# Orientation Determination of Cryo-EM Images Using Least Unsquared Deviations[*]

**Lanhui Wang**[†], **Amit Singer**[‡], and **Zaiwen Wen**[§]

Amit Singer: amits@math.princeton.edu; Zaiwen Wen: wenzw@math.pku.edu.cn

[‡]Department of Mathematics and PACM, Princeton University, Princeton, NJ 08544-1000

[§]Beijng International Center for Mathematical Research, Peking University, Beijing, China

## Abstract

A major challenge in single particle reconstruction from cryo-electron microscopy is to establish a reliable ab initio three-dimensional model using two-dimensional projection images with unknown orientations. Common-lines–based methods estimate the orientations without additional geometric information. However, such methods fail when the detection rate of common-lines is too low due to the high level of noise in the images. An approximation to the least squares global self-consistency error was obtained in [A. Singer and Y. Shkolnisky, *SIAM J. Imaging Sci.*, 4 (2011), pp. 543–572] using convex relaxation by semidefinite programming. In this paper we introduce a more robust global self-consistency error and show that the corresponding optimization problem can be solved via semidefinite relaxation. In order to prevent artificial clustering of the estimated viewing directions, we further introduce a spectral norm term that is added as a constraint or as a regularization term to the relaxed minimization problem. The resulting problems are solved using either the alternating direction method of multipliers or an iteratively reweighted least squares procedure. Numerical experiments with both simulated and real images demonstrate that the proposed methods significantly reduce the orientation estimation error when the detection rate of common-lines is low.

## Keywords

angular reconstitution; cryo-electron microscopy; single particle reconstruction; common-lines; least unsquared deviations; semidefinite relaxation; alternating direction method of multipliers; iteratively reweighted least squares

## 1. Introduction

Single particle reconstruction (SPR) from cryo-electron microscopy (cryo-EM) [12, 57] is an emerging technique for determining the three-dimensional (3D) structure of macromolecules. One of the main challenges in SPR is to attain a resolution of 4Å or better, thereby allowing interpretation of atomic coordinates of macromolecular maps [14, 60]. Although X-ray crystallography and NMR spectroscopy can achieve higher resolution levels (~ 1Å by X-ray crystallography and 2–5Å by NMR spectroscopy), these traditional methods

[†]Corresponding author: The Program in Applied and Computational Mathematics (PACM), Princeton University, Princeton, NJ 08544-1000 (lanhuiw@math.princeton.edu).

http://www.siam.org/journals/siims/6-4/91643.html

are often limited to relatively small molecules. In contrast, cryo-EM is typically applied to large molecules or assemblies with size ranging from 10 to 150 nm, such as ribosomes [13], protein complexes, and viruses.

Cryo-EM is used to acquire 2D projection images of thousands of individual, identical frozen-hydrated macromolecules at random unknown orientations and positions. The collected images are extremely noisy due to the limited electron dose used for imaging to avoid excessive beam damage. A 3D electron density map can be reconstructed computationally from the 2D projection images. However, a broad computational framework is required for the reconstruction.

## 1.1. The algorithmic pipeline of SPR

The main steps of a full SPR procedure can be summarized as follows:

1.  Particle selection: The individual molecules are detected and extracted from the micrograph by image segmentation.

2.  Class averaging: Images with similar projection directions are averaged together to improve their signal-to-noise ratio (SNR).

3.  Orientation estimation: This is done via the random-conical tilt technique, or common-lines–based approaches, or reprojection. This step may also include a shift estimation.

4.  Reconstruction: A 3D volume is generated by a tomographic inversion algorithm.

5.  Iterative refinement: The whole procedure is repeated from step 3, using the outcome of the most recent reconstruction until convergence.

Numerous technical and computational advances have contributed to the different processing steps in SPR. First, the particles can be automatically selected from micrographs using algorithms based on template matching, computer vision, neural network, or other approaches [67]. Then multivariate statistical data compression [22, 56] and classification techniques [54, 34, 49] can be used to sort and partition the large set of particle images by their viewing directions, producing "class averages" of enhanced SNR. After class averaging, an ab initio estimation of the pose parameters can be obtained using the random-conical tilt technique [40] or common-lines–based approaches [55, 47, 48]. Using the ab initio estimation of the pose parameters, a preliminary 3D map is reconstructed from the projection images by a tomographic inversion algorithm. The tomographic inversion is computed either by applying a weighted adjoint operator [40, 39, 19, 35] or by using an iterative approach for inversion [17, 28, 15]. We refer to the preliminary 3D map as the ab initio model or simply initial model. The initial model is then iteratively refined [33] in order to obtain a higher-resolution 3D reconstruction. In each iteration of the refinement process, the current 3D model is projected at several pre-chosen viewing directions, and the resulting reprojection images are matched with the particle images, giving rise to new estimates of their pose parameters. The new pose parameters are then used to produce a refined 3D model using a 3D reconstruction algorithm. This process is repeated for several iterations until convergence is achieved.

In this paper, we address the problem of orientation estimation using common-lines between cryo-EM images in order to obtain an improved initial model. This is a major challenge in SPR due to the low SNR of the raw images, especially for small particles.

## 1.2. Orientation determination using common-lines between images

The Fourier projection-slice theorem (see, e.g., [31]) plays a fundamental role in common-lines–based reconstruction methods. The theorem states that restricting the 3D Fourier transform of the volume to a planar central slice yields the Fourier transform of a 2D projection of the volume in a direction perpendicular to the slice (Figure 1). Thus, any two projections imaged from nonparallel viewing directions intersect at a line in Fourier space, which is called the common-line between the two images. The common-lines between any three images with linearly independent projection directions determine their relative orientation up to handedness. This is the basis of the "angular reconstitution" technique of van Heel [55], which was also developed independently by Vainshtein and Goncharov [53]. In this technique, the orientations of additional projections are determined in a sequential manner. Farrow and Ottensmeyer [10] used quaternions to obtain the relative orientation of a new projection in a least squares sense. The main problem with such techniques is that they are sensitive to false detection of common-lines, which leads to the accumulation of errors. Penczek, Zhu, and Frank [37] tried to obtain the rotations corresponding to all projections simultaneously by minimizing a global energy functional, which requires a brute force search in an exponentially large parametric space of all possible orientations for all projections. Mallick et al. [26] and Singer et al. [47] applied Bayesian approaches to use common-lines information from different groups of projections. Recently, Singer and Shkolnisky [48] developed two algorithms based on eigenvectors and semidefinite programming for estimating the orientations of all images. These two algorithms correspond to convex relaxations of the global self-consistency error minimization and can accurately estimate all orientations at relatively low common-line detection rates.

When the SNR of the images is significantly low, the detected common-lines consist of a modest number of noisy inliers, which are explained well by the image orientations, along with a large number of outliers, which have no structure. The standard common-lines–based methods, including those using least squares (LS) [10, 48], are sensitive to these outliers. In this paper we estimate the orientations using a different, more robust self-consistency error, which is the sum of unsquared residuals [32, 51] rather than the sum of squared residuals of the LS formulation. Convex relaxations of least unsquared deviations (LUD) have recently been proposed for other applications, such as robust principal component analysis [23, 64] and robust synchronization of orthogonal transformations [61]. Under certain noise models for the distribution of the outliers (e.g., the haystack model of [23, 64]), such convex relaxations enjoy proven guarantees for exact and stable recovery with high probability. Such theoretical and empirical improvements that LUD brings compared to LS serve as the main motivation for considering in this paper the application of LUD to the problem of orientation estimation from common-lines in SPR.

