Coherence retrieval using trace regularization*

Chenglong Bao†, George Barbastathis‡, Hui Ji†, Zuowei Shen†, and Zhengyun Zhang§

Abstract. The mutual intensity and its equivalent phase-space representations quantify an optical field's state of coherence and are important tools in the study of light propagation and dynamics, but they can only be estimated indirectly from measurements through a process called coherence retrieval, otherwise known as phase-space tomography. As practical considerations often rule out the availability of a complete set of measurements, coherence retrieval is usually a challenging high-dimensional ill-posed inverse problem. In this paper, we propose a trace-regularized optimization model for coherence retrieval and a provably-convergent adaptive accelerated proximal gradient algorithm for solving the resulting problem. Applying our model and algorithm to both simulated and experimental data, we demonstrate an improvement in reconstruction quality over previous models as well as an increase in convergence speed compared to existing first-order methods.

Key words. coherence retrieval, phase-space tomography, trace regularization, adaptive restart

AMS subject classifications. 78M30, 78M50, 90C22

1. Introduction.

Coherence retrieval is the unified mathematical treatment for two analogous physical measurement processes: estimating the density matrix of a quantum state, and reconstructing the mutual intensity of partially coherent light. Many, if not most coherence retrieval methods in optics originate from the seminal phase-space tomography approach [34, 27]. This reconstructed mutual intensity, or equivalent phase-space representations such as the Wigner distribution, is highly useful in that it can predict the three-dimensional distribution of light intensity after propagation through *any* known linear optical system [9, 22]. Likewise, knowledge of both the input and output coherence state of a system enables study of its dynamics [15]. Furthermore, the mutual intensity's phase-space representations enable intuitive understanding of light propagation [42] and form the physical basis for light field imaging in the field of computational photography [31, 30, 23, 50].

In a coherence retrieval experiment, stationary quasimonochromatic light, *i.e.,* narrow in temporal frequency with mean wavelength λ, in some unknown state of coherence is sent through one or more known optical systems, with the intensity measured at their outputs. The spatial coherence properties of this light source, *e.g.,* the mutual intensity, are then recovered from these measurements [43, 1, 4, 28, 24, 3, 34, 27, 18, 35, 44, 47, 45, 40]. More concretely, the electric field for quasimonochromatic light on some source plane is described by a complex-valued stochastic function $U(\boldsymbol{p})$, where $\boldsymbol{p} \in \mathbb{R}^2$ denotes a position on this plane.

*Submitted to the editors April 3, 2017.

Funding: The work of the authors was partially supported by Singapore MOE Research Grant MOE2014-T2-1-065 and MOE2012-T3-1-008. The research was supported by the National Research Foundation (NRF), Prime Ministers Office, Singapore, under its CREATE programme, Singapore-MIT Alliance for Research and Technology (SMART) BioSystems and Micromechanics (BioSyM) IRG.

†Department of Mathematics, National University of Singapore (matbc, matjh, matzuows@nus.edu.sg).

‡Singapore-MIT Alliance for Research and Technology (SMART) Centre, and Department of Mechanical Engineering, Massachusetts Institute of Technology, (gbarb@mit.edu).

§Singapore-MIT Alliance for Research and Technology (SMART) Centre, (zhengyun@smart.mit.edu).

For the special case of a fully coherent field, such as light from a laser, $U(\boldsymbol{p})$ is deterministic, yielding the amplitude and phase for the wave function. However, in this paper we consider the general case where the field is not necessarily coherent, such as one emanating from a light emitting diode (LED), and thus $U(\boldsymbol{p})$ can take on many possible values [25].[1] The mutual intensity is the correlation function describing this electric field, given by:

$$J(\boldsymbol{p}_1, \boldsymbol{p}_2) = \mathrm{E}\big[U(\boldsymbol{p}_1)U^*(\boldsymbol{p}_2)\big] = \iint \upsilon_1 \upsilon_2^* P_U(\upsilon_1, \upsilon_2; \boldsymbol{p}_1, \boldsymbol{p}_2)\,\mathrm{d}\upsilon_1\,\mathrm{d}\upsilon_2\,, \tag{1}$$

where $\mathrm{E}[\cdot]$ denotes the expected value (*i.e.,* the ensemble average), * denotes the complex conjugate, and $P_U(\upsilon_1, \upsilon_2; \boldsymbol{p}_1, \boldsymbol{p}_2)$ is the joint probability density function for $U(\boldsymbol{p}_1) = \upsilon_1$ and $U(\boldsymbol{p}_2) = \upsilon_2$. Other phase-space representations such as the Wigner distribution and ambiguity function can be obtained from the mutual intensity via Fourier transforms [42].

In typical experimental conditions, the optical field propagates linearly from a source point $\boldsymbol{p}$ to one of the M measurement points $\boldsymbol{q}_m \in \mathbb{R}^3$, traversing various parts of an experimental apparatus. This linear relationship is characterized by the possibly stochastic kernel $K(\boldsymbol{p}, \boldsymbol{q}_m)$, known as the transmission function [9] or the amplitude spread function [22]:

$$V(\boldsymbol{q}_m) = \int U(\boldsymbol{p})K(\boldsymbol{p}, \boldsymbol{q}_m)\,\mathrm{d}\boldsymbol{p}$$

where $V(\boldsymbol{q}_m)$ describes the output field at position $\boldsymbol{q}_m$. The resulting intensity is the expected value of the magnitude of the field squared, yielding the Hopkins integral:

$$I(\boldsymbol{q}_m) = \mathrm{E}\big[V(\boldsymbol{q}_m)V^*(\boldsymbol{q}_m)\big] = \iint \mathrm{E}\big[U(\boldsymbol{p}_1)U^*(\boldsymbol{p}_2)\big]\,\mathrm{E}\big[K(\boldsymbol{p}_1, \boldsymbol{q}_m)K^*(\boldsymbol{p}_2, \boldsymbol{q}_m)\big]\,\mathrm{d}\boldsymbol{p}_1\,\mathrm{d}\boldsymbol{p}_2$$

if we assume that the statistical fluctuations in the optical field are independent of the statistical fluctuations in K, typically the case in practical scenarios.

To enable numerical computation, we choose some basis $\{\,\xi_i\,\}_i$ and approximate the field on the source plane as $U(\boldsymbol{p}) \approx \sum_{i=1}^{N} u_i \xi_i(\boldsymbol{p})$ where $\{\,u_i\,\}_i$ are complex-valued random variables. This naturally yields a discretization for the mutual intensity (1) as a Hermitian positive semidefinite matrix $\boldsymbol{X} = \mathrm{E}\big(\boldsymbol{u}\boldsymbol{u}^{\mathsf{H}}\big)$ such that

$$J(\boldsymbol{p}_1, \boldsymbol{p}_2) \approx \boldsymbol{\xi}^{\mathsf{T}}(\boldsymbol{p}_1)\boldsymbol{X}\boldsymbol{\xi}^*(\boldsymbol{p}_2), \tag{2}$$

where $^{\mathsf{T}}$ denotes transpose, $^{\mathsf{H}}$ denotes conjugate transpose, $\boldsymbol{u} = \big[u_1, u_2, \ldots, u_N\big]^{\mathsf{T}} \in \mathbb{C}^N$, and $\boldsymbol{\xi}(\boldsymbol{p}) = \big[\xi_1(\boldsymbol{p}), \xi_2(\boldsymbol{p}), \ldots, \xi_N(\boldsymbol{p})\big]^{\mathsf{T}}$. The basis $\{\xi_i\}_i$ is chosen based on both ease of computation and efficiency in representing the field and mutual intensity; typical basis functions can be constructed from sinc functions, Hermite functions and the prolate spheroidal functions.

Using the definition of the mutual intensity (1) as well as our discretization scheme (2), we obtain:

$$I(\boldsymbol{q}_m) \approx \iint \boldsymbol{\xi}^{\mathsf{T}}(\boldsymbol{p}_1)\boldsymbol{X}\boldsymbol{\xi}^*(\boldsymbol{p}_2)\,\mathrm{E}\big[K(\boldsymbol{p}_1, \boldsymbol{q}_m)K^*(\boldsymbol{p}_2, \boldsymbol{q}_m)\big]\,\mathrm{d}\boldsymbol{p}_1\,\mathrm{d}\boldsymbol{p}_2 \tag{3}$$

[1] The analogous case in quantum tomography of a partially coherent field is a mixed state, whereas the fully coherent field corresponds to a pure state.

Now let us define a vector $\boldsymbol{k}_m \in \mathbb{C}^N$ with nth element $\boldsymbol{k}_m[n] = \int \xi_n(\boldsymbol{p}) K(\boldsymbol{p}, \boldsymbol{q}_m)\,\mathrm{d}\boldsymbol{p}$ being a random variable with probability density $P_{\boldsymbol{k}_m}\left(\hat{\boldsymbol{k}}_m\right)$ for each possible realization $\hat{\boldsymbol{k}}_m \in \mathbb{C}^N$. We can then construct a discretized measurement operator that characterizes propagation of light from the source to point $\boldsymbol{r}_m$:

$$\boldsymbol{K}_m = \mathrm{E}\left[\boldsymbol{k}_m^* \boldsymbol{k}_m^{\mathsf{T}}\right] = \int P_{\boldsymbol{k}_m}\left(\hat{\boldsymbol{k}}_m\right) \hat{\boldsymbol{k}}_m^* \hat{\boldsymbol{k}}_m^{\mathsf{T}}\,\mathrm{d}\hat{\boldsymbol{k}}_m\,.$$

Substituting these new definitions back into (3), we obtain that the intensity is an inner product of the source mutual intensity matrix $\boldsymbol{X}$ and the discretized measurement operator $\boldsymbol{K}_m$ (*i.e.,* the sum of the elementwise product of $\boldsymbol{X}$ with the complex conjugate of $\boldsymbol{K}_m$):

$$I(\boldsymbol{q}_m) \approx \langle \boldsymbol{K}_m, \boldsymbol{X} \rangle = \mathrm{tr}\left(\boldsymbol{K}_m^{\mathsf{H}} \boldsymbol{X}\right)$$

where $\langle \boldsymbol{A}, \boldsymbol{B} \rangle$ denotes the inner product of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, and $\mathrm{tr}(\cdot)$ denotes the trace. We note that the $\boldsymbol{K}_m$s are Hermitian, but we keep the conjugate transpose notation here to be consistent with later uses of inner products in the paper, which involve matrices which may or may not be Hermitian. In many situations, the $\boldsymbol{K}_m$s are low-rank; for example, rank-one operators were used in [44, 49].

We can model measurement noise as an additive term:

$$y_m = \mathrm{tr}(\boldsymbol{K}_m^{\mathsf{H}} \boldsymbol{X}) + n_m, \tag{4}$$

where y_m denotes the mth measured value, and n_m denotes the noise term. The n_ms can be well-approximated by zero-mean normal distributions with standard deviations σ_ms if measurements are from a standard camera sensor pixel with at least ~ 10 photons recorded and the noise level before quantization is much larger than a single quantization level, *i.e.,* if the noise is predominantly Poisson and the rate parameter is high enough.

The problem of coherence retrieval is then about recovering $\boldsymbol{X}$ from the measured intensities $\{y_m\}_m$, which can be formulated as solving a weighted least-squares semidefinite problem:

$$\underset{\boldsymbol{X}}{\text{minimize}} \quad \sum_{m=1}^{M} \frac{1}{2\sigma_m^2}\left(\mathrm{tr}(\boldsymbol{K}_m^{\mathsf{H}} \boldsymbol{X}) - y_m\right)^2 \qquad \text{subject to} \quad \boldsymbol{X} \in \mathcal{S}_+^N, \tag{5}$$

where $\mathcal{S}^N$ is the set of $N \times N$ Hermitian matrices, $\mathcal{S}_+^N$ is the set of $N \times N$ positive semidefinite matrices in $\mathcal{S}^N$, and M is the total number of measurements.

Unlike the phase retrieval problem [21, 20, 13, 12], the rank of $\boldsymbol{X}$ is generally bigger than one; only in the special case of a coherent field do we have that $\mathrm{rank}\,\boldsymbol{X} = 1$, and that is what is referred to as phase retrieval. With coherent light in phase retrieval, the goal is to reconstruct the field, represented by a *deterministic* vector $\boldsymbol{u}$; the lifting approach [13, 12] reformulates the problem as seeking the rank-one matrix $\boldsymbol{u}\boldsymbol{u}^{\mathsf{H}}$ instead, with low-rank promoters to rule out higher rank solutions. However, with coherence retrieval, we consider the more general partially coherent fields [25]—$\boldsymbol{u}$ is a *stochastic* vector, and we seek to reconstruct $\boldsymbol{X} = \mathrm{E}\left(\boldsymbol{u}\boldsymbol{u}^{\mathsf{H}}\right)$,

which can have rank greater than 1 due to $\boldsymbol{X}$ being an expectation across many possible rank-one matrices. Hence, our goal in coherence retreival is not to recover a single N-dimensional vector lifted to matrix form as was the case in phase retrieval, but rather to recover an $N \times N$ matrix, which is not necessarily rank-one and might not even be low rank.

In practice, coherence retrieval is usually an ill-posed inverse problem. Recovering the $N \times N$ coefficient matrix $\boldsymbol{X}$ requires $O(N^2)$ measurements and thus a forward operator of size $O(N^4)$, which would require prohibitively high computational and storage costs. Furthermore, even with smaller values of N, sometimes it is not straight-forward to take enough measurements for the forward operator to have a reasonable condition number. For example, in translation-only phase-space tomography, the camera would need to be translated infinitely far away as well as behind the source to capture a complete set of measurements.

To tackle these difficulties, one could introduce optical elements such as lenses [27, 26], but this can also introduce systematic error in the solution without accurate enough calibration of optical element positioning and aberrations. Another approach is to add a regularization term to (5); for example, nuclear norm regularization is used in [44] to promote low-rank solutions. However, the low-rank prior may not be reasonable in many coherence retrieval scenarios in practice, and a single scalar parameter for the regularizer may not be flexible enough, either.

This paper aims to develop an effective approach for coherence retrieval. We propose a trace-regularized model based on a penalty term physically analogous to the total intensity of light after passing through a chosen (virtual) linear system. The proposed regularization is motivated by the concept that for a set of solutions with similar likelihood, the one which is least energetic is likely to be closer to the truth—the extra intensity could simply be an artifact due to noise. Flexibility in choosing an arbitrary virtual system enables encoding additional *a priori* information about the solution as well. These concepts lead to the following trace-regularized optimization formulation for coherence retrieval:

$$\text{(6)} \qquad \underset{\boldsymbol{X}}{\text{minimize}} \quad \sum_{m=1}^{M} \frac{1}{2\sigma_m^2} \Big[\operatorname{tr}\big(\boldsymbol{K}_m^{\mathsf{H}} \boldsymbol{X}\big) - y_m \Big]^2 + \mu \operatorname{tr}\big(\boldsymbol{R}^{\mathsf{H}} \boldsymbol{X}\big) \qquad \text{subject to} \quad \boldsymbol{X} \in \mathcal{S}_+^N,$$

where $\mu > 0$ is the penalty parameter and $\boldsymbol{R} \in \mathcal{S}_+^N$ is a virtual measurement operator encoding our choice of virtual system. Given *a priori* information, we can set $\boldsymbol{R}$ to a value other than $\boldsymbol{I}$ to penalize unlikely values of $\boldsymbol{X}$. Candidates for the matrix $\boldsymbol{R}$ include diagonal weighting matrices as well as difference operators or high-pass filters constructed from wavelet tight frames. While previous literature [13, 12, 44] have used the $\boldsymbol{R} = \boldsymbol{I}$ special case to promote low-rank solutions, our goal here is not to specifically promote low-rank solutions, but rather to promote less noisy solutions and encode other *a priori* information such as smoothness or soft constraints on support.

