PHASE RETRIEVAL FOR SIGNALS IN UNION OF SUBSPACES

M. Salman Asif[*]

Electrical and Computer Engineering
University of California, Riverside

Chinmay Hegde[†]

Electrical and Computer Engineering
Iowa State University

ABSTRACT

We consider the phase retrieval problem for signals that belong to a union of subspaces. We assume that amplitude measurements of the signal of length n are observed after passing it through a random $m \times n$ measurement matrix. We also assume that the signal belongs to the span of a single d-dimensional subspace out of R subspaces, where $d \ll n$. We assume the knowledge of all possible subspaces, but the true subspace of the signal is unknown. We present an algorithm that jointly estimates the phase of the measurements and the subspace support of the signal. We discuss theoretical guarantees on the recovery of signals and present simulation results to demonstrate the empirical performance of our proposed algorithm. Our main result suggests that if properly initialized, then $O(d + \log R)$ random measurements are sufficient for phase retrieval if the unknown signal belongs to the union of R low-dimensional subspaces.

Index Terms— alternating minimization, subspace identification, block sparsity.

1. INTRODUCTION

Phase retrieval refers to a broad class of problems in which we seek to recover a real- or complex-valued signal from its amplitude measurements. We assume that an unknown signal, $\mathbf{x}^* \in \mathbb{R}^n$, is measured via (possibly noisy) observations of the form:

$$y_i = |\langle \mathbf{a}_i, \mathbf{x}^* \rangle| + e_i, \qquad i = 1, \ldots, m, \tag{1}$$

Throughout this paper, we assume that the measurements are *Gaussian*, i.e., each element of $\mathbf{a}_i \sim \mathcal{N}(0, 1)$. The task is to recover an estimate of $\mathbf{x}^*$ from the (phaseless) measurements $\mathbf{y} \in \mathbb{R}^m$.

Our focus is on the special case where $\mathbf{x}^*$ is assumed to belong to a *union of subspaces*. Formally, let us assume that we are given a collection of we represent the union of subspaces as

$$\mathcal{M} = \{\mathbf{x} \in \mathbb{R}^n | \mathbf{x} = \mathbf{U}_i \alpha \text{ for some } i \in \{1, \ldots, R\}\},$$

where $\mathbf{U}_i \in \{\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_R\}$ and each $\mathbf{U}_i \in \mathbb{R}^{n \times d}$ is an orthonormal basis for a d-dimensional subspace of $\mathbb{R}^n$ and $\alpha \in \mathbb{R}^d$ is a coefficient vector. Figure 1 is a depiction of a union of subspaces (planes). Throughout this paper, we assume that $\mathcal{M}$ is known, and that $\mathbf{x}^* \in \mathcal{M}$ (i.e., we consider no model mismatch). For the rest of this paper, we ignore noise (i.e., we set $e_i = 0$) while noting that most of our arguments carry over under sub-gaussian assumptions on the noise variables e_i.

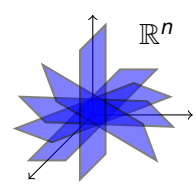

Fig. 1: Illustration of a union-of-subspaces.

We present a two-stage algorithm in which we first construct a coarse initial estimate of the signal using spectral initialization. Then we iteratively refine the signal estimate using an alternating minimization method in which we alternately solve a least-squares problem with union-of-subspace constraints and update the measurement phase.

We analyze the performance of our proposed algorithm with theoretical guarantees. Our main result is given in Theorem 4.1, which suggests that $m > C(d + \log R)$ i.i.d. Gaussian measurements are sufficient for the recovery of any signal $\mathbf{x}^* \in \mathcal{M}$, where $C > 0$ is a constant factor. We also analyze the empirical performance of our algorithm using simulated measurements. The results are summarized in Figure 2.

2. BACKGROUND

The phase retrieval problem has been extensively studied over the last few decades [1, 2, 3] and it appears in several applica-

[*]Email: sasif@ucr.edu. This work is supported in part by UCR Regents Faculty Fellowship and DARPA REVEAL Program.

