23rd International Symposium on Mathematical Theory of Networks and Systems
Hong Kong University of Science and Technology, Hong Kong, July 16-20, 2018

Convergence analysis of accelerated first-order methods for phase retrieval

Huaqing Xiong[1], Yuejie Chi[2], Bin Hu[3] and Wei Zhang[1]

***Abstract*— Phase retrieval finds many applications in signal processing, and recently, there has been a growing interest in directly solving it via nonconvex optimization. Under certain generic statistical models, the loss function satisfies the so-called Regularity Condition (RC) in a local neighborhood of the true signal, which guarantees the linear convergence of gradient descent if initialized properly. However, accelerated first-order methods (e.g. Nesterov's method and Heavy-ball method), despite the empirical success, currently lack similar performance guarantees as the unaccelerated counterpart. This paper studies the convergence of accelerated first-order methods in phase retrieval using tools from robust control. We derive a set of Linear Matrix Inequalities (LMIs) that can be used to numerically certify linear convergence of accelerated first-order methods under the Regularity Condition. For the Heavy-ball method, analytical conditions of algorithmic parameters for linear convergence are further obtained.**

I. Introduction

This paper focuses on the convergence analysis of some popular first-order methods in the phase retrieval problem, where the goal is to recover some real-valued signal $x \in \mathbb{R}^n$ from its measurements $\{g_i\}_{i=1}^m$:

$$g_i = |\langle a_i, x\rangle|, \quad \text{for} \quad i = 1, \cdots, m, \tag{1}$$

where $\{a_i\}_{i=1}^m$ are known sampling vectors. It is one of the most classical problems in signal processing and has a wide range of applications in various fields such as X-ray crystallography [1], microscopy [2], and optics [3]. The phase retrieval problem is typically solved through optimizing some nonconvex loss function that penalizes the deviation of the collected measurements from those of the estimated signal,

$$\hat{x} = \arg\min_{z\in\mathbb{R}^n} \ell(z).$$

For example, in [4], Candès et al. used the squared loss function of the intensity measurements:

$$\ell(z) := \frac{1}{4m}\sum_{i=1}^m (|a_i^T z|^2 - g_i^2)^2. \tag{2}$$

In [5], Zhang et al. adopted the squared loss function of the amplitude measurements:

$$\ell(z) := \frac{1}{2m}\sum_{i=1}^m (|a_i^T z| - g_i)^2. \tag{3}$$

Compared with (2), the loss function in (3) is nonsmooth. In general, optimizing these nonconvex functions is challenging. A popular two-step strategy is to first find a proper initial estimate z_0, and then update the estimate via local refinement, for example via the Gradient method,

$$z_{k+1} = z_k - \alpha\nabla\ell(z_k), \alpha > 0, \tag{4}$$

where α is a carefully chosen step size.

The spectral method [4] is often used to provide an initial estimate. It is demonstrated in [4] that with $m = \mathcal{O}(n \log n)$ measurements, the output of the spectral method lands in $\mathcal{N}_x(c) = \{z \in \mathbb{R}^n : \|z - x\| \le c\|x\|\}$, where $c = 1/10$ with high probability[1]. A few variants of the spectral method are also proposed in the literature, such as the truncated spectral method in [6], which provide a provably good initial estimate with $m = \mathcal{O}(n)$ measurements. In the remaining, we assume the initialization satisfies $z_0 \in \mathcal{N}_x(c)$ and focus on the local refinement stage.

Firstly we define the so-called Regularity Condition.

Definition 1: A function $\ell(\cdot)$ is said to satisfy the Regularity Condition RC(μ, λ, c) with positive constants μ, λ and c, if

$$\langle\nabla\ell(z), z - x\rangle \ge \frac{\mu}{2}\|\nabla\ell(z)\|^2 + \frac{\lambda}{2}\|z - x\|^2 \tag{5}$$

for all $z \in \mathcal{N}_x(c)$.

A nice consequence is that, if a loss function $\ell(z)$ satisfies RC(μ, λ, c), then as long as $z_0 \in \mathcal{N}_x(c)$, the Gradient method converges linearly provided $0 < \alpha \le \mu$.

