OUTLIER-ROBUST RECOVERY OF LOW-RANK POSITIVE SEMIDEFINITE MATRICES FROM MAGNITUDE MEASUREMENTS

Yue Sun[†], *Yuanxin Li*[‡] *and Yuejie Chi*[‡]

[†]Department of Electronic Engineering, Tsinghua University, China
[‡]Department of Electrical and Computer Engineering, The Ohio State University, USA

ABSTRACT

We address the problem of estimating a low-rank positive semidefinite (PSD) matrix from a set of magnitude measurements that are quadratic in the sensing vectors in the presence of arbitrary outliers. We propose a parameter-free algorithm that seeks the PSD matrix that minimizes the ℓ_1-norm of the measurement residual. It is shown that the algorithm can exactly recover a rank-r PSD matrix of size-n from $\mathcal{O}\left(nr^2\right)$ measurements with high probability, even when a fraction of the measurements is corrupted by arbitrary outliers. Furthermore, the recovery is also robust to bounded noise. When an upper bound of the rank of the PSD matrix is known a priori, we further propose a non-convex algorithm based on subgradient descent that demonstrates superior empirical performance.

Index Terms— matrix recovery, outlier-robust, low-rank, positive semidefinite

1. INTRODUCTION

In many emerging applications of science and engineering, we are interested in estimating a positive semidefinite (PSD) matrix $\boldsymbol{X}_0 \in \mathbb{R}^{n\times n}$ from a set of magnitude measurements that are quadratic in the sensing vectors $\boldsymbol{a}_i \in \mathbb{R}^n$:

$$z_i = \langle \boldsymbol{a}_i\boldsymbol{a}_i^T, \boldsymbol{X}_0\rangle = \boldsymbol{a}_i^T\boldsymbol{X}_0\boldsymbol{a}_i, \quad i = 1, \dots, m. \qquad (1)$$

These measurements either arise due to physical limitations in the capability of capturing phases, such as in phase retrieval [1], where $\boldsymbol{X}_0$ is a lifted rank-one matrix from the signal of interest; or arise by design, such as in covariance sketching or power spectrum estimation [2], where $\boldsymbol{X}_0$ corresponds to the covariance matrix of the data, typically of low-rank due to the intrinsic low dimensionality of the data.

Our goal is to recover a low-rank PSD matrix $\boldsymbol{X}_0$ from (1) using as small number of measurements as possible in a computationally efficient and robust manner. A popular convex relaxation algorithm is based on trace minimization [2], which seeks the PSD matrix with the smallest trace norm that satisfies the observation constraint. It is shown in [2] that this algorithm exactly recovers all rank-r PSD matrices as soon as the number of measurements exceeds the information-theoretic limit $\mathcal{O}\left(nr\right)$ in the absence of noise, and the recover is robust against bounded noise as well.

In this paper, we focus on robust recovery of the low-rank PSD matrix when the measurements are further corrupted by outliers, which are possibly adversarial with arbitrary amplitudes. In practice, outliers are somewhat inevitable, which may be caused by sensor failures, malicious attacks, or reading errors. Fortunately, the number of outliers is usually much smaller than the number of total measurements, so it is possible to leverage the sparsity of the outliers to still faithfully recover the low-rank PSD matrix of interest.

We first propose a convex relaxation algorithm that seeks the PSD matrix that minimizes the ℓ_1-norm of the measurement residual. The algorithm is free of tuning parameters and therefore is easy to implement. When the sensing vectors are composed of i.i.d. Gaussian entries, we establish that for a fixed rank-r PSD matrix, as long as the number of measurements exceeds $\mathcal{O}(nr^2)$, the algorithm can exactly recover it with high probability, even when a fraction of $\mathcal{O}(1/r)$ measurements are arbitrarily corrupted. Furthermore, the recovery is also robust to bounded noise. In the special case of phase retrieval, the proposed algorithm coincides with a variant of the Phaselift algorithm studied in [3, 4], which has been recently shown robust to outliers [5]. Our result generalizes [5] to the general low-rank setting.

