Sparse Phase Retrieval via Iteratively Reweighted Amplitude Flow

Gang Wang*,⋆, Liang Zhang⋆, Georgios B. Giannakis*, and Jie Chen⋆
* Dept. of ECE and Digital Tech. Center, Univ. of Minnesota, Minneapolis, MN 55455, USA
⋆ State Key Lab of Intelligent Control and Decision of Complex Systems,
Beijing Institute of Technology, Beijing 100081, China
E-mails: {gangwang, zhan3523, georgios}@umn.edu, chenjie@bit.edu.cn

***Abstract*—Sparse phase retrieval (PR) aims at reconstructing a sparse signal vector from a few phaseless linear measurements. It emerges naturally in diverse applications, but it is NP-hard in general. Drawing from advances in nonconvex optimization, this paper presents a new algorithm that is termed compressive reweighted amplitude flow (CRAF) for sparse PR. CRAF operates in two stages: Stage one computes an initial guess by means of a new spectral procedure, and stage two implements a few hard thresholding based iteratively reweighted gradient iterations on the amplitude-based least-squares cost. When there are sufficient measurements, CRAF reconstructs the true signal vector exactly under suitable conditions. Furthermore, its sample complexity coincides with that of the state-of-the-art approaches. Numerical experiments showcase improved performance of the proposed approach relative to existing alternatives.**

***Index Terms*—Sparse recovery, spectral initialization, model-based hard thresholding, linear convergence**

I. Introduction

Phase retrieval (PR) refers to the task of reconstructing a signal vector from its phaseless measured linearly transformed entries. It emerges naturally in diverse engineering and physics applications such as X-ray crystallography, microscopy, optics, and coherent diffraction imaging. In such setups however, the physical devices can only measure the density of the incoming light, but not its phase. This missing phase information renders general PR ill-posed.

Several PR approaches have been reported so far. Alternating projection methods were developed in [1], [2], iterative nonconvex schemes in [3]–[11], and convex approaches were suggested in [12]–[15]. We also recently developed a reweighted amplitude flow (RAF) algorithm that benchmarks numerical performance for PR of general signal vectors from random measurements [7]. The signal vectors in many practical scenarios however, are naturally sparse after certain known and deterministic linear transformations have been applied [2], [16]. This prior information can be critical in reducing the number of measurements required by general PR strategies, and has prompted the development of various (block) sparse PR solvers. To this end, the soft-thresholded Wirtinger flow (TWF) [17] and the sparse truncated amplitude flow (SPARTA) [16] were developed; see also [18] for the (block) compressive PR with alternating minimization (CoPRAM).

Building on our precursors in [7], [16], this paper presents a new algorithm that is called *compressive reweighted amplitude flow* (CRAF) for (block)-sparse PR. Generalizing [16], while further accounting for structured sparsity patterns, the amplitude-based (block)-sparse PR problem is formulated, for which a two-stage approach is suggested. Stage one obtains an accurate sparse initialization using a new spectral procedure, which judiciously assigns negative or positive weights to data samples. The second stage refines the initial guess through (model-based) iteratively reweighted hard thresholding iterations. CRAF provably recovers the true signal vector at a linear rate under suitable conditions. Finally, numerical tests corroborate the merits of the proposed approach. Regarding notation, lower- (upper-) case boldface letters stand for column vectors (matrices). Symbol $\|\cdot\|_2$ denotes the Euclidean norm, whereas $\|\cdot\|_0$ for the ℓ_0 (pseudo)-norm counting the number of nonzero entries in a vector. Operator $\lceil\cdot\rceil$ returns the smallest integer greater than or equal to the given scalar. The Gauss error function $\mathrm{erf}(x)$ is given by $\mathrm{erf}(x) := (1/\sqrt{\pi})\int_{-x}^{x} e^{-\tilde{x}^2} d\tilde{x}$.

II. Compressive Phase Retrieval

The compressive PR aims at recovering a sparse signal vector from a few magnitude-only measurements [16], [17]. Mathematically, it can be posed as follows: Given a small set of phaseless linear measurements

$$\psi_i = |\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle|, \quad 1 \le i \le m \tag{1}$$

where $\{\psi_i\}_{i=1}^m$ are the observed amplitude data, and $\{\boldsymbol{a}_i \in \mathbb{R}^n\}_{i=1}^m$ the known sampling vectors, the goal is to recover a (kB)-sparse solution $\boldsymbol{x} \in \mathbb{R}^n$, namely $\|\boldsymbol{x}\|_0 \le kB$ with kB being the known sparsity level. To accommodate also block-sparse signal vectors, the following terminology is useful. Suppose without loss of generality that $\boldsymbol{x}$ is split into N_B blocks $\{\boldsymbol{x}_b\}_{b=1}^{N_B}$, namely $\boldsymbol{x} := [\boldsymbol{x}_1^\mathcal{T} \cdots \boldsymbol{x}_{N_B}^\mathcal{T}]^\mathcal{T}$. For brevity, let $\mathcal{N}_B := \{1, \ldots, N_B\}$ denote the index set of all blocks, and $\mathcal{B}_b$ collect all indices of the entries of $\boldsymbol{x}$ corresponding to the b-th block. Hence, $\mathcal{B}_b \subseteq [n]$ for all $b \in \mathcal{N}_b$, where $[n] := \{1, \ldots, n\}$ consists of all indices of $\boldsymbol{x}$.