The goal of this paper is to further study the convergence of accelerated first-order methods for phase retrieval under the assumption that the loss function satisfies the Regularity Condition. The simplified Nesterov's method [7] is:

$$\begin{aligned} z_{k+1} &= y_k - \alpha\nabla f(y_k), \\ y_k &= (1+\beta)z_k - \beta z_{k-1}, \alpha > 0, 0 \le \beta < 1. \end{aligned} \tag{6}$$

The Heavy-ball method [8] can be expressed as:

$$z_{k+1} = z_k - \alpha\nabla\ell(z_k) + \beta(z_k - z_{k-1}), \alpha > 0, 0 \le \beta < 1. \tag{7}$$

Both Nesterov's method and Heavy-ball method have been used empirically for solving phase retrieval with encouraging performance [9], [10], however there are currently no theoretical guarantees on the linear convergence of accelerated

[1]H. Xiong and W. Zhang are with the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210. Email: {xiong.309, zhang.491}@osu.edu. The work of Xiong and Zhang is partly supported by the National Science Foundation under grant CNS-1552838.

[2]Y. Chi is with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213. The work of Y. Chi is supported in part by AFOSR under the grant FA9550-15-1-0205, by ONR under the grant N00014-18-1-2142, and by NSF under the grants CAREER ECCS-1818571 and CCF-1806154. Email: yuejiechi@cmu.edu

[3]B. Hu is with the Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison WI 53715. Email: bhu38@wisc.edu

[1]This means the probability approaches to 1 as the signal size goes to infinity. For example, the probability is at least $1 - 10e^{-\gamma n} - 8/n^2$ (γ is a fixed positive number) [4].

gradient descent under the Regularity Condition. Compared with other geometric properties [11], the Regularity Condition doesn't require the loss function to be strongly convex or smooth, and therefore is broadly applicable to analysis of other nonconvex problems sharing similar properties.

This paper provides a unified approach to analyze the convergence of first-order algorithms for phase retrieval under the Regularity Condition. We construct a dynamical system that can represent three first-order methods (Gradient, Nesterov, and Heavy-ball) and the Regularity Condition. Through stability analysis of the dynamical control system, we can numerically certify the convergence conditions of these algorithms. We also explore the analytical convergence conditions regarding the parameters of the Heavy-ball method. Although our results are stated in the context of the phase retrieval problem, they can also be used to analyze convergence of first-order methods in blind deconvolution [12] and matrix completion [13] whose loss functions also satisfy the Regularity Condition.

II. Convergence analysis

Motivated by the seminal work [14], we adopt the dynamical system viewpoint of first-order optimization algorithms. In particular, all the three first-order algorithms (4), (6), and (7) can be written as:

$$\begin{aligned} z_{k+1}^{(1)} &= (1+\beta_1) z_k^{(1)} - \beta_1 z_k^{(2)} - \alpha u_k, \\ z_{k+1}^{(2)} &= z_k^{(1)}, \\ y_k &= (1+\beta_2) z_k^{(1)} - \beta_2 z_k^{(2)}, \\ u_k &= \nabla \ell(y_k) \end{aligned} \tag{8}$$

where $(\beta_1, \beta_2) = (0,0)$ for the Gradient method, $(\beta_1, \beta_2) = (\beta, 0)$ for the Heavy-ball method and $(\beta_1, \beta_2) = (\beta, \beta)$ for the Nesterov's method.

If we define $\phi_k = \begin{bmatrix} z_k^{(1)} \\ z_k^{(2)} \end{bmatrix}$ as the state, u_k as the input and y_k as the output, the algorithms can be represented as a dynamical system shown in Figure 1, where the feedback $\nabla \ell(y_k)$ is a static nonlinearity that depends on the gradient of the loss function, and G denotes a linear system with the following state space representation:

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] = \left[\begin{array}{cc|c} (1+\beta_1) I_n & -\beta_1 I_n & -\alpha I_n \\ I_n & 0_n & 0_n \\ \hline (1+\beta_2) I_n & -\beta_2 I_n & 0_n \end{array}\right]. \tag{9}$$

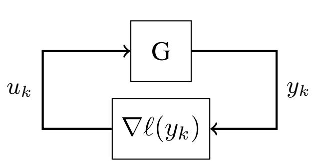

Fig. 1: Dynamical system representation of first-order optimization algorithms.

Denote the equilibrium of the dynamical system as $\phi_* = \begin{bmatrix} x \\ x \end{bmatrix}$, then $\phi_k \xrightarrow{k \to \infty} \phi_*$ implies $z_k \xrightarrow{k \to \infty} x$. In other words, the asymptotic stability of the dynamical system can indicate the convergence of the estimated signal to the ground truth.

