One-Bit Principal Subspace Estimation

Yuejie Chi
Electrical and Computer Engineering
The Ohio State University
Email: chi.97@osu.edu

***Abstract*—This paper proposes a simple sensing and estimation framework, called one-bit sketching, to faithfully recover the principal subspace of a data stream or dataset from a set of one-bit measurements collected at distributed sensors. Each bit indicates the comparison outcome between energy projections of the local sample covariance matrix over a pair of random directions. By leveraging low-dimensional structures, the top eigenvectors of a properly designed surrogate matrix is shown to recover the principal subspace as soon as the number of bit measurements exceeds certain threshold. The sample complexity to obtain reliable comparison outcomes is also obtained. We further develop a low-complexity algorithm to estimate the principal subspace in an online fashion when the bits arrive sequentially at the fusion center. Numerical examples on line spectrum estimation are provided to validate the proposed approach.**

***Index Terms*—one-bit measurements, principal subspace estimation, streaming data**

I. Introduction

Many practical datasets exhibit low-dimensional structures, such that a significant proportion of the variance can be captured by their top principal components; and that the subspace spanned by these principal components is the recovery object of interest rather than the datasets themselves. Consider a data stream which generates a zero-mean data sample $x_t \in \mathbb{C}^n$ at each time t. The covariance matrix of the data $\Sigma = \mathbb{E}[x_t x_t^H]$ is assumed to be of low rank with $\text{rank}(\Sigma) = r \ll n$. This assumption is widely applicable to data such as network traffic, wideband spectrum, images, and so on.

This paper is motivated by the challenge of estimating the principal subspace of the covariance matrix Σ in a networked sensing environment, where each sensor is distributed and has access to only a subset or substream of the whole data. Furthermore, the sensors are resource-limited to completely observe, store, and transmit any entire data sample. For example, the data samples may be processed on the fly at the sensor with only one pass or two passes without being stored. Therefore, it is necessary to first process the data locally at the sensor, and then transmit only enough information to the fusion center for inferring the principal subspace to minimize the communication overhead.

A. Our Contribution

We propose a simple yet efficient framework, called one-bit sketching, to estimate the principal subspace from a single bit per sensor with provable near-optimal performance guarantees, which has applications in wideband power spectrum estimation [1], [2], crowd-sensing [3] and many others. At the sensing stage, each sensor computes the accumulated energy projections of the local data samples onto *two* randomly selected directions with i.i.d. Gaussian entries, called *sketching vectors*. It then transmits a *one-bit* comparison outcome to the fusion center indicating the direction onto which it has a higher accumulated energy. This corresponds to comparing the energy projection of the *local sample covariance matrix* onto two randomly selected rank-one subspaces. A key observation is that, as long as the number of local samples is not too small, the comparison outcome will be exactly the same as if it is derived using the *population covariance matrix* with high probability, which is denoted as an *exact* bit measurement. By only transmitting the comparison outcome rather than the actual energy measurements or the data samples themselves, the communication overhead is minimized to a single bit and the bit measurements are more robust to communication channel errors. Moreover, as will be shown, the energy projections can be computed extremely simple without storing history data samples, and are always nonnegative, making them suitable for wideband and optical applications.

At the fusion center, the principal subspace of the covariance matrix is estimated as the top eigenvectors of a carefully designed surrogate matrix using the collected one-bit measurements, which can be efficiently computed via a truncated Eigen Value Decomposition (EVD) or power iteration with low complexity. Assuming all the bit measurements are exact, it is established that the principal subspace of a rank-r covariance matrix can be universally estimated with high accuracy as soon as the number of collected bits is on the order of $\Theta(nr^2 \log n)$. This is much smaller than the ambient dimension of the covariance matrix, and is order-wise near-optimal since at least nr measurements are required.

