SPARTA: SPARSE PHASE RETRIEVAL VIA TRUNCATED AMPLITUDE FLOW

Gang Wang[1,2], *Georgios B. Giannakis*[1], *Jie Chen*[2], *and Mehmet Akçakaya*[1]

[1] ECE Dept. and Digital Tech. Center, Univ. of Minnesota, Minneapolis, MN 55455, USA
[2] School of Automation, Beijing Institute of Technology, Beijing 100081, China
E-mails: {gangwang,georgios,akcakaya}@umn.edu; chenjie@bit.edu.cn

ABSTRACT

A linear-time algorithm termed *SPARse Truncated Amplitude flow* (SPARTA) is developed for the phase retrieval (PR) of sparse signals. Upon formulating the sparse PR as a nonconvex empirical loss minimization task, SPARTA emerges as an iterative solver consisting of two components: s1) a sparse orthogonality-promoting initialization leveraging support recovery and principal component analysis; and, s2) a series of refinements by hard thresholding based truncated gradient iterations. SPARTA is simple, scalable, and fast. It recovers any k-sparse n-dimensional signal ($k \ll n$) of large enough minimum (in modulus) nonzero entries from about $k^2 \log n$ measurements with high probability; this is achieved at computational complexity of order $k^2 n \log n$, improving upon the state-of-the-art by at least a factor of k. SPARTA is robust against bounded additive noise. Simulated tests corroborate the merits of SPARTA relative to existing alternatives.

***Index Terms*—** Nonconvex optimization, support recovery, hard thresholding, linear convergence.

1. INTRODUCTION

Phase retrieval (PR) refers to recovering a signal only from the magnitude of its Fourier (or any linear) transform. Such a task emerges in various science and engineering applications ranging from X-ray crystallography, microscopy, to optics as well as coherent diffraction imaging, where optical detectors record only the light intensity but not the phase. Oftentimes, the underlying signals are naturally sparse [1]. Enforcing sparsity constraints can also ensure uniqueness of the discretized one-dimensional PR [2]. Different types of measurement transforming systems have been employed, e.g., over-sampling Fourier, short-time Fourier, random Gaussian, and coded diffraction patterns, to name a few; see [1] for a contemporary review.

Past phase retrieval approaches can be grouped as convex and nonconvex ones. The latter include the alternating projection algorithms such as Gerchberg-Saxon [3] and Fienup [2], AltMinPhase [4], (S)TAF [5, 6, 7], PRIME [8, 9], and the Wirtinger flow (WF) variants [10, 11, 12] as well as trust-region methods [13]. The convex ones rely on matrix lifting to obtain semidefinite programming (SDP) based solvers such as PhaseLift [14]. The SDP, AltMinPhase, and WF recovery methods have been extended to PR of sparse inputs leading to solvers abbreviated with CPRL [15], SparseAltMinPhase (SAMP) [4], and thresholded WF (TWF) [16]. Specifically, CPRL accounts for the sparsity using the ℓ_1 convex proxy of ℓ_0-(pseudo) norm. The latter two are two-stage iterative alternatives involving a (sparse) initialization and successive refinements of the initialization. The greedy GESPAR approach is based on fast 2-opt local search [17]. Based on noiseless Gaussian random measurements, CPRL recovers any k-sparse n-dimensional signal exactly from $\mathcal{O}(k^2 \log n)$ measurements at computational complexity $\mathcal{O}(n^3)$ [18]; while SAMP and TWF require $\mathcal{O}(k^2 \log n)$ measurements and incur complexity $\mathcal{O}(k^3 n \log n)$ [4, 16].

Building on SAMP and TWF, we develop a novel linear-time sparse PR algorithm, which we call *SPARse Truncated Amplitude flow* (SPARTA). Adopting an amplitude-based nonconvex formulation of sparse PR, SPARTA is a two-stage iterative solver: Stage one first estimates the support of the underlying signal, and solves a PCA with power iterations restricted on the estimated support; while the second stage iteratively refines the initialization with successive hard thresholding based truncated gradient iterations. Both stages are conceptually simple, scalable, and fast. Further, SPARTA provably recovers any k-sparse signal $\boldsymbol{x} \in \mathbb{R}^n/\mathbb{C}^n$ ($k \ll n$) with minimum (in modulus) nonzero entries on the order of $(1/\sqrt{k})\|\boldsymbol{x}\|_2$ from $\mathcal{O}(k^2 \log n)$ measurements at computational complexity $\mathcal{O}(k^2 n \log n)$, which improves upon the state-of-the-art by at least a factor of k. This advantage is crucial in large-scale imaging applications, where the basis factor $n \log n$ is large typically on the order of millions. Simulated tests demonstrate markedly improved recovery performance and speedups over the state-of-the-art algorithms.