Although many convex optimization methods can be applied to solve our new model (6), we present an efficient first-order scheme tailored for this particular problem. It is based on the accelerated proximal gradient (APG) method [6] with adaptive restart [32, 41], and we introduce a new restart criterion so that under some mild conditions on the measurement operators $\boldsymbol{K}_m$s, we can prove that the proposed algorithm is globally convergent—the generated sequence converges to a global minimum of (6). Furthermore, we also propose a sufficient condition for when our algorithm is provably *linearly* convergent. Numerical results show

the advantage of the proposed model with both simulated and experimental data, and the proposed numerical algorithm also outperforms other state-of-the-art methods in terms of computational efficiency for these data sets.

The rest of the paper is organized as follows: we explain the principles guiding our trace-regularized approach in Section 2, give a numerical algorithm to solve the proposed model in Section 3, analyze its convergence in Section 4, present numerical experiments for both simulated and experimental data in Section 5, and discuss the results in Section 6.

2. The trace-regularized coherence retrieval model. For notational brevity, we first summarize our proposed convex problem for regularized coherence retrieval as follows:

$$\underset{\boldsymbol{X}}{\text{minimize}} \quad \tfrac{1}{2}\big\|\mathscr{A}(\boldsymbol{X}) - \boldsymbol{b}\big\|_2^2 + \mu \operatorname{tr}\big(\boldsymbol{R}^{\mathsf{H}}\boldsymbol{X}\big) \qquad \text{subject to} \quad \boldsymbol{X} \in \mathcal{S}_+^N, \tag{7}$$

where linear operator $\mathscr{A} : \mathcal{S}^N \to \mathbb{R}^M$ has mth element $\mathscr{A}(\boldsymbol{X})[m] = \operatorname{tr}\big(\boldsymbol{K}_m^{\mathsf{H}}\boldsymbol{X}\big)/\sigma_m$, and vector $\boldsymbol{b} \in \mathbb{R}^M$ has mth element $\boldsymbol{b}[m] = y_m/\sigma_m$.

Our choice of using regularizers of the form $\operatorname{tr}\big(\boldsymbol{R}^{\mathsf{H}}\boldsymbol{X}\big)$ derives from several motivations. The first is that $\operatorname{tr}\big(\boldsymbol{R}^{\mathsf{H}}\boldsymbol{X}\big)$ for positive semidefinte $\boldsymbol{R}$ corresponds to a physical quantity: the resulting intensity if the source is channeled through an optical apparatus defined by measurement operator $\boldsymbol{R}$. This interpretation allows us to pose the problem as seeking the least physically energetic solution that satisfies the measurements to an acceptable degree. Minimizing the energy is a common approach in many inverse problems, and this interpretation is more widely applicable than the rank-minimization interpretation [13, 12, 44]—not many partially coherent fields are exactly low-rank despite having decaying eigenvalues.

Furthermore, by not being restricted to setting $\boldsymbol{R} = \boldsymbol{I}$, we enable some flexibility in encoding other *a priori* information. For example, we can encode the unequal likelihood of the spatial basis functions by setting $\boldsymbol{R} = \boldsymbol{W}$, where $\boldsymbol{W}$ is a diagonal matrix whose entries give the relative negative log-likelihood of the spatial basis functions. This can be used to enforce a *soft* constraint on the spatial support of the solution (see for example [7]) if the basis functions are spatially localized, *e.g.,* sinc functions. Unlike a hard spatial support constraint, this soft constraint allows us to embed uncertainty into the solution—for example, imaging an aperture using a lens will likely not result in a sharp aperture due to aberrations, so forcing the solution to be zero outside the image of the aperture is overly restrictive.

The well-studied Gaussian Schell-model source and their relatives (see for example [39, 38, 33, 36]) have smooth underlying wave functions, and the statistics of natural images also suggest a decay property in amplitude of the Fourier transform of the intensity profile [10, 19, 46]. These optical fields will be more likely to have lower energy content at higher spatial frequencies, and this suggests setting $\boldsymbol{R}$ in such a way so that $\operatorname{tr}\big(\boldsymbol{R}^{\mathsf{H}}\boldsymbol{X}\big)$ is more sensitive to high frequency content. We can do this by setting $\boldsymbol{R} = \boldsymbol{D} = \boldsymbol{H}^{\mathsf{H}}\boldsymbol{H}$, where $\boldsymbol{H}$ is defined such that the continuous function $\boldsymbol{\xi}^{\mathsf{T}}(\boldsymbol{p})\boldsymbol{H}\boldsymbol{u}$ is equal to $\boldsymbol{\xi}^{\mathsf{T}}(\boldsymbol{p})\boldsymbol{u}$ with its high frequency components boosted. Therefore, $\operatorname{tr}\big(\boldsymbol{H}^{\mathsf{H}}\boldsymbol{H}\boldsymbol{X}\big) = \operatorname{tr}\big(\boldsymbol{H}\boldsymbol{X}\boldsymbol{H}^{\mathsf{H}}\big)$ is the trace of matrix $\boldsymbol{X}$ after boosting its high frequency components, resulting in a higher penalty value for nonsmooth solutions. In this paper, we obtain good results from the use of simple matrices for $\boldsymbol{H}$ such as one that extracts the high-pass component using the Haar wavelet, whereas a more powerful and flexible approach would be to design $\boldsymbol{R}$ based on the concept of high-pass filters of wavelet tight frames [17], which has been shown to have a close relationship to the difference operators [11].

We now present a numerical example that shows how trace regularization and choice of $\boldsymbol{R}$ affects reconstruction quality in an idealized coherence retrieval scenario wherein closed-form solutions exist, in order to avoid complications due to algorithmic differences. In this example, we seek to reconstruct a one-dimensional Gaussian Schell-model source [39] with parameter $\beta = 1$ from simulated noisy measurements through an ideal apparatus, *i.e,,* the linear operator $\mathscr{A}(\cdot)$ has unit singular values and $\boldsymbol{b}$ is drawn from an i.i.d. Gaussian ensemble. For simplicity, we chose a spatial basis consisting of the first 32 Hermite functions $\phi_n(x)$ given in [39], and thus the ground truth $\boldsymbol{X}_\star \in \mathcal{S}_+^N$ is a diagonal 32×32 matrix whose nth element is equal to $2^{n-1}\big(3+\sqrt{5}\big)^{1-n}$. Note that while $\boldsymbol{X}_\star$ has decaying eigenvalues, it is not low rank. To see the effect of regularization, we consider the following four closed-form solutions:

1. $\boldsymbol{X}_{\mathrm{U}} = \boldsymbol{X}_\star + \sigma\boldsymbol{G}$ is the solution to (5) where positivity is ignored, *i.e.,* where we replace $\mathcal{S}_+^N$ with $\mathcal{S}^N$; σ gives the noise level and $\boldsymbol{G}$ is drawn from a Gaussian unitary ensemble. We use this primarily as a point of reference for the other reconstruction approaches.
2. $\boldsymbol{X}_0 = \mathrm{Proj}_{\mathcal{S}_+^N}(\boldsymbol{X}_\star + \sigma\boldsymbol{G})$ is the solution to (5).
3. $\boldsymbol{X}_{\mathrm{I}} = \mathrm{Proj}_{\mathcal{S}_+^N}(\boldsymbol{X}_\star + \sigma\boldsymbol{G} - \mu\boldsymbol{I})$ is the solution to (7) with $\boldsymbol{R} = \boldsymbol{I}$.
4. $\boldsymbol{X}_{\mathrm{D}} = \mathrm{Proj}_{\mathcal{S}_+^N}(\boldsymbol{X}_\star + \sigma\boldsymbol{G} - \mu\boldsymbol{D})$ is the solution to (7) with $\boldsymbol{R} = \boldsymbol{D} = \hat{\boldsymbol{D}}^{\mathsf{T}}\hat{\boldsymbol{D}}$, where $\hat{\boldsymbol{D}}$ acting on a discretized field $\boldsymbol{u}$ is equivalent to performing a derivative on the continuous quantity that $\boldsymbol{u}$ represents. This choice of $\boldsymbol{R}$ is used to incorporate the idea that Gaussian Schell-modes are generally smooth, and hence contain less high frequency content. With Hermite functions as the spatial basis, the (i,j) entry of $\hat{\boldsymbol{D}}$ is given by:

$$\hat{D}_{i,j} = \begin{cases} -\sqrt{j+1}, & \text{if } i = j+1 \\ \sqrt{j}, & \text{if } i = j-1 \\ 0, & \text{otherwise.} \end{cases}$$

The noise level parameter σ was allowed to take on 11 different values exponentially equally spaced between $10^{-3}\|\boldsymbol{X}_\star\|_{\mathsf{F}}$ and $10^{-1}\|\boldsymbol{X}_\star\|_{\mathsf{F}}$. For each noise level, the regularization parameter μ that minimized the average reconstruction error was found using the bisection method, with different parameters for the $\boldsymbol{R} = \boldsymbol{I}$ and $\boldsymbol{R} = \boldsymbol{D}$ cases. The resulting spread of reconstruction error across 256 realizations for each method and noise level is shown in Figure 1. While adding a $\mathrm{tr}(\boldsymbol{X})$ term to the optimization results in an improvement in reconstruction quality over the unregularized result, using $\mathrm{tr}\big(\boldsymbol{D}^{\mathsf{H}}\boldsymbol{X}\big)$ results in even less error, due to incorporating *a priori* information about the smoothness of the solution. We also display a scatter of the reconstruction error as a function of both $\mathrm{tr}(\boldsymbol{X})$ and $\mathrm{tr}\big(\boldsymbol{D}^{\mathsf{H}}\boldsymbol{X}\big)$ in Figure 2. Note that the trace regularization terms and the reconstruction error are positively correlated in the unregularized case; while desiring a less energetic solution is a good physical rule of thumb, this correlation could explain why it is good mathematically with further study.

3. Numerical algorithm.

We will now briefly review the accelerated proximal gradient (APG) method [6] and then present a novel restart condition for solving (7).

3.1. The accelerated proximal gradient method.

Define

$$f(\boldsymbol{X}) = \tfrac{1}{2}\big\|\mathscr{A}(\boldsymbol{X}) - \boldsymbol{b}\big\|_2^2 + \mu\langle \boldsymbol{R}, \boldsymbol{X}\rangle, \quad g(\boldsymbol{X}) = \iota_{\mathcal{S}_+^N}(\boldsymbol{X}) \tag{8}$$

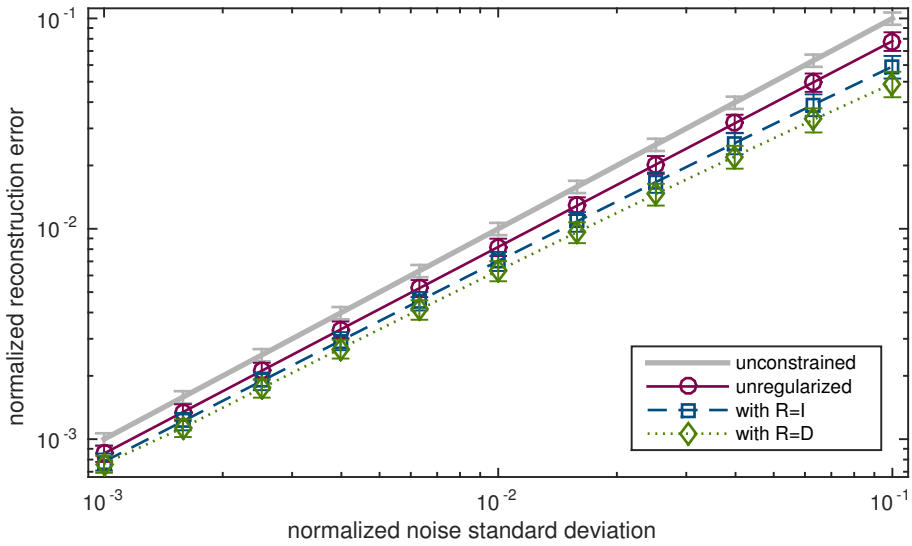

Figure 1. *Normalized reconstruction error* $\|\boldsymbol{X} - \boldsymbol{X}_\star\|_{\mathsf{F}}/\|\boldsymbol{X}_\star\|$ *as a function of the normalized noise level* $\sigma/\|\boldsymbol{X}_\star\|$. *The mean is plotted for the four different solution methods, where* $\boldsymbol{X}$ *is set to* $\boldsymbol{X}_U$, $\boldsymbol{X}_0$, $\boldsymbol{X}_I$ *and* $\boldsymbol{X}_D$ *for the "unconstrained", "unregularized", "with R=I" and "with R=D" cases, respectively. Error bars indicate three standard deviations for each case.*

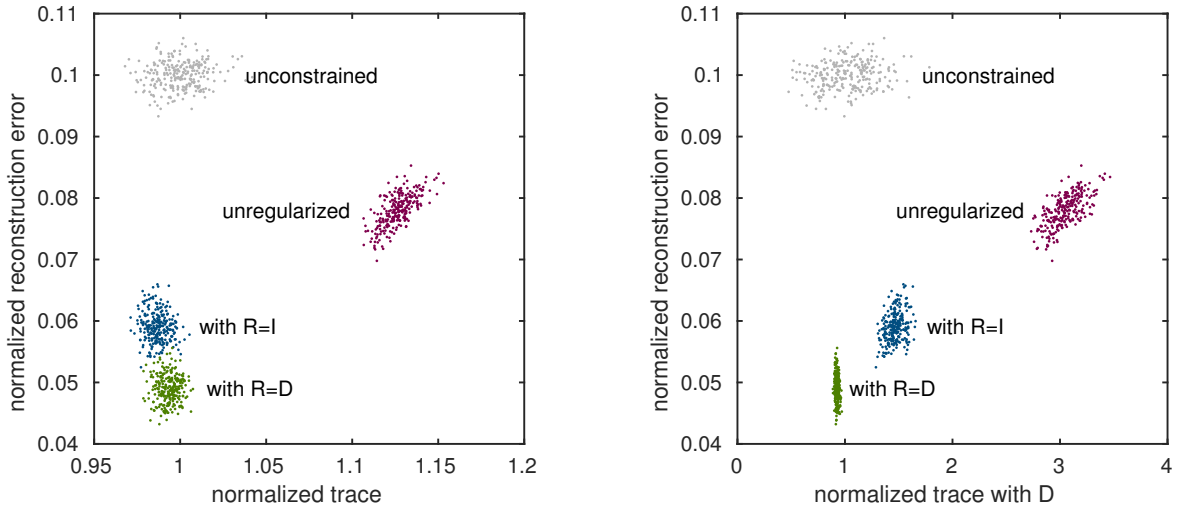

Figure 2. *Normalized reconstruction error* $\|\boldsymbol{X} - \boldsymbol{X}_\star\|_{\mathsf{F}}/\|\boldsymbol{X}_\star\|$ *versus normalized trace* $\operatorname{tr}(\boldsymbol{X})/\operatorname{tr}(\boldsymbol{X}_\star)$ *(left) and* $\operatorname{tr}(\boldsymbol{D}^{\mathsf{H}}\boldsymbol{X})/\operatorname{tr}(\boldsymbol{D}^{\mathsf{H}}\boldsymbol{X}_\star)$ *(right) for the four different reconstruction approaches, represented as a scatter plot showing each of the* 256 *realizations.*

where $\langle \boldsymbol{A}, \boldsymbol{B}\rangle = \operatorname{tr}(\boldsymbol{A}^{\mathsf{H}}\boldsymbol{B})$ is the inner product and $\iota_{\mathcal{S}_+^N}(\cdot)$ is the indicator function for the set of positive semidefinite Hermitian matrices, *i.e.*, $\iota_{\mathcal{S}_+^N}(\boldsymbol{X}) = 0$ if $\boldsymbol{X} \in \mathcal{S}_+^N$ and $\iota_{\mathcal{S}_+^N}(\boldsymbol{X}) = +\infty$ if $\boldsymbol{X} \notin \mathcal{S}_+^N$. Then, equation (7) is equivalent to the composite convex minimization problem

$$\underset{\boldsymbol{X}\in\mathcal{S}^N}{\text{minimize}}\ h(\boldsymbol{X}) = f(\boldsymbol{X}) + g(\boldsymbol{X}). \tag{9}$$