The main challenge for stability analysis of the dynamical system (Figure 1) lies in the nonlinear feedback term $\nabla \ell(\cdot)$. Our key observation is that the Regularity Condition is equivalent to a quadratic constraint imposed on the feedback block.

Lemma 1: Suppose the loss function ℓ satisfies RC(μ, λ, c). If the k^{th} iterate of the Gradient method and the Heavy-ball method satisfies $z_k \in \mathcal{N}_x(c)$ and for the Nesterov's method $z_{k-1}, z_k \in \mathcal{N}_x(c/6)$, then the nonlinear feedback $u_k = \nabla \ell(y_k)$ can be quadratically bounded by

$$\begin{bmatrix} y_k - y_* \\ u_k - u_* \end{bmatrix}^T \begin{bmatrix} -\lambda I_n & I_n \\ I_n & -\mu I_n \end{bmatrix} \begin{bmatrix} y_k - y_* \\ u_k - u_* \end{bmatrix} \geq 0. \tag{10}$$

Thus, we convert the convergence analysis of first-order methods to the stability analysis of a dynamical system with quadratically bounded feedback. Consider a quadratic Lyapunov function $V(\phi) = (\phi - \phi_*)^T P (\phi - \phi_*)$, with $P \succeq 0$. By letting $V(\phi_k)$ decay exponentially and combing with (10), we can use S-procedure [15] to derive sufficient conditions for convergence as stated in the following theorem.

Theorem 1: Let $x \in \mathbb{R}^n$ be the true signal, and suppose that the loss function satisfies RC(μ, λ, c) with some positive constants μ, λ, c. For a given first-order method characterized by $G(A, B, C, D)$ and a fixed $0 < \rho < 1$, if there exists a matrix $P \succeq 0$ such that the following LMI (11) holds,

$$\begin{aligned} &\begin{bmatrix} A^T P A - \rho^2 P & A^T P B \\ B^T P A & B^T P B \end{bmatrix} + \\ &\begin{bmatrix} C & 0 \\ \mathbf{0}_{1\times 2} & 1 \end{bmatrix}^T \begin{bmatrix} -\lambda & 1 \\ 1 & -\mu \end{bmatrix} \begin{bmatrix} C & 0 \\ \mathbf{0}_{1\times 2} & 1 \end{bmatrix} \preceq 0, \end{aligned} \tag{11}$$

then with an initial estimate $z_0 \in \mathcal{N}_x(c/c')$, where $c' = \mathrm{cond}(P)$ for the Gradient method and the Heavy-ball method and $c' = 6\mathrm{cond}(P)$ for the Nesterov's method, the algorithm characterized by $G(A, B, C, D)$ guarantees

$$\|\phi_k - \phi_*\| \leq \sqrt{\mathrm{cond}(P)} \rho^k \|\phi_0 - \phi_*\| \text{ for all } k,$$

and thus implies the estimated signals converge to the ground truth linearly.

Remark 1: By homogeneity of (11), one can check that we can remove all dimension-dependent identity I_n without affecting the results of the LMI. Therefore, the LMI is equivalent to a smaller-size one which is independent of the dimension of signals, and thus can be efficiently solved. For simplification purposes, in (11) and all equations in the remaining, A, B, C, P do not contain I_n.

III. Numerical experiment and insights

Numerically, it is straightforward to certify the convergence conditions by solving the LMI (11) for a given ρ. For a given $\rho \in (0, 1)$ and the Regularity Condition RC(μ, λ, c), one can search over the pair (α, β) over a fine grid, and find those that (11) have feasible solutions of P. Its application on the Heavy-ball method and the Nesterov's method are shown in Figure 2. As shown in Figure 2, the stability regions of the accelerated first-order algorithms can be obtained by the

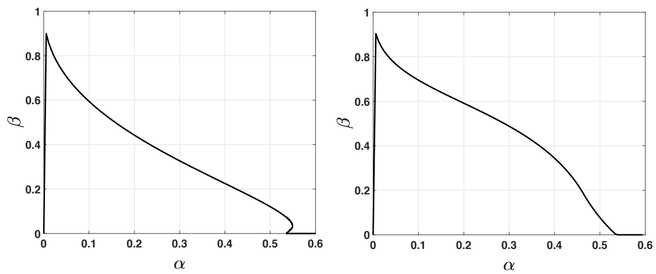

Fig. 2: Example of stability regions (all pairs of (α, β) beneath the curves) of the Heavy-ball method (left) and the Nesterov's method (right) when fixing $\rho = 0.997$ and taking the Regularity Condition parameters as $\mu = 0.5, \lambda = 0.5$.

grid search method. However, when the grid size is small, the computational complexity can be substantial. This motivates us to look for more insights to reduce complexity.