The rest of the paper is outlined as follows. Section 2 reviews the sparse PR and known necessary and sufficient conditions for uniqueness. Section 3 details the two stages of the SPARTA algorithm, which together with performance analysis is summarized in Section 4. Finally, numerical results are provided in Section 5, and conclusions are drawn in Section 6.

Work in this paper was supported in part by NSF grants 1500713 and 1514056.

2. SPARSE PHASE RETRIEVAL

Succinctly stated, the sparse PR (a.k.a., compressive PR) task amounts to solving for a sparse $\boldsymbol{x} \in \mathbb{R}^n$ (or $\mathbb{C}^n$) a system of phaseless quadratic equations of the form

$$\psi_i = |\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle|, \;\; i = 1, \ldots, m, \quad \text{subject to} \quad \|\boldsymbol{x}\|_0 = k \;\; (1)$$

where $\{\psi_i\}_{i=1}^m$ are the observed data, $\boldsymbol{a}_i$ known sampling vectors, and the sparsity level $k \ll n$ is assumed to be known. The data can be given in squared (a.k.a., intensity) form as $y_i := \psi_i^2 = |\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle|^2$. For concreteness, we focus on the real-valued Gaussian design vectors $\boldsymbol{a}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$, that are assumed independently and identically distributed (i.i.d.). It has been established that $m = 2k$ generic measurements as in (1) are necessary and sufficient for uniquely determining a k-sparse solution in the real case, and $m \geq 4k-2$ are sufficient in the complex case [19]. In the noisy case, $\mathcal{O}(k\log(n/k))$ measurements suffice [20]. Assuming existence of a unique k-sparse solution (up to a global sign), our goal is to design simple yet effective algorithms to provably reconstruct $\boldsymbol{x}$ from a small number (far less than n) of phaseless equations.

Adopting the least-squares criterion (which coincides with the maximum likelihood one when the additive noise is white Gaussian), the task of recovering a k-sparse solution from phaseless equations reduces to that of minimizing the following amplitude-based empirical loss function

$$\underset{\|\boldsymbol{z}\|_0 = k}{\text{minimize}} \;\; \frac{1}{2m}\sum_{i=1}^{m}\left(\psi_i - \left|\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{z}\right|\right)^2 \qquad (2)$$

where $(\cdot)^{\mathcal{T}}$ stands for transposition. It is clear that both the objective function and the constraint are nonconvex, rendering problem (2) *NP-hard* in general and hence computationally intractable. It is worth mentioning that TWF deals with the intensity-based counterpart of (2), which was experimentally shown to be less effective than the amplitude-based one even when no sparsity was exploited [6]. Although focusing on a formulation similar to (2), SAMP first estimates the support of the underlying signal, and performs standard PR of signals with reduced dimension $k \ll n$. Relying on alternating minimization, it solves a series of least-squares problems, hence requiring matrix inversion at each iteration. Further, numerical tests suggest that a very large number of measurements are required to estimate the support exactly. Once wrong, SAMP confining the PR on the estimated support would be impossible to recover the underlying signal. On the other hand, an adaptive thresholding procedure that maintains only certain largest entries per iteration during the gradient refinement stage turns out to be effective [16]. Both SAMP and TWF were based on the spectral initialization, which was later shown to be less accurate and robust than the orthogonality-promoting initialization [6].

Broadening our TAF approach and TWF, the present paper puts forth a novel linear-time solver of (2) that operates in two stages: 1) a sparse orthogonality-promoting initialization attainable by solving a PCA with simple power iterations on an estimated support of the underlying sparse signal; and, 2) which is refined by means of scalable truncated gradient iterations, followed by a hard thresholding per iteration to set all entries to zero except for the k ones of largest magnitudes.