[†]Email: chinmay@iastate.edu. This work is supported in part by grants CCF-1750920 and CCF-1815101 from the National Science Foundation, and a faculty fellowship from the Black and Veatch Foundation.

tions, including optical imaging [2, 4], microscopy [5, 6], and X-ray crystallography [7].

Phase retrieval is a non-convex problem and classical solution methods rely on alternating projection heuristics; examples include Gerchberg-Saxton [1] and Fienup algorithms [2]. In recent years, lifting-based methods were introduced that reformulate the phase retrieval as a rank-one matrix recovery problem and relax it to a nuclear norm minimization problem that can be solved as a semidefinite program [3]. Because of large computational complexity and memory requirements of semidefinite program-based approaches, they are infeasible for large-scale problems. More recently, a number of convex and non-convex methods have been proposed for solving phase retrieval problem with theoretical performance guarantees [8, 9, 10, 11, 12, 13]. Almost all these methods rely on estimating a good initial solution via the so-called *spectral initialization* method. In its simplest form, a spectral initializer computes the top singular vector of the following Hermitian matrix:

$$\mathbf{G} = \frac{1}{m}\sum_{i=1}^{m} y_i^2 \mathbf{a}_i \mathbf{a}_i^T, \tag{2}$$

which is equivalent to finding a unit norm vector $\mathbf{x}$ that maximizes the inner product between the observed amplitude vector $\mathbf{y}$ and $|\mathbf{A}\mathbf{x}|$. In our proposed algorithm, we use a similar approach for initialization in which we estimate an initial solution in every subspace and select the one that maximize the inner product.

3. ALGORITHM

Our algorithm follows a two-stage approach that has become common in the phase retrieval literature.

- *Initialization*: We first construct a coarse initial estimate of the signal. Specifically, we obtain a signal estimate $\mathbf{x}^0$ using spectral initialization process described in Algorithm 1. Such an estimate is expected to satisfy the following condition:

  $$\operatorname{dist}\left(\mathbf{x}^0, \mathbf{x}^*\right) \leq \delta \left\|\mathbf{x}^*\right\|_2$$

  for some small constant δ, where $\operatorname{dist}(\cdot,\cdot)$ is a suitably defined distance measure.

- *Refinement*: We then sequentially refine this estimate using a strategy based on *alternating minimization*, following a variant of [10, 14, 15]. We prove that our refinement strategy demonstrates linear convergence to $\mathbf{x}^*$, i.e., for $t = 0, 1, 2, \ldots$, our sequence of estimates satisfies:

  $$\operatorname{dist}\left(\mathbf{x}^{t+1}, \mathbf{x}^*\right) \leq \rho \operatorname{dist}\left(\mathbf{x}^t, \mathbf{x}^*\right).$$

Algorithm 1 PR-UoS: Spectral initialization

Input: $\mathbf{A}, \mathbf{y}, \{\mathbf{U}_1, \ldots, \mathbf{U}_R\}$

1: Compute top singular vector for every subspace as

$$\hat{\alpha}_r = \underset{\alpha:\|\alpha\|_2=1}{\operatorname{argmax}} \ \|\mathbf{y} \circ \mathbf{A}\mathbf{U}_r\alpha\|_2. \tag{3}$$

2: Select the subspace that maximizes the objective in (3):

$$r^* = \underset{r}{\operatorname{argmax}} \ \|\mathbf{y} \circ \mathbf{A}\mathbf{U}_r\hat{\alpha}_r\|_2.$$

Output: $\mathbf{x}^0 \leftarrow \mathbf{U}_{r^*}\alpha_{r^*}$.

Algorithm 2 PR-UoS: Descent

Input: $\mathbf{A}, \mathbf{y}, \mathcal{M}$

1: Initialize $\mathbf{x}^0$ according to Algorithm 1
2: **for** $t = 1, \cdots, T$ **do**
3: $\quad \mathbf{p}^t \leftarrow \operatorname{sign}\left(\mathbf{A}\mathbf{x}^{t-1}\right)$,
4: $\quad \mathbf{x}^t = \operatorname{argmin}_{\mathbf{x}\in\mathcal{M}} \left\|\mathbf{p}^t \circ \mathbf{y} - \mathbf{A}\mathbf{x}\right\|_2$
5: **end for**

Output: $\hat{\mathbf{x}} \leftarrow \mathbf{x}^T$.