One representative method to solve (9) is the so-called accelerated proximal gradient (APG) method [6]. Given a function g and $\alpha > 0$, define the proximal operator of g as

$$\mathrm{Prox}_g^\alpha(\boldsymbol{X}) = \arg\min_{\boldsymbol{Y}} \left\{ g(\boldsymbol{Y}) + \tfrac{1}{2\alpha}\|\boldsymbol{Y} - \boldsymbol{X}\|_{\mathsf{F}}^2 \right\}$$

where $\|\cdot\|_{\mathsf{F}}$ is the Frobenius norm. Then, the APG method constructs two sequences $\{\boldsymbol{X}_k\}$ and $\{\boldsymbol{Y}_k\}$ via the following steps:

$$t_k = \left(\sqrt{4(t_{k-1})^2 + 1} + 1\right)/2, \tag{10a}$$

$$\boldsymbol{Y}_k = \boldsymbol{X}_k + \tfrac{t_{k-1}-1}{t_k}(\boldsymbol{X}_k - \boldsymbol{X}_{k-1}), \tag{10b}$$

$$\boldsymbol{X}_{k+1} = \mathrm{Prox}_g^{\alpha_k}(\boldsymbol{Y}_k - \alpha_k \nabla f(\boldsymbol{Y}_k)), \tag{10c}$$

where $t_0 = 1$, $0 < \alpha_k \le 1/L$, and L is the maximal eigenvalue of $\mathscr{A}^{\mathsf{H}}\mathscr{A}$ with $\mathscr{A}^{\mathsf{H}}$ being the adjoint operator for $\mathscr{A}$. With $\mathcal{X}_\star$ being the set of minimizers of (9) and $h_\star$ being the minimal value, it is well known that the APG method has the following convergence property [6]:

Theorem 3.1. *Let $\boldsymbol{X}_k$ be the sequence generated by (10a)-(10c) and $\alpha_k = 1/L, \forall k$. Then for any $k \ge 1$*

$$h(\boldsymbol{X}_k) - h_\star \le \frac{2L\,\mathrm{dist}(\boldsymbol{X}_0, \mathcal{X}_\star)^2}{(k+1)^2},$$

where $\mathrm{dist}(\boldsymbol{X}, \mathcal{X}_\star) = \inf\{\|\boldsymbol{X} - \boldsymbol{Y}\|_{\mathsf{F}} : \boldsymbol{Y} \in \mathcal{X}_\star\}$.

3.2. The proposed algorithm. Indeed, the attractive $O(1/k^2)$ convergence rate of the APG method is optimal when solving a convex problem when only first order information is available [29]. However, it has been observed that the objective value (the value of f in this case) sequence generated by the APG method shows oscillations, which slows down convergence speed in practice [32]. To avoid these oscillations, restarting techniques were proposed in [32] wherein t_k is reset to 1 when certain criteria are met. Despite the numerical success of this adaptive restart APG method, it is still unknown whether this method is guaranteed to converge. We therefore propose a new restart criterion that leads to a globally convergent numerical algorithm for solving the trace-regularized coherence retrieval problem (7). Given per-iteration step size $\alpha_k > 0$, define

$$\boldsymbol{Z}_{k+1} = \mathrm{Prox}_g^{\alpha_k}(\boldsymbol{Y}_k - \alpha_k \nabla f(\boldsymbol{Y}_k)), \quad \boldsymbol{U}_k = \boldsymbol{Y}_k - \boldsymbol{Z}_{k+1}, \quad \boldsymbol{V}_k = \boldsymbol{X}_k - \boldsymbol{Z}_{k+1}. \tag{11}$$

When $\boldsymbol{X}_k \neq \boldsymbol{Y}_k$, we restart our APG method by setting $t_k = 1$ if the following does not hold:

$$\langle \boldsymbol{U}_k, \boldsymbol{V}_k \rangle - \alpha_k \langle \mathscr{A}(\boldsymbol{U}_k), \mathscr{A}(\boldsymbol{V}_k) \rangle \ge \gamma \|\boldsymbol{V}_k\|_{\mathsf{F}}^2 \tag{12}$$

where $\gamma > 0$ is a small constant. We note that no extra computation is needed for checking criterion (12) due to the linearity of $\mathscr{A}$ and the necessary quantities having been computed either during step size estimation or in a previous iteration. We summarize our proposed adaptive APG algorithm in Algorithm 1.

Remark 3.1. $\mathrm{Prox}_g^{\alpha}(\boldsymbol{X})$ *is equal to a projection onto* $\mathcal{S}_+^N$*, independent of the value of* α*:*

$$\mathrm{Prox}_g^{\alpha}(\boldsymbol{X}) = \mathrm{Proj}_{\mathcal{S}_+^N}(\boldsymbol{X}) = \sum\nolimits_{i=1}^{N} \max(\lambda_i, 0)\boldsymbol{q}_i\boldsymbol{q}_i^{\mathsf{H}},$$

where $\{\,\lambda_i\,\}_i$ *and* $\{\,\boldsymbol{q}_i\,\}_i$ *are the eigenvalues and eigenvectors, respectively, of* $\boldsymbol{X}$*. The gradient of quadratic function* $f(\boldsymbol{X})$ *is equal to:*

$$\boldsymbol{\nabla} f(\boldsymbol{X}) = \mathscr{A}^{\mathsf{H}}\big[\mathscr{A}(\boldsymbol{X}) - \boldsymbol{b}\big] + \mu\boldsymbol{R}^{\mathsf{H}}.$$

In practice, the majority of computation is spent on evaluating $\mathscr{A}$ *and* $\mathscr{A}^{\mathsf{H}}$ *as well as the eigenvalue decomposition in the proximal operator. Eigenvalue decomposition scales asymptotically as* $\mathcal{O}\big(N^3\big)$*, whereas evaluating* $\mathscr{A}$ *and* $\mathscr{A}^{\mathsf{H}}$ *scales as* $\mathcal{O}\big(N^2M\big)$ *in the worst case, with* $\mathcal{O}\big(N^3\big)$ *being the best possible complexity via reduced measurements [14] or exploiting the structure of* $\mathscr{A}$ *using fast Fourier transforms or separable tensor contractions [37].*

Algorithm 1 Adaptive APG algorithm

Initialize $\boldsymbol{X}_1 = \boldsymbol{Y}_1 = \boldsymbol{Y}_0 = \boldsymbol{X}_0$, $t_1 = 1$, $k_{\max}, k_{\mathrm{maxres}} \in \mathbb{N}$, $k = 1$, $k_{\mathrm{res}} = 0$, and $\rho \in (0, 1)$
while $k \leq k_{\max}$ **do**
  Estimate step size α_k using Algorithm 2 and obtain $\boldsymbol{Z}_{k+1} = \mathrm{Prox}_g^{\alpha_k}(\boldsymbol{Y}_k - \alpha_k \boldsymbol{\nabla} f(\boldsymbol{Y}_k))$.
  if { $\boldsymbol{X}_k = \boldsymbol{Y}_k$ **or** (12) holds } and $k - k_{\mathrm{res}} \leq k_{\mathrm{maxres}}$ **then**
    Set $\boldsymbol{X}_{k+1} = \boldsymbol{Z}_{k+1}$.
    Set $t_{k+1} = \frac{1}{2}\Big(\sqrt{4t_k^2 + 1} + 1\Big)$.
    Set $\boldsymbol{Y}_{k+1} = \boldsymbol{X}_{k+1} + \frac{t_k - 1}{t_{k+1}}(\boldsymbol{X}_{k+1} - \boldsymbol{X}_k)$.
    Set $k = k + 1$.
  else
    Reset $t_k = 1$ and update $k_{\mathrm{res}} = k$.
    Set $\boldsymbol{Y}_k = \boldsymbol{X}_k$.
  end if
end while

3.2.1. Step size estimation. Setting $\alpha_k = 1/L$ might be too conservative when the Lipschitz constant L is large. We would like to adaptively choose α_k by first initializing with the Barzilai-Borwein (BB) method [5]; our choice of the form with the squared norm in the denominator is motivated by numerical considerations—an inner product is sensitive to cancellation errors and placing it in the denominator can result in numerical instability. After initialization, we then use a standard backtracking technique and adopt the step size α_k at $\boldsymbol{Y}_k$ whenever the following inequality holds:

$$h(\boldsymbol{Y}_k) - h(\boldsymbol{Z}_{k+1}) \geq \delta\|\boldsymbol{Y}_k - \boldsymbol{Z}_{k+1}\|_{\mathsf{F}}^2, \tag{13}$$

where $\delta > 0$ is a small constant. It is noted that (13) holds whenever $\alpha_k \leq 1/(L + \delta)$ and hence backtracking must terminate. In practice, we find that BB initialization gives a good estimate for the step size and that backtracking is rare. A detailed listing for this step size estimation algorithm is given in Algorithm 2.

Algorithm 2 Estimation of step size α_k

Inputs: $\boldsymbol{X}_k$, $\boldsymbol{Y}_k$, $\boldsymbol{Y}_{k-1}$, $\boldsymbol{\nabla} f(\boldsymbol{Y}_k)$, $\boldsymbol{\nabla} f(\boldsymbol{Y}_{k-1})$, $\rho < 1$ and $\alpha_{\min}, \alpha_{\max} > 0$.
Outputs: Step size α_k, proximal point $\boldsymbol{Z}_{k+1}$
if k=1 **then**
    Initialize $\beta = \|\boldsymbol{b} - \mathscr{A}(\boldsymbol{Y}_k)\|_{\mathsf{F}}^2 / \left\|\mathscr{A}^{\mathsf{H}}[\boldsymbol{b} - \mathscr{A}(\boldsymbol{Y}_k)]\right\|_{\mathsf{F}}^2$.
else
    Calculate $\boldsymbol{S}_k = \boldsymbol{Y}_k - \boldsymbol{Y}_{k-1}$ and $\boldsymbol{T}_k = \nabla f(\boldsymbol{Y}_k) - \nabla f(\boldsymbol{Y}_{k-1})$.
    Initialize $\beta = |\langle \boldsymbol{S}_k, \boldsymbol{T}_k \rangle| / \|\boldsymbol{T}_k\|_{\mathsf{F}}^2$.
end if
for $j = 1, 2 \ldots$ **do**
    Calculate $\boldsymbol{Z}_{k+1} = \mathrm{Prox}_g^{\beta}[\boldsymbol{Y}_k - \beta \boldsymbol{\nabla} f(\boldsymbol{Y}_k)]$.
    if { $\boldsymbol{X}_k \neq \boldsymbol{Y}_k$ **and** (12) fails to hold } **or** (13) holds **or** $\beta < \alpha_{\min}$ **then**
        break
    else
        Backtrack $\beta = \rho\beta$.
    end if
end for
Set $\alpha_k = \min[\max(\alpha_{\min}, \beta), \alpha_{\max}]$.

4. Convergence analysis.

In this section, we focus on convergence analysis for Algorithm 1 including an analysis regarding global convergence as well as on the convergence rate. Before proceeding, we first introduce some notation and definitions to be used in the analysis.

4.1. Notation and definitions.

Denote $\mathcal{X}_\star$ to be the solution set of (7). Given a point $\boldsymbol{x}$ and $\epsilon > 0$, we define $\mathbb{B}(\boldsymbol{x}, \epsilon) = \{\, \boldsymbol{y} : \|\boldsymbol{x} - \boldsymbol{y}\|_2 \leq \epsilon \,\}$. Given a set $\mathcal{X}$, the relative interior of $\mathcal{X}$, denoted by $\mathrm{ri}(\mathcal{X})$, is defined as

$$\mathrm{ri}(\mathcal{X}) := \{\, \boldsymbol{x} \in \mathcal{X} : \exists \epsilon > 0, \mathbb{B}(\boldsymbol{x}, \epsilon) \cap \mathrm{aff}(\mathcal{X}) \subseteq \mathcal{X} \,\},$$

where $\mathrm{aff}(\mathcal{X})$ is the affine hull of $\mathcal{X}$. Given a set $\mathcal{X}$ and a member $\boldsymbol{x}$, the distance from $\boldsymbol{x}$ to $\mathcal{X}$, denoted by $\mathrm{dist}(\boldsymbol{x}, \mathcal{X})$, is defined as $\mathrm{dist}(\boldsymbol{x}, \mathcal{X}) = \inf \{\, \|\boldsymbol{x} - \boldsymbol{y}\|_2 : \boldsymbol{y} \in \mathcal{X} \,\}$. Given a function f and $b \in \mathbb{R}$, we define the sub-level set $[f(\boldsymbol{x}) \leq b]$ to be $\{\, \boldsymbol{x} : f(\boldsymbol{x}) \leq b \,\}$. We use ∂f to denote the (limiting) subgradient of f.

Let $\mathcal{A}$ and $\mathcal{B}$ be finite dimensional Euclidean spaces and $\Gamma : \mathcal{A} \rightrightarrows \mathcal{B}$ be a *set-valued mapping*. The *graph*, *domain* and *inverse* of Γ are defined by

$$\mathrm{gph}(\Gamma) := \{\, (\boldsymbol{u}, \boldsymbol{v}) : \boldsymbol{v} \in \Gamma(\boldsymbol{u}) \,\}, \ \mathrm{dom}(\Gamma) := \{\, \boldsymbol{u} : \Gamma(\boldsymbol{u}) \neq \emptyset \,\}, \ \Gamma^{-1}(v) := \{\, \boldsymbol{u} : \Gamma(\boldsymbol{u}) = \boldsymbol{v} \,\}.$$

In the following context, we define a useful property for set-valueed mappings.

Definition 4.1. *A set-valued mapping $\Gamma : \mathcal{A} \rightrightarrows \mathcal{B}$ is said to be metrically subregular at $\bar{\boldsymbol{x}} \in \mathcal{A}$ for $\bar{\boldsymbol{y}} \in \mathcal{B}$ if $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}) \in \mathrm{gph}(\Gamma)$ and there exist $\kappa, \epsilon > 0$ such that*

$$\mathrm{dist}(\boldsymbol{x}, \Gamma^{-1}(\bar{\boldsymbol{y}})) \leq \kappa \, \mathrm{dist}(\bar{\boldsymbol{y}}, \Gamma(\boldsymbol{x})), \ \forall \boldsymbol{x} \in \mathbb{B}(\bar{\boldsymbol{x}}, \epsilon). \tag{14}$$

4.2. Global convergence. We first show that Algorithm 1 converges to a global minimum, provided the generated sequence is bounded.

Theorem 4.2. *Let $\{\boldsymbol{X}_k\}$ be the sequence generated by Algorithm 1. If $\{\boldsymbol{X}_k\}$ is bounded, then $\{\boldsymbol{X}_k\}$ converges to a global minimum of (7), denoted as $\bar{\boldsymbol{X}}$.*

Proof. See Appendix A. ∎

Remark 4.1. *The proof is based on the recently established Kurdyka-Łojasiewicz (KL) property [2, 8], which provides a framework for typical descent algorithms whose generated sequences are bounded. However, the boundedness condition might not hold for semidefinite programming problems such as (7).*

In coherence retrieval, we will show that the boundedness condition on $\{\boldsymbol{X}_k\}$ holds when the measurement operator $\mathscr{A}$ satisfies a certain mild condition. Let $\boldsymbol{A} = (\hat{\boldsymbol{K}}_1, \ldots, \hat{\boldsymbol{K}}_M)^{\mathsf{H}} \in \mathbb{C}^{MN \times N}$ where the $\hat{\boldsymbol{K}}_m$s $\in \mathbb{C}^{N \times N}$ are such that $\boldsymbol{K}_m = \hat{\boldsymbol{K}}_m \hat{\boldsymbol{K}}_m^{\mathsf{H}}$ for all $m \in \{1, \ldots, M\}$. We make the following assumption on $\boldsymbol{A}$:

Assumption 4.1. *The matrix $\boldsymbol{A}$ has full column rank.*

In practice, Assumption 4.1 always holds because either the measurements are designed to ensure $\boldsymbol{A}$ is full rank, or a smaller set of basis functions are used to remove the ambiguity, *e.g.,* we can choose a larger sampling interval in the case of sinc basis functions, or we can use the nonzero right singular vectors of $\boldsymbol{A}$ to set a new basis based on our original basis.