The above batch computation requires the fusion center to construct an $n \times n$ surrogate matrix with the knowledge of all sketching vectors, taking a memory space of at least $\Theta(n^2)$. To ease this requirement as well as handle streaming data, we further develop an online subspace estimation algorithm at the fusion center, which updates its estimate as new bit measurements and their associated sketching vectors (or their seeds) arrive sequentially. The online algorithm is based on incremental SVD [4], [5], and has a reduced memory space of $\Theta(nr)$ without storing all the sketching vectors.

B. Related Work

Parameter estimation from one-bit quantizations has been considered for *scalar* parameter estimation problems, such as the sample mean [6]–[9], sigma-delta quantization [10], and a single direction-of-arrival from direct one-bit quantizations of the data samples [11]. Recent work on frugal sensing [2], [12] has considered estimation of the power spectral density, which is a vector-valued functional, from one-bit measurements. Our approach is fundamentally different, which aims to estimate *matrix functionals*, i.e. a low-dimensional subspace, from binary measurements with *finite-sample guarantees*. The design of pairwise comparisons through aggregations to produce one-bit measurements is critical in our proposed framework, as direct one-bit quantizations of the data samples in general fail to inform an underlying subspace without additional structures [13].

The proposed framework is inspired by our recent work in [14]–[16], where a quadratic sampling scheme is designed for low-rank covariance estimation. It is shown in [14] that a number of $\Theta(nr)$ quadratic (energy) measurements suffices to exactly recover a rank-r covariance matrix via nuclear norm minimization, assume the measurement vectors are composed of i.i.d. sub-Gaussian entries. However, communicating these energy measurements with high precision may cause unwanted overhead in practice. Moreover, in the presence of finite samples, it is usually difficult to estimate the noise level which is an input to the convex optimization algorithm in [14].

A related line of research is on one-bit compressed sensing [17]–[20], where the authors aim to recover the signal with low-dimensional structures from the signs of random linear measurements up to a scaling factor. In contrast, our measurements are signs of random *quadratic* measurements. Mroueh and Rosasco first proposed the one-bit phase retrieval framework in [21], which is a highly non-trivial extension of one-bit compressed sensing with phaseless measurements. Our estimation algorithm subsumes the scenario in [21] as a special case when the covariance matrix is assumed to be of rank one, in which case the number of local samples can also be reduced to one.

Paper Organization: The rest of this paper is organized as follows. Section II describes the sensing framework and formulates the subspace estimation problem using one-bit measurements. Section III presents the algorithm, and its performance guarantees regarding sample complexities. Section IV presents an online algorithm to estimate the low-rank principal subspace with sequential one-bit measurements. The numerical examples are given in Section V. Finally, we conclude in Section VI.

II. One-Bit Sketching Scheme

Let $x_t \in \mathbb{C}^n$ be a data stream that satisfies $\mathbb{E}[x_t] = 0$ and its covariance matrix $\Sigma = \mathbb{E}[x_t x_t^H] \in \mathbb{C}^{n\times n}$ is a low-rank positive semidefinite (PSD) matrix with rank$(\Sigma) = r \ll n$. Consider a collection of m sensors that are deployed distributively to measure the data stream. Each sensor can either access the complete data stream or a substream. At the ith sensor, define a pair of sketching vectors $a_i \in \mathbb{C}^n$ and $b_i \in \mathbb{C}^n$, $1 \le i \le m$, where their entries are i.i.d. standard Gaussian. Without loss of generality, we assume the local data samples accessed by the ith sensor are indexed by $t = 1, \ldots, T$. At the tth sample time, the ith sensor takes the quadratic measurements:

$$u_{i,t} = |a_i^H x_t|^2, \quad v_{i,t} = |b_i^H x_t|^2,$$

which are phaseless and more reliable in optical and high frequency applications [22]. These quadratic measurements are then averaged over T samples to obtain