4. ANALYSIS

Our overall algorithm is described in pseudocode form in Algorithm 2. In this section, we discuss some theoretical guarantees for subspace identification and signal reconstruction using our proposed two-step method.

4.1. Analysis of Descent

This part of the algorithm is described in Lines 2-5 of Algorithm 2. Once we obtain a good enough initial estimate $\mathbf{x^0} \in \mathcal{M}$ such that $\operatorname{dist}\left(\mathbf{x^0}, \mathbf{x}^*\right) \leq \delta_0 \left\|\mathbf{x}^*\right\|_2$, we construct a method to refine this estimate.

The intuition is as follows. The observation model in (1) can be restated as follows:

$$\operatorname{sign}\left(\langle \mathbf{a}_i, \mathbf{x}^*\rangle\right) \circ y_i = \langle \mathbf{a}_i, \mathbf{x}^*\rangle,$$

for all $i = \{1, 2, \ldots, m\}$. To ease notation, denote the *phase vector* $\mathbf{p} \in \mathbb{R}^m$ as a vector that contains the unknown signs of the measurements, i.e., $p_i = \operatorname{sign}\left(\langle \mathbf{a}_i, \mathbf{x}\rangle\right)$ for all $i = \{1, 2, \ldots, m\}$. Let $\mathbf{p}^*$ denote the true phase vector and let $\mathcal{P}$ denote the set of all phase vectors, i.e. $\mathcal{P} = \{\mathbf{p} : p_i = \pm 1, \forall i\}$. Then our measurement model gets modified as:

$$\mathbf{p}^* \circ \mathbf{y} = \mathbf{A}\mathbf{x}^*.$$

Therefore, the recovery of $\mathbf{x}^*$ can be posed as a (non-convex) optimization problem:

$$\min_{\mathbf{x}\in\mathcal{M},\mathbf{p}\in\mathcal{P}} \|\mathbf{A}\mathbf{x} - \mathbf{p}\circ\mathbf{y}\|_2 \tag{4}$$

To solve this problem, we alternate between estimating $\mathbf{p}$ and $\mathbf{x}$. We perform two estimation steps:

(a) if we fix the signal estimate $\mathbf{x}$, then the minimizer $\mathbf{p} \in \mathcal{P}$ is given in closed form as:

$$\mathbf{p} = \operatorname{sign}(\mathbf{A}\mathbf{x}), \tag{5}$$

(b) and if we fix the phase vector $\mathbf{p}$, the signal vector $\mathbf{x} \in \mathcal{M}_s$ can be obtained by solving a sparse recovery problem,

$$\min_{\mathbf{x}\in\mathcal{M}} \|\mathbf{A}\mathbf{x} - \mathbf{p}\circ\mathbf{y}\|_2. \tag{6}$$

We now analyze our proposed descent scheme. We obtain the following theoretical result:

Theorem 4.1. *Given an initialization $\mathbf{x^0} \in \mathcal{M}$ satisfying $\operatorname{dist}\left(\mathbf{x}^0, \mathbf{x}^*\right) \leq \delta \|\mathbf{x}^*\|_2$, for $0 < \delta_0 < 1$, if the number of (Gaussian) measurements,*

$$m > C\left(d + \log R\right),$$

then with high probability, the iterates $\mathbf{x}^{t+1}$ of Algorithm 2, satisfy:

$$\operatorname{dist}\left(\mathbf{x}^{t+1}, \mathbf{x}^*\right) \leq \rho \operatorname{dist}\left(\mathbf{x}^t, \mathbf{x}^*\right), \tag{7}$$

where $\mathbf{x}^t, \mathbf{x}^{t+1}, \mathbf{x}^* \in \mathcal{M}$, and $0 < \rho < 1$ is a constant.