Definition 1 (k-block-sparse vectors [19]): The k-block sparse vectors refer to vectors $\boldsymbol{x} = [\boldsymbol{x}_1^\mathcal{T} \cdots \boldsymbol{x}_{N_B}^\mathcal{T}]^\mathcal{T}$ such that $\boldsymbol{x}_b = \boldsymbol{0}$ for all $b \notin \mathcal{S}_B$, where $\mathcal{S}_B$ is a subset of $\mathcal{N}_B$ with cardinality $|\mathcal{S}_B| = k$.

For simplicity, we consider that each block of the signal vector has equal length, that is, $|\mathcal{B}_b| = B$ for all $b \in \mathcal{N}_b$ with

ISBN 978-90-827970-1-5 © EURASIP 2018

$BN_B = n$. Clearly, when $B = 1$, the block-sparse PR boils down to the ordinary or unstructured sparse PR. Accordingly, we will henceforth focus on developing recovery algorithms for a block-sparse signal vector.

Adopting the least-squares criterion, the task of recovering a k-block sparse vector from m amplitude-only measurements can be formulated as [6], [16]

$$\underset{\boldsymbol{z}\in\mathcal{M}_B^k}{\text{minimize}} \;\; \ell(\boldsymbol{z}) := \frac{1}{2m}\sum_{i=1}^{m}\big(\psi_i - |\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{z}|\big)^2 \tag{2}$$

where $\mathcal{M}_B^k$ denotes the set of all k-block-sparse vectors of dimension n. Because of the modulus operator as well as the combinatorial constraint, problem (2) is in general NP-hard, hence computationally intractable.

For concreteness, the real Gaussian model is adopted, which assumes $\boldsymbol{x} \in \mathbb{R}^n$, and independent and identically distributed (i.i.d.) sensing vectors $\boldsymbol{a}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$. When there are enough measurements, it is reasonable to assume existence of a unique (up to a global sign) k-block-sparse solution $\{\pm\boldsymbol{x}\}$ to the quadratic system in (2) [16]. The goal of this paper is to devise simple scalable algorithms to provably recover $\boldsymbol{x}$ from as few intensity-only measurements as possible.

III. Compressive Reweighted Amplitude Flow

This section introduces the two stages of CRAF, namely the initialization and the refinement stages. To begin with , the distance from any estimate $\boldsymbol{z} \in \mathbb{R}^n$ to the solution set $\{\pm\boldsymbol{x}\} \subseteq \mathbb{R}^n$ is defined as $\mathrm{dist}(\boldsymbol{z}, \boldsymbol{x}) := \min\{\|\boldsymbol{z}+\boldsymbol{x}\|_2, \|\boldsymbol{z}-\boldsymbol{x}\|_2\}$.

A. Sparse spectral initialization

A modified spectral initialization that utilizes the information from all available data is discussed first. Relative to past PR initializations in [4]–[7], enhanced numerical performance is achieved by assigning judicious weights to all sampling vectors. Subsequently, the generalization of the new initialization procedure to sparse PR settings is justified.

1) Spectral initialization: Obtaining a good initialization is instrumental in establishing convergence of iterative nonconvex optimization procedures to the global optimum. Consider first the general PR, in which the sparse prior information is not exploited. Similar to prior art, the new initialization entails estimating the norm $\|\boldsymbol{x}\|_2$ as well as the directional vector $\boldsymbol{d} := \boldsymbol{x}/\|\boldsymbol{x}\|_2$ [9]. Regarding the former, it is well documented that $\hat{r} := \sqrt{m^{-1}\sum_{i=1}^{m}\psi_m^2}$ is an unbiased and tightly concentrated estimate of $r := \|\boldsymbol{x}\|_2$ when $m \approx n$ [9]. The challenge remains to estimate $\boldsymbol{d}$, namely obtain a unit vector $\hat{\boldsymbol{d}}$ that is maximally correlated with $\boldsymbol{d}$.

Among different initialization strategies, the procedure proposed in [6] proves successful in achieving excellent numerical performance in estimating $\boldsymbol{d}$; see robustified alternatives in [9]. Nonetheless, the truncation therein discards the useful information carried over in a non-negligible portion of samples. To leverage all data samples, the new procedure obtains the wanted approximation vector as

$$\hat{\boldsymbol{d}} := \arg\max_{\|\boldsymbol{z}\|_2=1} \; \boldsymbol{z}^{\mathcal{T}}\Big(\frac{\lambda^-}{|\mathcal{I}^-|}\sum_{i\in\mathcal{I}^-}\boldsymbol{a}_i\boldsymbol{a}_i^{\mathcal{T}} + \frac{\lambda^+}{|\mathcal{I}^+|}\sum_{i\in\mathcal{I}^+}\boldsymbol{a}_i\boldsymbol{a}_i^{\mathcal{T}}\Big)\boldsymbol{z} \tag{3}$$

where $\lambda^- < 0$ and $\lambda^+ > 0$ are preselected coefficients, and the index sets $\mathcal{I}^- := \{i \in [m] : \psi_i^2 \le \hat{r}^2/2\}$, and $\mathcal{I}^+ := \{i \in [m] : \psi_i^2 \ge \hat{r}^2/2\}$. It is worth pointing out that the judiciously devised index sets satisfy $\mathcal{I} = \mathcal{I}^- \cup \mathcal{I}^+$, suggesting full use of the available data samples. With $\hat{r}$ and $\hat{\boldsymbol{d}}$ at hand, the initial estimate of $\boldsymbol{x}$ can be obtained conveniently as $\boldsymbol{z}^0 := \hat{r}\hat{\boldsymbol{d}}$.