$$U_{i,T} = \frac{1}{T}\sum_{t=1}^{T} u_{i,t} = a_i^H \Sigma_T a_i, \; V_{i,T} = \frac{1}{T}\sum_{t=1}^{T} v_{i,t} = b_i^H \Sigma_T b_i,$$

where $\Sigma_T = \frac{1}{T}\sum_{t=1}^{T} x_t x_t^H$ is the local sample covariance matrix. It is easy to see that $U_{i,T} = \frac{T-1}{T} U_{i,T-1} + \frac{1}{T} u_{i,T}$, which can be updated adaptively without storing all the history data. At the end of the sample time T, the ith sensor compares the average energy $U_{i,T}$ and $V_{i,T}$, and transmit to the fusion center a single bit $y_{i,T}$ to indicate the outcome:

$$y_{i,T} = \begin{cases} 1, & \text{if } U_{i,T} > V_{i,T} \\ -1, & \text{otherwise} \end{cases}. \tag{1}$$

Depending on whether the fusion center has prior knowledge of the sketching vectors, the sensors may also need to transmit a_i and b_i, or their seeds. The communication overhead is minimized when only the bit indicating the comparison outcome is transmitted.

Since the sample covariance matrix Σ_T converges to the true covariance matrix as T approaches infinity, the bit measurement at the ith sensor also approaches to the following

$$y_i = \begin{cases} 1, & \text{if } a_i^H \Sigma a_i > b_i^H \Sigma b_i \\ -1, & \text{otherwise} \end{cases}, \tag{2}$$

when T goes to infinity. Let $W_i = a_i a_i^H - b_i b_i^H$, we can then compactly represent (2) as

$$y_i = \mathrm{sign}(\langle W_i, \Sigma\rangle), \tag{3}$$

where sign$(\cdot)$ is the sign function. The above operation can be regarded as comparing the energy projections of the covariance matrix onto two random selected directions. We denote the ith bit measurement is exact if $y_i = y_{i,T}$, which happens with high probability as soon as T is not too small, as will shown later.

Apparently not all information about Σ can be recovered from one-bit measurements, such as its Frobenius norm $\|\Sigma\|_{\mathrm{F}}$. Assuming all the bit measurements are exact, i.e. $\{y_{i,T}\}_{i=1}^m$ given in (1) agree with $\{y_i\}_{i=1}^m$, our goal is to recover the principal subspace of Σ from the one-bit measurements $\{y_i\}_{i=1}^m$ given in (3) and characterize the reconstruction accuracies in terms of the number of bit measurements m. The sample complexity T to guarantee that $\{y_{i,T}\}_{i=1}^m$ given in (1) agree with $\{y_i\}_{i=1}^m$ with high probability is also derived. Together they guarantee the performance of the proposed scheme.

III. Principal Subspace Estimation From One-Bit Measurements and Performance Guarantees

We propose an extremely simple and low-complexity estimator whose complexity amounts to computing a few top eigenvectors of a surrogate matrix. Specifically, define a symmetric matrix $J_m \in \mathbb{R}^{n\times n}$ as

$$J_m = \frac{1}{m}\sum_{i=1}^{m} y_i W_i = \frac{1}{m}\sum_{i=1}^{m} y_i (a_i a_i^H - b_i b_i^H). \tag{4}$$

Our estimate is the subspace spanned by the top r eigenvectors of J_m, denoted as $\hat{U}$. This can be found via computing a truncated EVD or using the power iteration methods [23] with very low computational cost.

A. Performance Guarantee with respect to m

Let the covariance matrix $\Sigma = \sum_{k=1}^{r} \lambda_k u_k u_k^H$, with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r$ and $U = [u_1, \ldots, u_r]$ denoting its principal subspace. We have the following theorem.