Proof sketch: The high level idea behind the proof is that with a δ-ball around the true signal $\mathbf{x}^*$, the "phase noise" can be suitably bounded in terms of a constant times the signal estimation error. To be more precise, suppose that $\mathbf{z}^* = \mathbf{A}\mathbf{x}^* = \mathbf{p}^* \circ \mathbf{y}$. Then, at any iteration t, we have:

$$\begin{aligned}\mathbf{z}^t &= \mathbf{p}^t \circ \mathbf{y} \\ &= (\mathbf{p}^t - \mathbf{p}^*) \circ \mathbf{y} + p^* \circ \mathbf{y} \\ &= \mathbf{z}^* + \mathbf{e}^t,\end{aligned}$$

where $\mathbf{e}^t$ can be viewed as the "phase noise". Now, examining Line 4 of the above algorithm, we have:

$$\begin{aligned}\left\|\mathbf{A}\mathbf{x}^t - \mathbf{p}^t \circ \mathbf{y}\right\|_2 &\leq \left\|\mathbf{A}\mathbf{x}^* - \mathbf{p}^t \circ \mathbf{y}\right\|_2 \quad \text{(feasibility of } \mathbf{x}^*) \\ &= \left\|\mathbf{z}^* - \mathbf{z}^t\right\|_2 \\ &= \left\|\mathbf{e}^t\right\|_2.\end{aligned}$$

On the other hand, we have:

$$\begin{aligned}\left\|\mathbf{A}\mathbf{x}^t - \mathbf{p}^t \circ \mathbf{y}\right\|_2 &= \left\|\mathbf{A}\mathbf{x}^t - \mathbf{z}^t\right\|_2 \\ &\geq \left\|\mathbf{A}\mathbf{x}^t - \mathbf{z}^*\right\|_2 - \left\|\mathbf{z}^* - \mathbf{z}^t\right\|_2 \\ &= \left\|\mathbf{A}(\mathbf{x}^t - \mathbf{x}^*)\right\|_2 - \left\|\mathbf{e}^t\right\|_2 \\ &\geq (1-\delta_0)\left\|\mathbf{x}^t - \mathbf{x}^*\right\|_2 - \left\|\mathbf{e}^t\right\|_2,\end{aligned}$$

where the second inequality follows from the triangle inequality and the last inequality follows from the restricted isometry property of $\mathbf{A}$ (with constant δ_0) over the union of subspaces $\mathcal{M}$; for Gaussian $\mathbf{A}$, this holds provided:

$$M \geq \frac{C}{\delta_0^2}(d + \log R),$$

for any finite union of subspaces $\mathcal{M}$ [16]. Rearranging the inequalities provide us the following bound:

$$\left\|\mathbf{x}^t - \mathbf{x}^*\right\|_2 \leq \frac{2}{1-\delta_0}\left\|\mathbf{e}^t\right\|_2.$$

It remains to show that $\|e^t\|_2$ can be bounded in terms of $\left\|\mathbf{x}^{t-1} - \mathbf{x}^*\right\|_2$. We do this through Lemma 4.2 below. Consequently, we get

$$\left\|\mathbf{x}^t - \mathbf{x}^*\right\|_2 \leq \frac{2\rho'}{1-\delta_0}\left\|\mathbf{x}^{t-1} - \mathbf{x}^*\right\|_2,$$

where ρ' is a small enough constant.

We therefore achieve a per-step error reduction scheme of the form:

$$\left\|\mathbf{x}^{t+1} - \mathbf{x}^*\right\|_2 \leq \rho_0\left\|\mathbf{x}^t - \mathbf{x}^*\right\|_2,$$

if the initial estimate $\mathbf{x^0}$ satisfies $\left\|\mathbf{x^0} - \mathbf{x}^*\right\|_2 \leq \delta_0 \|\mathbf{x}^*\|_2$, and this result can be trivially extended to the case where the initial estimate $\mathbf{x^0}$ satisfies $\left\|\mathbf{x^0} + \mathbf{x}^*\right\|_2 \leq \delta_0 \|\mathbf{x}^*\|_2$, hence giving the convergence criterion of the form (for $\rho < 1$):

$$\operatorname{dist}\left(\mathbf{x}^{t+1}, \mathbf{x}^*\right) \leq \rho \operatorname{dist}\left(\mathbf{x}^t, \mathbf{x}^*\right).$$

We now state Lemma 4.2 without proof. A proof will be provided in an extended version of this paper [17]. The proof follows an adaptation of the approach in [14, 15], which itself is based on the approach of [18].