Intuitively, the initialization strategy in (3) can be justified as follows. Using the rotational invariance of $\boldsymbol{a} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, it holds for any thresholds τ_1, $\tau_2 \in [0,\, 1]$ that

$$\begin{aligned}\mathbb{E}\big[\boldsymbol{a}\boldsymbol{a}^{\mathcal{T}}|\langle\boldsymbol{a},\boldsymbol{d}\rangle^2 \le \tau_1\big] \\ = \boldsymbol{I}_n - \boldsymbol{d}\boldsymbol{d}^{\mathcal{T}} + \mathbb{E}[\langle\boldsymbol{a},\boldsymbol{d}\rangle^2|\langle\boldsymbol{a},\boldsymbol{d}\rangle^2 \le \tau_1]\boldsymbol{d}\boldsymbol{d}^{\mathcal{T}}\end{aligned} \tag{4}$$

$$\begin{aligned}\mathbb{E}\big[\boldsymbol{a}\boldsymbol{a}^{\mathcal{T}}|\langle\boldsymbol{a},\boldsymbol{d}\rangle^2 \ge \tau_2\big] \\ = \boldsymbol{I}_n - \boldsymbol{d}\boldsymbol{d}^{\mathcal{T}} + \mathbb{E}[\langle\boldsymbol{a},\boldsymbol{d}\rangle^2|\langle\boldsymbol{a},\boldsymbol{d}\rangle^2 \ge \tau_2]\boldsymbol{d}\boldsymbol{d}^{\mathcal{T}}.\end{aligned} \tag{5}$$

Furthermore, [9, Lemma 3.2] asserts that

$$\mathbb{E}\big[\langle\boldsymbol{a},\boldsymbol{d}\rangle^2|\langle\boldsymbol{a},\boldsymbol{d}\rangle^2 \le \tau_1\big] \le \tau_1/3.$$

Therefore, the smallest eigenvalue of $\mathbb{E}[\boldsymbol{a}\boldsymbol{a}^{\mathcal{T}}|\langle\boldsymbol{a},\boldsymbol{d}\rangle^2 \le \tau_1]$ satisfies

$$\lambda_n\big(\mathbb{E}[\boldsymbol{a}\boldsymbol{a}^{\mathcal{T}}|\langle\boldsymbol{a},\boldsymbol{d}\rangle^2 \le \tau_1]\big) \le \tau_1/3$$

whereas all other eigenvalues are

$$\lambda_i\big(\mathbb{E}[\boldsymbol{a}\boldsymbol{a}^{\mathcal{T}}|\langle\boldsymbol{a},\boldsymbol{d}\rangle^2 \le \tau_1]\big) = 1, \quad 1 \le i \le n-1.$$

Similarly, one can establish the following lower bound for the second term $\mathbb{E}[\langle\boldsymbol{a},\boldsymbol{d}\rangle^2|\langle\boldsymbol{a},\boldsymbol{x}\rangle^2 \ge \tau_2]$ in (5).

Lemma 1: Consider any nonzero signal vector $\boldsymbol{d} \in \mathbb{R}^n$ with $\|\boldsymbol{d}\|_2 = 1$. If $\boldsymbol{a} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, then for any $\tau \ge 0$, it holds that

$$\mathbb{E}\big[\langle\boldsymbol{a},\boldsymbol{d}\rangle^2\,\big|\langle\boldsymbol{a},\boldsymbol{d}\rangle^2 \ge \tau\big] \ge \frac{6 - \tau\,\mathrm{erf}(\sqrt{\tau})}{6 - 3\,\mathrm{erf}(\sqrt{\tau})}. \tag{6}$$

To help understanding the assertion of Lemma 1, taking $\tau = 0.5$ as an example, we find $\mathbb{E}[\langle\boldsymbol{a},\boldsymbol{d}\rangle^2|\langle\boldsymbol{a},\boldsymbol{d}\rangle^2 \ge 0.5] \ge 1.42$ by substituting the inequality $\mathrm{erf}(\sqrt{0.5}) \ge 0.68$ into (6). Hence, it holds that $\lambda_1(\mathbb{E}[\langle\boldsymbol{a},\boldsymbol{d}\rangle^2|\langle\boldsymbol{a},\boldsymbol{d}\rangle^2 \ge 0.5]) \ge 1.42$, and $\lambda_i\big(\mathbb{E}[\boldsymbol{a}\boldsymbol{a}^{\mathcal{T}}|\langle\boldsymbol{a},\boldsymbol{x}\rangle^2 \ge 0.5]\big) = 1,\ 1 \le i \le n-1$. Subsequently, it can be deduced that for $\lambda^- < 0$ and $\lambda^+ > 0$, the largest eigenvalue of $\lambda^-\mathbb{E}[\boldsymbol{a}\boldsymbol{a}^{\mathcal{T}}|\langle\boldsymbol{a},\boldsymbol{d}\rangle^2 \le 0.5] + \lambda^+\mathbb{E}[\boldsymbol{a}\boldsymbol{a}^{\mathcal{T}}|\langle\boldsymbol{a},\boldsymbol{d}\rangle^2 \ge 0.5]$ is greater than or equal to $1.42\lambda^+ + 0.166\lambda^-$, and all others are $\lambda^+ + \lambda^-$. Hence, once the sample average matrix $\frac{\lambda^-}{|\mathcal{I}^-|}\sum_{i\in\mathcal{I}^-}\boldsymbol{a}_i\boldsymbol{a}_i^{\mathcal{T}} + \frac{\lambda^+}{|\mathcal{I}^+|}\sum_{i\in\mathcal{I}^+}\boldsymbol{a}_i\boldsymbol{a}_i^{\mathcal{T}}$ is sufficiently close to its mean, it becomes possible to estimate $\boldsymbol{d}$ with high accuracy according to the matrix perturbation lemma in [20, Corollary 1]. These arguments speak for the effectiveness of the suggested initialization procedure, whereas the next theorem quantifies rigorously the estimation error $\mathrm{dist}(\boldsymbol{z}^0, \boldsymbol{x})$.