Theorem 1. *Let $0 < \delta < 1$. With probability at least $1-\delta$, there exists a $r \times r$ orthogonal matrix Q such that*

$$\begin{aligned}\|\hat{U} - UQ\|_{\mathrm{F}} &\leq \alpha^{-1}\sqrt{\frac{c_1 n}{m}\log\left(\frac{2n}{\delta}\right)} \\ &\leq \min\left\{(1+\kappa(\Sigma))^{r-1}, 9e^{\kappa(\Sigma)} r\right\}\sqrt{\frac{c_1 n}{m}\log\left(\frac{2n}{\delta}\right)},\end{aligned}$$

where c_1 is an absolute constant, and $\kappa(\Sigma) = \lambda_1/\lambda_r$ is the conditioning number of Σ.

For not too small r and small $\kappa(\Sigma)$, the second term in the upper bound dominates and we have that as soon as the number of one-bit measurements exceeds the order of $\Theta(nr^2 \log n)$, it is sufficient to recover to recover the principal subspace U with high accuracy. Since there are at least nr degrees of freedom to describe the subspace U, our bound is near-optimal up to a polynomial factor with respect to r and a logarithmic factor with respect to n. It is worth emphasizing that a highlight of our result indicates that a number of *bit measurements* that is much smaller than the ambient dimension of the covariance matrix could still yield reliable subspace estimate.

B. Sample Complexity for Exact Bit Measurements

In practice, instead of directly measuring the energy projection of the covariance matrix Σ, we measure the energy projection of the sample covariance matrix in (1). Therefore it is necessary to derive the sample complexity required to obtain faithful bit measurements through (1). Note that the measured energy difference before taking its sign is

$$z_i = \langle W_i, \Sigma_T\rangle = \frac{1}{T}\sum_{t=1}^{T}\langle W_i, x_t x_t^H\rangle, \tag{5}$$

Under additional assumptions on the data samples, we have the following proposition to bound the local sample size to guarantee exact bit measurements with high probability.

Proposition 1. *Let $0 < \delta \leq 1$. Assume x_t are i.i.d. Gaussian satisfying $x_t \sim \mathcal{CN}(0, \Sigma)$. Then*

$$\mathbb{P}\left[\mathit{sign}(z_i) \neq \mathit{sign}(\langle W_i, \Sigma\rangle)\right] \leq \delta$$

as soon as $T > c\frac{\mathrm{Tr}(\Sigma)}{\|\Sigma\|_{\mathrm{F}}}\log^2(1/\delta)$ for some sufficiently large constant c.

In order to guarantee that all the bits are exact, we need to further apply a union bound on Proposition 1, which yields the following theorem.

Theorem 2. *Let x_t be zero-mean with $x_t \sim \mathcal{CN}(0,\Sigma)$. Let $0 < \delta \leq 1$. With probability at least $1-\delta$, all bit measurements are exact, given that the number of samples observed by each sensor satisfies*

$$T > c\frac{\mathrm{Tr}(\Sigma)}{\|\Sigma\|_{\mathrm{F}}}\log^2\left(\frac{m}{\delta}\right)$$

for some sufficiently large constant c.

Recall that $\mathrm{Tr}(\Sigma) \leq \sqrt{r}\|\Sigma\|_{\mathrm{F}}$, then as soon as T is on the order of $\sqrt{r}\log^2 m$ all bit measurements are accurate with high probability.

IV. Online Principal Subspace Estimation From One-Bit Energy Measurements

In this section, we develop an online subspace estimation algorithm for the fusion center to update the principal subspace estimate as new bit measurements arrive sequentially. This is particularly useful when the fusion center is also *memory limited* so that it is incapable to store all sketching vectors up to a memory space of $\Theta(n^2)$. The proposed online algorithm has a memory space of $\Theta(nr)$ to implement, which is comparable to the size of the principal subspace, and does not require storing all the sketching vectors. In the online setting, the sensors transmit both the bit measurement and the sketching vectors or their seeds (hence resulting in an increase in communications overhead), while the fusion center assumes no prior knowledge about the sketching vectors.