The LUD minimization problem is solved here via semidefinite relaxation. When the detection rate of common-lines is extremely low, the estimated viewing directions of the projection images are observed to cluster together (see Figure 4). This artificial clustering can be explained by the fact that images that share the same viewing direction also share more than one common-line (see more details in section 5). In order to mitigate this spurious clustering of the estimated viewing directions, we add to the minimization formulation a spectral norm term, either as a constraint or as a regularization term. The resulting minimization problem is solved by the alternating direction method of multipliers (ADMM). We also consider the application of the iteratively reweighted least squares (IRLS) procedure, which is not guaranteed to converge to the global minimizer but performs well in our numerical experiments. We demonstrate that the ab initio models resulting from our new methods are more accurate and require fewer refinement iterations compared to LS-based methods.

The paper is organized as follows. In section 2 we review the detection procedure of common-lines between images. Section 3 presents the LS and LUD global self-consistency cost functions. Section 4 introduces the semidefinite relaxation and rounding procedure for the LUD formulation. The additional spectral norm constraint is considered in section 5. The ADMM method for obtaining the global minimizer is detailed in section 6, and the IRLS procedure is described in section 7. Numerical results for both simulated and real data are provided in section 8. Finally, section 9 is a summary with discussion.

## 2. Detection of common-lines between images

Typically, the first step for detecting common-lines is to compute the 2D Fourier transform of each image on a polar grid using, e.g., the nonuniform fast Fourier transform (NUFFT) [9, 11, 18]. The transformed images have resolution $n_r$ in the radial direction and resolution $n_\theta$ in the angular direction; that is, the radial resolution $n_r$ is the number of equi-spaced samples along each ray in the radial direction, and the angular resolution $n_\theta$ is the number of angularly evenly distributed Fourier rays computed for each image (Figure 1). For simplicity, we let $n_\theta$ be an even number. The transformed images are denoted by $(\vec{l}^k_0, \vec{l}^k_1, \ldots, \vec{l}^k_{n_\theta-1})$, where $\vec{l}^k_m = (l^k_{m,1}, l^k_{m,2}, \ldots, l^k_{m,n_r})$ is an $n_r$-dimensional vector, $m \in \{0, 1, \ldots, n_\theta-1\}$ is the index of a ray, $k \in \{1, 2, \ldots, K\}$ is the index of an image, and $K$ is the number of images. The zero frequency term is shared by all lines independently of the image and is therefore excluded from comparison. To determine the common-line between two images $P_i$ and $P_j$, the similarity between all $n_\theta$ radial lines $\vec{l}^i_0, \vec{l}^i_1, \ldots, \vec{l}^i_{n_\theta-1}$ from the first image and all $n_\theta$ radial lines $\vec{l}^j_0, \vec{l}^j_1, \ldots, \vec{l}^j_{n_\theta-1}$ from the second image is measured (overall, $n_\theta^2$ comparisons), and the pair of radial lines $\vec{l}^i_{m_{i,j}}$ and $\bar{l}^j_{m_{j,i}}$ with the highest similarity is declared as the common-line pair between the two images. However, as a radial line is the complex conjugate of its antipodal line, the similarity measure between $\vec{l}^i_{m_1}$ and $\vec{l}^j_{m_2}$ has the same value as that between their antipodal lines $\vec{l}^i_{m_1+n_\theta/2}$ and $\vec{l}^j_{m_2+n_\theta/2}$ (where addition of indices is taken modulo $n_\theta$). Thus the number of distinct similarity measures that need to be computed is $n_\theta^2/2$, obtained by restricting the index $m_1$ to take values between 0 and $n_\theta/2$ and letting $m_2$ take any of the $n_\theta$ possibilities (see also [55] and [36, p. 255]). Equivalently, it is possible to compare real-valued 1D line projections of the 2D projection images, instead of comparing radial Fourier lines that are complex-valued. According to the Fourier projection-slice theorem, each 1D projection is obtained by the inverse Fourier transform of the corresponding Fourier radial line $\vec{l}^k_m$ and its antipodal line $\vec{l}^k_{m+n_\theta/2}$, and is denoted as $\vec{s}^k_m$. The 1D projection lines of a cryo-EM image can be displayed as a 2D image known as a "sinogram" (see [55, 45]).

Traditionally, the pair of radial lines (or sinogram lines) that has the maximum normalized cross correlation is declared as the common line, that is,

$$(m_{i,j}, m_{j,i}) = \underset{0 \le m_1 < n_\theta/2, 0 \le m_2 < n_\theta}{\arg\max} \frac{\left\langle \vec{l}^i_{m_1}, \vec{l}^j_{m_2} \right\rangle}{\|\vec{l}^i_{m_1}\| \|\vec{l}^j_{m_2}\|} \quad \text{for all } i \ne j, \quad (2.1)$$

where $m_{i,j}$ is a discrete estimate for where the $j$th image intersects with the $i$th image. In practice, a weighted correlation, which is equivalent to applying a combination of high-pass

and low-pass filters, is used to determine proximity. As noted in [55], the normalization is performed so that the correlation coefficient becomes a more reliable measure of similarity between radial lines. Note that, even with clean images, this estimate will have a small deviation from its ground truth (unknown) value due to discretization errors. With noisy images, large deviations of the estimates from their true values (say, errors of more than 10°) are frequent, and their frequency increases with the level of noise. We refer to common lines whose $m_{i,j}$ and $m_{j,i}$ values were estimated accurately (up to the discretization error tolerance 10°) as "correctly detected" common lines, or "inliers," and to the remaining common lines as "falsely detected" or "outliers."

A simple inliers-outliers model was introduced and analyzed in [48, sections 5.1 and 7]. That model assumes that each common-line pair is an inlier with probability $p$, independent of the remaining common-lines, while the directions of the outliers are uniformly distributed. It was noted, however, already in [48, p. 560] that this model is too simplistic as it assumes independence while there must be dependency: The number of common-line pairs is $\mathcal{O}(K^2)$, and since there are only $\mathcal{O}(Kn_\theta)$ radial lines, there must be dependency among the pairs for sufficiently large $K$. It is therefore difficult to provide a theoretical analysis for the evaluation of the practical performance of the different common-lines–based orientation determination algorithms.