Proposition 4.1. *Let $\{\boldsymbol{X}_k\}$ to be the sequence generated by Algorithm 1. Suppose Assumption 4.1 holds. Then, $\{\boldsymbol{X}_k\}$ is bounded.*

Proof. See Appendix B. ∎

Combining Theorem 4.2 and Proposition 4.1, we obtain that Algorithm 1 is globally convergent under Assumption 4.1.

Corollary 4.1. *Let $\{\boldsymbol{X}_k\}$ to be the sequence generated by Algorithm 1. Suppose Assumption 4.1 holds. Then, $\{\boldsymbol{X}_k\}$ converges to a global minimum of (7), denoted as $\bar{\boldsymbol{X}}$.*

4.3. Convergence rate analysis. In this section, we first impose a reasonable condition on the solution set $\mathcal{X}_\star$. Using this condition, we prove that Algorithm 1 converges linearly.

Assumption 4.2. *We make the following assumptions on $\mathcal{X}_\star$: a) $\mathcal{X}_\star \neq \emptyset$. b) There exists an $\boldsymbol{X}_\star \in \mathcal{X}_\star$ satisfying*

$$\mathbf{0} \in \boldsymbol{\nabla} f(\boldsymbol{X}_\star) + \mathrm{ri}\Big(\mathcal{N}_{\mathcal{S}_+^N}(\boldsymbol{X}_\star)\Big) \tag{15}$$

where $\mathcal{N}_{\mathcal{S}_+^N}(\boldsymbol{X})$ denotes the normal cone of $\mathcal{S}_+^N$ at $\boldsymbol{X}$.

Remark 4.2. *The first order optimality condition of h implies*

$$0 \in \boldsymbol{\nabla} f(\boldsymbol{X}) + \mathcal{N}_{\mathcal{S}_+^N}(\boldsymbol{X}),\ \forall \boldsymbol{X} \in \mathcal{X}_\star. \tag{16}$$

Condition (15) is slightly more restrictive than (16), but (15) only needs to hold at one point of $\mathcal{X}_\star$. Moreover, from the proof of Proposition 4.2, one sufficient condition for ensuring (15) is that there exists some $\boldsymbol{X}_\star \in \mathcal{X}_\star$ such that $\mathrm{rank}(\boldsymbol{X}_\star) = N$, i.e., *$\boldsymbol{X}_\star$ is full rank.*

Based on recent work [16], we ensure ∂h is metrically sub-regular at any $\bar{\boldsymbol{X}} \in \mathcal{X}_\star$ for $\mathbf{0}$ in the next proposition.

Proposition 4.2. *Let $h = f + g$ be defined in (8). Suppose Assumption 4.2 holds. Then, for any $\bar{\boldsymbol{X}} \in \mathcal{X}_\star$, ∂h is metrically sub-regular at $\bar{\boldsymbol{X}}$ for 0.*

Proof. See Appendix C. ∎

Remark 4.3. *Proposition 3.2 in [16] is a characterization of metric sub-regularity for real symmetric positive semidefinite matrices. However, the analysis in [16] can easily be extended for Hermitian positive semidefinite matrices over $\mathbb{R}$.*

Now, we can establish local linear convergence for Algorithm 1 via the following:

Theorem 4.3. *Let $h = f + g$ be defined in (8) and the sequence $\{\boldsymbol{X}_k\}$ be generated by Algorithm 1. Suppose Assumption 4.1 and Assumption 4.2 holds. Then, there exists some $\bar{\boldsymbol{X}} \in \mathcal{X}_\star$ such that one of the following assertions holds:*
1. $\{\boldsymbol{X}_k\}$ converges to $\bar{\boldsymbol{X}}$ in finite steps.
2. $\{h(\boldsymbol{X}_k)\}$ and $\{\boldsymbol{X}_k\}$ linearly converge to $h(\bar{\boldsymbol{X}})$ and $\bar{\boldsymbol{X}}$, respectively, i.e., *there exist $c_1, c_2 > 0$, $w_1, w_2 \in (0,1)$ and $k_\ell > 0$ such that*

$$h(\boldsymbol{X}_k) - h(\bar{\boldsymbol{X}}) \le c_1 w_1^k, \text{ and } \|\boldsymbol{X}_k - \bar{\boldsymbol{X}}\|_{\mathsf{F}} \le c_2 w_2^k, \ \forall k > k_\ell.$$

Proof. See Appendix D. ∎

5. Results. We now apply our algorithm to two data sets from translation-only one-dimensional phase-space tomography [35, 44]. One is a simulation with realistic noise of two coherent Gaussian beams that are slightly decohered with respect to each other. Another is experimental data from [49] consisting of a Schell-model source imaged by a single positive lens. In both cases, the data consists of intensity profiles captured at 250 µm axial intervals by a single row of pixels in a 3.2 µm pitch camera with wavelength λ equal to 532 nm.

While our approach can be applied to more complicated tomographic apparatuses [27, 26] as long as the amplitude transfer function $K(\boldsymbol{p}, \boldsymbol{q}_m)$ is known, we focus on the translation-only one-dimensional phase-space tomography example as it is well-studied and a good base from which to extrapolate insights.

With one-dimensional phase-space tomography, the optical field as well as the mutual intensity are assumed to be constant along one of the spatial axes, *i.e.,* $U(\boldsymbol{p})$ can be written as a one-dimensional function $U(x)$, being constant along y, and thus $J(\boldsymbol{p}_1, \boldsymbol{p}_2)$ can also be written as $J(x_1, x_2)$. We assume that the optical field can be Nyquist sampled at intervals of Δ and has negligible energy when $|x| > N\Delta/2$. Thus, we employ the following sinc basis functions:

$$\xi_n(x) = \sqrt{\Delta} \sin\big[\pi(x/\Delta - n + n_0)\big] \big/ \big[\pi(x/\Delta - n + n_0)\big]$$

where $n = 1, \ldots, N$ and $n_0 = (N+1)/2$.

For translation-only one-dimensional phase-space tomography, the light from the source propagates through free space towards the measurement points $\boldsymbol{r}_m$, each of which can be fully specified by transverse position x_m and axial position z_m; a camera sensor is translated to

the various axial positions and intensity measurements are obtained from the pixel values in a single row. We assume $\Delta \gg \lambda$ and thus use the Fresnel diffraction integral to compute the forward model $\boldsymbol{k}_m$s and thus the measurement operators $\boldsymbol{K}_m$s:

$$\begin{aligned}\boldsymbol{k}_m[n] &= \int \xi_n(x) \exp\left[\frac{\mathrm{i}2\pi}{\lambda z_m}(x_m - x)^2\right] \Big/ \sqrt{\mathrm{i}\lambda z_m}\,\mathrm{d}x \\ &= \frac{\exp(\alpha_{m,n})}{2\sqrt{\mathrm{i}\lambda z_m/\Delta}} \left\{ \operatorname{erf}\left[-\sqrt{\alpha_{m,n}} + \frac{\sqrt{\mathrm{i}\pi\lambda z_m}}{2\Delta}\right] - \operatorname{erf}\left[-\sqrt{\alpha_{m,n}} - \frac{\sqrt{\mathrm{i}\pi\lambda z_m}}{2\Delta}\right] \right\},\end{aligned}$$

where i is the imaginary constant, $\sqrt{\mathrm{i}} = (1+\mathrm{i})/\sqrt{2}$, $\operatorname{erf}(\zeta) = \frac{2}{\sqrt{\pi}} \int_0^\zeta \exp(-t^2)\,\mathrm{d}t$ is the error function, and $\alpha_{m,n} = \mathrm{i}\pi(x_m - n\Delta)^2/(\lambda z_m)$. Since these $\boldsymbol{k}_m$s are deterministic, $\boldsymbol{K}_m = \boldsymbol{k}_m^{\mathsf{T}} \boldsymbol{k}_m^*$.

The constant parameters we chose for our numerical algorithm were:

$$\begin{array}{llll} \boldsymbol{X}_0 = \boldsymbol{0} & \delta = 10^{-8} & \gamma = 10^{-5} & \rho = \frac{1}{2} \\ \alpha_{\min} = 10^{-8} & \alpha_{\max} = 10^{8} & k_{\max} = 1000 & k_{\mathrm{maxres}} = 250 \end{array}$$

For comparison, we also run standard proximal gradients (PG), accelerated proximal gradients (APG) and adaptive restart accelerated proximal gradients (restart APG) using publicly available code accompanying [32]. For instrumentation, we added code to the official source code in order to record the time taken and acceleration parameter at each iteration. Furthermore, each algorithm was run a second time wherein the value of $\boldsymbol{X}$ at each iteration was recorded in order to compute the value of the merit function at each iteration. All computations were performed using `MATLAB` running on an Intel Xeon E5-2630 CPU.

5.1. Simulated Data. We simulate a mostly coherent sum of two parallel Gaussian beams with coplanar waists, as given by the following mutual intensity function:

$$\begin{aligned} J(x_1, x_2) = {} & G(x_1; x_0)G(x_2; x_0) + G(x_1; -x_0)G(x_2; -x_0) \\ & + \chi\big[G(x_1; x_0)G(x_2; -x_0) + G(x_1; -x_0)G(x_2; x_0)\big] \end{aligned}$$

where $G(x; a) = \exp\Big[-(x-a)^2/(2\sigma^2)\Big]$, $x_0 = 64\,\mu\mathrm{m}$, $\sigma = 32\,\mu\mathrm{m}$ and $\chi = 0.9$. This partially coherent field, discretized into a 51×51 mutual intensity matrix ($N = 51$) using a sampling interval of $\Delta = 6.4\,\mu\mathrm{m}$, is then propagated to 201 axial positions spaced $250\,\mu\mathrm{m}$ apart. For each axial position, we consider 101 intensity point samples spaced $3.2\,\mu\mathrm{m}$ apart for the measurements. This set of true intensities is shown in Figure 3 under the heading "noiseless". The smallest and largest singular values of matrix $\boldsymbol{A}$ as it is defined in Subsection 4.2 were $a_{\min} = 3.094$ and $a_{\max} = 7.925$, respectively.

To emulate real world conditions where the σ_ms are unknown, we simulate collecting 16 measurements for each intensity point sample, with their mean treated as the y_ms in our model and the standard deviation across these 16 samples used as an approximation to $\sigma_m\sqrt{16}$. The noisy data for each measurement is generated by first drawing from a Poisson distribution whose rate parameter is proportional to the true intensity at that point, with the sum of all the rate parameters made to equal 1.02×10^5 photons. We then add Gaussian noise with standard deviation equal to 0.01 times the maximum of all the rate parameters; this is to

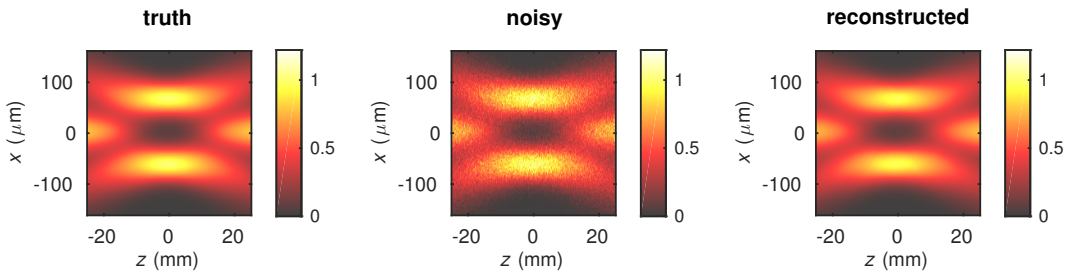

Figure 3. *Noisy measurements of the intensity are shown on left, with the true intensity shown on right. Propagated intensity of a regularized reconstructed result is shown in center. In all images, light propagates from left to right, and the intensity is in arbitrary units.*

simulate readout and quantization noise that would be typically expected of an 8-bit sensor. This set of simulated measurements is shown in Figure 3 under the heading "noisy". We then run our algorithm with the following inputs:

- `noiseless`: The y_ms are set to the ideal noiseless intensity, with the σ_ms set to 1 and $\mu = 0$ (*i.e.,* no regularization).
- `unregularized`: The y_ms and σ_ms are set to our simulated noisy data, with $\mu = 0$ (*i.e.,* no regularization).
- `nuclear`: This uses the simulated noisy data and $\boldsymbol{R} = \boldsymbol{I}$ (*i.e.,* nuclear norm regularization). μ is set such that $\frac{1}{2}\|\mathscr{A}(\boldsymbol{X}) - \boldsymbol{b}\|_2^2 \approx \alpha M/2$ upon convergence. The $\alpha M/2$ threshold for $\frac{1}{2}\|\mathscr{A}(\boldsymbol{X}) - \boldsymbol{b}\|_2^2$ arises from the fact that the norm-squared expression follows a chi-squared distribution with M degrees of freedom assuming that the σ_ms are correct values for the standard deviations of the Gaussian noise. $M/2$ would be the mean for such a distribution, and α is used to adjust the threshold to take into account how well estimated the σ_ms are and how much of the distribution we want to include. For this particular set of data, we are fairly confident of our estimates of σ_m and have thus set $\alpha = 1.5$.
- `gradient`: This uses the simulated noisy data and $\boldsymbol{R} = \boldsymbol{D}$, a tridiagonal matrix with all the elements in the diagonal equal to 1 and all the off-diagonal elements equal to $-\frac{1}{2}$. μ is set such that $\frac{1}{2}\|\mathscr{A}(\boldsymbol{X}) - \boldsymbol{b}\|_2^2 \approx \alpha M/2$ upon convergence. The penalty term $\mathrm{tr}(\boldsymbol{D}^{\mathsf{H}}\boldsymbol{X})$ is thus equivalent to applying a Haar wavelet to both the rows and columns of $\boldsymbol{X}$, keeping only the high frequency components and then taking the trace. The application of $\boldsymbol{D}$ is a simple approximation for a derivative operator on x_1 and x_2 in the continuous domain, and it physically corresponds to a desire to reduce the energy present in the first order spatial derivative of the optical wave function. Mathematically, it penalizes nonsmooth solutions to our problem, and it is motivated by our *a priori* knowledge that the true solution is smooth.
- `early stop`: This is the same as the unregularized case, except we terminate our algorithm when $\frac{1}{2}\|\mathscr{A}(\boldsymbol{X}) - \boldsymbol{b}\|_2^2$ drops below $\alpha M/2$ where $\alpha = 1.5$. This result is used as a point of reference for comparison with the regularized results, since these results all have approximately the same level of measurement mismatch, and therefore the differences are due to the presence of and choice of regularizer.