3. SPARTA ALGORITHM

In this section, the two stages of SPARTA will be depicted in detail. To start, define the distance from any estimate $\boldsymbol{z}$ to the solution set $\{\pm\boldsymbol{x}\}$ as follows: $\text{dist}(\boldsymbol{z}, \boldsymbol{x}) := \min \|\boldsymbol{z} \pm \boldsymbol{x}\|_2$ for real-valued signals, where $\|\cdot\|_2$ is the Euclidean norm. Define the indistinguishable global phase constant in the real case as

$$\phi(\boldsymbol{z}) := \begin{cases} 0, & \|\boldsymbol{z} - \boldsymbol{x}\|_2 \leq \|\boldsymbol{z} + \boldsymbol{x}\|_2, \\ \pi, & \text{otherwise.} \end{cases} \qquad (3)$$

Hereafter, assume $\boldsymbol{x}$ to be the solution to (1) with $\phi(\boldsymbol{z}) = 0$; otherwise, one can replace $\boldsymbol{z}$ by $\boldsymbol{z}\mathrm{e}^{i\phi}$, but the constant phase shift shall be dropped for notational brevity. Assume without loss of generality $\|\boldsymbol{x}\|_2 = 1$.

3.1. Sparse Orthogonality-promoting Initialization

The orthogonality-promoting initialization in [5, 6] builds upon a fundamental characteristic in high-dimensional spaces, where random vectors are almost always nearly orthogonal to each other. The key idea is approximating the unknown $\boldsymbol{x}$ by a vector $\boldsymbol{z}^0$ most orthogonal to a judiciously selected subset of sensing vectors $\{\boldsymbol{a}_i\}_{i\in\mathcal{I}^0}$, where $\mathcal{I}^0 \subseteq [m] := \{1, 2, \ldots, m\}$ is some index set to be designed. It is known that the orthogonality between vectors can be captured by their (squared normalized) inner-product given as $(\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{x})^2/(\|\boldsymbol{a}_i\|^2\|\boldsymbol{x}\|^2)$. Thus, $\mathcal{I}^0$ collects the indices of $\boldsymbol{a}_i$'s having the $|\mathcal{I}^0|$-smallest normalized inner-products with $\boldsymbol{x}$ [6]. Mathematically, the orthogonality-promoting initialization method amounts to solving a smallest eigenvalue problem

$$\underset{\|\boldsymbol{z}\|_2=1}{\text{minimize}} \;\; \boldsymbol{z}^{\mathcal{T}}\boldsymbol{Y}\boldsymbol{z} := \boldsymbol{z}^{\mathcal{T}}\Big(\frac{1}{|\mathcal{I}^0|}\sum_{i\in\mathcal{I}^0}\frac{\boldsymbol{a}_i\boldsymbol{a}_i^{\mathcal{T}}}{\|\boldsymbol{a}_i\|_2^2}\Big)\boldsymbol{z} \qquad (4)$$

which entails a full eigen-decomposition of complexity $\mathcal{O}(n^3)$. Upon letting $\overline{\mathcal{I}}^0$ be the complement of $\mathcal{I}^0$ in $[m]$, it holds that $\sum_{i\in\mathcal{I}^0}\frac{\boldsymbol{a}_i\boldsymbol{a}_i^{\mathcal{T}}}{\|\boldsymbol{a}_i\|_2^2} = \sum_{i\in[m]}\frac{\boldsymbol{a}_i\boldsymbol{a}_i^{\mathcal{T}}}{\|\boldsymbol{a}_i\|_2^2} - \sum_{i\in\overline{\mathcal{I}}^0}\frac{\boldsymbol{a}_i\boldsymbol{a}_i^{\mathcal{T}}}{\|\boldsymbol{a}_i\|_2^2}$. Appealing to the following standard concentration result

$$\sum_{i\in[m]}\frac{\boldsymbol{a}_i\boldsymbol{a}_i^{\mathcal{T}}}{\|\boldsymbol{a}_i\|_2^2} \approx \frac{m}{n}\boldsymbol{I}_n, \qquad (5)$$

the smallest eigenvalue problem in (4) can be approximated as a largest eigenvalue (PCA-type) problem

$$\tilde{\boldsymbol{z}}^0 := \arg\max_{\|\boldsymbol{z}\|_2=1} \; \boldsymbol{z}^{\mathcal{T}}\overline{\boldsymbol{Y}}\boldsymbol{z} := \frac{1}{|\overline{\mathcal{I}}^0|}\boldsymbol{z}^{\mathcal{T}}\Big(\sum_{i\in\overline{\mathcal{I}}^0}\frac{\boldsymbol{a}_i\boldsymbol{a}_i^{\mathcal{T}}}{\|\boldsymbol{a}_i\|_2^2}\Big)\boldsymbol{z} \quad (6)$$

which can be solved with a few (e.g., 100) power iterations. When $\|\boldsymbol{x}\| \neq 1$, $\tilde{\boldsymbol{z}}^0$ can be scaled by the norm of $\boldsymbol{x}$ estimated to be $\sqrt{\sum_{i=1}^m y_i/m}$, thus yielding $\boldsymbol{z}^0 = \sqrt{\sum_{i=1}^m y_i/m}\,\tilde{\boldsymbol{z}}^0$.