## 3. Weighted LS and LUD

We define the directions of detected common-lines between the Fourier transformed image $\hat{P_i}$ and the Fourier transformed image $\hat{P_j}$ as the following unit vectors (Figure 1):

$$\overrightarrow{c}_{i,j}=(c_{ij}^1, c_{ij}^2)=(\cos\,(2\pi m_{ij}/n_\theta), \sin\,(2\pi m_{ij}/n_\theta)),\quad (3.1)$$

$$\overrightarrow{c}_{ji}=(c_{ji}^1, c_{ji}^2)=(\cos\,(2\pi m_{ji}/n_\theta), \sin\,(2\pi m_{ji}/n_\theta)),\quad (3.2)$$

where $\overrightarrow{c_{ij}}$ and $\overrightarrow{c_{ji}}$ are on the Fourier transformed images $\hat{P_i}$ and $\hat{P_j}$, respectively, and $m_{ij}$ and $m_{ji}$ are discrete estimates for the common lines' positions using (2.1). Let the rotation matrices $R_i \in \mathbf{SO}(3)$, $i = 1, \ldots, K$, represent the orientations of the $K$ images. According to the Fourier projection-slice theorem (Figure 1), the common lines on every two images should be the same after the 2D Fourier transformed images are inserted in the 3D Fourier space using the corresponding rotation matrices; that is,

$$R_i \begin{pmatrix} \bar{c}_{ij}^T \\ 0 \end{pmatrix} = R_j \begin{pmatrix} \bar{c}_{ji}^T \\ 0 \end{pmatrix} \quad \text{for } 1 \leq i < j \leq K. \quad (3.3)$$

These can be viewed as $\begin{pmatrix} K \\ 2 \end{pmatrix}$ linear equations for the $6K$ variables corresponding to the first two columns of the rotation matrices. (The third column of each rotation matrix does not contribute in (3.3) since it is multiplied by the zero third entries in $(\overrightarrow{c_{ij}}, 0)$ and $(\overrightarrow{c_{ji}}, 0)$.) The weighted LS approach for solving this system can be formulated as the minimization problem

$$\min_{R_1,\ldots,R_K \in \mathbf{SO}(3)} \sum_{i \neq j} w_{ij} \| R_i\,(\overrightarrow{c}_{ij}, 0)^T - R_j\,(\overrightarrow{c}_{ji}, 0)^T \|^2, \quad (3.4)$$

where the weights $w_{ij}$ indicate the confidence in the detections of common-lines between pairs of images. Since $(\vec{c}_{ij}, 0)^T$ and $(\vec{c}_{ji}, 0)^T$ are 3D unit vectors, their rotations are also unit vectors; that is, $\|R_i(\vec{c}_{ij}, 0)^T\| = \|R_j(\vec{c}_{ji}, 0)^T\| = 1$. It follows that the minimization problem (3.4) is equivalent to the maximization problem of the sum of dot products:

$$\max_{R_1,\ldots,R_K \in \mathbf{SO}(3)} \sum_{i \neq j} w_{ij} \langle R_i(\vec{c}_{ij}, 0)^T, R_j(\vec{c}_{ji}, 0)^T \rangle. \quad (3.5)$$

When the weight $w_{ij} = 1$ for each pair $i$ $j$, (3.5) is equivalent to the LS problem that was considered in [37], and more recently in [48] using convex relaxation of the nonconvex constraint set. The solution to the LS problem may not be optimal due to the typically large proportion of outliers (Figure 2).

To guard the estimation of the orientations from outliers, we replace the sum of weighted squared residuals in (3.4) with the more robust sum of unsquared residuals and obtain

$$\min_{R_1,\ldots,R_K \in \mathbf{SO}(3)} \sum_{i \neq j} \|R_i(\vec{c}_{ij}, 0)^T - R_j(\vec{c}_{ji}, 0)^T\|, \quad (3.6)$$

or equivalently,

$$\min_{R_1,\ldots,R_K \in \mathbf{SO}(3)} \sum_{i \neq j} \|(\vec{c}_{ij}, 0)^T - R_i^T R_j(\vec{c}_{ji}, 0)^T\|. \quad (3.7)$$

We refer to the minimization problem (3.6) as the least unsquared deviation (LUD) problem. The self-consistency error given in (3.6) reduces the contribution from large residuals that may result from outliers. We note that it is also possible to consider the weighted version of (3.6), namely,

$$\min_{R_1,\ldots,R_K \in \mathbf{SO}(3)} \sum_{i \neq j} w_{ij} \|R_i(\vec{c}_{ij}, 0)^T - R_j(\vec{c}_{ji}, 0)^T\|.$$

For simplicity, we focus here on the unweighted version.

## 4. Semidefinite programming relaxation (SDR) and the rounding procedure

Both the weighted LS problem (3.4) and the LUD problem (3.6) are nonconvex and therefore extremely difficult to solve if one requires the matrices $R_i$ to be rotations, that is, when adding the constraints

$$R_i R_i^T = I_3, \ \det(R_i) = 1 \quad \text{for } i = 1, \ldots, K, \quad (4.1)$$

where $I_3$ is the $3 \times 3$ identity matrix. A relaxation method that neglects the constraints (4.1) will simply collapse to the trivial solution $R_1 = \cdots = R_K = 0$, which obviously does not satisfy the constraint (4.1).

The relaxation in [48] that uses semidefinite programming (SDP) can be modified in a straightforward manner in order to deal with nonunity weights $w_{ij}$ in (3.5). We present this modification here for three reasons. First, the weighted version is required by the IRLS

procedure (see section 7). Second, the rounding procedure after the SDP employed here is slightly different than the deterministic rounding procedure presented in [48] and is closer in spirit to the randomized rounding procedure of Goemans and Williamson for the MAX-CUT problem [16]. Finally, in the appendix we prove exact recovery of the rotations by the SDR procedure when the detected common-lines are all correct.

### 4.1. Constructing the Gram matrix *G* from the rotations *R$_i$*

We denote the columns of the rotation matrix $R_i$ by $R_i^1, R_i^2$, and $R_i^3$, and write the rotation matrices as

$$R_i = \begin{pmatrix} | & | & | \\ R_i^1 & R_i^2 & R_i^3 \\ | & | & | \end{pmatrix}, \quad i = 1, \ldots, K.$$

We define a $3 \times 2K$ matrix $R$ by concatenating the first two columns of all rotation matrices:

$$R = \begin{pmatrix} | & | & & | & | & & | & | \\ R_1^1 & R_1^2 & \cdots & R_k^1 & R_k^2 & \cdots & R_K^1 & R_K^2 \\ | & | & & | & | & & | & | \end{pmatrix}. \quad (4.2)$$

The Gram matrix $G$ for the matrix $R$ is a $2K \times 2K$ matrix of inner products between the 3D column vectors of $R$; that is,

$$G = R^T R. \quad (4.3)$$

Clearly, $G$ is a rank-3 semidefinite positive matrix ($G \succeq 0$), which can be conveniently written as a block matrix

$$G = (G_{ij})_{i,j=1,\ldots,K},$$

where $G_{ij}$ is the $2 \times 2$ upper left block of the rotation matrix $R_i^T R_j$, that is,

$$G_{ij} = \begin{pmatrix} (R_i^1)^T \\ (R_i^2)^T \end{pmatrix} \begin{pmatrix} R_i^1 & R_i^2 \end{pmatrix}.$$

In addition, the orthogonality of the rotation matrices ($R_i^T R_i = I$) implies that

$$G_{ii} = I_2, \quad i = 1, 2, \ldots, K, \quad (4.4)$$

where $I_2$ is the $2 \times 2$ identity matrix.