We extend the LMI (11) as

$$\begin{bmatrix} A & B \end{bmatrix}^T P \begin{bmatrix} A & B \end{bmatrix} - \rho^2 \begin{bmatrix} P & 0 \\ 0 & 0 \end{bmatrix} + M \preceq 0, \quad (12)$$

where $M = \begin{bmatrix} C & 0 \\ \mathbf{0}_{1\times 2} & 1 \end{bmatrix}^T \begin{bmatrix} -\lambda & 1 \\ 1 & -\mu \end{bmatrix} \begin{bmatrix} C & 0 \\ \mathbf{0}_{1\times 2} & 1 \end{bmatrix}$. Observe that P is PSD and M is deterministic when considering the Heavy-ball method. By the Shur complement, (12) can be equivalently stated as:

$$\begin{bmatrix} P^{-1} & B' \\ B'^T & C' \end{bmatrix} \succeq 0, \quad (13)$$

where $B' = \begin{bmatrix} A & B \end{bmatrix}, C' = \rho^2 \begin{bmatrix} P & 0 \\ 0 & 0 \end{bmatrix} - M$. Now, (13) is linear with respect to α and β. The following proposition can help reduce the computational complexity for numerical characterization of the stability region.

Proposition 1: If we fix P, ρ in (13), and S denotes the set of feasible pairs of parameters for the Heavy-ball method, that is, $S = \left\{\left(\alpha^{(i)}, \beta^{(i)}\right) : (13) \text{ is feasible}\right\}$, then all the points in the convex hull of S are also feasible solutions.

In the numerical search for the stability region of (α, β), every single pair in the sampled grid needs to be used to solve the LMI (11). The previous proposition, however, can help to reduce the work. When we fix some P and get some feasible solutions $\left\{\left(\alpha^{(i)}, \beta^{(i)}\right)\right\}_i$ corresponding to the same P, then all points in the convex hull determined by the existing points are feasible. Unfortunately, this property does not help the numerical convergence analysis of the Nesterov's method.

IV. Analytical Result

A. Convergence condition of the Gradient method

In the remaining, some analytical results are explored. In [4], the authors proved that by their Wirtinger flow algorithm, the linear convergence of the Gradient method is guaranteed when taking the step size as $\alpha \leq \mu$ (μ is determined by RC(μ, λ, c)). In our method, without changing their initialization step, we can get a less conservative stability region than that in [4].

Theorem 2: Let $x \in \mathbb{R}^n$ be the true signal and let z_0 be located in the basin where RC(μ, λ, c) holds with spectral initialization. If the step size $\alpha < \frac{2-2\sqrt{1-\mu\lambda}}{\lambda}$, then the Gradient method can guarantee the estimate z_k linearly converges to x as $k \to \infty$.

By simple calculation, we can find that $\frac{2-2\sqrt{1-\mu\lambda}}{\lambda} \geq \mu$ for all μ, λ satisfying $\mu\lambda \in (0, 1)$, which means this bound can cover the existing one.

B. Convergence condition of the Heavy-ball method

The stability region of the Heavy-ball method is more complicated to derive. (11) corresponding to the Heavy-ball method is a 3×3 matrix instead of a 2×2 one for the Gradient method. This is mainly due to that the Lyapunov function parameter P for the Heavy-ball method is of higher dimension and thus introduces more uncertain parameters. To reduce unknown parameters, we transfer the time domain matrix inequality into a frequency domain inequality (FDI) through the KYP lemma [16].