When $\boldsymbol{x}$ is known to be k-sparse *a priori*, the same rationale as the orthogonality-promoting initialization would lead to

$$\underset{\|\boldsymbol{z}\|_2=1}{\text{minimize}} \;\; \boldsymbol{z}^{\mathcal{T}}\boldsymbol{Y}\boldsymbol{z} \quad \text{subject to } \; \|\boldsymbol{z}\|_0 = k \tag{7}$$

which, however, could not be readily converted to a (sparse) PCA due to a limited number of data samples hardly to validate (5). Instead of coping with (7), our approach is to first estimate the support of the underlying signal with a well-justified rule; followed by power iterations for the PCA problem [cf. (6)] on the estimated support to produce a k-sparse $\tilde{\boldsymbol{z}}^0 \in \mathbb{R}^n$; and, subsequently scaling $\tilde{\boldsymbol{z}}^0$ by $\sqrt{\sum_{i=1}^m y_i/m}$ to yield the k-sparse orthogonality-promoting initialization $\boldsymbol{z}^0$.

The support recovery step is detailed next. To this end, assume that $\boldsymbol{x}$ has support $S \subseteq [n] := \{1, \ldots, n\}$ with $|S| = k$. Consider the random variables $Z_{i,j} = \psi_i^2 a_{i,j}^2$, $j = 1, \ldots, n$. With $\mathbb{E}[a_{i,j}^4] = 3$, $\mathbb{E}[a_{i,j}^2] = 1$, and using the rotational invariance property of Gaussian distributions, one arrives at

$$\begin{aligned}\mathbb{E}[Z_{i,j}] = \mathbb{E}[(\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{x})^2 a_{i,j}^2] &= \mathbb{E}[a_{i,j}^4 x_j^2 + (\boldsymbol{a}_{/j}^{\mathcal{T}}\boldsymbol{x}_{/j})^2 a_{i,j}^2] \\ &= 3x_j^2 + \|\boldsymbol{x}_{/j}\|_2^2, \quad \forall j \in [n] \end{aligned} \tag{8}$$

where $\boldsymbol{x}_{/j} \in \mathbb{R}^{n-1}$ is obtained by removing the j-th entry from $\boldsymbol{x}$; and likewise for $\boldsymbol{a}_{/j} \in \mathbb{R}^{n-1}$. If $j \in S$, then $x_j \neq 0$ and $\mathbb{E}[Z_{i,j}] = \|\boldsymbol{x}\|_2^2 + 2x_j^2$; and $\mathbb{E}[Z_{i,j}] = \|\boldsymbol{x}_{/j}\|_2^2 = \|\boldsymbol{x}\|_2^2$ otherwise. It is now clear that there is a separation in the expected values of $Z_{i,j}$ for $j \in S$ and $j \notin S$. If all $\mathbb{E}[Z_{i,j}]$ values are available, the indices corresponding to the k-largest $\mathbb{E}[Z_{i,j}]$ values recover exactly the support of $\boldsymbol{x}$. Nevertheless, $\mathbb{E}[Z_{i,j}]$ are not available. One has available a number of their realizations instead. Leveraging the strong law of large numbers, the sample average approaches the ensemble one, namely, $\hat{Z}_{i,j} := (1/m)\sum_{i=1}^m Z_{i,j} \to \mathbb{E}[Z_{i,j}]$ as m increases. Hence, the support can be estimated as

$$\hat{S} := \left\{1 \leq i \leq n \middle| \text{indices of top-}k \text{ instances in } \{\hat{Z}_{i,j}\}_{j=1}^n \right\}$$

which will be shown to recover S exactly with high probability when $\mathcal{O}(k^2 \log n)$ measurements are taken and the minimum (in modulus) nonzero entry $x_{\min} := \min_{j \in S} |x_j|$ is on the order of $(1/\sqrt{k})\|\boldsymbol{x}\|_2$.