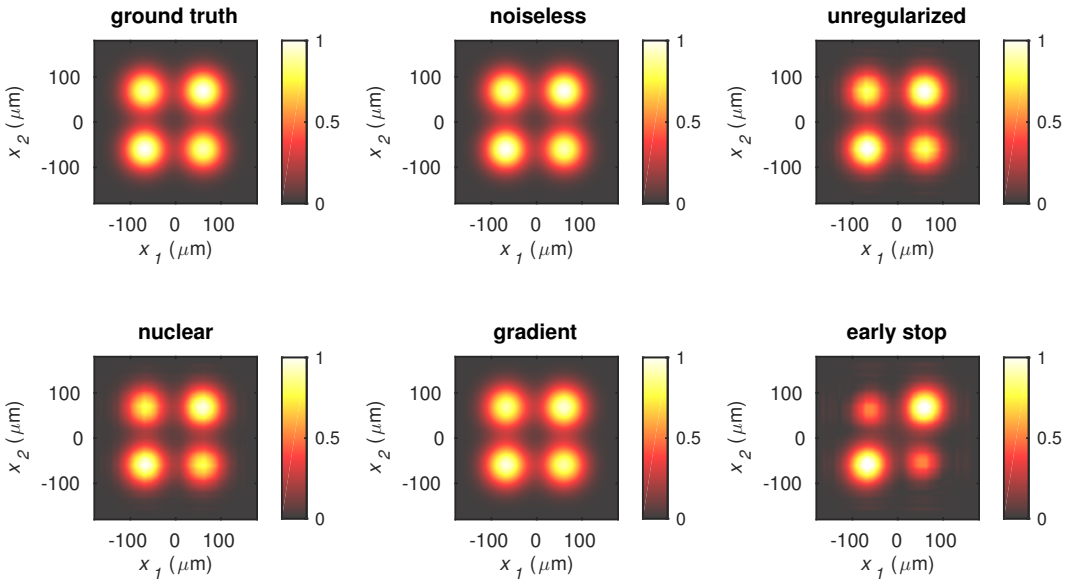

Figure 4. *Magnitude of the mutual intensity functions for the simulated data. They are all drawn using the same scale, normalized to the maximum magnitude of the ground truth.*

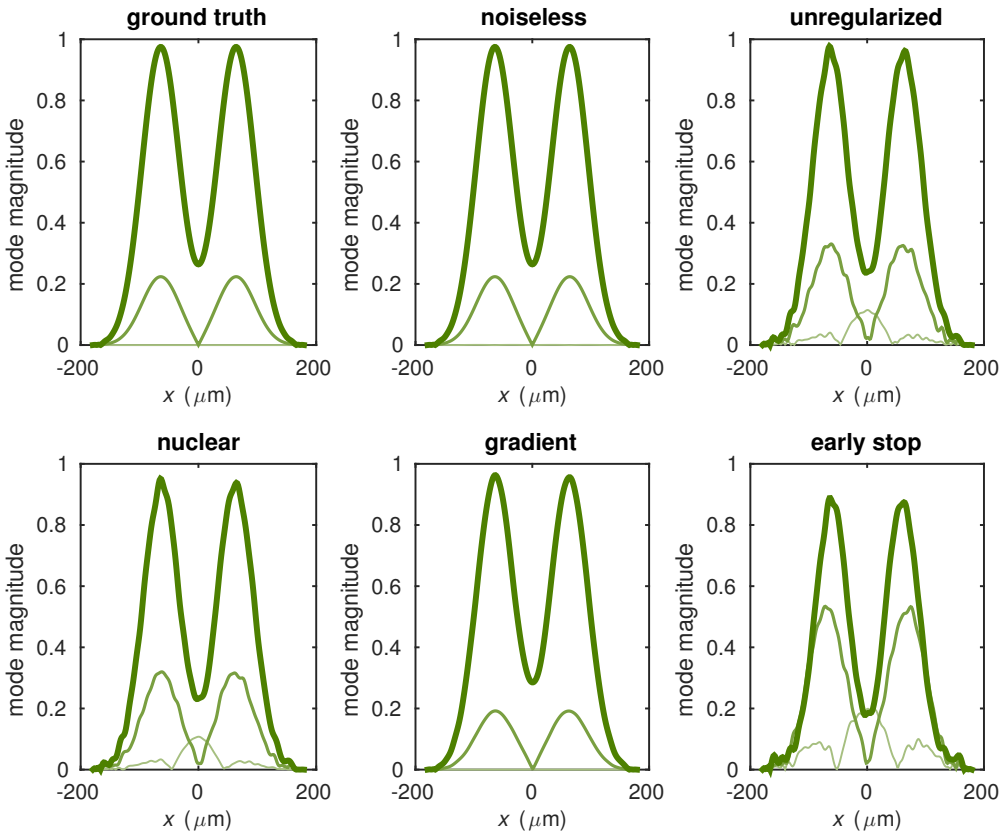

Figure 5. *Amplitude plots of the 3 highest energy coherent modes for each mutual intensity function.*

	noiseless	unregularized	nuclear	gradient	early stop
normalized error	7.524×10^{-5}	0.1147	0.1438	0.039 79	0.3977
trace distance	6.103×10^{-5}	0.079 08	0.076 58	0.025 34	0.2887

Table 1

Quantitative measurements of error for the different reconstructed mutual intensity matrices. The normalized error is defined as $\|\boldsymbol{X} - \boldsymbol{X}_{true}\|_{\mathsf{F}}/\|\boldsymbol{X}_{true}\|_{\mathsf{F}}$. *The trace distance, a quantity used to describe the difference between two quantum states, is equal to half the sum of the singular values of* $\rho - \rho_{true}$, *where* ρ *and* ρ_{true} *are equal to* $\boldsymbol{X}/\operatorname{tr}(\boldsymbol{X})$ *and* $\boldsymbol{X}_{true}/\operatorname{tr}(\boldsymbol{X}_{true})$ *respectively.*

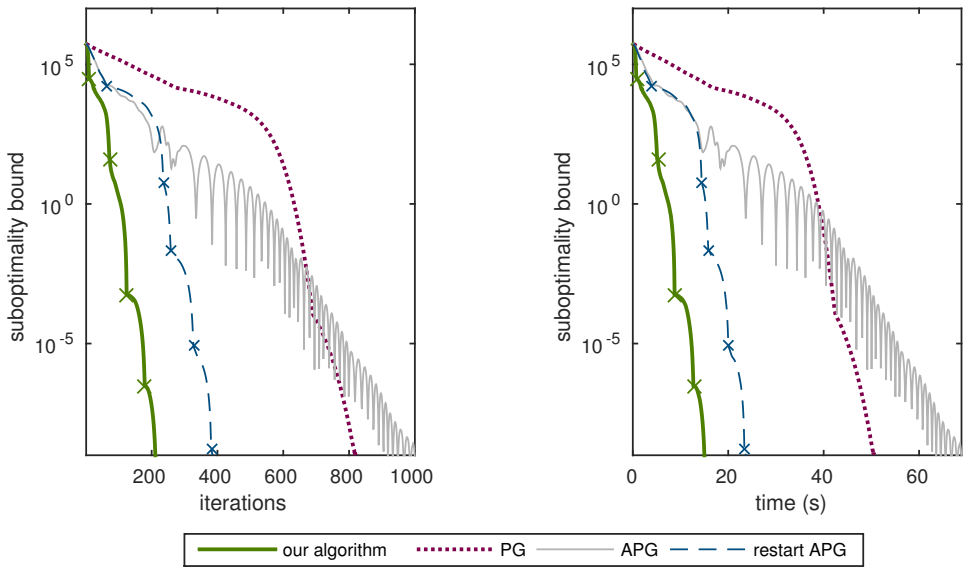

Figure 6. *Comparison of convergence across different algorithms for the simulated data. The vertical axis is the difference between the value of* $f(\boldsymbol{X})$ *and the lowest attained value of* $f(\boldsymbol{X})$ *across all algorithms, which in turn gives an upper bound on the suboptimality. On left, the horizontal axis is the number of iterations. On right, the horizontal axis is time taken. The* $\times$*s mark where restarts occurred.*

The magnitude of the reconstructed mutual intensity functions are shown in Figure 4, and the magnitude of their coherent modes [48] are given in Figure 5.[2] The propagated intensity using the `gradient` result is shown in Figure 3. Since the ground truth is known, we give a summary of the reconstruction error using various metrics in Table 1. A convergence comparison between algorithms for the `gradient` input is shown in Figure 6.

As is evident from the results, especially Figure 5, noise introduces additional energy (as seen in the increased amplitude for the second mode) as well as an additional mode, resulting in a reconstructed mutual intensity that appears *less* coherent than the ground truth. Furthermore, noise also induces additional high frequency content in the reconstructed result. For the specific value of α used, regularization using the nuclear norm only yields

[2]The ith coherent mode is $\sqrt{\lambda_i}\boldsymbol{v}_i^{\mathsf{T}}\boldsymbol{\xi}(x)$ where $\boldsymbol{X} = \sum_i \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^{\mathsf{H}}$ is $\boldsymbol{X}$'s eigenvalue decomposition.

marginal improvements over the unregularized reconstruction, whereas regularization using the smoothness-inducing regularizer both smooths the result and reduces the number of significant modes back down to the correct number of modes. Of course, we can increase the μ parameter for the nuclear norm case to reduce the number of modes, but it does not fully smooth the result and results in a mutual intensity that does not match the measurements as well as the smooth-prior regularized result. We note that the `early stop` result has the same amount of measurement mismatch as the two regularized solutions, showing how ill-posed the problem is and the necessity for regularization.

Our algorithm also converges faster than the three other methods given in Figure 6, albeit restart APG converges asymptotically as fast. Non-restarting APG oscillates due to excess critical momentum, as described in [32]. The advantages of our algorithm are that: (1) it is provably convergent, whereas to the best of our knowledge no proof of convergence has been given for the algorithm given in [32], and (2) it converges much faster than the provably convergent non-restarting APG (*i.e.,* FISTA [6]).

5.2. Experimental Data. We use the experimental data from [49], wherein the intensity profile of a partially coherent beam is imaged at 201 positions along the optical axis. The partially coherent beam was generated by focusing an LED light source through a 532 nm bandpass filter onto a 100 µm slit located at the front focal plane of a 100 mm focal length cylindrical lens. A 500 µm slit placed at the back focal plane is illuminated by light passing through the cylindrical lens, and this slit is imaged using a 50 mm cylindrical lens placed 150 mm after the slit. Based on visual inspection, the axial positions captured were located between $z = -30.25\,\text{mm}$ and $z = 19.75\,\text{mm}$ relative to the image of the slit.

The experimental data y_ms are taken from a single row on a camera with 3.2 µm pitch, and the standard deviation σ_ms were estimated from the neighboring 16 rows. A visualization of the measured and theoretical intensities are shown in Figure 7. We use a sinc basis with sampling interval $\Delta = 6.4\,\mu\text{m}$ to discretize the mutual intensity into a 101×101 matrix ($N = 101$). We note that the field does not necessarily have a spatial band-limit compatible with the sampling interval, but since we only measure the intensity at intervals of 3.2 µm, it would be very difficult to recover any information at a higher sampling rate and hence we ignore the higher frequency components. The smallest and largest singular values of the matrix $\boldsymbol{A}$ as it is defined in Subsection 4.2 were $a_{\min} = 7.625$ and $a_{\max} = 14.18$.

An aberration-free theoretical estimate of the mutual intensity is:

$$\begin{aligned} J(x_1, x_2) \propto\ & \operatorname{rect}(\beta_1 x_1)\operatorname{rect}(\beta_1 x_2)\operatorname{sinc}\big[\beta_2(x_1 - x_2)\big]\exp\big[\mathrm{i}\beta_3(x_1^2 - x_2^2)\big] \\ & \otimes \operatorname{sinc}(x_1/\Delta)\operatorname{sinc}(x_2/\Delta) \end{aligned}$$

where $\beta_1 = 4\,\text{mm}^{-1}$, $\beta_2^{-1} = 532\,\mu\text{m}$, $\beta_3 = 1.9684 \times 10^{-4}\,\mu\text{m}^{-2}$, and $\otimes$ denotes convolution. Since ground truth is not available, we use this aberration-free estimate as a rough point of reference; it is not intended to be interpreted as the ground truth, which may be slightly blurred or distorted by aberrations.

We again use several different sets of parameters for our algorithm, although we replaced two of the input sets and used a different value of α for the target value of $\frac{1}{2}\|\mathscr{A}(\boldsymbol{X}) - \boldsymbol{b}\|_2^2$ in the regularized reconstructions. Instead of the `early stop` and `noiseless` input parameter sets, we added the `nuclear+support` and `window` input parameter sets. The former uses the

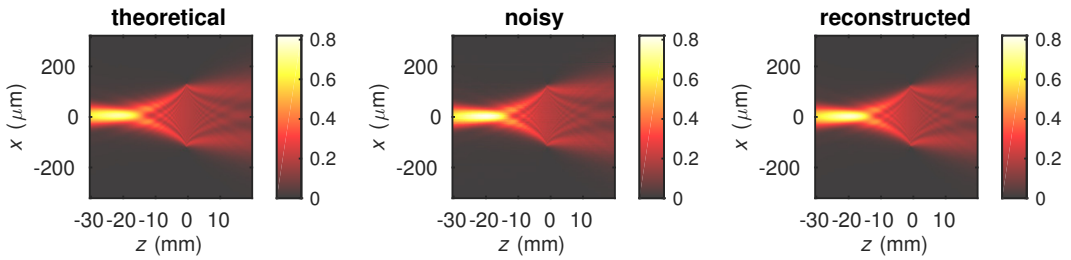

Figure 7. *Noisy measurements of the intensity are shown on left, with the theoretical aberration-free intensity profile shown on the right. Propagated intensity of a regularized reconstructed result is shown in center. In all images, light propagates from left to right, and the intensity is in arbitrary units. The theoretical intensity is provided as a point of reference and is not necessarily the ground truth.*

same regularizer as the `nuclear` data set, but additionally forces coefficients of basis functions whose centers lie outside the center 250 µm region to be zero, as a way to demonstrate the state of the art nuclear norm regularizer combined with a *hard* support constraint. The `window` input parameter set uses $\boldsymbol{R} = \boldsymbol{W}$, a diagonal matrix whose entries are unity in the center 250 µm region and increase linearly away from this region, up to a maximum of 391 at the ends. The idea with these additional input sets is to demonstrate how one can incorporate *a priori* information about the support of the solution—since we know our slit should be imaged to a region that wide, we would like to penalize any contributions outside of this region. While `nuclear+support` uses a hard constraint, it is not necessarily appropriate because the field may not actually be exactly zero outside the region, due to the presence of possible aberrations in the system. The `window` approach is a gentler way of finding a less energetic solution while at the same time preferring to remove energy from areas where we do not expect much energy to be present, *i.e.,* it is a *soft* support constraint. Coincidentally, this regularizer is also equivalent to imposing a smoothness constraint on the intensity in the far field of the partially coherent beam. We also chose a value of $\alpha = 5$ to account for additional possible errors in $\mathscr{A}$ (due to imperfect equipment and calibration) and standard deviation estimation.

	`unregularized`	`nuclear+support`	`nuclear`	`gradient`	`window`
normalized RMSE	0.2363	0.2336	0.2317	0.2189	0.2176
trace distance	0.1795	0.1717	0.1666	0.1743	0.1486

Table 2

Quantitative measurements of the difference between the mutual intensity matrices reconstructed from experimental data and the aberration-free mutual intensity estimate. The quantities are defined in the same way as Table 1, with the aberration-free mutual intensity estimate taken as the "truth".