### 4.2. SDR for weighted LS

We first define two $2K \times 2K$ matrices $S = (S_{ij})_{i,j=1,\ldots,K}$ and $W = (W_{ij})_{i,j=1,\ldots,K}$, where the $2 \times 2$ subblocks $S_{ij}$ and $W_{ij}$ are given by

$$S_{ij} = \overrightarrow{c}_{ji}^{T} \overrightarrow{c}_{ij}$$

and

$$W_{ij} = w_{ij} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Both matrices $S$ and $W$ are symmetric, and they store all available common-line information and weight information, respectively. It follows that the objective function (3.5) is the trace of the matrix $(W \circ S) G$:

$$\sum_{i \neq j} w_{ij} \langle R_i (\overrightarrow{c}_{ij}, 0)^T, R_j (\overrightarrow{c}_{ji}, 0)^T \rangle = \mathrm{trace}\left( (W \circ S) G \right), \quad (4.5)$$

where the symbol $\circ$ denotes the Hadamard product between two matrices. A natural relaxation of the optimization problem (3.5) is thus given by the SDP problem

$$\max_{G \in \mathbb{R}^{2K \times 2K}} \mathrm{trace}\left( (W \circ S) G \right) \quad (4.6)$$

$$\text{s.t.} \quad G_{ii} = I_2, \ i = 1, 2, \ldots, K, \quad (4.7)$$

$$G \succcurlyeq 0. \quad (4.8)$$

The nonconvex rank-3 constraint on the Gram matrix $G$ is missing from this SDR [25]. The problem (4.6)–(4.8) is an SDP that can be solved by standard SDP solvers. In particular, it can be well solved by the solver SDPLR [4], which takes advantage of the low-rank property of $G$. SDPLR is a first-order algorithm via low-rank factorization and hence can provide approximate solutions for large-scale problems. Moreover, the iterations of SDPLR are extremely fast.

### 4.3. SDR for LUD

Similar to defining the Gram matrix $G$ in (4.3), we define a $3K \times 3K$ matrix $\tilde{G}$ as $\tilde{G} = (\tilde{G}_{ij})_{i,j=1,\ldots,K}$, where each $\tilde{G}_{ij}$ is a $3 \times 3$ block defined as $\tilde{G}_{ij} = R_i^T R_j$. Then, a natural SDR for (3.7) is given by

$$\min_{\tilde{G} \succcurlyeq 0} \sum_{i \neq j} \| (\overrightarrow{c}_{ij}, 0)^T - \tilde{G}_{ij} (\overrightarrow{c}_{ji}, 0)^T \| \quad \text{s.t.} \ \tilde{G}_{ii} = I_3. \quad (4.9)$$

The constraints missing in this SDP formulation are the nonconvex rank-3 constraint and the determinant constraints $\det(\tilde{G}_{ij}) = 1$ on the Gram matrix $\tilde{G}$. However, the solution $\tilde{G}$ to (4.9) is not unique. Note that if a set of rotation matrices $\{R_i\}$ is a solution to (3.7), then the set of

conjugated rotation matrices $\{JR_iJ\}$ is also a solution to (3.7), where the matrix $J$ is defined as

$$J = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Thus, another solution to (4.9) is the Gram matrix $\tilde{G}^J = (\tilde{G}_{ij}^J)_{i,j=1,\ldots,K}$ with the $3 \times 3$ sub-blocks given by $\tilde{G}_{ij}^J = JR_i^T JJR_j J = JR_i^T R_j J$. It can be verified that $\frac{1}{2}(\tilde{G} + \tilde{G}^J)$ is also a solution to (4.9). Using the fact that

$$\frac{1}{2}(\tilde{G}_{ij} + \tilde{G}_{ij}^J) = \begin{pmatrix} & & 0 \\ G_{ij} & & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

the problem (4.9) is reduced to

$$\min_{G \succcurlyeq 0} \sum_{i \neq j} \| \vec{c}_{ij}^T - G_{ij} \vec{c}_{ji}^T \| \quad \text{s.t. } G_{ii} = I_2. \quad (4.10)$$

This is a SDR for the LUD problem (3.6). The problem (4.10) can be solved using ADMM (see details in section 6.2).

## 4.4. The randomized rounding procedure

The matrix $R$ is recovered from a random projection of the solution $G$ of the SDP (4.6). We randomly draw a $2K \times 3$ matrix $P$ from the Stiefel manifold $V_3(\mathbb{R}^{2K})$. The random matrix $P$ is computed using the orthogonal matrix $Q$ and the upper triangular matrix $R$ from QR factorization of a random matrix with standard i.i.d. Gaussian entries; that is, $P = Q$ sign (diag $(R)$), where sign stands for the entrywise sign function and diag$(R)$ is a diagonal matrix whose diagonal entries are the same as those of the matrix $R$. The matrix $P$ is shown to be drawn uniformly from the Stiefel manifold in [29]. We compute the Cholesky decomposition of $G = LL^T$ and then project the lower triangular matrix $L$ onto the subspace spanned by the columns of the $2K \times 3$ matrix $LP$.[1]

The $2K \times 3$ matrix $LP$ is a proxy for the matrix $R^T$ (up to a global $3 \times 3$ orthogonal transformation). In other words, we can regard the $3 \times 2K$ matrix $(LP)^T$ as composed from $K$ matrices of size $3 \times 2$, denoted $A_i$ ($i = 1, \ldots, K$), namely,

$$(LP)^T = (\begin{array}{cccc} A_1 & A_2 & \cdots & A_K \end{array}).$$

---

[1]The 3D subspace can also be spanned by the eigenvectors associated with the top three eigenvalues of $G$, while the fourth largest eigenvalue is expected to be significantly smaller; see also [48].

The two columns of each $A_i$ correspond to $R_i^1$ and $R_i^2$ (compare to (4.2)). We therefore estimate the matrix $R_i^{[1,2]} = (\ R_i^1 \quad R_i^2\ )$ as the closest matrix to $A_i$ on the Stiefel manifold $V_2(\mathbb{R}^3)$ in the Frobenius matrix norm. The closest matrix is given by (see, e.g., [1]) $R_i^{[1,2]} = U_i V_i^T$, where $A_i = U_i \sum_i V_i^T$ is the singular value decomposition of $A_i$. We note that, except for the orthogonality constraint (4.7), the SDP (4.6)–(4.8) is identical to the Goemans–Williamson SDP for finding the maximum cut in a weighted graph [16], where the SDR and the randomized rounding procedure [50, 25] for maximum cut problem is proved to have a 0.87 performance guarantee. Currently we do not have a similar performance guarantee for the SDP relaxation of the LS problem (3.5). From the complexity point of view, SDP can be solved in polynomial time to any given precision. The idea of using SDP for determining image orientations in cryo-EM was originally proposed in [48].

## 5. The spectral norm constraint

In our numerical experiments (see section 8) we observed that in the presence of many "outliers" (i.e., a large proportion of misidentified common-lines) the estimated viewing directions[2] that are obtained by solving either (4.6)–(4.8) or (4.10) are highly clustered (Figure 4). This empirical behavior of the solutions can be explained by the fact that images whose viewing directions are parallel share many common-lines, which results from the overlapping Fourier slices. In other words, when the viewing directions of $R_i$ and $R_j$ are nearby, the fidelity term $\|R_i(\vec{c_{ij}}, 0)^T - R_j(\vec{c_{ji}}, 0)^T\|$ (that appears in all cost functions) can become small (i.e., close to 0), even when the common-line pair $(\vec{c_{ij}}, \vec{c_{ji}})$ is misidentified. When the proportion of outliers in the detected common-lines is large enough, the solution to either the LS problem or the LUD problem can consist of many rotations with similar viewing directions, so that most of the fidelity terms have small values. In such cases, the Gram matrix $G$ has just two dominant eigenvalues, instead of three.