The above convex algorithm may still pose significant computational burden when facing large-scale problems. Motivated by [6, 7], we next develop a non-convex algorithm when the rank of the PSD matrix, or an estimate of it, is known a priori. Since any rank-r PSD matrix can be uniquely decomposed as $\boldsymbol{X}_0 = \boldsymbol{U}_0\boldsymbol{U}_0^T$, where $\boldsymbol{U}_0 \in \mathbb{R}^{n\times r}$ up to orthonormal transformations, it is sufficient to recover $\boldsymbol{U}_0 \in \mathbb{R}^{n\times r}$ without constructing the PSD matrix explicitly. The algorithm iteratively updates the estimate by descending along the subgradient of the ℓ_1 norm of the measurement residual using a properly selected step-size and initialization. Numerical experiments are provided to validate its superior empirical performance.

This work is supported in part by NSF under grant CCF-1422966, ECCS-1462191 and AFOSR under grant FA9550-15-1-0205. Part of the work was done while Y. Sun was visiting The Ohio State University.

The rest of the paper is organized as below. Section 2 details the problem formulation and presents the performance guarantee of the proposed convex relaxation algorithm. Section 3 describes the proposed non-convex subgradient descent algorithm. Numerical examples are provided in Section 4. Finally, we conclude in Section 5.

2. PSD-CONSTRAINED CONVEX RELAXATION

Let $\boldsymbol{X}_0 \in \mathbb{R}^{n\times n}$ be a PSD matrix of rank-r, whose measurements are given as

$$\boldsymbol{z} = \mathcal{A}\left(\boldsymbol{X}_0\right) + \boldsymbol{\beta} + \boldsymbol{w}, \tag{2}$$

where $\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w} \in \mathbb{R}^m$. The linear mapping $\mathcal{A}$: $\mathbb{R}^{n\times n} \to \mathbb{R}^m$ is defined as $\mathcal{A}\left(\boldsymbol{X}_0\right) = \left\{\boldsymbol{a}_i^T \boldsymbol{X}_0 \boldsymbol{a}_i\right\}_{i=1}^m$, where $\boldsymbol{a}_i \in \mathbb{R}^n$'s are the sensing vectors that are composed of i.i.d. standard Gaussian entries. The vector $\boldsymbol{\beta}$ denotes the outlier vector, which is assumed sparse with the fraction of nonzero entries given as $s = \|\boldsymbol{\beta}\|_0 / m$. Moreover, the vector $\boldsymbol{w}$ denotes the additive noise, which is assumed bounded as $\|\boldsymbol{w}\|_1 \leq \epsilon$.

In this paper, we consider the following algorithm for recovery of $\boldsymbol{X}_0$, which seeks the PSD matrix that minimizes the ℓ_1-norm of the measurement residual to motivate outlier sparsity:

$$\hat{\boldsymbol{X}} = \operatorname{argmin}_{\boldsymbol{X}\in\mathbb{R}^{n\times n}} \|\boldsymbol{z} - \mathcal{A}(\boldsymbol{X})\|_1 \quad \text{subject to} \quad \boldsymbol{X} \succeq 0. \tag{3}$$

This algorithm coincides with the Phaselift algorithm studied in [3–5]. The advantage of this formulation is that it does not require any knowledge of the noise bound, the rank of $\boldsymbol{X}_0$, or the sparsity level of the outliers, and is free of tuning parameters.

2.1. Main Theorem

Encouragingly, we prove that the algorithm (3) admits exact recovery of a rank-r PSD matrix as soon as the number of measurements is large enough, even with a fraction of arbitrary outliers. Our main theorem is given as below.