Theorem 1: Let $\boldsymbol{z}_0 = \hat{r}\hat{\boldsymbol{d}}$ with $\hat{\boldsymbol{d}}$ obtained from (3). For any given constant $\delta_0 \in (0, 1)$, there exists numerical constants $c_0 > 0$ and C_0 such that the following holds

$$\mathrm{dist}(\boldsymbol{z}_0, \boldsymbol{x}) \le \delta_0\|\boldsymbol{x}\|_2$$

ISBN 978-90-827970-1-5 © EURASIP 2018

with probability exceeding $1-10\exp(-c_0 m)$ when $m \geq C_0 n$.

Proof of Thm. 1 can be found in our journal version [21]. It is worth pointing out that the weak recovery performance of similar procedures has been studied in [22], which only provides guarantee for the case of $n \to \infty$.

2) Support recovery: The initialization procedure in (3) is designed for general vectors $\boldsymbol{x}$, without exploiting the structural information that is present in diverse applications. When the vector is sparse, the required number of data samples to yield an accurate initialization can be reduced [16]. Next, we show how to obtain a sparse initialization based on the procedure discussed in Sec. III-A. Similar to [16], obtaining a sparse initialization entails first estimating the (block)-support of the underlying (block)-sparse signal vectors.

Specifically, define random variables $Z_{i,j} := \psi_i^2 a_{i,j}^2$, $\forall j \in [n]$. According to [16, Eq. (16)], the following holds

$$\mathbb{E}\Big[\sum_{j\in\mathcal{B}_b} Z_{i,j}^2\Big] = \mathbb{E}\Big[\sum_{j\in\mathcal{B}_b} (\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{x})^4 a_{i,j}^4\Big] = 9B\|\boldsymbol{x}\|_2^4 + 24\sum_{j\in\mathcal{B}_b} x_j^4 + 72\|\boldsymbol{x}_b\|^2\|\boldsymbol{x}\|_2^2. \quad (7)$$

If $b \in \mathcal{S}_B$, then $\boldsymbol{x}_b \neq \boldsymbol{0}$, yielding $\mathbb{E}\big[\sum_{j\in\mathcal{B}_b} Z_{i,j}^2\big] > 9B\|\boldsymbol{x}\|_2^4 + 72\|\boldsymbol{x}_b\|^2\|\boldsymbol{x}\|_2^2$ in (7). On the contrary, if $b \notin \mathcal{S}_B$, one has $\boldsymbol{x}_b = \boldsymbol{0}$, yielding $\mathbb{E}\big[\sum_{j\in\mathcal{B}_b} Z_{i,j}^2\big] = 9B\|\boldsymbol{x}\|_2^4$. It is evident that there is a separation of at least $72\|\boldsymbol{x}_b\|^2\|\boldsymbol{x}\|_2^2$ in the expected values of $\sum_{j\in\mathcal{B}_b} Z_{i,j}^2$ for $b \in \mathcal{S}_B$ and $b \notin \mathcal{S}_B$. As long as the gap $72\|\boldsymbol{x}_b\|^2\|\boldsymbol{x}\|_2^2$ is large enough, the (block)-support set $\mathcal{S}_B$ can be recovered exactly in this way.

To estimate the (block)-support $\mathcal{S}_B$ in practice, compute first the so-called block marginals

$$\zeta_b := \sum_{j\in\mathcal{B}_b}\Big(\frac{1}{m}\sum_{i=1}^m \psi_i^2 |a_{i,j}|^2\Big)^2, \quad \forall b \in \mathcal{N}_b$$

which serves as an empirical estimate of $\mathbb{E}\big[\sum_{j\in\mathcal{B}_b} Z_{i,j}^2\big]$. As explained earlier, the larger ζ_b is, the more likely is for the block to be nonzero, namely $\|\boldsymbol{x}_b\|_2 > 0$ [18]. Upon collecting $\{\zeta_b\}_{b=1}^{N_B}$, one can pick the indices associated with the k-largest values in $\{\zeta_b\}_{b=1}^{N_B}$, which form the estimated block-support set $\hat{\mathcal{S}}_B$. Subsequently, an estimate of the support of $\boldsymbol{x}$ denoted as $\hat{\mathcal{S}}$ can be determined as $\hat{\mathcal{S}} := \{i \in \mathcal{B}_b \,|\, \forall b \in \hat{\mathcal{S}}_B\}$.

The support estimation procedure is summarized in Steps 2-4 of Alg. 1. Appealing to [18, Thm. 5.1], Steps 2-4 recover the support of $\boldsymbol{x}$ exactly with probability at least $1 - \frac{6}{m}$ provided that $m \geq C_0' k^2 B \log(mn)$ for some positive constant C_0' and the minimum block

$$x_{\min}^B := \min_{b\in\mathcal{S}_B} \|\boldsymbol{x}_b\|_2^2$$

is on the order of $(1/k)\|\boldsymbol{x}\|_2^2$, namely, $x_{\min}^B = (C_0''/k)\|\boldsymbol{x}\|_2^2$ for some number $C_0'' > 0$.