In order to prevent the viewing directions from clustering and the Gram matrix $G$ from collapsing to a rank-2 matrix, we add the following constraint on the spectral norm of the Gram matrix $G$ to the optimization problem (4.6)–(4.8) or (4.10):

$$G \preccurlyeq \alpha K I_{2K}, \quad \text{(5.1)}$$

where $I_{2K}$ is the $2K \times 2K$ identity matrix, or equivalently,

$$\|G\|_2 \leq \alpha K, \quad \text{(5.2)}$$

where $\|G\|_2$ is the spectral norm of the matrix $G$, and the parameter $\alpha \in \left[\frac{2}{3}, 1\right]$ controls the spread of the viewing directions. If the true imaging orientations are uniformly sampled from the rotation group **SO**(3), then by the law of large numbers and the symmetry of the distribution of orientations, the spectral norm of the true Gram matrix $G_{true}$ is approximately $\frac{2}{3}K$. (To see this, notice that $\mathrm{Tr}(G) = \mathrm{Tr}(KI_2) = 2K$. Thus, the sum of eigenvalues of $G$ is $2K$. Recall that $G$ is of rank 3, so if the rotations are uniformly distributed, then each of its three nontrivial eigenvalues equals $\frac{2K}{3}$.) On the other hand, if the true viewing directions are highly clustered, then the spectral norm of the true Gram matrix $G_{true}$ is close to $K$. For a known distribution of orientations, we can compute the spectral norm of the true Gram matrix $G_{true}$ accordingly, which can be verified to be a number between $\frac{2}{3}$ and 1. In practice, however, the distribution of the viewing directions is usually unknown a priori, and often it cannot be assumed to be uniform. To prevent a solution with clustered viewing directions,

---

[2]The viewing direction is the third column of the underlying rotation matrix.

we fix the parameter $a$ to some number satisfying $\frac{2}{3} \leq \alpha < 1$, and perhaps even try a few possible values for $a$ and choose the best value by examining the resulting reconstructions. Another method for choosing the best value for $a$ is to compare the likelihood scores. We defer the definition of the likelihood score to section 9.

## 6. The alternating direction method of multipliers (ADMM) for SDRs with spectral norm constraint

The application of ADMM to SDP problems was considered in [62]. Here we generalize the application of ADMM to the optimization problems considered in the previous sections. ADMM is a multiple-splitting algorithm that minimizes the augmented Lagrangian function in an alternating fashion such that in each step it minimizes over one block of the variables with all other blocks fixed and then updates the Lagrange multipliers. We apply ADMM to the dual problems since the linear constraint (6.2) satisfies $\mathcal{A}\mathcal{A}^* = I$, which simplifies the computation of subproblems. The strong duality theorem, which is known as Slater's theorem, guarantees that in the presence of a strictly feasible solution, a primal problem can be solved by solving its dual problem. To obtain a strictly feasible solution to the primal problems with a positive semidefinite constraint, the linear constraint (6.2), and the spectral norm constraint (6.3), we can construct a Gram matrix $G$ in (4.3) using rotations sampled from a uniform distribution over the rotation group. Therefore, strong duality holds for the primal problems, and the primal problems can be solved by applying ADMM to their corresponding dual problems.

### 6.1. The relaxed weighted LS problem

The weighted LS problem after SDR (4.6)–(4.8) can be efficiently solved using SDPLR [4]. However, SDPLR is not suitable for the problem once the spectral norm constraint on $G$ (5.2) is added to (4.6)–(4.8). This is because the constraint (5.2) can be written as $aKI - G \succeq 0$, but the matrix $aKI - G$ does not have a low-rank structure. Moreover, SDP solvers using polynomial-time primal-dual interior point methods are designed for small to medium-sized problems. Therefore, they are not suitable for our problem. Instead, here we devise a version of the ADMM which takes advantage of the low-rank property of $G$.

After the spectral norm constraint (5.2) is added, the problem (4.6)–(4.8) becomes

$$\min_{G \succeq 0} -\langle C, G \rangle \quad (6.1)$$

$$\text{s.t. } \mathcal{A}(G) = \mathbf{b}, \quad (6.2)$$

$$\|G\|_2 \leq \alpha K, \quad (6.3)$$

where

$$\mathcal{A}(G) = \begin{pmatrix} G_{ii}^{11} \\ G_{ii}^{22} \\ \frac{\sqrt{2}}{2}G_{ii}^{12} + \frac{\sqrt{2}}{2}G_{ii}^{21} \end{pmatrix}_{i=1,2,\ldots,K}, \quad \mathbf{b} = \begin{pmatrix} b_i^1 \\ b_i^2 \\ b_i^3 \end{pmatrix}_{i=1,2,\ldots,K}, \quad (6.4)$$

$$b_i^1 = b_i^2 = 1, \ b_i^3 = 0 \quad \text{for all } i,$$

$G_{ij}^{pq}$ denotes the $(p, q)$th element in the $2 \times 2$ subblock $G_{ij}$, $C = W \circ S$ is a symmetric matrix, and $\langle C, G \rangle = \text{trace}\,(CG)$. Following the equality $\langle \mathcal{A}(G), \mathbf{y} \rangle = \langle G, \mathcal{A}^*(\mathbf{y}) \rangle$ for arbitrary

$$\mathbf{y} = \begin{pmatrix} y_i^1 \\ y_i^2 \\ y_i^3 \end{pmatrix}_{i=1,2,\dots,K},$$

the adjoint of the operator $\mathcal{A}$ is defined as

$$\mathcal{A}^*(\mathbf{y}) = Y = \begin{pmatrix} Y^{11} & Y^{12} \\ Y^{21} & Y^{22} \end{pmatrix},$$

where for $i = 1, 2, \dots, K$

$$Y_{ii}^{11} = y_i^1, \quad Y_{ii}^{22} = y_i^2, \quad \text{and} \quad Y_{ii}^{12} = Y_{ii}^{21} = y_i^3 / \sqrt{2}.$$

It can be verified that $\mathcal{A}\mathcal{A}^* = I$. The dual problem of problem (6.1)–(6.3) is

$$\max_{\mathbf{y}, X \succcurlyeq 0} \min_{\|G\|_2 \leq \alpha K} - \langle C, G \rangle - \langle \mathbf{y}, \mathcal{A}(G) - \mathbf{b} \rangle - \langle G, X \rangle. \quad (6.5)$$

By rearranging terms in (6.5), we obtain

$$\max_{\mathbf{y}, X \succcurlyeq 0} \min_{\|G\|_2 \leq \alpha K} - \langle C + X + \mathcal{A}^*(\mathbf{y}), G \rangle + \mathbf{y}^T \mathbf{b}. \quad (6.6)$$