When the estimated support is exact, one can rewrite $\psi_i = |\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{x}| = |\boldsymbol{a}_{i,\hat{S}}^{\mathcal{T}}\boldsymbol{x}_{\hat{S}}|$, where $\boldsymbol{a}_{i,\hat{S}} \in \mathbb{R}^k$ includes the j-the entry of $\boldsymbol{a}_i$ if and only if $j \in \hat{S}$; and likewise for $\boldsymbol{x}_{\hat{S}} \in \mathbb{R}^k$. Instead of seeking an n-dimensional initialization directly, one can first apply the orthogonality-promoting initialization steps in (4)-(6) on the data $\{(\boldsymbol{a}_{i,\hat{S}}, \psi_i)\}_{i=1}^m$ to produce a k-dimensional vector

$$\tilde{\boldsymbol{z}}_{\hat{S}}^0 := \arg \max_{\|\boldsymbol{z}_{\hat{S}}\|_2=1} \; \frac{1}{|\bar{\mathcal{I}}^0|}\boldsymbol{z}_{\hat{S}}^{\mathcal{T}}\Big(\sum_{i \in \bar{\mathcal{I}}^0} \frac{\boldsymbol{a}_{i,\hat{S}}\boldsymbol{a}_{i,\hat{S}}^{\mathcal{T}}}{\|\boldsymbol{a}_{i,\hat{S}}\|_2^2}\Big)\boldsymbol{z}_{\hat{S}} \tag{9}$$

and reconstruct a k-sparse n-dimensional initialization as $\tilde{\boldsymbol{z}}^0$ by zero-padding $\tilde{\boldsymbol{z}}_{\hat{S}}^0$ at entries with indices not in $\hat{S}$. Likewise for the case of $\|\boldsymbol{x}\| \neq 1$, set $\boldsymbol{z}^0 = \sqrt{\sum_{i=1}^m y_i/m}\,\tilde{\boldsymbol{z}}^0$.

Algorithm 1 SPARse Truncated Amplitude flow (SPARTA)

1: **Input:** Data $\{(\boldsymbol{a}_i; \psi_i)\}_{i=1}^m$ and sparsity level k; maximum number of iterations $T = 1,000$; step size $\mu = 1$, truncation thresholds $|\bar{\mathcal{I}}^0| = \lceil \frac{1}{6}m \rceil$, [1] and $\gamma = 0.7$.
2: **Set** $\hat{S}$ to include indices corresponding to the k-largest instances in $\{(1/m)\sum_{i=1}^m \psi_i^2 a_{i,j}^2\}_{j=1}^n$.
3: **Evaluate** $\bar{\mathcal{I}}^0$ to consist of indices of the top-$|\bar{\mathcal{I}}^0|$ values in $\Big\{\psi_i/\|\boldsymbol{a}_{i,\hat{S}}\|_2\Big\}_{i=1}^m$; and compute the principal eigenvector $\tilde{\boldsymbol{z}}_{\hat{S}}^0$ of $\frac{1}{|\bar{\mathcal{I}}^0|}\sum_{i \in \bar{\mathcal{I}}^0} \frac{\boldsymbol{a}_{i,\hat{S}}\boldsymbol{a}_{i,\hat{S}}^{\mathcal{T}}}{\|\boldsymbol{a}_{i,\hat{S}}\|_2^2}$ using 100 power iterations.
4: **Initialize** $\boldsymbol{z}^0$ as $\sqrt{\sum_{i=1}^m \psi_i^2/m}\,\tilde{\boldsymbol{z}}^0$, where $\tilde{\boldsymbol{z}}^0$ is obtained by augmenting $\tilde{\boldsymbol{z}}_{\hat{S}}^0$ with zeros at entries not in $\hat{S}$.
5: **Loop: For** $t = 0$ **to** $T - 1$

$$\boldsymbol{z}^{t+1} = \mathcal{T}_k\bigg(\boldsymbol{z}^t - \frac{\mu}{m}\sum_{i \in \mathcal{I}^{t+1}}\Big(\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{z}^t - \psi_i\frac{\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{z}^t}{|\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{z}^t|}\Big)\boldsymbol{a}_i\bigg)$$

where $\mathcal{I}^{t+1} := \big\{1 \leq i \leq m \big| |\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{z}^t| \geq \psi_i/(1+\gamma)\big\}$.
6: **Output:** $\boldsymbol{z}^T$.