We now describe an online algorithm to estimate the principal subspace based on a modification of the incremental SVD algorithm in [4], [5] to allow fast rank-two updates of the EVD of a symmetric matrix. Denote the rank-two update of the surrogate matrix J_m in (4) as

$$J_m = \frac{m-1}{m}J_{m-1} + \frac{y_m}{m}\left(a_m a_m^H - b_m b_m^H\right), \tag{6}$$

where y_m is a new bit measurement. We can rewrite (6) using more general notations as

$$J_m = \eta_m J_{m-1} + K_m \Lambda_m K_m^H, \tag{7}$$

by letting $\eta_m = \frac{m-1}{m}$, $K_m = [a_m, b_m] \in \mathbb{C}^{n\times 2}$, $\Lambda_m = \mathrm{diag}([y_m/m, -y_m/m])$.

A key difference from [4], [5] is that we fix the size of the principal subspace as r. In the update we first compute an expanded principal subspace of rank $(r+2)$ for J_m and then only keep its r largest principal components. Assume the EVD of J_{m-1} as $J_{m-1} = U_{m-1}\Pi_{m-1}U_{m-1}^H$, where $U_{m-1} \in \mathbb{C}^{n\times r}$ is orthonormal and $\Pi_{m-1} \in \mathbb{R}^{r\times r}$ is diagonal. The goal is to find the best rank-r approximation of J_m by updating U_{m-1} and Π_{m-1}. Let $R_m = (I - U_{m-1}U_{m-1}^H)K_m$ and $P_m = \mathrm{orth}(R_m)$ be the orthonormal columns spanning R_m. We can rewrite J_m as

$$J_m := \begin{bmatrix} U_{m-1} & P_m \end{bmatrix} \Gamma'_m \begin{bmatrix} U_{m-1}^T \\ P_m^T \end{bmatrix},$$

where Γ'_m is a small $(r+2)\times(r+2)$ matrix whose EVD can be computed easily and it yields $\Gamma_m = U'_m \Pi'_m U'_m$. Set Π_m be the top $r \times r$ sub-matrix of Π'_m assuming the eigenvalues are given in a descending order in absolute values, the principal subspace of J_m can be updated correspondingly as

$$U_m := \begin{bmatrix} U_{m-1} & P_m \end{bmatrix} U'_m I_r,$$

where I_r is the first r columns of the $(r+2)\times(r+2)$ identity matrix.

V. Numerical Experiments

In the numerical experiments, we first examine the performance of the proposed framework in a batch setting in terms of reconstruction accuracy as a function of the number of bits. We then examine the performance of the online subspace estimation algorithm in Section IV and apply it to the problem of line spectrum estimation.

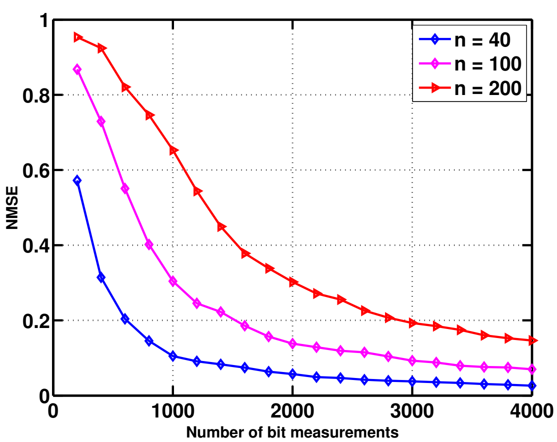

Fig. 1. NMSE with respect to the number of bit measurements for estimating the principal subspace of the covariance matrix with $n = 40, 100, 200$ when $r = 3$.