Theorem 1. *Suppose that $\|\boldsymbol{w}\|_1 \leq \epsilon$. Assume the support of $\boldsymbol{\beta}$ is selected uniformly at random with the fraction of outliers given as $s = \|\boldsymbol{\beta}\|_0 / m$, and the signs of $\boldsymbol{\beta}$ are generated from a symmetric Bernoulli distribution as $\mathbb{P}\left\{\operatorname{sgn}\left(\beta_i\right) = -1\right\} = \mathbb{P}\left\{\operatorname{sgn}\left(\beta_i\right) = 1\right\} = 1/2$ for each $i \in \operatorname{supp}(\boldsymbol{\beta})$. Then for a fixed rank-$r$ PSD matrix $\boldsymbol{X}_0 \in \mathbb{R}^{n\times n}$, there exist some absolute constants $C_1 > 0$ and $0 < s_0 < 1$ such that as long as*

$$m > C_1 n r^2, \quad s \leq \frac{s_0}{r},$$

the solution to (3) *satisfies*

$$\left\|\hat{\boldsymbol{X}} - \boldsymbol{X}_0\right\|_{\mathrm{F}} \leq C_2 \frac{r\epsilon}{m},$$

with probability exceeding $1 - \exp(-\gamma m / r^2)$ for some constants C_2 and γ.

The proof of Theorem 1 is based on the construction of a dual certificate that certifies the optimality of (3) with high probability. Due to space limitations, we refer to the technical report [8] for the complete proof. Theorem 1 has the following consequences:

• **Exact Recovery with Outliers:** When $\epsilon = 0$, Theorem 1 suggests the recovery is exact, even when a fraction of $\mathcal{O}(1/r)$ measurements is corrupted, as long as the number of measurements m exceeds $\mathcal{O}(nr^2)$. Given there are at least nr unknowns, our measurement complexity is near-optimal up to a factor of r. However, our bound is slightly worse than the guarantee in the outlier-free case [2], which requires only $\mathcal{O}(nr)$ measurements.

• **Stable Recovery with Bounded Noise:** In the presence of bounded noise, Theorem 1 suggests that the recovery performance decreases gracefully with the increase of ϵ, where the Frobenius norm of the reconstruction error is proportional to the per-entry noise level of the measurements up to a factor of r.

• **Phase Retrieval:** When $r = 1$, the problem degenerates to the case of phase retrieval. Theorem 1 recovers existing results in [5], where the sample complexity for recovery is on the order of n, optimal up to a scaling factor.

Our theorem also suggests that recovery of PSD matrices from quadratic sampling may be as effective by only exploiting the PSD constraint, without the usual wisdom of trace minimization [2], which is also empirically validated in Section 4.1.

2.2. Comparisons to Related Work

We note that [9] also considers a regularization-free algorithm for PSD matrix estimation that minimizes the ℓ_2 norm of the residual, which unfortunately, is not robust to outliers as our algorithm (3) that minimizes the ℓ_1 norm of the residual. Another standard approach is based on convex decomposition of low-rank and sparse components [10–13], given as

$$\min_{\boldsymbol{X}\succeq 0, \boldsymbol{\beta}} \operatorname{Tr}(\boldsymbol{X}) + \lambda \|\boldsymbol{\beta}\|_1, \quad \text{subject to } \|\boldsymbol{z} - \mathcal{A}(\boldsymbol{X}) - \boldsymbol{\beta}\|_1 \leq \epsilon,$$

where λ is a regularization parameter that requires to be tuned properly. In contrast, the formulation (3) is parameter-free, making it easy to implement.

3. A NON-CONVEX SUBGRADIENT DESCENT ALGORITHM

When the rank of the PSD matrix $\boldsymbol{X}_0$ is known a priori as r, it is possible to decompose $\boldsymbol{X}_0$ as $\boldsymbol{X}_0 = \boldsymbol{U}_0 \boldsymbol{U}_0^T$ where $\boldsymbol{U}_0 \in \mathbb{R}^{n\times r}$. Instead of directly recovering $\boldsymbol{X}_0$, instead, we may aim at recovering $\boldsymbol{U}_0$ up to orthogonal transforms, since $(\boldsymbol{U}_0\boldsymbol{Q})(\boldsymbol{U}_0\boldsymbol{Q})^T = \boldsymbol{U}_0\boldsymbol{U}_0^T$ for any orthonormal matrix $\boldsymbol{Q} \in \mathbb{R}^{r\times r}$. Since any rank-$r$ PSD matrix $\boldsymbol{X}$ can be written $\boldsymbol{X} =$