3.2. Thresholded Truncated Gradient Stage

Upon obtaining a sparse orthogonality-promoting initialization $\boldsymbol{z}^0$, our approach to solving problem (2) amounts to iteratively refining $\boldsymbol{z}^0$ with truncated gradient iterations followed by a k-sparse hard thresholding per iteration, namely,

$$\boldsymbol{z}^{t+1} := \mathcal{T}_k\big(\boldsymbol{z}^t - \mu\nabla\ell_{\mathrm{tr}}(\boldsymbol{z}^t)\big), \quad t = 0, 1, \ldots \tag{10}$$

where t is the iteration index, $\mu > 0$ a constant step size, and $\mathcal{T}_k(\boldsymbol{u})$ a hard thresholding operation setting all entries in $\boldsymbol{u} \in \mathbb{R}^n$ to zero except for k entries of the largest magnitudes. The truncated gradient $\nabla\ell_{\mathrm{tr}}(\boldsymbol{z}^t)$ is given by

$$\nabla\ell_{\mathrm{tr}}(\boldsymbol{z}^t) := \frac{1}{m}\sum_{i \in \mathcal{I}^{t+1}}\Big(\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{z}^t - \psi_i\frac{\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{z}^t}{|\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{z}^t|}\Big)\boldsymbol{a}_i \tag{11}$$

where the index set is defined as

$$\mathcal{I}^{t+1} := \big\{1 \leq i \leq m \big| |\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{z}^t|/|\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{x}| \geq 1/(1+\gamma)\big\} \tag{12}$$

for some $\gamma > 0$ depending on the sparsity level k. The truncation rule (12) developed in [6] was shown capable of removing most 'bad' gradient components involving wrongly estimated signs, i.e., $\frac{\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{z}^t}{|\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{z}^t|} \neq \frac{\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{x}}{|\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{x}|}$. Moreover, this regularization maintains only large enough $|\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{z}^t|$ terms, hence protecting the objective function in (2) from being non-differentiable at $\boldsymbol{z}^t$ and simplifying the theoretical analysis [6].

4. MAIN RESULTS

The SPARTA solver is summarized in Algorithm 1 with suggested parameter values. Assuming independent data samples $\{(\boldsymbol{a}_i; \psi_i)\}$, the following result establishes theoretical performance of SPARTA.

[1] The symbol $\lceil\cdot\rceil$ is the ceiling operation returning the smallest integer greater than or equal to the given number.

Theorem 1 (Exact recovery). *Let $\boldsymbol{x} \in \mathbb{R}^n$ be any k-sparse ($k \ll n$) signal vector with the minimum nonzero entry on the order of $(1/\sqrt{k})\|\boldsymbol{x}\|_2$. Consider the measurements $\psi_i = |\boldsymbol{a}_i^T \boldsymbol{x}|$, where i.i.d. $\{\boldsymbol{a}_i\}_{i=1}^m \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_n)$. With a constant step size $\mu \in [\underline{\mu}, \overline{\mu}]$, and a truncation threshold $\gamma = 0.7$, successive estimates of SPARTA in Algorithm 1 satisfy*

$$\text{dist}\left(\boldsymbol{z}^t, \boldsymbol{x}\right) \leq (1/10)\,(1-\nu)^t\,\|\boldsymbol{x}\|_2\,, \quad t = 0, 1, \ldots \quad (13)$$

which holds with probability exceeding $1 - c_1 m \mathrm{e}^{-c_0 k} - 7/m$ provided that $m \geq C_0|\overline{\mathcal{I}}^0| \geq C_1 k^2 \log n$. Here, c_0, c_1, C_0, C_1, $\mu_0 > 0$, and $0 < \nu < 1$ are some universal constants.

Proof of Thm. 1 can be found in our journal version [**?**]. Regarding Thm. 1, three observations are in order.

Remark 1. SPARTA recovers any k-sparse solution $\boldsymbol{x}$ of large enough minimum nonzero entries exactly when $m \geq C_1 k^2 \log n$ for some constant $C_1 > 0$, which coincides with those required by the CPRL, SAMP, and TWF approaches.

Remark 2. SPARTA converges exponentially fast to $\boldsymbol{x}$ with convergence rate independent of n. In other words, fixing any $\epsilon > 0$, after running at most $\log(1/\epsilon)$ iterations (10), the current estimate $\boldsymbol{z}^t$ is at most $\epsilon\|\boldsymbol{x}\|_2$ away from $\boldsymbol{x}$.