The reconstructed mutual intensity functions are shown in Figure 8 with modes shown in Figure 9. We give quantitative comparisons between the theoretical mutual intensity and the reconstructed mutual intensity in Table 2, and a comparison of algorithm convergence for the `window` input set is given in Figure 10. All regularized solutions remove some amount of noise, as evident in the increased smoothness and reduction of high order mode energy in the coherent modes visualization. The `nuclear` result cleans up the reconstruction compared to

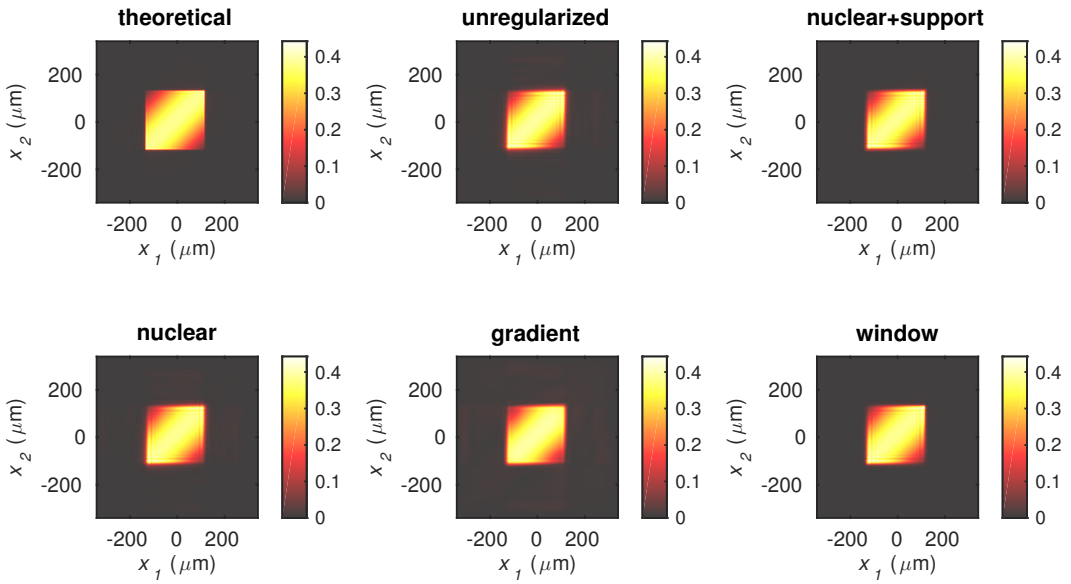

Figure 8. *Magnitude of the mutual intensity functions for the experimental data. They are all drawn using the same scale, with values in arbitrary units.*

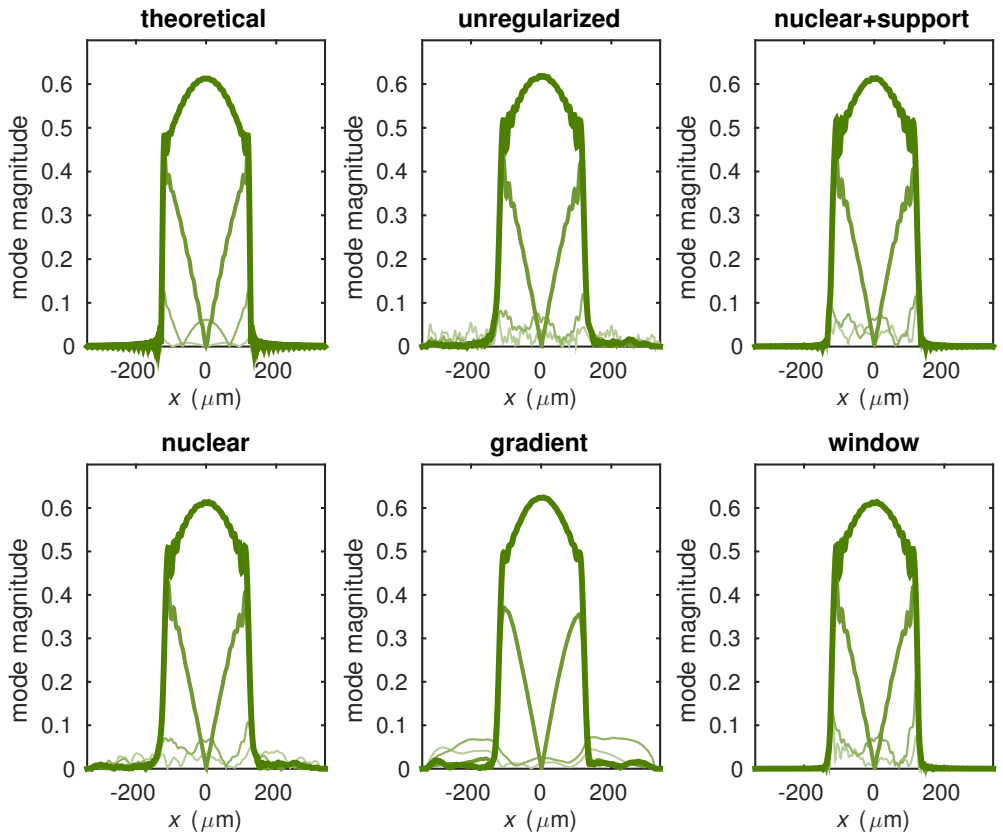

Figure 9. *Magnitude plots of the 4 highest energy coherent modes for each mutual intensity function.*

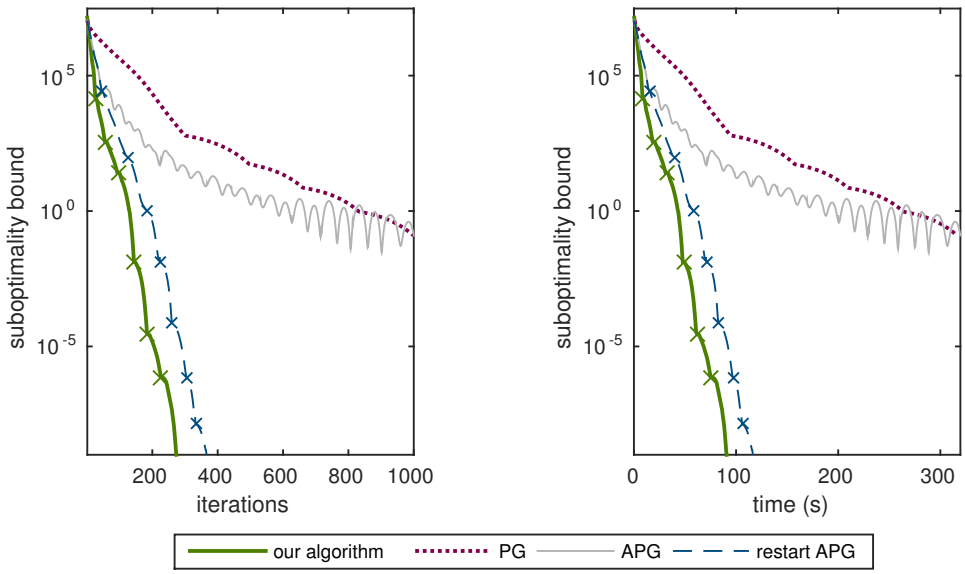

Figure 10. *Comparison of convergence across different algorithms for the experimental data. The vertical axis is the difference between the value of $f(\boldsymbol{X})$ and the lowest attained value of $f(\boldsymbol{X})$ across all algorithms, which in turn gives an upper bound on the suboptimality. On left, the horizontal axis is the number of iterations. On right, the horizontal axis is time taken. The $\times$s mark where restarts occurred.*

the noisy reconstruction, but it still leaves a lot of excess energy in the higher order modes, especially energy outside the region occupied by the slit. The `gradient` regularizer is good at reducing the number of modes, but it oversmoothes the result—the third mode spills outside of the region occupied by the slit and resembles neither the third mode from the theoretical results nor that of the noisy results, and the sharp edges of the second mode are gone. The image of the magnitude of the mutual intensity in Figure 8 is also quite blurry. While the `nuclear` result is an improvement over the unregularized result, it is obvious that applying additional prior information about the support yields a much better reconstruction, as can be seen in the `nuclear+support` and `window` cases. However, the application of a hard support constraint might not be suitable in this particular situation, as we do not know about the extent of aberrations in the imaging system. Furthermore, `window` does manage to reconstruct the third mode better than all of the other methods; `nuclear+support` still does not perform as well, and leaves excess energy in the fourth mode.

The experimental data convergence results in Figure 10 are quite similar to the one for the simulated data—our algorithm has an asymptotic convergence rate comparable to gradient restart APG, albeit we again reach the fast convergence regime faster.

6. Conclusion. We have demonstrated that trace-regularized coherence retrieval can be a powerful tool in recovering the mutual intensity when the inverse problem is ill-conditioned. The generalization of the nuclear-norm enables flexibility in applying *a priori* information, leading to higher quality reconstructions. Furthermore, we have demonstrated an efficient

numerical scheme for our coherence retrieval model, with performance at worst comparable to the state-of-the-art adaptive restart APG scheme while simultaneously being provably globally convergent, with mild conditions required for linear convergence.

This work uses very simple $\boldsymbol{R}$ matrices for regularization, with good results, but more flexibility and power can be attained by leveraging the framework of tight frames through further study. Furthermore, a method for exploiting redundant information in the measurements to calibrate real-world $\boldsymbol{K}_m$s as well as methods to reduce the memory and computational footprint for high dimensional structured mutual intensity matrices are all potential avenues for future exploration.

REFERENCES

[1] T. Asakura, H. Fujii, and K. Murata, *Measurement of spatial coherence using speckle patterns*, Optica Acta: Int. J. Opt., 19 (1972), pp. 273–290.

[2] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, *Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality*, Math. Oper. Res., 35 (2010), pp. 438–457.

[3] J. C. Barreiro and J. Ojeda-Castañeda, *Degree of coherence: a lensless measuring technique*, Opt. Lett., 18 (1993), pp. 302–304.

[4] H. O. Bartelt, K.-H. Brenner, and A. W. Lohmann, *The Wigner distribution function and its optical production*, Opt. Commun., 32 (1980), pp. 32–38.

[5] J. Barzilai and J. M. Borwein, *Two-point step size gradient methods*, IAM J. Numer. Anal., 8 (1988), pp. 141–148.

[6] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202.

[7] G. L. Besnerais, S. Lacour, L. M. Mugnier, E. Thiebaut, G. Perrin, and S. Meimon, *Advanced imaging methods for long-baseline optical interferometry*, IEEE Journal of Selected Topics in Signal Processing, 2 (2008), pp. 767–780.

[8] J. Bolte, S. Sabach, and M. Teboulle, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Math. Prog., 146 (2014), pp. 459–494.

[9] M. Born and E. Wolf, *Principles of Optics, 7th. ed.*, Cambridge University Press, Cambridge, UK, 2005.

[10] G. J. Burton and I. R. Moorhead, *Color and spatial structure in natural scenes*, Appl. Opt., 26 (1987), pp. 157–170.

[11] J.-F. Cai, B. Dong, S. Osher, and Z. Shen, *Image restoration: total variation, wavelet frames, and beyond*, J. Am. Math. Soc., 25 (2012), pp. 1033–1089.

[12] E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski, *Phase retrieval via matrix completion*, SIAM J. Imaging Sci., 6 (2013), pp. 199–225.

[13] A. Chai, M. Moscoso, and G. Papanicolaou, *Array imaging using intensity-only measurements*, Inverse Probl., 27 (2011), p. 015005.

[14] J. Chen, H. Dawkins, Z. Ji, N. Johnston, D. Kribs, F. Shultz, and B. Zeng, *Uniqueness of quantum states compatible with given measurement results*, Phys. Rev. A, 88 (2013), p. 012109.

[15] J. N. Clark, X. Huang, R. J. Harder, and I. K. Robinson, *Dynamic imaging using ptychography*, Phys. Rev. Lett., 112 (2014), p. 113901.

[16] Y. Cui, D. Sun, and K.-C. Toh, *On the asymptotic superlinear convergence of the augmented lagrangian method for semidefinite programming with multiple solutions*, arXiv preprint arXiv:1610.00875, (2016).

[17] B. Dong and Z. Shen, *MRA-based wavelet frames and applications*, IAS/Park city mathematics series: the mathematics of image processing, 19 (2010), pp. 7–158.

[18] D. Dragoman, *Unambiguous coherence retrieval from intensity measurements*, J. Opt. Soc. Am. A, 20 (2003), pp. 290–295.

[19] D. J. Field, *Relations between the statistics of natural images and the response properties of cortical*

cells, J. Opt. Soc. Am. A, 4 (1987), pp. 2379–2394.
[20] J. R. Fienup, *Phase retrieval algorithms: A comparison*, Appl. Opt., 21 (1982), pp. 2758–2769.
[21] R. W. Gerchberg and W. O. Saxton, *A practical algorithm for the determination of the phase from image and diffraction plane pictures*, Optik, 35 (1972), pp. 237–246.
[22] J. W. Goodman, *Statistical Optics*, Wiley Interscience, New York, 1985.
[23] M. Levoy, Z. Zhang, and I. McDowall, *Recording and controlling the 4D light field in a microscope*, J. Microsc., 235 (2009), pp. 144–162.
[24] Y. Li, G. Eichmann, and M. Conner, *Optical Wigner distribution and ambiguity function for complex signals and images*, Opt. Commun., 67 (1988), pp. 177–179.
[25] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics*, Cambridge University Press, Sept. 1995.
[26] D. M. Marks, R. A. Stack, and D. J. Brady, *Astigmatic coherence sensor for digital imaging*, Opt. Lett., 25 (2000), pp. 1726–1728.
[27] D. F. McAlister, M. Beck, L. Clarke, A. Mayer, and M. G. Raymer, *Optical phase retrieval by phase-space tomography and fractional-order Fourier transforms*, Opt. Lett., 20 (1995), pp. 1181–1183.
[28] M. Michalski, E. E. Sicre, and H. J. Rabal, *Display of the complex degree of coherence due to quasi-monochromatic spatially incoherent sources*, Opt. Lett., 10 (1985), pp. 585–587.
[29] Y. Nesterov, *Introductory lectures on convex optimization: A basic course*, vol. 87, Springer Science & Business Media, 2013.
[30] R. Ng, *Fourier slice photography*, in Proc. ACM SIGGRAPH, 2005.
[31] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, *Light field photography with a hand-held plenoptic camera*, Tech. Report CTSR 2005-02, Stanford University, 2005.
[32] B. O'Donoghue and E. Candès, *Adaptive restart for accelerated gradient schemes*, Found. Comp. Math., 15 (2015), pp. 715–732.
[33] Y. Qiu, H. Guo, and Z. Chen, *Paraxial propagation of partially coherent hermitegauss beams*, Optics Communications, 245 (2005), pp. 21 – 26.
[34] M. G. Raymer, M. Beck, and D. McAlister, *Complex wave-field reconstruction using phase-space tomography*, Phys. Rev. Lett., 72 (1994), pp. 1137–1140.
[35] C. Rydberg and J. Bengtsson, *Numerical algorithm for the retrieval of spatial coherence properties of partially coherent beams from transverse intensity measurements*, Opt. Express, 15 (2007), pp. 13613–13623.
[36] S. Sahin and O. Korotkova, *Light sources generating far fields with tunable flat profiles*, Opt. Lett., 37 (2012), pp. 2970–2972.
[37] J. Shang, Z. Zhang, and H. K. Ng, *Superfast maximum-likelihood reconstruction for quantum tomography*, Phys. Rev. A, 95 (2017), p. 062336.
[38] R. Simon and N. Mukunda, *Twisted gaussian schell-model beams*, J. Opt. Soc. Am. A, 10 (1993), pp. 95–109.
[39] A. Starikov and E. Wolf, *Coherent-mode representation of Gaussian Schell-model sources and of their radiation fields*, J. Opt. Soc. Am., 72 (1982), pp. 923–928.
[40] B. Stoklasa, L. Motka, J. Rehacek, Z. Hradil, and L. Sánchez-Soto, *Wavefront sensing reveals optical coherence*, Nat. Commun., 5 (2014).
[41] W. Su, S. Boyd, and E. Candes, *A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights*, in NIPS, 2014, pp. 2510–2518.
[42] M. Testorf, B. Hennelly, and J. Ojeda-Castañeda, *Phase-Space Optics: Fundamentals and Applications*, McGraw-Hill Education, 2009.
[43] B. J. Thompson and E. Wolf, *Two-beam interference with partially coherent light*, J. Opt. Soc. Am., 47 (1957), p. 895.
[44] L. Tian, J. Lee, S. B. Oh, and G. Barbastathis, *Experimental compressive phase space tomography*, Opt. Express, 20 (2012), pp. 8296–8308.
[45] L. Tian, Z. Zhang, J. C. Petruccelli, and G. Barbastathis, *Wigner function measurement using a lenslet array*, Opt. Express, 21 (2013), pp. 10511–10525.
[46] D. J. Tolhurst, Y. Tadmor, and T. Chao, *Amplitude spectra of natural images*, Ophthalmic and Physiological Optics, 12 (1992), pp. 229–232.