$\boldsymbol{U}\boldsymbol{U}^T$ for some $\boldsymbol{U} \in \mathbb{R}^{n\times r}$, we can equivalently rewrite (3) as

$$\hat{\boldsymbol{U}} = \mathrm{argmin}_{\boldsymbol{U}\in\mathbb{R}^{n\times r}} f(\boldsymbol{U}), \tag{4}$$

where we denote

$$f(\boldsymbol{U}) = \|\boldsymbol{z} - \mathcal{A}(\boldsymbol{U}\boldsymbol{U}^T)\|_1 = \frac{1}{m}\sum_{i=1}^{m}\left| z_i - \|\boldsymbol{U}^T\boldsymbol{a}_i\|_2^2 \right|.$$

The algorithm (4) is no longer convex, since $f(\boldsymbol{U})$ is quadratic in $\boldsymbol{U}$. Motivated by the recent non-convex approaches [6, 7, 14] of solving quadratic systems, we propose a subgradient descent algorithm to solve (4) effectively. Note that a subgradient of $f(\boldsymbol{U})$ with respect to $\boldsymbol{U}$ can be given as

$$\partial f(\boldsymbol{U}) = -\frac{1}{m}\sum_{i=1}^{m}\mathrm{sgn}(z_i - \|\boldsymbol{U}^T\boldsymbol{a}_i\|_2^2)\boldsymbol{a}_i\boldsymbol{a}_i^T\boldsymbol{U}, \tag{5}$$

where the sign function $\mathrm{sgn}(\cdot)$ is defined as

$$\mathrm{sgn}(x) = \begin{cases} +1, & x > 0 \\ 0, & x = 0 \\ -1, & x < 0 \end{cases}.$$

Our subgradient descent algorithm proceeds as below. Denote the estimate in the tth iteration as $\boldsymbol{U}^{(t)}$. First, we initialize $\boldsymbol{U}^{(0)}$ as the best rank-r approximation of the following matrix:

$$\boldsymbol{U}^{(0)}(\boldsymbol{U}^{(0)})^T = \mathcal{P}_r\left(\frac{1}{m}\sum_{i=1}^{m} z_i\boldsymbol{a}_i\boldsymbol{a}_i^T\right), \tag{6}$$

where $\mathcal{P}_r(\boldsymbol{Z}) := \min_{\boldsymbol{X}:\mathrm{rank}(\boldsymbol{X})=r}\|\boldsymbol{X} - \boldsymbol{Z}\|_F^2$ denotes the projection of $\boldsymbol{Z}$ to the closest rank-r matrix in Frobenius norm. Secondly, at the $(t+1)$th iteration, $t \geq 0$, we apply subgradient descent to refine our estimate as

$$\boldsymbol{U}^{(t+1)} = \boldsymbol{U}^{(t)} - \mu_t \cdot \partial f(\boldsymbol{U}^{(t)}),$$

where the step size μ_t is adaptively set as $\mu_t = \mu f(\boldsymbol{U}^{(t)})$ with μ being some constant. This is because the subgradient only depends on the sign of the residuals, but not their amplitudes. The step size is selected to reflect the magnitude of the current residual. The procedure is summarized in Alg. 1. In the numerical simulations, the default value of μ is set as 0.1. The stopping rule in Alg. 1 is simply put as a maximum number of iterations, while in practice, we can also examine the difference of the residuals between consecutive iterations, and stop when the difference is negligible.