Lemma 4.2. *As long as the initial estimate is a small distance away from the true signal $\mathbf{x}^* \in \mathcal{M}$, $\operatorname{dist}\left(\mathbf{x^0}, \mathbf{x}^*\right) \leq \delta \|\mathbf{x}^*\|_2$, and subsequently, $\operatorname{dist}\left(\mathbf{x}^t, \mathbf{x}^*\right) \leq \delta \|\mathbf{x}^*\|_2$, where $\mathbf{x}^t$ is the tth update of Algorithm 2, then the following bound holds for any $t \geq 0$:*

$$\left\|\mathbf{e}^{t+1}\right\|_2 \leq \rho_1\left\|\mathbf{x}^t - \mathbf{x}^*\right\|_2,$$

with high probability, as long as $m > C(d + \log R)$ and $\rho_1 < 1$ is a constant.

5. SIMULATION RESULTS

In this section, we present some simulation results to demonstrate the performance of our proposed algorithm for different values of signal length (n), number of subspaces (R), and the dimension of each subspace (d). The number of measurements (m) in all our simulations is same as n. Figure 2 summarizes our simulation results, where each point represents

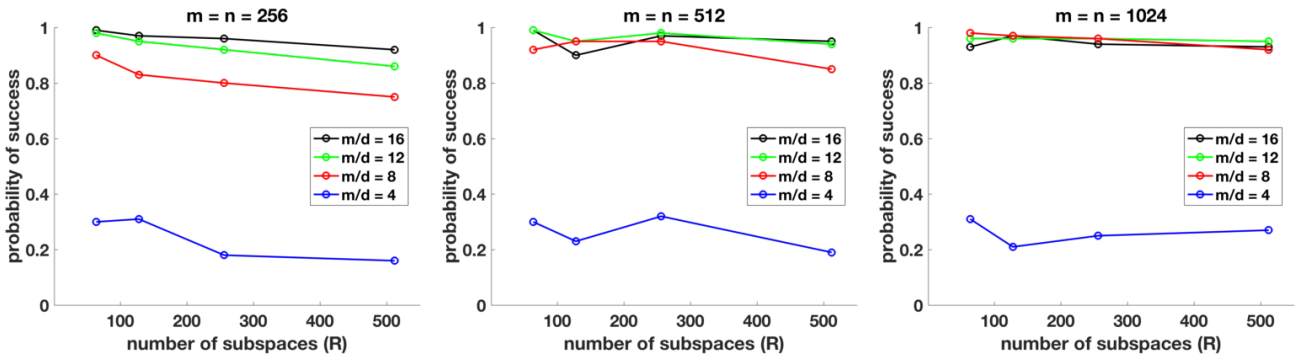

Fig. 2: Probability of success in signal recovery for different values of n, d, R. $m = n$ in all the simulations.

the probability of success in signal recovery for the given values of n, d, R.

Our simulation setup is as follows. For given values of n, R, d, we first generated an $n \times n$ random matrix $\mathbf{A}$ with i.i.d. Gaussian entries. Then we selected an $n \times n$ matrix $\mathbf{Q}$ with orthogonal columns and generated $\{\mathbf{U}_1, \ldots \mathbf{U}_R\}$ by selecting d columns from $\mathbf{Q}$ at random for each $\mathbf{U}_i$. We generated $\mathbf{x}^* = \mathbf{U}_i \alpha$ by picking a $\mathbf{U}_i$ at random and creating α with i.i.d. Gaussian entries. We simulated measurements as $\mathbf{y} = |\mathbf{A}\mathbf{x}^*|$ and estimated $\hat{\mathbf{x}}$ using the two-stage phase retrieval process described in Algorithm 2. We performed 100 independent trial for each tuple n, R, d. In every trial, we considered the recovery successful if the normalized difference between the true and estimated signals

$$\frac{\|\mathbf{x}^* - \hat{\mathbf{x}}\|_2}{\|\mathbf{x}^*\|_2} < \tau,$$

where $\tau = 10^{-5}$ is a threshold.