If the support has been exactly estimated, that is, $\hat{\mathcal{S}} = \mathcal{S}$, one can rewrite $\psi_i = |\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{x}| = |\boldsymbol{a}_{i,\hat{\mathcal{S}}}^{\mathcal{T}}\boldsymbol{x}_{\hat{\mathcal{S}}}|$ for all $i \in [m]$, where $\boldsymbol{a}_{i,\hat{\mathcal{S}}} \in \mathbb{R}^{kB}$ contains entries of $\boldsymbol{a}_i$ whose indices belong to $\hat{\mathcal{S}}$; and similarly for $\boldsymbol{x}_{\hat{\mathcal{S}}} \in \mathbb{R}^k$. Then, the proposed initialization procedure in (3) is applied to the dimensionality-reduced data $\{(\boldsymbol{a}_{i,\hat{\mathcal{S}}}, \psi_i)\}_{i=1}^m$ to obtain

$$\hat{\boldsymbol{d}}_{\hat{\mathcal{S}}} := \max_{\boldsymbol{z}\in\mathbb{R}^{kB}} \boldsymbol{z}^{\mathcal{T}}\left(\frac{\lambda^-}{|\mathcal{I}^-|}\sum_{i\in\mathcal{I}^-}\boldsymbol{a}_{i,\hat{\mathcal{S}}}\boldsymbol{a}_{i,\hat{\mathcal{S}}}^{\mathcal{T}} + \frac{\lambda^+}{|\mathcal{I}^+|}\sum_{i\in\mathcal{I}^+}\boldsymbol{a}_{i,\hat{\mathcal{S}}}\boldsymbol{a}_{i,\hat{\mathcal{S}}}^{\mathcal{T}}\right)\boldsymbol{z}.$$

Subsequently, an estimate of the n-dimensional vector $\boldsymbol{d}$ can be constructed by zero-padding entries of $\hat{\boldsymbol{d}}_{\hat{\mathcal{S}}}$ whose indices are not in $\hat{\mathcal{S}}$.

Algorithm 1 Compressive Reweighted Amplitude Flow

1: **Input:** Data $\{(\boldsymbol{a}_i; \psi_i)\}_{i=1}^m$, block length B, and block sparsity level k; parameters $\lambda^- = -3$, $\lambda^+ = 1$, $\{\beta_i = 0.6\}_{i=1}^m$, $\tau_w = 0.1$; and stepsize $\mu = 1$.

2: **For** $b = 1$ **to** N_B, compute

$$\zeta_b := \sum_{j\in\mathcal{B}_b}\Big(\frac{1}{m}\sum_{i=1}^m \psi_i^2|a_{i,j}|^2\Big)^2.$$

3: **Set** $\hat{\mathcal{S}}_B$ to include indices associated with the k-largest instances in $\{\zeta_b\}_{b=1}^{N_B}$.

4: **Set** $\hat{\mathcal{S}}$ to consist of indices of $\mathcal{B}_b$ for $b \in \hat{\mathcal{S}}_B$.

5: **Compute** the principal eigenvector $\hat{\boldsymbol{d}}_{\hat{\mathcal{S}}} \in \mathbb{R}^{kB}$ of

$$\frac{\lambda^-}{|\mathcal{I}^-|}\sum_{i\in\mathcal{I}^-}\boldsymbol{a}_{i,\hat{\mathcal{S}}}\boldsymbol{a}_{i,\hat{\mathcal{S}}}^{\mathcal{T}} + \frac{\lambda^+}{|\mathcal{I}^+|}\sum_{i\in\mathcal{I}^+}\boldsymbol{a}_{i,\hat{\mathcal{S}}}\boldsymbol{a}_{i,\hat{\mathcal{S}}}^{\mathcal{T}}$$

where $\mathcal{I}^- := \{i \in [m] : \psi_i^2 \leq \hat{r}^2/2\}$ and $\mathcal{I}^+ := \{i \in [m] : \psi_i^2 \geq \hat{r}^2/2\}$ with $\hat{r} := \sqrt{\sum_{i=1}^m \psi_i^2/m}$.

6: **Initialize** $\boldsymbol{z}^0$ as $\hat{r}\tilde{\boldsymbol{d}}$, where $\tilde{\boldsymbol{d}} \in \mathbb{R}^n$ is given by augmenting $\hat{\boldsymbol{d}}_{\hat{\mathcal{S}}}$ in Step 5 with $\tilde{d}_i = 0$ for $i \notin \hat{\mathcal{S}}$.

7: **Loop: For** $t = 0$ **to** $T-1$

$$\boldsymbol{z}^{t+1} = \mathcal{H}_k^B\left(\boldsymbol{z}^t - \frac{\mu}{m}\sum_{i\in[m]} w_i^t\left(\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{z}^t - \psi_i\frac{\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{z}^t}{|\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{z}^t|}\right)\boldsymbol{a}_i\right)$$

where $w_i^t := \max\Big\{\tau_w, \frac{|\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{z}^t|}{|\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{z}^t| + \psi_i\beta_i}\Big\}$.

8: **Output:** $\boldsymbol{z}^T$.

B. Refinement via iteratively reweighted gradient iterations

Upon obtaining an accurate initial point, successive refinements based on reweighted gradient iterations are effected. To account for the block-sparsity structure of the wanted signal vector $\boldsymbol{x}$, the model-based iterative hard thresholding (M-IHT) [19] is invoked. To start, recall that the generalized gradient of the objective function in (2) is [6]

$$\nabla\ell(\boldsymbol{z}) := \frac{1}{m}\sum_{i\in[m]}\Big(\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{z} - \psi_i\frac{\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{z}}{|\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{z}|}\Big)\boldsymbol{a}_i \quad (8)$$

in which the convention $\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{z}/|\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{z}| := 0$ for $|\boldsymbol{a}_i^{\mathcal{T}}\boldsymbol{z}| = 0$ is adopted.