The main advantage of Alg. 1 is its low memory and computational complexity. Given that it is not necessary to construct the full PSD matrix, the memory complexity is simply the size of $\boldsymbol{U}^{(t)}$, which is $O(nr)$[1]. The computational complexity per iteration is also low, which is on the order of $O(mnr)$, that is linear in all the parameters. We demonstrate the excellent empirical performance of Alg. 1 in Section 4.2.

[1] We do not count the storage complexity of the sensing vectors here.

Algorithm 1: Non-convex subgradient descent for solving (4)

Parameters: Rank r, the number of iterations T, step-size μ_t.
Input: measurements $\boldsymbol{z}$, and sensing vectors $\{\boldsymbol{a}_i\}_{i=1}^m$
Initialization: Initialize $\boldsymbol{U}^{(0)} \in \mathbb{R}^{n\times r}$ via (6);
for $t = 0 : T-1$ **do**

$$\boldsymbol{U}^{(t+1)} = \boldsymbol{U}^{(t)} + \frac{\mu_t}{m}\sum_{i=1}^{m}\mathrm{sgn}(z_i - \|(\boldsymbol{U}^{(t)})^T\boldsymbol{a}_i\|_2^2)\boldsymbol{a}_i\boldsymbol{a}_i^T\boldsymbol{U}^{(t)},$$

end for
Output: $\hat{U} = \boldsymbol{U}^{(T)}$.

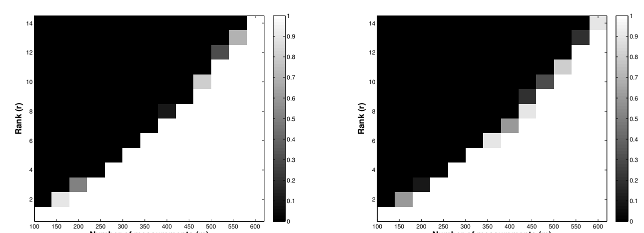
---

4. NUMERICAL EXAMPLES

In this section, we present numerical experiments to demonstrate the performance of the convex algorithm (3) and the non-convex algorithm in Alg. 1.

4.1. Performance of Convex Relaxation

We first consider the performance of (3). Let $n = 40$. We randomly generate a low-rank PSD matrix of rank-r as $\boldsymbol{X}_0 = \boldsymbol{U}\boldsymbol{U}^T$, where $\boldsymbol{U} \in \mathbb{R}^{n\times r}$ is composed of i.i.d. standard Gaussian variables. The sensing vectors are also composed of i.i.d. standard Gaussian variables. Denote the solution of (3) as $\hat{\boldsymbol{X}}$. Each Monte Carlo simulation is called successful if $\|\hat{\boldsymbol{X}} - \boldsymbol{X}_0\|_{\mathrm{F}}/\|\boldsymbol{X}_0\|_{\mathrm{F}} \leq 10^{-3}$. For each cell, the success rate is calculated by averaging over 10 Monte Carlo simulations.

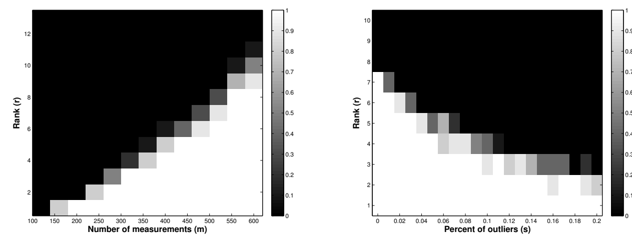


(a) With trace minimization (b) Without trace minimization

Fig. 1: Phase transitions for PSD matrix recovery with respect to the number of measurements and the rank, (a) with trace minimization; and (b) without trace minimization of noise-free measurements, where $n = 40$.

Fig. 1 shows the success rate of exact recovery with respect to the number of measurements and the rank, (a) with the trace minimization as in [1, 3, 15]; and (b) without the trace minimization as proposed in (3). It can be seen that the performance of the two algorithms are almost equivalent, confirming a similar numerical observation for the phase retrieval problem [16] also holds in the low-rank setting. This also suggests there is possible room for improvements of our theoretical guarantee, where the sample complexity depends quadratically in r.