Lemma 2: (Theorem 2 in [16]) Given A, B, M, with $\det(e^{j\omega} I - A) \neq 0$ for $\omega \in \mathbb{R}$, A is Schur stable and the left upper corner of M is positive semidefinite (PSD), the following two statements are equivalent:

1) $\forall \omega \in \mathbb{R}$,
$$\begin{bmatrix} (e^{j\omega} I - A)^{-1} B \\ I \end{bmatrix}^* M \begin{bmatrix} (e^{j\omega} I - A)^{-1} B \\ I \end{bmatrix} \prec 0. \quad (14)$$
2) There exists a matrix $P \in \mathbb{R}^{n\times n}$ such that $P \succeq 0$ and
$$M + \begin{bmatrix} A^T P A - P & A^T P B \\ B^T P A & B^T P B \end{bmatrix} \prec 0. \quad (15)$$

One can easily check, however, that A and M in (11) do not satisfy these conditions. Therefore, we need to rewrite the dynamical system (9) representing the Heavy-ball method. A way to generalize the representation without changing the equilibrium is to introduce all possible uncertain parameters, such as a_{11}, a_{12}, b in the dynamics and s, t in the feedback.

$$\begin{cases} z_{k+1}^{(1)} = a_{11} z_k^{(1)} + a_{12} z_k^{(2)} + b u_k, \\ z_{k+1}^{(2)} = z_k^{(1)}, \\ \quad y_k = z_k^{(1)}, \\ \quad u_k = s\nabla\ell(y_k) + t y_k \end{cases} \quad (16)$$
$$\Leftrightarrow \left[\begin{array}{c|c} A' & B' \\ \hline C' & D' \end{array}\right] = \left[\begin{array}{cc|c} a_{11} & a_{12} & b \\ 1 & 0 & 0 \\ \hline 1 & 0 & 0 \end{array}\right],$$

By letting $z_k^{(1)} = z_k$ and $z_k^{(2)} = z_{k-1}$, the new system characterized by $G(A', B', C', D')$ can be simplified as

$$z_{k+1} = (a_{11} + bt) z_k + a_{12} z_{k-1} + bs\nabla\ell(z_k). \quad (17)$$

To make sure (16) corresponds to the Heavy-ball method, we compare (17) and (7), then we have

$$a_{11} + bt = 1 + \beta, \quad a_{12} = -\beta, \quad bs = -\alpha. \quad (18)$$

Correspondingly, the original Regularity Condition (10) is now shifted as:

$$\begin{bmatrix} y_k - y_* \\ u_k - u_* \end{bmatrix}^T M' \begin{bmatrix} y_k - y_* \\ u_k - u_* \end{bmatrix} \geq 0. \quad (19)$$

where

$$M' = \begin{bmatrix} -\left(2st + s^2\lambda + t^2\mu\right) & s + t\mu \\ s + t\mu & -\mu \end{bmatrix}.$$

By Theorem 1, we can solve the stability of the new system (16) by finding some $P \succeq 0$ to make the following time domain matrix inequality feasible.

$$\begin{bmatrix} A'^T PA' - P & A'^T PB' \\ B'^T PA' & B'^T PB' \end{bmatrix} + \begin{bmatrix} C' & 0 \\ \mathbf{0}_{1\times 2} & 1 \end{bmatrix}^T M' \begin{bmatrix} C' & 0 \\ \mathbf{0}_{1\times 2} & 1 \end{bmatrix} \prec 0. \tag{20}$$

To apply the KYP lemma to solve (20), we first figure out the region of the algorithm parameters (α, β) such that the system (16) satisfy the conditions in Lemma 2.

Lemma 3: Let ℓ be a loss function satisfying RC(μ, λ, c). To ensure that the dynamical system (16) representing the Heavy-ball method can satisfy the following conditions: $\det(e^{j\omega} I - A') \neq 0$ for $\omega \in \mathbb{R}$; A' is Schur stable; the left upper corner of M' is PSD, the step size α and the momentum parameter β should obey the following restriction:

$$0 < \alpha < \frac{2(1+\beta)(1+\sqrt{1-\mu\lambda})}{\lambda}. \tag{21}$$

Then we can solve (20) by the KYP lemma, which leads to the following analytical convergence condition for the Heavy-ball method.