With $t \geq 0$ being the iteration count and $\boldsymbol{z}^0$ being the initial point, the M-IHT algorithm proceeds with the following k-block-sparse hard thresholding, namely

$$\boldsymbol{z}^{t+1} = \mathcal{H}_k^B\Big(\boldsymbol{z}^t - \frac{\mu}{m}\nabla\ell(\boldsymbol{z}^t)\Big) \quad (9)$$

ISBN 978-90-827970-1-5 © EURASIP 2018

where $\mu > 0$ is a preselected stepsize, and the block-sparse hard thresholding operator $\mathcal{H}_k^B(\bar{\boldsymbol{u}}) : \mathbb{R}^n \to \mathbb{R}^n$ converts an n-dimensional vector $\bar{\boldsymbol{u}} := [\bar{\boldsymbol{u}}_1^\mathcal{T} \ \ldots \ \bar{\boldsymbol{u}}_{N_B}^\mathcal{T}]^\mathcal{T}$ into a k-block-sparse one $\boldsymbol{u} := [\boldsymbol{u}_1^\mathcal{T} \ \ldots \ \boldsymbol{u}_{N_B}^\mathcal{T}]^\mathcal{T}$ such that

$$\boldsymbol{u}_b = \begin{cases} \bar{\boldsymbol{u}}_b, & \text{if } b \in \mathcal{U}_B \\ \boldsymbol{0}, & \text{if } b \notin \mathcal{U}_B \end{cases}$$

where $\mathcal{U}_B$ consists of indices associated with the k-largest entities in $\{\|\bar{\boldsymbol{u}}_b\|_2\}_{b=1}^{N_B}$.

Unfortunately, the negative gradient $-\nabla\ell(\boldsymbol{z})$ may not drag the iterate sequence $\{\boldsymbol{z}^t\}$ to the global optimum $\boldsymbol{x}$ because the estimated sign $\boldsymbol{a}_i^\mathcal{T}\boldsymbol{z}/|\boldsymbol{a}_i^\mathcal{T}\boldsymbol{z}|$ in $\nabla\ell(\boldsymbol{z})$ may not coincide with the true one $\boldsymbol{a}_i^\mathcal{T}\boldsymbol{x}/|\boldsymbol{a}_i^\mathcal{T}\boldsymbol{x}|$ [6]. As such, the update in (9) may not always reduce the distance of the iterate to the global optimum. To alleviate the negative influence of the mistakenly estimated signs, the following truncated gradient $\nabla\ell_{\rm tr}(\boldsymbol{z}^t)$ is adopted in SPARTA [16]

$$\nabla\ell_{\rm tr}(\boldsymbol{z}^t) := \frac{1}{m}\sum_{i\in\mathcal{I}^t}\Big(\boldsymbol{a}_i^\mathcal{T}\boldsymbol{z}^t - \psi_i\frac{\boldsymbol{a}_i^\mathcal{T}\boldsymbol{z}^t}{|\boldsymbol{a}_i^\mathcal{T}\boldsymbol{z}^t|}\Big)\boldsymbol{a}_i \qquad (10)$$

where

$$\mathcal{I}^t := \Big\{1 \le i \le m \Big| \frac{|\boldsymbol{a}_i^\mathcal{T}\boldsymbol{z}^t|}{|\boldsymbol{a}_i^\mathcal{T}\boldsymbol{x}|} \ge \tau_{\rm g}\Big\}$$

for some preselected truncation parameter $\tau_g > 0$. It is clear that $\nabla\ell_{\rm tr}(\boldsymbol{z})$ is based on data whose associated $|\boldsymbol{a}_i^\mathcal{T}\boldsymbol{z}|$ are of relatively large sizes. The reason behind this truncation is that gradients (summands in (10)) of large $|\boldsymbol{a}_i^\mathcal{T}\boldsymbol{z}|/|\boldsymbol{a}_i^\mathcal{T}\boldsymbol{x}|$ provably point toward the global optimum $\boldsymbol{x}$ with high probability [6]. As also highlighted in [7], this truncation may reject meaningful samples, which hampers the efficacy of $\nabla\ell_{\rm tr}$ especially when the sample size is limited.

An alternative to the gradient trimming technique is to assign different weights for different gradients [7], which helps fusing useful information from all gradients. Specifically, the following reweighted gradient pursued in [7] proves successful in PR of general signal vectors

$$\nabla\ell_{\rm rw}(\boldsymbol{z}^t) := \frac{1}{m}\sum_{i\in[m]} w_i^t\Big(\boldsymbol{a}_i^\mathcal{T}\boldsymbol{z}^t - \psi_i\frac{\boldsymbol{a}_i^\mathcal{T}\boldsymbol{z}^t}{|\boldsymbol{a}_i^\mathcal{T}\boldsymbol{z}^t|}\Big)\boldsymbol{a}_i \qquad (11)$$

where the weights are suggested as

$$w_i^t := \max\left\{\tau_w, \ \frac{|\boldsymbol{a}_i^\mathcal{T}\boldsymbol{z}^t|}{|\boldsymbol{a}_i^\mathcal{T}\boldsymbol{z}^t| + \psi_i\beta_i}\right\}, \quad \forall i \in [m] \qquad (12)$$

for certain preselected parameters $\tau_w > 0$ and $\beta_i > 0$ for all $i \in [m]$. Evidently, it holds that $\tau_w \le w_i^t \le 1$ for all $i \in [m]$, and the larger the ratio $|\boldsymbol{a}_i^\mathcal{T}\boldsymbol{z}|/|\boldsymbol{a}_i^\mathcal{T}\boldsymbol{x}|$, the larger the weight w_i^t. In this aspect, w_i^t reflects the confidence in the i-th negative gradient pointing toward the global optimum $\boldsymbol{x}$.