Fig. 2 further shows the success rate of the proposed algorithm (a) with respect to the number of measurements and the rank, when 5% of measurements are selected uniformly at random and corrupted by arbitrary standard Gaussian variables; and (b) with respect to the percent of outliers and the rank, for a fixed number of measurements $m = 400$.



(a) (b)

Fig. 2: Phase transitions of PSD matrix recovery with respect to (a) the number of measurements and the rank, with 5% of measurements corrupted by arbitrary standard Gaussian variables; (b) the percent of outliers and the rank, when the number of measurements is $m = 400$, where $n = 40$.

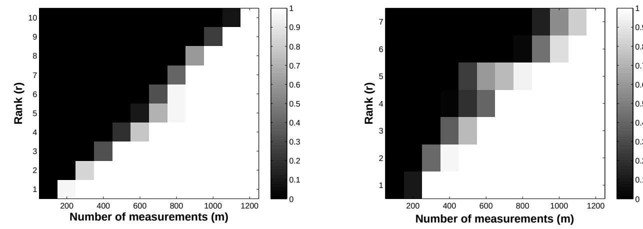
4.2. Performance of Non-Convex Subgradient Descent

We next consider the performance of the non-convex sugradient descent algorithm in Alg. 1 under the same setup of Fig. 1. In Alg. 1, the number of iterations is set as $T = 6 \times 10^4$ (T is set at a large value to guarantee convergence when terminated) and $\mu = 0.1$. Denote the solution to Alg. 1 as $\hat{\boldsymbol{U}}$, and each Monte Carlo simulation is deemed successful if $\|\hat{\boldsymbol{U}}\hat{\boldsymbol{U}}^T - \boldsymbol{X}_0\|_\mathrm{F}/\|\boldsymbol{X}_0\|_\mathrm{F} \leq 10^{-6}$. For each cell, the success rate is calculated by averaging over 50 Monte Carlo simulations. Fig. 3 (a) shows the success rate of Alg. 1 with respect to the number of measurements and the rank, when $n = 100$. Indeed, empirically the algorithm succeeds as soon as the number of measurements is on the order of nr. We also compare against the Wirtinger Flow (WF) algorithm in [6,14] that minimizes the squared ℓ_2-norm of the residual, where the update rule is given as

$$\boldsymbol{U}^{(t+1)} = \boldsymbol{U}^{(t)} + \frac{\mu_t^{\mathrm{WF}}}{m} \sum_{i=1}^{m} (z_i - \|(\boldsymbol{U}^{(t)})^T \boldsymbol{a}_i\|_2^2) \boldsymbol{a}_i \boldsymbol{a}_i^T \boldsymbol{U}^{(t)},$$

where μ_t^{WF} is set as 0.1 using the same initialization (6). Fig. 3 (b) depicts the success rate of the WF algorithm under the same condition of Fig. 3 (a). Both algorithms achieve comparable performance with noise-free observations.

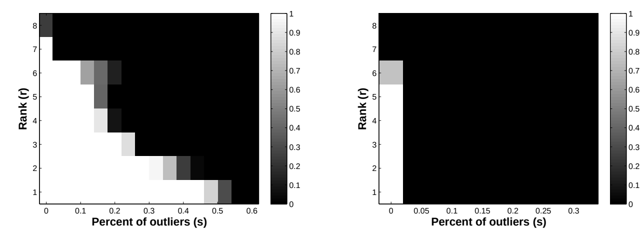
However, the proposed Alg. 1 allows perfect recovery even in the presence of outliers, while the WF algorithm fails.



(a) (b)

Fig. 3: Phase transitions of PSD matrix recovery with respect to the number of measurements and the rank by (a) Alg. 1 with ℓ_1-norm objective, and (b) the WF algorithm with ℓ_2-norm objective, when $n = 100$.