Using the fact that the dual norm of the spectral norm is the nuclear norm (Proposition 2.1 in [41]), we can obtain from (6.6) the dual problem

$$\max_{\mathbf{y}, X \succcurlyeq 0} \mathbf{y}^T \mathbf{b} - \alpha K \| C + X + \mathcal{A}^*(\mathbf{y}) \|_*, \quad (6.7)$$

where $\|\cdot\|_*$ denotes the nuclear norm. By introducing a variable $Z = C + X + \mathcal{A}^*(\mathbf{y})$, we obtain from (6.7) that

$$\min_{\mathbf{y}, X \succcurlyeq 0} \quad -\mathbf{y}^T \mathbf{b} + \alpha K \| Z \|_* \quad (6.8)$$

$$\text{s.t.} \quad Z = C + X + \mathcal{A}^*(\mathbf{y}). \quad (6.9)$$

Since $Z$ is a symmetric matrix, $\|Z\|_*$ is the summation of the absolute values of the eigenvalues of $Z$. The augmented Lagrangian function of (6.8)–(6.9) is defined as

$$\mathscr{L}(\mathbf{y}, Z, X, G) = -\mathbf{y}^T \mathbf{b} + \alpha K \|Z\|_* + \langle G, C + X + \mathscr{A}^*(\mathbf{y}) - Z \rangle + \frac{\mu}{2} \|C + X + \mathscr{A}^*(\mathbf{y}) - Z\|_F^2, \quad (6.10)$$

where $\mu > 0$ is a penalty parameter. Using the augmented Lagrangian function (6.10), we devise an ADMM that minimizes (6.10) with respect to $\mathbf{y}$, $Z$, $X$, and $G$ in an alternating fashion; that is, given some initial guess, in each iteration the following three subproblems are solved sequentially:

$$\mathbf{y}^{k+1} = \arg \min_{\mathbf{y}} \mathscr{L}\left(\mathbf{y}, Z^k, X^k, G^k\right), \quad (6.11)$$

$$Z^{k+1} = \arg \min_{Z} \mathscr{L}\left(\mathbf{y}^{k+1}, Z, X^k, G^k\right), \quad (6.12)$$

$$X^{k+1} = \arg \min_{X \geqslant 0} \mathscr{L}\left(\mathbf{y}^{k+1}, Z^{k+1}, X, G^k\right), \quad (6.13)$$

and the Lagrange multiplier $G$ is updated by

$$G^{k+1} = G^k + \gamma \mu \left(C + X^{k+1} + \mathscr{A}^*\left(\mathbf{y}^{k+1}\right) - Z^{k+1}\right), \quad (6.14)$$

where $\gamma \in (0, \frac{1+\sqrt{5}}{2})$ is an appropriately chosen step length.

To solve the subproblem (6.11), we use the first-order optimality condition

$$\nabla_{\mathbf{y}} \mathscr{L}\left(\mathbf{y}, Z^k, X^k, G^k\right) = 0$$

and the fact that $\mathscr{A}\mathscr{A}^* = I$, and we obtain

$$\mathbf{y}^{k+1} = -\mathscr{A}\left(C + X^k - Z^k\right) - \frac{1}{\mu}(\mathscr{A}(G) - \mathbf{b}).$$

By rearranging the terms of $\mathscr{L}(\mathbf{y}^{k+1}, Z, X^k, G^k)$, it can be verified that the subproblem (6.12) is equivalent to

$$\min_{Z} \frac{\alpha K}{\mu} \|Z\|_* + \frac{1}{2} \|Z - B^k\|_F^2,$$

where $B^k = C + X^k + \mathscr{A}^*(\mathbf{y}^{k+1}) + \frac{1}{\mu} G^k$. Let $B^k = U \Lambda U^T$ be the spectral decomposition of the matrix $B^k$, where $\Lambda = \text{diag}(\boldsymbol{\lambda}) = \text{diag}(\lambda_1, \ldots, \lambda_{2K})$. Then $Z^{k+1} = U \text{diag}(\hat{\mathbf{z}}) U^T$, where $\hat{\mathbf{z}}$ is the optimal solution of the problem

$$\min_{\mathbf{z}} \frac{\alpha K}{\mu} \|\mathbf{z}\|_1 + \frac{1}{2} \|\mathbf{z} - \boldsymbol{\lambda}\|_2^2. \quad (6.15)$$

It can be shown that the unique solution of (6.15) admits a closed-form solution called the soft-thresholding operator, following a terminology introduced by Donoho and Johnstone [8]; it can be written as

$$\widehat{z}_i = \begin{cases} 0 & \text{if } |\lambda_i| \leq \frac{\alpha K}{\mu}, \\ (1 - \frac{\alpha}{\mu}/|\lambda_i|)\lambda_i & \text{otherwise.} \end{cases}$$

The problem (6.13) can be shown to be equivalent to

$$\min_{X} \|X - H^k\|_F^2 \quad \text{s.t. } X \succeq 0,$$

where $H^k = Z^{k+1} - C - \mathscr{A}^*(\mathbf{y}^{k+1}) - \frac{1}{\mu} G^k$. The solution $X^{k+1} = V_+ \sum_+ V_+^T$ is the Euclidean projection of $H^k$ onto the semidefinite cone (section 8.1.1 in [3]), where

$$V \sum V^T = \begin{pmatrix} V_+ & V_- \end{pmatrix} \begin{pmatrix} \Sigma_+ & 0 \\ 0 & \Sigma_- \end{pmatrix} \begin{pmatrix} V_+^T \\ V_-^T \end{pmatrix}$$

is the spectral decomposition of the matrix $H^k$, and $\Sigma_+$ and $\Sigma_-$ are the positive and negative eigenvalues of $H^k$.

It follows from the update rule (6.14) that

$$G^{k+1} = (1-\gamma)G^k + \gamma\mu \left( C + X^{k+1} + \mathscr{A}^* \left( \mathbf{y}^{k+1} \right) - Z^{k+1} + \frac{1}{\mu} G^k \right)$$
$$= (1-\gamma)G^k + \gamma\mu \left( X^{k+1} - H^k \right).$$

## 6.2. The relaxed LUD problem

Consider the LUD problem after SDR:

$$\min_{G \succeq 0} \sum_{i<j} \|\overrightarrow{c}_{ij}^T - G_{ij} \overrightarrow{c}_{ji}^T\| \quad \text{s.t. } \mathscr{A}(G) = \mathbf{b}, \quad (6.16)$$

where $G$, $\mathscr{A}$, and $\mathbf{b}$ are defined in (4.3) and (6.4). The ADMM devised to solve (6.16) is similar to and simpler than the ADMM devised to solve the version with the spectral norm constraint. We focus on the more difficult problem with the spectral norm constraint. By introducing $\mathbf{x}_{ij} = \overrightarrow{c}_{ij}^T - G_{ij} \overrightarrow{c}_{ji}^T$ and adding the spectral norm constraint $\|G\|_2 \quad \alpha K$, we obtain

$$\min_{\mathbf{x}_{ij}, G \geqslant 0} \sum_{i<j} \|\mathbf{x}_{ij}\| \quad \text{s.t. } \mathscr{A}(G) = \mathbf{b}, \ \mathbf{x}_{ij} = \overrightarrow{c}^{T}_{ij} - G_{ij} \overrightarrow{c}^{T}_{ji}, \ \|G\|_2 \leq \alpha K. \quad (6.17)$$