A. Performance Evaluation with Batch Measurements

We generate the covariance matrix as $\Sigma = XX^T$, where $X \in \mathbb{R}^{n\times r}$ is composed of standard Gaussian entries. The sketching vectors a_i's and b_i's are also generated with standard Gaussian entries. Assuming all the bit measurements are collected perfectly, we compute the top-r eigenvectors of the surrogate matrix J_m in (4) and obtain an estimate of the principal subspace $\hat{X}$. The error metric is calculated as the Normalized Mean Squared Error (NMSE) given as $\|(I - \hat{X}\hat{X}^T)X\|_F^2/\|X\|_F^2$. Fig. 1 shows the average NMSE with respect to the number of bit measurements for different dimensions $n = 40, 100, 200$ when $r = 3$, averaged over 10 Monte Carlo runs. We can see that the recovery accuracy increases gracefully as more bit measurements become available.

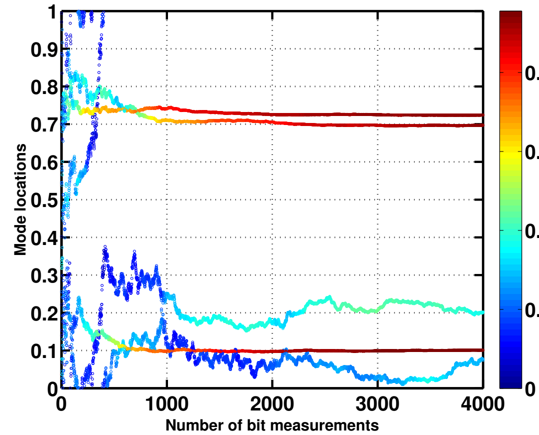

Fig. 2. Online line spectrum estimation: the estimated frequency values against the number of bit measurements when $n = 40$ and $r = 3$. Each time 5 frequencies are estimated with the color bar indicates their amplitudes. The true frequencies are $\mathcal{F} = [0.1, 0.7, 0.725]$ with unit variance.

B. Online Line Spectrum Estimation

We consider the problem of line spectrum estimation from its power spectral density, where we assume the covariance matrix Σ is a low-rank Toeplitz PSD matrix with $n = 40$ and $r = 3$. Let the set of frequencies be $\mathcal{F} = [0.1, 0.7, 0.725]$ with unit variance $\sigma^2 = 1$. At each new one-bit measurement, we first use the online subspace estimation algorithm proposed in Section IV to estimate a principal subspace of rank $r_{est} = 5$, then apply ESPRIT [24] to the reconstructed principal subspace to recover the set of frequencies. Fig. 2 shows the estimates of frequency locations, which are plotted vertically with respect to the index of each new bit measurement. The color indicates the amplitudes of the estimated frequencies. The algorithm successfully estimated all the frequencies even when the frequencies are closely located from a moderate number of one-bit measurements.

VI. Concluding Remarks

In this paper, we presented a simple distributed sensing and central estimation framework to recover the principal subspace of a low-rank covariance matrix from a small number of one-bit energy comparisons, and described the sample complexities for its guaranteed performance. We also develop an online subspace estimation algorithm to ease the memory requirement at the fusion center with a slight increase in communication overhead. Numerical examples are provided to validate the proposed approach. In the future work, we will examine the performance of the proposed approach to more general scenarios with noise and imperfect model assumptions, as well as develop subspace tracking algorithms [25] from one-bit measurements.

Acknowledgment

This work was partially supported by the Air Force Summer Faculty Fellowship Program, NSF under the grant CCF-1422966 and a Google Faculty Research Award.

References

[1] D. Romero and G. Leus, "Wideband spectrum sensing from compressed measurements using spectral prior information," *Signal Processing, IEEE Transactions on*, vol. 61, no. 24, pp. 6232–6246, Dec 2013.

[2] O. Mehanna and N. Sidiropoulos, "Frugal sensing: Wideband power spectrum sensing from few bits," *Signal Processing, IEEE Transactions on*, vol. 61, no. 10, pp. 2693–2703, May 2013.

[3] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell, "A survey of mobile phone sensing," *Communications Magazine, IEEE*, vol. 48, no. 9, pp. 140–150, 2010.