Fig. 4 (a) shows the success rate of Alg. 1 with respect to the percent of outliers and the rank, under the same setup of Fig. 2 (b), where the performance is similar to the convex counterpart in (3). In contrast, the WF algorithm performs poorly even with a single outlier, as shown in its success rate plot in Fig. 4 (b).



(a) (b)

Fig. 4: Phase transitions of PSD matrix recovery with respect to the number of measurements and rank by (a) Alg. 1, and (b) the WF algorithm, when $n = 40$ and $m = 400$.

5. CONCLUSION

In this paper, we address the problem of estimating a low-rank PSD matrix from quadratic magnitude measurements that are possibly corrupted by arbitrary outliers and bounded noise. This problem has many applications in covariance sketching, phase space tomography, and noncoherent detection in communications. It is shown that with $\mathcal{O}(nr^2)$ random Gaussian sensing vectors, a PSD matrix of rank-r can be robustly recovered by the PSD matrix minimizing the ℓ_1-norm of the measurement residual with high probability, even when a fraction of the measurements are adversarially corrupted. This convex formulation eliminates the need for trace minimization and tuning of parameters. Moreover, a non-convex subgradient descent algorithm is proposed with excellent empirical performance with the additional information of the rank of the PSD matrix.

6. REFERENCES

[1] E. J. Candes, T. Strohmer, and V. 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.

[2] Y. Chen, Y. Chi, and A. Goldsmith, "Exact and stable covariance estimation from quadratic sampling via convex programming," *Information Theory, IEEE Transactions on*, vol. 61, no. 7, pp. 4034–4059, July 2015.

[3] E. J. Candes and X. Li, "Solving quadratic equations via phaselift when there are about as many equations as unknowns," *Foundations of Computational Mathematics*, vol. 14, no. 5, pp. 1017–1026, 2014.

[4] L. Demanet and P. Hand, "Stable optimizationless recovery from phaseless linear measurements," *Journal of Fourier Analysis and Applications*, vol. 20, no. 1, pp. 199–221, 2014.

[5] P. Hand, "Phaselift is robust to a constant fraction of arbitrary errors," *arXiv preprint arXiv:1502.04241*, 2015.

[6] E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval via wirtinger flow: Theory and algorithms," *Information Theory, IEEE Transactions on*, vol. 61, no. 4, pp. 1985–2007, 2015.

[7] Y. Chen and E. J. Candès, "Solving random quadratic systems of equations is nearly as easy as solving linear systems," *arXiv:1505.05114*, May 2015.

[8] Y. Li, Y. Sun, and Y. Chi, "Robust recovery of low-rank positive semidefinite matrices from quadratic magnitude measurements with outliers," *Technical Report*, 2015.

[9] M. Kabanava, R. Kueng, H. Rauhut, and U. Terstiege, "Stable low-rank matrix recovery via null space properties," *arXiv preprint arXiv:1507.07184*, 2015.

[10] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of ACM*, vol. 58, no. 3, pp. 11:1–11:37, Jun 2011.

[11] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, "Rank-sparsity incoherence for matrix decomposition," *SIAM Journal on Optimization*, vol. 21, no. 2, pp. 572–596, 2011.

[12] J. Wright, A. Ganesh, K. Min, and Y. Ma, "Compressive principal component pursuit," *Information and Inference*, vol. 2, no. 1, pp. 32–68, 2013.

[13] X. Li, "Compressed sensing and matrix completion with constant proportion of corruptions," *Constructive Approximation*, vol. 37, pp. 73–99, 2013.

[14] C. D. White, R. Ward, and S. Sanghavi, "The local convexity of solving quadratic equations," *arXiv preprint arXiv:1506.07868*, 2015.

[15] E. J. Candes, Y. C. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," *SIAM Journal on Imaging Sciences*, vol. 6, no. 1, pp. 199–225, 2013.

[16] I. Waldspurger, A. d'Aspremont, and S. Mallat, "Phase recovery, maxcut and complex semidefinite programming," *Mathematical Programming*, vol. 149, no. 1-2, pp. 47–81, 2015.