In the context of PR of block-sparse vectors, it is therefore reasonable to implement the M-IHT based iterations using reweighted gradients, namely

$$\boldsymbol{z}^{t+1} := \mathcal{H}_k^B\big(\boldsymbol{z}^t - \mu\nabla\ell_{\rm rw}(\boldsymbol{z}^t)\big). \qquad (13)$$

The proposed block-sparse PR solver is summarized in Alg. 1. Its exact recovery is established in the next theorem, whose proof is provided in our journal version [21].

Theorem 2: Let $\boldsymbol{x} \in \mathbb{R}^n$ be any k-block-sparse ($kB \ll n$) vector with $x_{\min}^B := (C_0''/k)\|\boldsymbol{x}\|_2^2$. Consider noiseless measurements $\{\psi_i = |\boldsymbol{a}_i^\mathcal{T}\boldsymbol{x}|\}_{i=1}^m$ from the real Gaussian model. If $m \ge C_1k^2B\log(mn)$, there exists a constant learning rate $\mu > 0$, such that the successive estimates $\boldsymbol{z}^t$ in Algorithm 1 obey

$$\|\boldsymbol{z}^t - \boldsymbol{x}\|_2 \le \delta_0\rho^t\|\boldsymbol{x}\|_2, \quad t = 0, 1, \ldots \qquad (14)$$

with probability at least $1 - c_2\exp(-c_1m) - 6/m$. Here, $0 < \delta_0 < 1$, $0 < \rho < 1$, μ, $c_1 > 0$, $c_2 > 0$, C_0'', and C_1 are certain numerical constants.

Regarding the implications of Thm. 2, two observations are in order. As soon as $m \ge C_1k^2B\log(mn)$, CRAF recovers exactly k-block-sparse vectors $\boldsymbol{x}$ of non-negligible blocks. This sample complexity is consistent with the Block CoPRAM method in [18]. Moreover, CRAF converges linearly to the global optimum. Expressed differently, it takes CRAF at most $T := \mathcal{O}(\log(1/\epsilon))$ iterations to reach a solution of ϵ-relative accuracy.

IV. Numerical Tests

This section compares CRAF with the state-of-the-art sparse PR procedures, including SPARTA [16] and CoPRAM [18]. In all experiments, the support $\mathcal{S}$ of the true signal vectors $\boldsymbol{x} \in \mathbb{R}^{3,000}$ was randomly chosen. The nonzero entries were generated using $\boldsymbol{x}_\mathcal{S} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. The obtained $\boldsymbol{x}$ was subsequently normalized such that $\|\boldsymbol{x}\|_2 = 1$. The sampling vectors were generated using $\boldsymbol{a}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, $1 \le i \le m$. For SPARTA, its suggested parameters were used. The parameters of CRAF were set as $\lambda^- = -3$, $\lambda^+ = 1$, $\{\beta_i = 0.6\}_{i=1}^m$, $\tau_w = 0.1$, and $\mu = 1$. For all simulated schemes, the maximum number of iterations was fixed to $T = 1,000$, and all reported results are averaged over 100 Monte Carlo runs.

The first experiment evaluates the performance of our initialization relative to that in SPARTA [16] and CoPRAM [18] for block length $B = 1$. Figure 1 depicts the average relative error of the three initialization schemes with the sparsity level k varying from 25 to 35, and m/k fixed to 30. Clearly, the new initialization outperforms the other two with large margins.

The second experiment examines the empirical success rates of CRAF, SPARTA, and CoPRAM for solving the ordinary compressive PR with $B = 1$. Each of the 100 Monte Carlo trials is declared a success if the relative error $\mathrm{dist}(\boldsymbol{z}^T, \boldsymbol{x})/\|\boldsymbol{x}\|_2$ is less than 10^{-5}. The empirical success rates of CRAF, SPARTA, and CoPRAM are reported in Fig. 2 with m varying from 400 to 1,800. Notably, the curves showcase improved exact recovery performance of CRAF relative to its competing alternatives.

The last experiment evaluates the robustness of CRAF against additive noise in the amplitude measurements

$$\psi_i = |\boldsymbol{a}_i^\mathcal{T}\boldsymbol{x}| + \eta_i, \quad 1 \le i \le m$$

where $\{\eta_i\}_{i=1}^m$ are independently sampled from $\mathcal{N}(0, \sigma^2)$. In this test, the parameters $k = 30$, $B = 1$, and $m = 1,600$

ISBN 978-90-827970-1-5 © EURASIP 2018

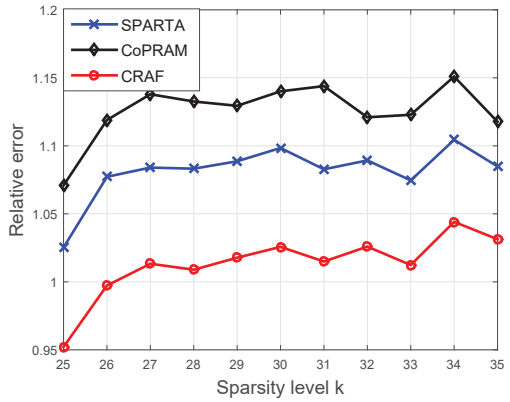

Fig. 1: Average relative error for $B = 1$.