[47] L. Waller, G. Situ, and J. W. Fleischer, *Phase-space measurement and coherence synthesis of optical beams*, Nat. Photon., 6 (2012), pp. 474–479.
[48] E. Wolf, *New theory of partial coherence in the space-frequency domain. Part I: spectra and cross spectra of steady-state sources*, J. Opt. Soc. Am., 72 (1982), pp. 343–351.
[49] Z. Zhang, Z. Chen, S. Rehman, and G. Barbastathis, *Factored form descent: a practical algorithm for coherence retrieval*, Opt. Express, 21 (2013), pp. 5759–5780.
[50] Z. Zhang and M. Levoy, *Wigner distributions and how they relate to the light field*, in ICCP, 2009, pp. 1–10.

Appendix A. Proof of Theorem 4.2.

Given $\eta \in (0, +\infty]$, define Φ_η be the class of all concave and continuous functions $\phi : [0, \eta) \to \mathbb{R}_+$ that satisfy $\phi(0) = 0$, ϕ is C^1 on $(0, \eta)$ and continuous at 0, and $\phi^{'}(s) > 0,\ \forall s \in (0, \eta)$.

Definition A.1. *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be proper and lower semicontinuous. The function f is said to satisfy the* Kurdyka-Łojasiewicz *(KL) inequality at $\bar{\boldsymbol{x}} \in \mathrm{dom}(\partial f)$ if there exist $\eta \in (0, +\infty]$, a neighborhood $\mathcal{U}$ of $\bar{\boldsymbol{x}}$ and a function $\phi \in \Phi_\eta$, such that for all $\boldsymbol{x} \in \mathcal{U} \cap [f(\bar{\boldsymbol{x}}) < f(\boldsymbol{x}) < f(\bar{\boldsymbol{x}}) + \eta]$, the following inequality holds:*

$$\phi^{'}(f(\boldsymbol{x}) - f(\bar{\boldsymbol{x}}))\,\mathrm{dist}(\boldsymbol{0}, \partial f(\boldsymbol{x})) \geq 1. \tag{17}$$

A function $f(\boldsymbol{x})$ is called a KL function if f satisfies the KL property at every $\boldsymbol{x} \in \mathrm{dom}(\partial f)$.

We will now show some basic properties of Algorithm 1 in the following lemmas and use them to prove the global convergence of Algorithm 1.

Lemma A.1. *Let $\{\boldsymbol{X}_k\}$ be the sequence generated by Algorithm 1. Then, there exists $a, b > 0$ and $\bar{w} \in [0, 1)$ such that for all $k \geq 0$, we have*

$$h(\boldsymbol{X}_k) - h(\boldsymbol{X}_{k+1}) \geq a\|\boldsymbol{X}_k - \boldsymbol{X}_{k+1}\|_{\mathsf{F}}^2, \tag{18}$$

$$\mathrm{dist}(0, \partial h(\boldsymbol{X}_{k+1})) \leq b(\|\boldsymbol{X}_{k+1} - \boldsymbol{X}_k\|_{\mathsf{F}} + \bar{w}\|\boldsymbol{X}_k - \boldsymbol{X}_{k-1}\|_{\mathsf{F}}). \tag{19}$$

Proof. Define

$$\boldsymbol{W}_{k+1} = \mathrm{Prox}_g^{\alpha_k}(\boldsymbol{X}_k - \alpha_k \nabla f(\boldsymbol{X}_k)). \tag{20}$$

Let $\Omega_1 = \{\, k : \boldsymbol{X}_{k+1} = \boldsymbol{W}_{k+1}, k > 1 \,\}$ and $\Omega_2 = \{\, k : \boldsymbol{X}_{k+1} = \boldsymbol{Z}_{k+1} \,\}$ where $\boldsymbol{Z}_{k+1}$ is defined in (11). Then, $\Omega_1 \cap \Omega_2 = \emptyset$ and $\Omega_1 \cup \Omega_2 = \mathbb{N}$. We consider two cases: (I) $k \in \Omega_1$ and (II) $k \in \Omega_2$. **Case I.** $k \in \Omega_1$ implies that $\boldsymbol{X}_k = \boldsymbol{Y}_k$, and hence $h(\boldsymbol{X}_k) - h(\boldsymbol{X}_{k+1}) \geq \delta\|\boldsymbol{X}_k - \boldsymbol{X}_{k+1}\|_{\mathsf{F}}^2$ due to inequality (13) from the step size estimation process. From the optimality condition of (20), we know

$$\tfrac{1}{\alpha_k}(\boldsymbol{X}_k - \boldsymbol{X}_{k+1}) - \nabla f(\boldsymbol{X}_k) \in \partial g(\boldsymbol{X}_{k+1}). \tag{21}$$

The inequality (21) implies

$$\begin{aligned}\mathrm{dist}(0, \partial h(\boldsymbol{X}_{k+1})) &\leq \left\|\nabla f(\boldsymbol{X}_{k+1}) + \tfrac{1}{\alpha_k}(\boldsymbol{X}_k - \boldsymbol{X}_{k+1}) - \nabla f(\boldsymbol{X}_k)\right\|_{\mathsf{F}} \\ &\leq (L + 1/\alpha_{\min})\|\boldsymbol{X}_{k+1} - \boldsymbol{X}_k\|_{\mathsf{F}}.\end{aligned}$$

since $\alpha_k \geq \alpha_{\min} > 0, \forall k$.
Case II: $k \in \Omega_2$**.** From the first order optimality condition of (11), we have

$$\tfrac{1}{\alpha_k}(\boldsymbol{Y}_k - \boldsymbol{Z}_{k+1}) - \nabla f(\boldsymbol{Y}_k) \in \partial g(\boldsymbol{Z}_{k+1}). \tag{22}$$

Then, together with the convexity of h and the fact that $\boldsymbol{X}_{k+1} = \boldsymbol{Z}_{k+1}$, we have

$$\begin{aligned} h(\boldsymbol{X}_k) - h(\boldsymbol{X}_{k+1}) &\geq \left\langle \boldsymbol{\nabla} f(\boldsymbol{Z}_{k+1}) + \tfrac{1}{\alpha_k}(\boldsymbol{Y}_k - \boldsymbol{Z}_{k+1}) - \boldsymbol{\nabla} f(\boldsymbol{Y}_k), \boldsymbol{X}_k - \boldsymbol{Z}_{k+1} \right\rangle \\ &= \left\langle \left(\tfrac{1}{\alpha_k}\mathbf{I} - \mathscr{A}^{\mathsf{H}}\mathscr{A}\right)(\boldsymbol{Y}_k - \boldsymbol{Z}_{k+1}), \boldsymbol{X}_k - \boldsymbol{Z}_{k+1} \right\rangle \\ &\geq \frac{\gamma}{\alpha_{\max}} \|\boldsymbol{X}_k - \boldsymbol{X}_{k+1}\|_{\mathsf{F}}^2, \end{aligned}$$

since the inequality (12) holds and $0 < \alpha_{\min} \leq \alpha_k \leq \alpha_{\max}$. Moreover, the inequality

$$\begin{aligned} \operatorname{dist}(0, \partial h(\boldsymbol{X}_{k+1})) &\leq \left\| \boldsymbol{\nabla} f(\boldsymbol{Z}_{k+1}) + \tfrac{1}{\alpha_k}(\boldsymbol{Y}_k - \boldsymbol{Z}_{k+1}) - \boldsymbol{\nabla} f(\boldsymbol{Y}_k) \right\|_{\mathsf{F}} \\ &\leq (L + 1/\alpha_{\min})(\|\boldsymbol{X}_{k+1} - \boldsymbol{X}_k\|_{\mathsf{F}} + \bar{w}\|\boldsymbol{X}_k - \boldsymbol{X}_{k-1}\|_{\mathsf{F}}) \end{aligned}$$

holds where $\bar{w} \in [0, 1)$ since $(t_k - 1)/t_{k+1} < 1$ for all $k \leq k_{\text{maxres}}$.

Consequently, the two cases yield that the inequality (18) holds with $a = \min(\delta, \gamma/\alpha_{\max}) > 0$ and the inequality (19) holds with $b = L + 1/\alpha_{\min} > 0$. ∎

Denote $w(\boldsymbol{X}_0)$ to be the limiting points of $\{\boldsymbol{X}_k\}$. The next lemma shows that Algorithm 1 is subsequence convergent, *i.e.,* all the convergent subsequences converge to a minimal point when the generated sequence is bounded.

Lemma A.2. *Let $\{\boldsymbol{X}_k\}$ be the sequence generated by Algorithm 1 starting from $\boldsymbol{X}_0$. If $\{\boldsymbol{X}_k\}$ is bounded, then $w(\boldsymbol{X}_0)$ is a non-empty compact set, and $w(\boldsymbol{X}_0) \subseteq \mathcal{X}_\star \neq \emptyset$.*

Proof. Since the sequence $\{\boldsymbol{X}_k\}$ is bounded, $w(\boldsymbol{X}_0)$ is non-empty. Meanwhile, $w(\boldsymbol{X}_0)$ is compact since it is the intersection of compact sets, *i.e.,* $w(\boldsymbol{X}_0) = \cap_{q\in\mathbb{N}} \overline{\cup_{k\geq q}\{\boldsymbol{X}_k\}}$, where $\overline{\mathcal{A}}$ denotes the closure of set $\mathcal{A}$. Let $\bar{\boldsymbol{X}}$ be any point in $w(\boldsymbol{X}_0)$ and $\{\boldsymbol{X}_{k_j}\}$ be a convergent subsequence of $\{\boldsymbol{X}_k\}$ such that $\boldsymbol{X}_{k_j} \to \bar{\boldsymbol{X}}$ as $j \to +\infty$. Since $\boldsymbol{R} \in \mathcal{S}_+^N$, $h(\boldsymbol{X}) \geq 0, \forall \boldsymbol{X}$. Together with the inequality (18), we know there exists some $\bar{h}$ such that $h(\boldsymbol{X}_k) \to \bar{h}$ as $k \to +\infty$. Moreover, (18) implies

$$a\sum_{k=0}^{\infty} \|\boldsymbol{X}_k - \boldsymbol{X}_{k+1}\|_{\mathsf{F}}^2 \leq \sum_{k=0}^{\infty} (h(\boldsymbol{X}_k) - h(\boldsymbol{X}_{k+1})) \leq h(\boldsymbol{X}_0) - \bar{h} < +\infty.$$

So, we have $\|\boldsymbol{X}_{k_j+1} - \boldsymbol{X}_{k_j}\|_{\mathsf{F}} \to 0$ and $\|\boldsymbol{X}_{k_j} - \boldsymbol{X}_{k_j-1}\|_{\mathsf{F}} \to 0$ as $k \to +\infty$. Moreover, from the above facts, it is easy to prove that $\{\boldsymbol{X}_{k_j+1}\}$ also converges to $\bar{X}$. Together with (19), we know $\operatorname{dist}(\mathbf{0}, \partial h(\boldsymbol{X}_{k_j+1})) \to 0$ as $j \to +\infty$. Let $\mathcal{U}$ be a neighborhood of $\bar{\boldsymbol{X}}$ with radius M. Then, we have

$$h(\boldsymbol{X}) \geq h(\boldsymbol{X}_{k_j+1}) - \operatorname{dist}(\mathbf{0}, \partial h(\boldsymbol{X}_{k_j+1}))\|\boldsymbol{X} - \boldsymbol{X}_{k_j+1}\|_{\mathsf{F}}, \ \forall \boldsymbol{X} \in \mathcal{U}. \tag{23}$$

Since $\lim_{j\to+\infty} h(\boldsymbol{X}_{k_j+1}) = h(\bar{\boldsymbol{X}})$, taking the limit $j \to +\infty$ in (23), we have $h(\boldsymbol{X}) \geq h(\bar{\boldsymbol{X}})$ for all $\boldsymbol{X} \in \mathcal{U}$. By the convexity of h, $\bar{\boldsymbol{X}}$ is a global minima of (7), i.e. $\mathbf{0} \in \partial h(\bar{\boldsymbol{X}})$. ∎

Based on the proof of Theorem 1 in [8], we can prove Theorem 4.2 as follows.

Proof. It is noted that the objective function h in (7) is a KL function as both f and g are semi-algebraic. Since $\{\boldsymbol{X}_k\}$ is bounded, $w(\boldsymbol{X}_0)$ is not empty. Denote $h(\boldsymbol{X}) = \bar{h}$ for

all $\boldsymbol{X} \in w(\boldsymbol{X}_0)$ as $\mathbf{0} \in \partial h(\boldsymbol{X}), \forall \boldsymbol{X} \in w(\boldsymbol{X}_0)$ from Lemma A.2. Let $\{\boldsymbol{X}_{k_j}\}$ be a convergent subsequence of $\{\boldsymbol{X}_k\}$ such that $\boldsymbol{X}_{k_j} \to \bar{\boldsymbol{X}}$ as $j \to +\infty$. Since $\boldsymbol{X}_k \in \mathcal{S}_+^N, \forall k$ the decreasing property (18) yields $\lim_{k\to+\infty} h(\boldsymbol{X}_k) = \lim_{j\to+\infty} h(\boldsymbol{X}_{k_j}) = \bar{h}$. Moreover, we assume that $h(\boldsymbol{X}_k) > \bar{h}$ for all k. Otherwise, if there exists some k_0 such that $h(\boldsymbol{X}_{k_0}) = \bar{h}$, from the decreasing property (18) and $h(\boldsymbol{X}_k) \geq \bar{h}$, we know $\boldsymbol{X}_k = \boldsymbol{X}_{k_0}$ for all $k \geq k_0$. By the definition of $w(\boldsymbol{X}_0)$, we have $\lim_{k\to+\infty} \operatorname{dist}(\boldsymbol{X}_k, w(\boldsymbol{X}_0)) = 0$. Applying the uniformized KL property ([8], Lemma 6) of h on $w(\boldsymbol{X}_0)$, there exist $k_\ell > 0$, $\eta > 0$ and $\phi \in \boldsymbol{\Phi}_\eta$ such that for all $\bar{\boldsymbol{X}} \in w(\boldsymbol{X}_0)$, we have

$$\phi'(h(\boldsymbol{X}_k) - \bar{h}) \operatorname{dist}(\mathbf{0}, \partial h(\boldsymbol{X}_k)) \geq 1, \ \forall k > k_\ell. \tag{24}$$

By the inequality (19), (24) implies

$$\phi'(h(\boldsymbol{X}_k) - \bar{h}) \geq \frac{1}{b(\|\boldsymbol{X}_k - \boldsymbol{X}_{k-1}\|_{\mathsf{F}} + \bar{w}\|\boldsymbol{X}_{k-1} - \boldsymbol{X}_{k-2}\|_{\mathsf{F}})}, \ \forall k > k_\ell. \tag{25}$$

By the concavity of ϕ (18) and (25), we know that

$$\begin{aligned} \Delta_{k,k+1} &:= \phi(h(\boldsymbol{X}_k) - \bar{h}) - \phi(h(\boldsymbol{X}_{k+1}) - \bar{h}) \\ &\geq \phi'(h(\boldsymbol{X}_k) - \bar{h})(h(\boldsymbol{X}_k) - h(\boldsymbol{X}_{k+1})) \geq \frac{a\|\boldsymbol{X}_{k+1} - \boldsymbol{X}_k\|_{\mathsf{F}}^2}{b(\|\boldsymbol{X}_k - \boldsymbol{X}_{k-1}\|_{\mathsf{F}} + \bar{w}\|\boldsymbol{X}_{k-1} - \boldsymbol{X}_{k-2}\|_{\mathsf{F}})} \end{aligned} \tag{26}$$