Remark 3. SPARTA enjoys low computational complexity $\mathcal{O}(k^2 n \log n)$, and running time of $\mathcal{O}(k^2 n \log n \log(1/\epsilon))$ required to achieve an ϵ-accurate solution, which is proportional to the time $\mathcal{O}(k^2 n \log n)$ taken to read the data.

Besides exact recovery in the absence of noise, it is also worth mentioning that SPARTA can be shown to be robust to additive noise, especially when it has bounded values.

5. SIMULATED TESTS

Numerical tests evaluating performance of SPARTA relative to TAF [6] (no sparsity is exploited) and TWF [16] are presented next. For fairness, all parameters involved in each scheme are set to their default values. The initialization was obtained with 100 power iterations, and was refined by $T = 1,000$ gradient iterations. For reproducibility, the Matlab implementation of SPARTA is publicly available. [2]

The first experiment evaluates the exact recovery performance of various approaches in terms of the empirical success rate over 100 Monte Carlo trials, where a success is declared for a trial if the returned estimate incurs a $\text{Relative error} := \text{dist}(\boldsymbol{z}^T, \boldsymbol{x})/\|\boldsymbol{x}\|_2$ less than 10^{-5}. Curves in Fig. 1 clearly demonstrate markedly improved performance of SPARTA over state-of-the-art alternatives. Even when the exact number of nonzeros in $\boldsymbol{x}$ is unknown, taking k as an upper limit on the theoretically affordable sparsity level (e.g., $\lceil\sqrt{n}\rceil$ when m is about n according to Thm. 1) works well too (see the blue curve). Comparison between TAF and SPARTA shows the advantage of exploiting sparsity in sparse PR settings.

[2] http://www.tc.umn.edu/~gangwang/SPARTA/.

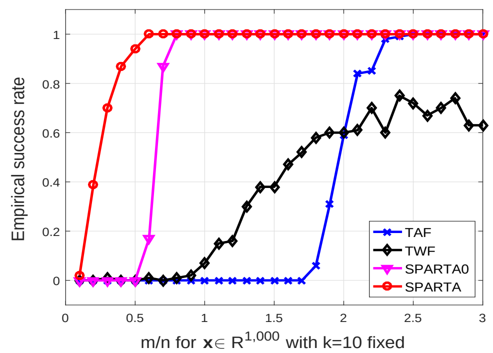

Fig. 1: Empirical success rate versus m/n for $\boldsymbol{x} \in \mathbb{R}^{1,000}$ with 10 nonzero entries using: i) TAF without exploiting sparsity [6]; ii) TWF [16]; iii) SPARTA0 with the exact number of nonzeros unknown, and k taken as an upper limit $\lceil\sqrt{n}\rceil = 32$; and iv) SPARTA with $k = 10$.

The second experiment tests the capability of SPARTA in recovering signals of various sparsity levels. Figure 2 depicts the empirical success rate versus the sparsity level k/n, where k equals the exact number of nonzeros in $\boldsymbol{x}$. Apparently, using $m = n$ magnitude-only measurements (in which TAF would fail), SPARTA significantly outperforms TWF, and it ensures exact recovery of sparse signals with up to about $25 < \sqrt{n} \approx 32$ nonzero entries, hence justifying our analytic results. Regarding running times, SPARTA converges much faster than TWF and TAF in all reported experiments.

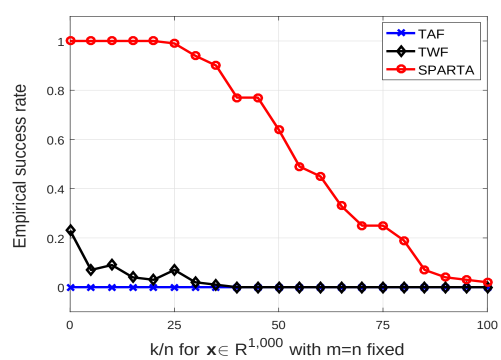

Fig. 2: Empirical success rate versus k/n for $\boldsymbol{x} \in \mathbb{R}^{1,000}$ with $m = n$ fixed using: i) TAF; ii) TWF; and iii) SPARTA.