Theorem 3: Let $x \in \mathbb{R}^n$ be the true signal. If there exists a neighborhood $\mathcal{N}_x(c)$ of the ground truth x such that RC(μ, λ, c) holds in the basin, then by proper initialization and taking the step size α and the momentum parameter β in the region as

$$\{(\alpha, \beta) : SR_1(\beta) \leq \alpha \leq SR_2(\beta) \text{ or } 0 \leq \alpha \leq SR_3(\beta)\},$$

where

$$SR_1(\beta) = \frac{\mu\beta^2 + 6\mu\beta + \mu}{\beta + 1}, SR_2(\beta) = \frac{2(\beta+1)(1-\sqrt{1-\mu\lambda})}{\lambda},$$
$$SR_3(\beta) = \min\left\{ SR_1(\beta), \frac{P_2 - \sqrt{P_2^2 - 4P_1P_3}}{2P_1} \right\} \tag{22}$$

and $P_1 = 4\mu\lambda\beta - \beta^2 - 1 - 2\beta$, $P_2 = 2\mu\beta + 2\mu\beta^2 - 2\mu\beta^3 - 2\mu$, $P_3 = 4\mu^2\beta^3 + 4\mu^2\beta - 6\mu^2\beta^2 - \mu^2\beta^4 - \mu^2$, the Heavy-ball method can guarantee the estimate z_k linearly converges to x as $k \to \infty$.

The KYP lemma successfully simplifies the stability region analysis of the Heavy-ball method. Similar analysis of the Nesterov's method is much more complicated due to its complex output in the dynamical system. This part of analysis will appear in our future work.

V. CONCLUSIONS

In this paper, we have developed a systematic approach to analyze the convergence of three popular first-order methods in the phase retrieval problem. A set of LMIs have been derived to numerically certify the stability regions of three first-order methods. Based on that, we also establish an efficient way to reduce the computational complexity in numerical analysis of the Heavy-ball method. Analytical conditions have been derived for the Gradient method and the Heavy-ball method.

REFERENCES

[1] R. P. Millane, "Phase retrieval in crystallography and optics," *JOSA A*, vol. 7, no. 3, pp. 394–411, 1990.
[2] J. Miao, P. Charalambous, J. Kirz, and D. Sayre, "Extending the methodology of x-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens," *Nature*, vol. 400, no. 6742, p. 342, 1999.
[3] A. Walther, "The question of phase retrieval in optics," *Journal of Modern Optics*, vol. 10, no. 1, pp. 41–49, 1963.
[4] E. J. Candes, X. Li, and M. Soltanolkotabi, "Phase retrieval via wirtinger flow: Theory and algorithms," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1985–2007, 2015.
[5] H. Zhang, Y. Liang, and Y. Chi, "A nonconvex approach for phase retrieval: Reshaped wirtinger flow and incremental algorithms," *Journal of Machine Learning Research*, vol. 18, no. 141, pp. 1–35, 2017.
[6] Y. Chen and E. Candes, "Solving random quadratic systems of equations is nearly as easy as solving linear systems," in *Advances in Neural Information Processing Systems*, 2015, pp. 739–747.
[7] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*. Springer Science & Business Media, 2003, vol. 87.
[8] B. T. Polyak, "Some methods of speeding up the convergence of iteration methods," *USSR Computational Mathematics and Mathematical Physics*, vol. 4, no. 5, pp. 1–17, 1964.
[9] E. Pauwels, A. Beck, Y. C. Eldar, and S. Sabach, "On fienup methods for sparse phase retrieval," *IEEE Transactions on Signal Processing*, 2017.
[10] Y. Zhou, H. Zhang, and Y. Liang, "Geometrical properties and accelerated gradient solvers of non-convex phase retrieval," in *Communication, Control, and Computing (Allerton), 2016 54th Annual Allerton Conference on*. IEEE, 2016, pp. 331–335.
[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 preprint arXiv:1711.10467*, 2017.
[12] X. Li, S. Ling, T. Strohmer, and K. Wei, "Rapid, robust, and reliable blind deconvolution via nonconvex optimization," *arXiv preprint arXiv:1606.04933*, 2016.
[13] S. Tu, R. Boczar, M. Simchowitz, M. Soltanolkotabi, and B. Recht, "Low-rank solutions of linear matrix equations via procrustes flow," in *International Conference on Machine Learning*, 2016, pp. 964–973.
[14] L. Lessard, B. Recht, and A. Packard, "Analysis and design of optimization algorithms via integral quadratic constraints," *SIAM Journal on Optimization*, vol. 26, no. 1, pp. 57–95, 2016.
[15] I. Pólik and T. Terlaky, "A survey of the s-lemma," *SIAM review*, vol. 49, no. 3, pp. 371–418, 2007.
[16] A. Rantzer, "On the kalman-yakubovich-popov lemma," *Systems & control letters*, vol. 28, no. 1, pp. 7–10, 1996.