathbb{C}^{729\times512}$ is a level-3 block Hankel matrix, $p = 0.05$, $r = 5$, and the stopping criterion is chosen as $(\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\| / \|\boldsymbol{x}_k\|) \leq 10^{-3}$. FIHT is not shown in the table as it does not converge.

| $\beta$ | 1 | | | 10 | | |
|---|---|---|---|---|---|---|
| Criterion | RMSE | Iter | Time(s) | RMSE | Iter | Time(s) |
| PGD [20] | 9.5e-2 | 166 | 8.5e-1 | 5.6e-1 | 710 | 3.8 |
| PG [15] | 9.5e-1 | 108 | 6.7e-1 | 9.5e-1 | 124 | 7.3e-1 |
| LPG | 6.4e-2 | 265 | 8.1 | 9.4e-1 | 37 | 2.2 |
| PGG | 2.9e-1 | 10 | 3.1e-1 | 3.0e-1 | 11 | 3.8e-1 |
| PPG | 2.6e-1 | 8 | 3.6e-1 | 2.8e-1 | 8 | 3.9e-1 |

## 5. CONCLUSIONS

In this paper, a projected proximal gradient (PPG) method is developed for spectrally sparse signal recovery. This method is based on a nonconvex optimization formulation involving a low-rank Hankel matrix. The key novelty is an iterative process involving two steps per iteration: the first is with respect to all variables except the column and row sub-spaces of the Hankel matrix; and the second is to update the column and row sub-spaces together with all other variables. Each step is carefully designed by employing the underlying structures. The PPG method is guaranteed to converge and is numerically demonstrated to be faster than several benchmark algorithms, making it more suitable for large scale problems in practice.

## 6. REFERENCES


[1] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[2] L. Schermelleh, R. Heintzmann, and H. Leonhardt, "A guide to super-resolution fluorescence microscopy," *Journal of Cell Biology*, vol. 190, no. 2, pp. 165–175, 2010.

[3] L. C. Potter, E. Ertin, J. T. Parker, and M. Cetin, "Sparsity and compressed sensing in radar imaging," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1006–1020, 2010.

[4] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, 2009.

[5] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.

[6] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.

[7] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, 1986.

[8] G. R. de Prony, "Essai experimental et analytique: Sur les lois de la dilatabilite des fluides elastique et sur celles de la force expansive de la vapeur de l'eau et de la vapeur de l'alkool, a differentes temperatures," *J. de l'Ecole Polytechnique*, 1795.

[9] Y. Hua, "Estimating two-dimensional frequencies by matrix enhancement and matrix pencil," *IEEE Trans. Signal Process.*, vol. 40, no. 9, pp. 2267–2280, 1992.

[10] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on pure and applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.

[11] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7465–7490, 2013.

[12] Y. Chen and Y. Chi, "Spectral compressed sensing via structured matrix completion," in *International Conference on Machine Learning*. PMLR, 2013, pp. 414–422.

[13] M. Fazel, T. K. Pong, D. Sun, and P. Tseng, "Hankel matrix rank minimization with applications to system identification and realization," *SIAM J. Matrix Anal. Appl*, vol. 34, no. 3, pp. 946–977, 2013.

[14] W. Xu, J. Yi, S. Dasgupta, J.-F. Cai, M. Jacob, and M. Cho, "Seperation-free super-resolution from compressed measurements is possible: an orthonormal atomic norm minimization approach," in *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2018, pp. 76–80.

[15] N. Parikh, S. Boyd *et al.*, "Proximal algorithms," *Foundations and trends® in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[16] S. Burer and R. D. Monteiro, "Local minima and convergence in low-rank semidefinite programming," *Mathematical programming*, vol. 103, no. 3, pp. 427–444, 2005.

[17] X. Zhang, Y. Liu, and W. Cui, "Spectrally sparse signal recovery via Hankel matrix completion with prior information," *IEEE Trans. Signal Process.*, vol. 69, pp. 2174–2187, 2021.

[18] J. A. Cadzow, "Signal enhancement-a composite property mapping algorithm," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 1, pp. 49–62, 1988.

[19] J. Ying, H. Lu, Q. Wei, J.-F. Cai, D. Guo, J. Wu, Z. Chen, and X. Qu, "Hankel matrix nuclear norm regularized tensor completion for $n$-dimensional exponential signals," *IEEE Trans. Signal Process.*, vol. 65, no. 14, pp. 3702–3717, 2017.

[20] J.-F. Cai, T. Wang, and K. Wei, "Spectral compressed sensing via projected gradient descent," *SIAM Journal on Optimization*, vol. 28, no. 3, pp. 2625–2653, 2018.

[21] ——, "Fast and provable algorithms for spectrally sparse signal reconstruction via low-rank hankel matrix completion," *Appl. Comput. Harmon. Anal.*, vol. 46, no. 1, pp. 94–121, 2019.

[22] H. Wang, J.-F. Cai, T. Wang, and K. Wei, "Fast cadzow's algorithm and a gradient variant," *Journal of Scientific Computing*, vol. 88, no. 2, pp. 1–21, 2021.

[23] J. Gillard and K. Usevich, "Hankel low-rank approximation and completion in time series analysis and forecasting: a brief review," *arXiv preprint arXiv:2206.05103*, 2022.

[24] L. Lu, W. Xu, and S. Qiao, "A fast SVD for multilevel block Hankel matrices with minimal memory storage," *Numerical Algorithms*, vol. 69, no. 4, pp. 875–891, 2015.

[25] G. H. Golub and C. F. Van Loan, *Matrix computations*. JHU press, 2013.

[26] N. Antonello, L. Stella, P. Patrinos, and T. van Waterschoot, "Proximal gradient algorithms: Applications in signal processing," *arXiv preprint arXiv:1803.01621*, 2018.

[27] H. Li and Z. Lin, "Accelerated proximal gradient methods for nonconvex programming," *Advances in neural information processing systems*, vol. 28, 2015.