The dual problem of problem (6.17) is

$$\max_{\boldsymbol{\theta}_{ij}, \mathbf{y}, X \geqslant 0} \ \min_{\mathbf{x}_{ij}, \|G\|_2 \leq \alpha K} \sum_{i<j} (\|\mathbf{x}_{ij}\| - \langle \boldsymbol{\theta}_{ij}, \mathbf{x}_{ij} - \overrightarrow{c}^{T}_{ij} + G_{ij} \overrightarrow{c}^{T}_{ji} \rangle) - \langle \mathbf{y}, \mathscr{A}(G) - \mathbf{b} \rangle - \langle G, X \rangle. \quad (6.18)$$

By rearranging terms in (6.18), we obtain

$$\max_{\boldsymbol{\theta}_{ij}, \mathbf{y}, X \geqslant 0} \ \min_{\mathbf{x}_{ij}, \|G\|_2 \leq \alpha K} -\langle Q(\boldsymbol{\theta}) + X + \mathscr{A}^*(\mathbf{y}), G \rangle + \mathbf{y}^T \mathbf{b} + \sum_{i<j} (\|\mathbf{x}_{ij}\| - \langle \boldsymbol{\theta}_{ij}, \mathbf{x}_{ij} \rangle + \langle \boldsymbol{\theta}_{ij}, \overrightarrow{c}^{T}_{ij} \rangle), \quad (6.19)$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_{ij})_{i,j=1,\dots,K}$, $\boldsymbol{\theta}_{ij} = (\theta^1_{ij}, \theta^2_{ij})^T$, $\overrightarrow{c}_{ij} = (c^1_{ij}, c^2_{ij})$,

$$Q(\boldsymbol{\theta}) = \frac{1}{2} \begin{pmatrix} Q^{11}(\boldsymbol{\theta}) & Q^{12}(\boldsymbol{\theta}) \\ Q^{21}(\boldsymbol{\theta}) & Q^{22}(\boldsymbol{\theta}) \end{pmatrix} \quad \text{and} \quad Q^{pq}(\boldsymbol{\theta}) = \begin{pmatrix} 0 & \theta^p_{12} c^p_{21} & \cdots & \theta^p_{1K} c^q_{K1} \\ c^q_{21} \theta^p_{12} & 0 & \cdots & \theta^p_{2K} c^q_{K2} \\ \vdots & \vdots & \ddots & \vdots \\ c^q_{K1} \theta^p_{1K} & c^q_{K2} \theta^p_{2K} & \cdots & 0 \end{pmatrix}$$

for $p, q = 1, 2$. It is easy to verify that for $1 \leq i < j \leq K$

$$\min_{\mathbf{x}_{ij}} (\|\mathbf{x}_{ij}\| - \langle \boldsymbol{\theta}_{ij}, \mathbf{x}_{ij} \rangle) = \begin{cases} 0 & \text{if } \|\boldsymbol{\theta}_{ij}\| \leq 1, \\ -\infty & \text{otherwise.} \end{cases} \quad (6.20)$$

In fact, (6.20) is obtained using the inequality

$$\begin{aligned} \|\mathbf{x}_{ij}\| - \langle \boldsymbol{\theta}_{ij}, \mathbf{x}_{ij} \rangle &= \|\mathbf{x}_{ij}\| - \|\boldsymbol{\theta}_{ij}\| \, \|\mathbf{x}_{ij}\| \langle \boldsymbol{\theta}_{ij}/\|\boldsymbol{\theta}_{ij}\|, \mathbf{x}_{ij}/\|\boldsymbol{x}_{ij}\| \rangle \\ &\geq \|\mathbf{x}_{ij}\| - \|\boldsymbol{\theta}_{ij}\| \, \|\boldsymbol{x}_{ij}\| = (1 - \|\boldsymbol{\theta}_{ij}\|) \, \|\boldsymbol{x}_{ij}\|, \end{aligned} \quad (6.21)$$

and the inequality in (6.21) holds when $\boldsymbol{\theta}_{ij}$ and $\mathbf{x}_{ij}$ have the same direction. Using (6.20) and the fact that the dual norm of the spectral norm is the nuclear norm, we can obtain from (6.19) the dual problem

$$\min_{\boldsymbol{\theta}_{ij}, \mathbf{y}, X \geqslant 0} -\mathbf{y}^T \mathbf{b} - \sum_{i<j} \langle \boldsymbol{\theta}_{ij}, \overrightarrow{c}^{T}_{ij} \rangle + \alpha K \|Z\|_* \quad (6.22)$$

$$\text{s.t. } Z = Q(\boldsymbol{\theta}) + X + \mathscr{A}^*(\mathbf{y}) \text{ and } \|\boldsymbol{\theta}_{ij}\| \leq 1. \quad (6.23)$$

The augmented Lagrangian function of problem (6.22)–(6.23) is defined as

$$\mathscr{L}(\mathbf{y}, \boldsymbol{\theta}, Z, X, G) = -\mathbf{y}^T \mathbf{b} + \alpha K \|Z\|_* - \sum_{i<j} \langle \boldsymbol{\theta}_{ij}, \overrightarrow{c}^{T}_{ij} \rangle + \langle G, Q(\boldsymbol{\theta}) + X + \mathscr{A}^*(\mathbf{y}) - Z \rangle + \frac{\mu}{2} \|Q(\boldsymbol{\theta}) + X + \mathscr{A}^*(\mathbf{y}) - Z\|^2_F \quad (6.24)$$

for $\|\boldsymbol{\theta}_{ij}\|$ 1, where $\mu > 0$ is a penalty parameter. Similar to section 6.1, using the augmented Lagrangian function (6.24), ADMM is used to minimize (6.24) with respect to $\mathbf{y}$, $\boldsymbol{\theta}$, $Z$, $X$, and $G$ alternatively; that is, given some initial guess, in each iteration the following four subproblems are solved sequentially:

$$\mathbf{y}^{k+1} = \arg\min_{\mathbf{y}} \mathscr{L}\left(\mathbf{y}, \boldsymbol{\theta}^k, Z^k, X^k, G^k\right), \quad (6.25)$$

$$\boldsymbol{\theta}_{ij}^{k+1} = \arg\min_{\|\boldsymbol{\theta}_{ij}\| \leq 1} \mathscr{L}\left(\mathbf{y}^{k+1}, \boldsymbol{\theta}, Z^k, X^k, G^k\right), \quad (6.26)$$

$$Z^{k+1} = \arg\min_{Z} \mathscr{L}\left(\mathbf{y}^{k+1}, \boldsymbol{\theta}^{k+1}, Z, X^k, G^k\right), \quad (6.27)$$

$$X^{k+1} = \arg\min_{X \succeq 0} \mathscr{L}\left(\mathbf{y}^{k+1}, \boldsymbol{\theta}^{k+1}, Z^{k+1}, X, G^k\right), \quad (6.28)$$

and the Lagrange multiplier $G$ is updated by

$$G^{k+1} = G^k + \gamma\mu\left(Q\left(\boldsymbol{\theta}^{k+1}\right) + X^{k+1} + \mathscr{A}^*\left(\mathbf{y}^{k+1}\right) - Z^{k+1}\right), \quad (6.29)$$