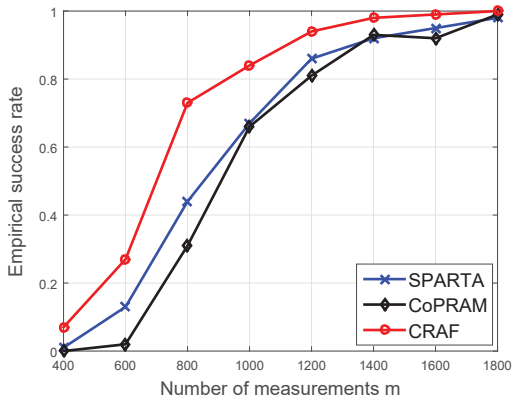

Fig. 2: Empirical success rate versus m for $B = 1$, $k = 30$.

were simulated. Figure 3 depicts the relative errors of the three approaches versus varying σ^2 from 0.1 to 0.6 by 0.1, from which it is clear that CRAF offers the most accurate estimates for all noise levels.

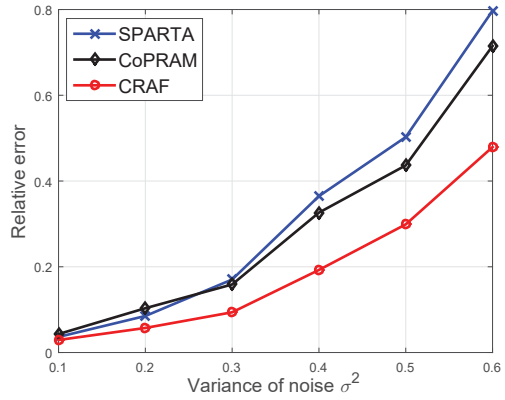

Fig. 3: Average relative error of signal recovery versus the variance of noise σ^2 for $B = 1$, $k = 30$, and $m = 1,600$.

V. Conclusions

This paper advocated the compressive reweighted amplitude flow (CRAF) algorithm for PR of sparse signal vectors. CRAF starts by estimating the support of the sparse signal vector, followed by a new spectral procedure to obtain an effective sparse initialization. To enhance the initial guess, CRAF proceeds with hard thresholding based iteratively reweighted gradient iterations. CRAF recovers the sparse signal vectors exponentially fast when a sufficient number of measurements become available. Simulated tests showcase the merits of CRAF relative to state-of-the-art sparse PR solvers.

References

[1] 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.

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

[3] 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.

[4] 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.

[5] Y. Chen and E. J. Candès, "Solving random quadratic systems of equations is nearly as easy as solving linear systems," *Comm. Pure Appl. Math.*, vol. 70, no. 5, pp. 822–883, Dec. 2017.

[6] G. Wang, G. B. Giannakis, and Y. C. Eldar, "Solving systems of random quadratic equations via truncated amplitude flow," *IEEE Trans. Inf. Theory*, vol. 64, no. 2, pp. 773–794, Feb. 2018.

[7] G. Wang, G. B. Giannakis, Y. Saad, and J. Chen, "Solving most systems of random quadratic equations," in *Adv. Neural Inf. Process. Syst.*, Long Beach, CA, USA, December 4-9 2017.

[8] J. Ma, J. Xu, and A. Maleki, "Optimization-based AMP for phase retrieval: The impact of initialization and L2-regularization," *arXiv:1801.01170*, 2018.

[9] J. C. Duchi and F. Ruan, "Solving (most) of a set of quadratic equalities: Composite optimization for robust phase retrieval," *arXiv:1705.02356*, May 2017.

[10] 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.

[11] C. Ma, K. Wang, Y. Chi, and Y. Chen, "Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval, matrix completion and blind deconvolution," *arXiv:1711.10467*, 2017.

[12] T. Goldstein and S. Studer, "PhaseMax: Convex phase retrieval via basis pursuit," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, Apr. 2018.

[13] P. Hand and V. Voroninski, "Compressed sensing from phaseless gaussian measurements via linear programming in the natural parameter space," *arXiv:1611.05985*, 2016.

[14] O. Dhifallah, C. Thrampoulidis, and Y. M. Lu, "Phase retrieval via polytope optimization: Geometry, phase transitions, and new algorithms," *arXiv:1805.09555*, 2018.

[15] R. Ghods, A. S. Lan, T. Goldstein, and C. Studer, "Linear spectral estimators and an application to phase retrieval," *arXiv:1806.03547*, 2018.

[16] G. Wang, L. Zhang, G. B. Giannakis, M. Akçakaya, and J. Chen, "Sparse phase retrieval via truncated amplitude flow," *IEEE Trans. Signal Process.*, vol. 66, no. 2, pp. 479–491, Jan. 2018.

[17] T. Cai, X. Li, and Z. Ma, "Optimal rates of convergence for noisy sparse phase retrieval via thresholded Wirtinger flow," *Ann. Stat.*, vol. 44, no. 5, pp. 2221–2251, 2016.

[18] G. Jagatap and C. Hedge, "Phase retrieval using structured sparsity: A sample efficient algorithmic framework," *arXiv:1705.06412*, 2017.

[19] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.

[20] Y. Yu, T. Wang, and R. J. Samworth, "A useful variant of the Davis–Kahan theorem for statisticians," *Biometrika*, vol. 102, no. 2, pp. 315–323, Jun. 2015.

[21] L. Zhang, G. Wang, G. B. Giannakis, and J. Chen, "Compressive phase retrieval via reweighted amplitude flow," *aXiv:1712.02426*, 2017.

[22] M. Mondelli and A. Montanari, "Fundamental limits of weak recovery with applications to phase retrieval," *arXiv:1708.05932*, Sep. 2017.

ISBN 978-90-827970-1-5 © EURASIP 2018