Define $C = b/a > 0$ in (26) and from the geometric inequality, we have

$$2\|\boldsymbol{X}_k - \boldsymbol{X}_{k+1}\|_{\mathsf{F}} \leq \|\boldsymbol{X}_k - \boldsymbol{X}_{k-1}\|_{\mathsf{F}} + \bar{w}\|\boldsymbol{X}_{k-1} - \boldsymbol{X}_{k-2}\|_{\mathsf{F}} + C\Delta_{k,k+1} \tag{27}$$

For any $k > k_\ell$, summing up (27) from $i = k_\ell + 1, \ldots, k$, we have

$$\begin{aligned} 2\sum_{i=k_\ell+1}^{k} \|\boldsymbol{X}_i - \boldsymbol{X}_{i+1}\|_{\mathsf{F}} &\leq \sum_{i=k_\ell+1}^{k} \|\boldsymbol{X}_i - \boldsymbol{X}_{i-1}\|_{\mathsf{F}} + \bar{w}\|\boldsymbol{X}_{i-1} - \boldsymbol{X}_{i-2}\|_{\mathsf{F}} + C\Delta_{i,i+1} \\ &\leq (1+\bar{w}) \sum_{i=k_\ell-1}^{k} \|\boldsymbol{X}_{i+1} - \boldsymbol{X}_i\|_{\mathsf{F}} + C\Delta_{k_\ell+1,k+1} \end{aligned} \tag{28}$$

where the last inequality is from the fact that $\Delta_{p,q} + \Delta_{q,r} = \Delta_{p,r}$ for all $p, q, r \in \mathbb{N}$. Since $\phi \geq 0$ and $\bar{w} \in [0, 1)$, the inequality (28) implies

$$(1-\bar{w}) \sum_{i=k_\ell+1}^{k} \|\boldsymbol{X}_i - \boldsymbol{X}_{i+1}\|_{\mathsf{F}} \leq (1+\bar{w}) \sum_{i=k_\ell-1}^{k_\ell} \|\boldsymbol{X}_{i+1} - \boldsymbol{X}_i\|_{\mathsf{F}} + C\phi\big(h(\boldsymbol{X}_{k_\ell}) - \bar{h}\big) \tag{29}$$

Let $k \to +\infty$ in (29), and thus $\sum_{k=1}^{+\infty} \|\boldsymbol{X}_k - \boldsymbol{X}_{k-1}\|_{\mathsf{F}} < +\infty$, which implies that $\{\boldsymbol{X}_k\}$ converges to an $\bar{\boldsymbol{X}}$, the sole member of $w(\boldsymbol{X}_0)$. Hence, $\mathbf{0} \in \partial h(\bar{\boldsymbol{X}})$. ∎

Appendix B. Proof of Proposition 4.1. By the decreasing property (18), it is sufficient to show that the sub-level set $[h(\boldsymbol{X}) \le h_0]$ is bounded where $h_0 = h(\boldsymbol{X}_0)$. Given any $\boldsymbol{X} \in [h(\boldsymbol{X}) \le h_0]$, we have $\boldsymbol{X} \in \mathcal{S}_+^N$ and $f(\boldsymbol{X}) \le h_0$. Then, there exists some $\boldsymbol{Y} \in \mathbb{C}^{N\times N}$ such that $\boldsymbol{X} = \boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}}$ with

$$\tfrac{1}{2}\left\|\mathscr{A}\left(\boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}}\right) - \boldsymbol{b}\right\|_2^2 + \mu\langle \boldsymbol{R}, \boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}}\rangle \le h_0. \tag{30}$$

By the triangle inequality, (30) implies

$$\left\|\mathscr{A}\left(\boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}}\right)\right\|_2 \le \sqrt{2h_0} + \|\boldsymbol{b}\|_2.$$

as $\boldsymbol{R} \in \mathcal{S}_+^N$ and $\mu \ge 0$. Standard norm inequalities gives:

$$\left\|\mathscr{A}\left(\boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}}\right)\right\|_1 \le \sqrt{M}\left(\sqrt{2h_0} + \|\boldsymbol{b}\|_2\right).$$

We note here that we can also write

$$\mathscr{A}\left(\boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}}\right) = \boldsymbol{C}\Big[(\boldsymbol{A}\boldsymbol{Y}) \odot (\boldsymbol{A}\boldsymbol{Y})^*\Big]\mathbf{1}$$

where $\odot$ denotes the element-wise (Hadamard) product, $\mathbf{1} \in \mathbb{R}^N$ of all 1s, and $\boldsymbol{C} \in \mathbb{R}^{M\times MN}$ is a blockwise diagonal matrix whose mth block is equal to $\mathbf{1}^{\mathsf{T}}/\sigma_m$. Let $\sigma_{\max}$ denote the minimal value among the σ_ms. Since the bracketed expression is the element-wise magnitude squared of $\boldsymbol{A}\boldsymbol{Y}$ and hence nonnegative, we then have

$$\|\boldsymbol{A}\boldsymbol{Y}\|_{\mathsf{F}}^2 \le \sigma_{\max}\sqrt{M}\left(\sqrt{2h_0} + \|\boldsymbol{b}\|_2\right).$$

Since $\boldsymbol{A}$ has full column rank, we obtain that

$$\|\boldsymbol{Y}\|_{\mathsf{F}}^2 \le \frac{\sigma_{\max}\sqrt{M}}{a_{\min}^2}\left(\sqrt{2h_0} + \|\boldsymbol{b}\|_2\right),$$

where $a_{\min} > 0$ the smallest nonzero singular value of $\boldsymbol{A}$. Since $\boldsymbol{X} = \boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}}$, the above inequality implies that the nuclear norm of $\boldsymbol{X}$ is bounded:

$$\|\boldsymbol{X}\|_* \le \frac{\sigma_{\max}\sqrt{M}}{a_{\min}^2}\left(\sqrt{2h_0} + \|\boldsymbol{b}\|_2\right)$$

and hence $\boldsymbol{X}$ must be bounded assuming bounded $\boldsymbol{b}$ and finite $\sigma_{\max}$, and hence the sub-level set must also be bounded.

Appendix C. Proof of Proposition 4.2. The dual problem of (7) is

$$\min_{\boldsymbol{w},\boldsymbol{S}} \ \tfrac{1}{2}\|\boldsymbol{w} - \boldsymbol{b}\|_2^2 + \iota_{\mathcal{S}_+^N}(\boldsymbol{S}), \ \text{s.t.} \ \mathscr{A}^{\mathsf{H}}(\boldsymbol{w}) + \boldsymbol{S} = \boldsymbol{R}. \tag{31}$$

The Lagrangian function ℓ associated with the problem (31) is

$$\ell(\boldsymbol{w}, \boldsymbol{S}; \boldsymbol{X}) = \tfrac{1}{2}\|\boldsymbol{w} - \boldsymbol{b}\|_2^2 + \iota_{\mathcal{S}_+^N}(\boldsymbol{S}) + \langle \boldsymbol{X}, \mathscr{A}^{\mathsf{H}}(\boldsymbol{w}) + \boldsymbol{S} - \boldsymbol{R}\rangle.$$

Since (31) is strongly convex with respect to $\boldsymbol{w}$ and $\boldsymbol{S}$ is uniquely determined by $\boldsymbol{w}$, thus (31) admits a unique solution, denoted by $(\bar{\boldsymbol{w}}, \bar{\boldsymbol{S}})$. By the Slater's condition and $\boldsymbol{X}_\star \in \mathcal{X}_\star$, the point $(\bar{\boldsymbol{w}}, \bar{\boldsymbol{S}}, \boldsymbol{X}_\star)$ satisfies the KKT equations:

$$0 = \bar{\boldsymbol{w}} - \boldsymbol{b} + \mathscr{A}(\boldsymbol{X}_\star),\ 0 \in \boldsymbol{X}_\star + \mathcal{N}_{\mathcal{S}_+^N}(\bar{\boldsymbol{S}}),\ 0 = \mathscr{A}^{\mathsf{H}}(\bar{\boldsymbol{w}}) + \bar{\boldsymbol{S}} - \boldsymbol{R}. \tag{32}$$

Combining the first and the last equalities in (32), we know $\bar{\boldsymbol{S}} = (\mathscr{A}^{\mathsf{H}}(\mathscr{A}(\boldsymbol{X}_\star) - \boldsymbol{b})) + \boldsymbol{R} = \boldsymbol{\nabla} f(\boldsymbol{X}_\star)$. It is from (15) and the Proposition 3.2 in [16] that $\operatorname{rank}(\boldsymbol{X}_\star) + \operatorname{rank}(\bar{\boldsymbol{S}}) = N$. Then, applying the Corollary 3.1 in [16], we obtain that for any $\bar{\boldsymbol{X}} \in \mathcal{X}_\star$, ∂h is metrically subregular at $\bar{\boldsymbol{X}}$ for 0.

Appendix D. Proof of Theorem 4.3. We first present a lemma that establishes the relationship between metric sub-regularity of ∂h and the KL inequality at critical points.

Lemma D.1. *Let $h = f + g$ be defined as they are in (8) and assume $\mathcal{X}_\star \neq \emptyset$. If ∂h is metrically subregular at $\bar{\boldsymbol{X}}$ for $\mathbf{0}$, then h satisfies the KL inequality at $\bar{\boldsymbol{X}}$ with $\phi(x) = c\sqrt{x}$ for some $c > 0$.*

Proof. Since h is convex, we know $\mathcal{X}_\star = \partial h^{-1}(\mathbf{0})$. If ∂h is metrically subregular at $\bar{\boldsymbol{X}}$ for $\mathbf{0}$, then there exist κ and $\epsilon > 0$ such that

$$\operatorname{dist}(\boldsymbol{X}, \mathcal{X}_\star) = \operatorname{dist}\big(\boldsymbol{X}, \partial h^{-1}(\mathbf{0})\big) \leq \kappa \operatorname{dist}(\mathbf{0}, \partial h(\boldsymbol{X})),\ \forall \boldsymbol{x} \in \mathbb{B}\big(\bar{\boldsymbol{X}}, \epsilon\big). \tag{33}$$

Thus, for any $\boldsymbol{X} \in \mathbb{B}(\bar{\boldsymbol{X}}, \epsilon) \cap [h(\boldsymbol{X}) > h(\bar{\boldsymbol{X}})]$, we have

$$h(\boldsymbol{X}) - h(\bar{\boldsymbol{X}}) = h(\boldsymbol{X}) - h(\boldsymbol{X}_\star) \leq \|\boldsymbol{U}\|_{\mathsf{F}} \|\boldsymbol{X}_\star - \boldsymbol{X}\|_{\mathsf{F}},\ \forall \boldsymbol{X}_\star \in \mathcal{X}_\star,\ \forall \boldsymbol{U} \in \partial h(\boldsymbol{X}), \tag{34}$$

where the inequality is a consequence of the convexity of h and the Cauchy-Schwartz inequality. Taking the infimum over all $\boldsymbol{X}_\star \in \mathcal{X}_\star$ and over all $\boldsymbol{U} \in \partial h(\boldsymbol{X})$ in (34) and then using (33), we have

$$h(\boldsymbol{X}) - h(\bar{\boldsymbol{X}}) \leq \kappa \operatorname{dist}(\mathbf{0}, \partial h(\boldsymbol{X}))^2,$$

which implies h satisfies the KL property with $\phi(x) = 2\sqrt{\kappa x}$. ∎

Now, inspired by the analysis in [2], we are ready to present the proof for Theorem 4.3.

Proof. Let $\bar{\boldsymbol{X}} \in \mathcal{X}_\star$ such that $\lim_{k \to +\infty} \boldsymbol{X}_k = \bar{\boldsymbol{X}}$. Assume that $h(\boldsymbol{X}_k) > h(\bar{\boldsymbol{X}})$ for all k. Otherwise, by the decrease property (18), it is easy to know $h(\boldsymbol{X}_k) = h(\bar{\boldsymbol{X}})$ and $\boldsymbol{X}_k = \bar{\boldsymbol{X}}$ for all $k > k_0$ whenever $h(\boldsymbol{X}_{k_0}) = h(\bar{\boldsymbol{X}})$. Define $r_k = h(\boldsymbol{X}_k) - h(\bar{\boldsymbol{X}})$. The fact that $\boldsymbol{X}_k \to \bar{\boldsymbol{X}}$ as $k \to +\infty$ combined with Proposition 4.2 and Lemma D.1 establishes the existence of some $k_\ell > 0$ such that the following inequality holds:

$$h(\boldsymbol{X}_k) - h(\bar{\boldsymbol{X}}) \leq \kappa \operatorname{dist}(\mathbf{0}, \partial h(\boldsymbol{X}_k))^2,\ \forall k > k_\ell. \tag{35}$$

Applying (19) and (18) to (35), we obtain that

$$\begin{aligned} r_k &\leq \kappa b^2 (\|\boldsymbol{X}_k - \boldsymbol{X}_{k-1}\|_{\mathsf{F}} + \bar{w} \|\boldsymbol{X}_{k-1} - \boldsymbol{X}_{k-2}\|_{\mathsf{F}})^2 \\ &\leq 2\kappa b^2 \Big(\|\boldsymbol{X}_k - \boldsymbol{X}_{k-1}\|_{\mathsf{F}}^2 + \bar{w}^2 \|\boldsymbol{X}_{k-1} - \boldsymbol{X}_{k-1}\|_{\mathsf{F}}^2 \Big) \\ &\leq 2\kappa a^{-1} b^2 \Big\{ F(\boldsymbol{X}_{k-1}) - F(\boldsymbol{X}_k) + \bar{w}^2 \Big[F(\boldsymbol{X}_{k-2}) - F(\boldsymbol{X}_{k-1}) \Big] \Big\} \\ &= c\big(r_{k-1} - r_k + \bar{w}(r_{k-2} - r_{k-1})\big) \end{aligned} \tag{36}$$

where $c = 2\kappa a^{-1}b^2$ and the second inequality is from the geometric inequality. Since $r_k \le r_{k-1}$ for all k, the inequality (36) implies

$$r_k \le \frac{c}{1+c}\Big[(1-\bar{w})r_{k-1} + \bar{w}r_{k-2}\Big] \le \frac{cr_{k-2}}{1+c} \le r_{k_\ell}\left(\frac{c}{1+c}\right)^{\frac{k-k_\ell-1}{2}}, \ \forall k > k_\ell.$$

Furthermore, using (29), for all $\hat{k} > k > k_\ell$, we have

$$\begin{aligned}(1-\bar{w})\|\boldsymbol{X}_{\hat{k}} - \boldsymbol{X}_k\|_{\mathsf{F}} &\le (1-\bar{w})\sum\nolimits_{i=k}^{\hat{k}-1}\|\boldsymbol{X}_{i+1} - \boldsymbol{X}_i\|_{\mathsf{F}} \\ &\le (1+\bar{w})\sum\nolimits_{i=k-2}^{k}\|\boldsymbol{X}_{i+1} - \boldsymbol{X}_i\|_{\mathsf{F}} + b\sqrt{r_k}/a \\ &\le (1+\bar{w})\sum_{i=k-2}^{k}\sqrt{\frac{h(\boldsymbol{X}_{i+1}) - h(\boldsymbol{X}_i)}{a}} + b\sqrt{r_k}/a \le \tilde{\nu}\sqrt{r_{k-2}}\end{aligned} \tag{37}$$

where $\tilde{\nu} = (1+\bar{w})\sqrt{2/a} + b/a$. Letting $\hat{k} \to +\infty$ in (37), we thus know that for all $k > k_\ell + 2$

$$\|\boldsymbol{X}_k - \bar{\boldsymbol{X}}\|_{\mathsf{F}} \le \nu\sqrt{r_{k_\ell}}\left(\frac{c}{1+c}\right)^{\frac{k-k_\ell-3}{4}},$$

where $\nu = (1-\bar{w})^{-1}\tilde{\nu}$. ∎