6. CONCLUDING SUMMARY

This paper developed SPARTA for solving PR of sparse signals, building on two main components: A sparse orthogonality-promoting initialization attainable by solving a PCA with simple power iterations on the estimated support; followed by successive refinements of the initialization via scalable hard thresholding based truncated gradient iterations. Simulated tests corroborate markedly improved performance of SPARTA relative to state-of-the-art algorithms.

7. REFERENCES

[1] K. Jaganathan, Y. C. Eldar, and B. Hassibi, "Phase retrieval: An overview of recent developments," *arXiv:1510.07713*, 2015.

[2] J. R. Fienup, "Phase retrieval algorithms: A comparison," *Appl. Opt.*, vol. 21, no. 15, pp. 2758–2769, Aug. 1982.

[3] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction," *Optik*, vol. 35, pp. 237–246, Nov. 1972.

[4] P. Netrapalli, P. Jain, and S. Sanghavi, "Phase retrieval using alternating minimization," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4814–4826, Sept. 2015.

[5] G. Wang and G. B. Giannakis, "Solving random systems of quadratic equations via truncated generalized gradient flow," in *Adv. Neural Inf. Process. Syst.*, Barcelona, Spain, 2016.

[6] G. Wang, G. B. Giannakis, and Y. C. Eldar, "Solving systems of random quadratic equations via truncated amplitude flow," *arXiv:1605.08285*, 2016.

[7] G. Wang, G. B. Giannakis, and J. Chen, "Solving large-scale systems of random quadratic equations via stochastic truncated amplitude flow," *arXiv:1610.09540*, 2016.

[8] T. Qiu, P. Babu, and D. P. Palomar, "PRIME: Phase retrieval via majorization-minimization," *IEEE Trans. Signal Process.*, vol. 64, no. 19, pp. 5174–5186, Oct. 2016.

[9] T. Qiu and D. Palomar, "Undersampled phase retrieval via majorization-minimization," *arXiv:1609.02842*, 2016.

[10] E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: Theory and algorithms," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1985–2007, Apr. 2015.

[11] Y. Chen and E. J. Candès, "Solving random quadratic systems of equations is nearly as easy as solving linear systems," *Comm. Pure Appl. Math.*, 2016 (to appear).

[12] X. Jiang, S. Rajan, and X. Liu, "Wirtinger flow method with optimal stepsize for phase retrieval," *IEEE Signal Process. Lett.*, vol. 23, no. 11, pp. 1627–1631, Nov 2016.

[13] J. Sun, Q. Qu, and J. Wright, "A geometric analysis of phase retrieval," *arXiv:1602.06664*, 2016.

[14] E. J. Candès, T. Strohmer, and V. Voroninski, "PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming," *Appl. Comput. Harmon. Anal.*, vol. 66, no. 8, pp. 1241–1274, Nov. 2013.

[15] H. Ohlsson, A. Y. Yang, R. Dong, and S. S. Sastry, "CPRL–An extension of compressive sensing to the phase retrieval problem," in *Adv. Neural Inf. Process. Syst.*, Stateline, NV, 2012, pp. 1367–1375.

[16] T. Cai, X. Li, and Z. Ma, "Optimal rates of convergence for noisy sparse phase retrieval via thresholded Wirtinger flow," *Ann. Stat.*, 2016 (to appear).

[17] Y. Shechtman, A. Beck, and Y. C. Eldar, "GESPAR: Efficient phase retrieval of sparse signals," *IEEE Trans. Signal Process.*, vol. 62, no. 4, pp. 928–938, Feb. 2014.

[18] X. Li and V. Voroninski, "Sparse signal recovery from quadratic measurements via convex programming," *SIAM J. Appl. Math.*, vol. 45, no. 5, pp. 3019–3033, Sept. 2013.

[19] M. Akcakaya and V. Tarokh, "Sparse signal recovery from a mixture of linear and magnitude-only measurements," *IEEE Signal Process. Lett.*, vol. 22, no. 9, pp. 1220–1223, Sept. 2015.

[20] Y. C. Eldar and S. Mendelson, "Phase retrieval: Stability and recovery guarantees," *Appl. Comput. Harmon. Anal.*, vol. 36, no. 3, pp. 473–494, May 2014.

[21] G. Wang, L. Zhang, G. B. Giannakis, J. Chen, and M. Akçakaya, "Sparse phase retrieval via truncated amplitude flow," *arXiv:1611.07641*, 2016.