[4] M. Brand, "Incremental singular value decomposition of uncertain data with missing values," in *ECCV 2002*. Springer, 2002, pp. 707–720.

[5] ——, "Fast low-rank modifications of the thin singular value decomposition," *Linear Algebra and its Applications*, pp. 20–30, 2006.

[6] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks-part i: Gaussian case," *Signal Processing, IEEE Transactions on*, vol. 54, no. 3, pp. 1131–1143, 2006.

[7] ——, "Bandwidth-constrained distributed estimation for wireless sensor networks-part ii: unknown probability density function," *Signal Processing, IEEE Transactions on*, vol. 54, no. 7, pp. 2784–2796, 2006.

[8] Z.-Q. Luo, "Universal decentralized estimation in a bandwidth constrained sensor network," *Information Theory, IEEE Transactions on*, vol. 51, no. 6, pp. 2210–2219, 2005.

[9] ——, "An isotropic universal decentralized estimation scheme for a bandwidth constrained ad hoc sensor network," *Selected Areas in Communications, IEEE Journal on*, vol. 23, no. 4, pp. 735–744, 2005.

[10] C. S. Güntürk, "One-bit sigma-delta quantization with exponential accuracy," *Communications on Pure and Applied Mathematics*, vol. 56, no. 11, pp. 1608–1630, 2003.

[11] O. Bar-Shalom and A. Weiss, "Doa estimation using one-bit quantized measurements," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 38, no. 3, pp. 868–884, 2002.

[12] O. Mehanna and N. Sidiropoulos, "Adaptive thresholding for distributed power spectrum sensing," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, May 2013, pp. 4459–4463.

[13] M. A. Davenport, Y. Plan, E. v. d. Berg, and M. Wootters, "1-bit matrix completion," *arXiv preprint arXiv:1209.3672*, 2012.

[14] Y. Chen, Y. Chi, and A. Goldsmith, "Exact and stable covariance estimation from quadratic sampling via convex programming," *arXiv preprint arXiv:1310.0807*, 2013.

[15] Y. Chen, Y. Chi, and A. J. Goldsmith, "Robust and universal covariance estimation from quadratic measurements via convex programming," in *IEEE International Symposium on Information Theory*, July 2014.

[16] ——, "Estimation of simultaneously structured covariance matrices from quadratic measurements," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, May 2014, pp. 7669–7673.

[17] P. T. Boufounos and R. G. Baraniuk, "1-bit compressive sensing," in *Information Sciences and Systems, 2008. CISS 2008. 42nd Annual Conference on*. IEEE, 2008, pp. 16–21.

[18] L. Jacques, J. Laska, P. Boufounos, and R. Baraniuk, "Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors," *Information Theory, IEEE Transactions on*, vol. 59, no. 4, pp. 2082–2102, April 2013.

[19] Y. Plan and R. Vershynin, "One-bit compressed sensing by linear programming," *Communications on Pure and Applied Mathematics*, vol. 66, no. 8, pp. 1275–1297, 2013.

[20] ——, "Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach," *Information Theory, IEEE Transactions on*, vol. 59, no. 1, pp. 482–494, 2013.

[21] Y. Mroueh and L. Rosasco, "Quantization and greed are good: One bit phase retrieval, robustness and greedy refinements," *arXiv preprint arXiv:1312.1830*, 2013.

[22] J. R. Fienup, "Reconstruction of an object from the modulus of its Fourier transform," *Optics letters*, vol. 3, no. 1, pp. 27–29, 1978.

[23] C. Lanczos, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*. United States Governm. Press Office, 1950.

[24] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 7, pp. 984 –995, Jul 1989.

[25] Y. Chi, Y. C. Eldar, and R. Calderbank, "Petrels: Parallel subspace estimation and tracking by recursive least squares from partial observations," *Signal Processing, IEEE Transactions on*, vol. 61, no. 23, pp. 5947–5959, 2013.