We observe that as m/d increases, the probability of success approaches 1. The reconstruction is almost always possible if $\frac{m}{d} > 8$. We also observe that increasing the number of subspaces (R) has very little effect on the recovery performance once $\frac{m}{d}$ increases beyond a certain value.

6. REFERENCES

[1] R.W. Gerchberg and A Saxton W. O., "A practical algorithm for the determination of phase from image and diffraction plane pictures," vol. 35, pp. 237–250, 11 1971.

[2] James R Fienup, "Phase retrieval algorithms: a comparison," *Applied optics*, vol. 21, no. 15, pp. 2758–2769, 1982.

[3] Emmanuel J Candes, Thomas Strohmer, and Vladislav Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," *Communications on Pure and Applied Mathematics*, vol. 66, no. 8, pp. 1241–1274, 2013.

[4] Jason Holloway, M. Salman Asif, Manoj Kumar Sharma, Nathan Matsuda, Roarke Horstmeyer, Oliver Cossairt, and Ashok Veeraraghavan, "Toward long-distance subdiffraction imaging using coherent camera arrays," *IEEE Transactions on Computational Imaging*, vol. 2, no. 3, pp. 251–265, 2016.

[5] Lei Tian, Xiao Li, Kannan Ramchandran, and Laura Waller, "Multiplexed coded illumination for fourier ptychography with an led array microscope," *Biomedical optics express*, vol. 5, no. 7, pp. 2376–2389, 2014.

[6] JM Rodenburg, "Ptychography and related diffractive imaging methods," *Advances in Imaging and Electron Physics*, vol. 150, pp. 87–184, 2008.

[7] Jianwei Miao, Pambos Charalambous, Janos Kirz, and David Sayre, "Extending the methodology of x-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens," *Nature*, vol. 400, no. 6742, pp. 342–344, 1999.

[8] S. Bahmani and J. Romberg, "Efficient compressive phase retrieval with constrained sensing vectors," in *Adv. Neural Inf. Proc. Sys. (NIPS)*, 2015, pp. 523–531.

[9] T. Goldstein and C. Studer, "Phasemax: Convex phase retrieval via basis pursuit," *arXiv preprint arXiv:1610.07531*, 2016.

[10] P. Netrapalli, P. Jain, and S. Sanghavi, "Phase retrieval using alternating minimization," in *Adv. Neural Inf. Proc. Sys. (NIPS)*, 2013, pp. 2796–2804.

[11] G. Wang, G. Giannakis, Y. Saad, and J. Chen, "Solving most systems of random quadratic equations," in *Adv. Neural Inf. Proc. Sys. (NIPS)*, 2017.

[12] Emmanuel J Candes, Xiaodong Li, and Mahdi Soltanolkotabi, "Phase retrieval from coded diffraction patterns," *Applied and Computational Harmonic Analysis*, vol. 39, no. 2, pp. 277–299, 2015.

[13] Gauri Jagatap and Chinmay Hegde, "Fast, sample-efficient algorithms for structured phase retrieval," in *Advances in Neural Information Processing Systems*, 2017, pp. 4924–4934.

[14] G. Jagatap and C. Hegde, "Fast, sample efficient algorithms for structured phase retrieval," in *Adv. Neural Inf. Proc. Sys. (NIPS)*, 2017, pp. 4924–4934.

[15] G. Jagatap and C. Hegde, "Sample efficient algorithms for recovering structured signals from magnitude-only measurements," *arXiv preprint arXiv:1705.06412*, 2017.

[16] Thomas Blumensath and Mike E Davies, "Iterative hard thresholding for compressed sensing," *Applied and computational harmonic analysis*, vol. 27, no. 3, pp. 265–274, 2009.

[17] M. Salman Asif and Chinmay Hegde, "Phase retrieval for signals in union of subspaces," *preprint*, https://arxiv.org/abs/1807.06222.

[18] H. Zhang and Y. Liang, "Reshaped wirtinger flow for solving quadratic system of equations," in *Adv. Neural Inf. Proc. Sys. (NIPS)*, 2016, pp. 2622–2630.