where $\gamma \in (0, \frac{1+\sqrt{5}}{2})$ is an appropriately chosen step length. The methods for solving subproblems (6.25), (6.27), and (6.28) are similar to those used in (6.11), (6.12), and (6.13). To solve subproblem (6.26), we rearrange the terms of $\mathcal{L}\left(\mathbf{y}^{k+1}, \boldsymbol{\theta}, Z^k, X^k, G^k\right)$ and obtain an equivalent problem,

$$\min_{\boldsymbol{\theta}_{ij}} -\langle\boldsymbol{\theta}_{ij}, \overrightarrow{c}_{ij}^T\rangle + \frac{\mu}{2}\|\boldsymbol{\theta}_{ij}\overrightarrow{c}_{ji} + \boldsymbol{\Phi}_{ij}\|_F^2 \quad \text{s.t.}\|\boldsymbol{\theta}_{ij}\| \leq 1, \quad (6.30)$$

where $\Phi = X^k + \mathscr{A}^*(\mathbf{y}^{k+1}) - Z^k + \frac{1}{\mu}G^k$, $\Phi = \begin{pmatrix} \Phi^{11} & \Phi^{12} \\ \Phi^{21} & \Phi^{22} \end{pmatrix}$, and $\Phi_{ij} = \begin{pmatrix} \Phi_{ij}^{11} & \Phi_{ij}^{12} \\ \Phi_{ij}^{21} & \Phi_{ij}^{22} \end{pmatrix}$. Problem (6.30) is further simplified as

$$\min_{\boldsymbol{\theta}_{ij}} \langle\boldsymbol{\theta}_{ij}, \mu\Phi_{ij}\overrightarrow{c}_{ji}^T - \overrightarrow{c}_{ij}^T\rangle + \frac{\mu}{2}\|\boldsymbol{\theta}_{ij}\|^2 \quad \text{s.t.}\|\boldsymbol{\theta}_{ij}\| \leq 1,$$

whose solution is

$$\boldsymbol{\theta}_{ij} = \begin{cases} \frac{1}{\mu}\overrightarrow{c}_{ij}^T - \Phi_{ij}\overrightarrow{c}_{ij}^T & \text{if } \|\frac{1}{\mu}\overrightarrow{c}_{ij}^T - \Phi_{ij}\overrightarrow{c}_{ij}^T\| \leq 1, \\ \frac{\overrightarrow{c}_{ij}^T - \mu\Phi_{ij}\overrightarrow{c}_{ij}^T}{\|\overrightarrow{c}_{ij}^T - \mu\Phi_{ij}\overrightarrow{c}_{ij}^T\|} & \text{otherwise.} \end{cases}$$

The practical issues related to how to take advantage of low-rank assumptions of $G$ in the eigenvalue decomposition performed at each iteration, strategies for adjusting the penalty parameter $\mu$, the use of a step size $\gamma$ for updating the primal variable $X$, and termination rules using the infeasibility measures are discussed in detail in [62]. The convergence analysis of

ADMM using more than two blocks of variables can be found in [20]. However, there are some conditions of Assumption A in [20, p. 5] that cannot be satisfied for our problem. For example, Assumption A requires that the feasible set be polyhedral, whereas the SDP cone in our problem is not a polyhedral. To generalize the convergence analysis in [20] to our problem, we will need to show that the local error bounds (see [20, pp. 8–9]) hold for the SDP cone. Currently we do not have a rigorous convergence proof for ADMM for our problem. In practice, we observe that the primal infeasibility, the dual infeasibility,[3] and the relative gap between the primal and dual problems converge simultaneously to zero, quickly during first hundreds of iterations, then the rate of convergence slows down as the iterates approach an optimal solution.

## 7. The iterative reweighted least squares (IRLS) procedure

Since $\vec{c_{ij}}$ and $\vec{c_{ji}}$ are unit vectors, it is tempting to replace the LUD problem (3.7) with the following SDR:

$$\min_{G\in\mathbb{R}^{2K\times 2K}} F(G)= \sum_{i,j=1,2,\dots,K} \sqrt{2-2\sum_{p,q=1,2} G_{ij}^{pq} S_{ij}^{pq}} \quad (7.1)$$

$$\text{s.t.} \quad G_{ii}=I_2,\ i=1,2,\dots,K, \quad (7.2)$$

$$G \succcurlyeq 0, \quad (7.3)$$

$$\|G\|_2 \leq \alpha K \text{ (optional)}, \quad (7.4)$$

where $a$ is a fixed number between $\frac{2}{3}$ and 1, and the spectral norm constraint on $G$ (7.4) is added when the solution to the problem (7.1)–(7.3) is a set of highly clustered rotations. Notice that this relaxed problem is, however, not convex since the objective function (7.1) is concave. We propose to solve (7.1)–(7.3) (possibly with (7.4)) by a variant of the IRLS procedure [6, 23, 64, 5], which at best converges to a local minimizer. With a good initial guess for $G$ it can be hoped that the global minimizer is obtained. Such an initial guess can be taken as the LS solution. In section 7.2, we prove that the sequence of IRLS converges to a stationary point. However, our convergence analysis cannot exclude the possibility of convergence to saddle points that are not stable.

### 7.1. Main idea

Before the rounding procedure, the IRLS procedure finds an approximate solution to the optimization problem (7.1)–(7.3) (possibly with (7.4)) by solving its smoothing version

$$\min_{G\in\mathbb{R}^{2K\times 2K}} F(G,\varepsilon)= \sum_{i,j=1,2,\dots,K} \sqrt{2-2\sum_{p,q=1,2} G_{ij}^{pq} S_{ij}^{pq}+\varepsilon^2} \quad (7.5)$$

$$\text{s.t.} \quad G_{ii}=I_2,\ i=1,2,\dots,K, \quad (7.6)$$

---

[3]The primal infeasibility and dual infeasibility are the residuals in the primal and dual equality constraints.

$$G \succcurlyeq 0, \quad \text{(7.7)}$$

$$\|G\|_2 \leq \alpha K \text{ (optinal),} \quad \text{(7.8)}$$

where $\varepsilon > 0$ is a small number. The solution to the smoothing version is close to the solution to the original problem. In fact, let $G_\varepsilon^* = \arg\min F(G, \varepsilon)$ and $G^* = \arg\min F(G)$; then we shall verify that

$$|F(G_\varepsilon^*) - F(G^*)| \leq 4K^2 \varepsilon. \quad \text{(7.9)}$$

Using the definition of $F(G, \varepsilon)$ and $F(G)$, we obtain that

$$0 \leq F(G, \varepsilon) - F(G) < 4K^2 \varepsilon,$$

and thus

$$(F(G^*, \varepsilon) - F(G_\varepsilon^*, \varepsilon)) + (F(G_\varepsilon^*) - F(G^*)) = (F(G^*, \varepsilon) - F(G^*)) - (F(G_\varepsilon^*, \varepsilon) - F(G_\varepsilon^*)) \leq 4K^2 \varepsilon.$$

Since $F(G^*, \varepsilon) - F(G_\varepsilon^*, \varepsilon) \geq 0$ and $F(G_\varepsilon^*) - F(G^*) \geq 0$, inequality (7.9) holds.

In each iteration, the IRLS procedure solves the problem

$$G^{k+1} = \arg\min_{G \succcurlyeq 0} \sum_{i \neq j} w_{ij}^k \left( 2 - 2 \langle G_{ij}, S_{ij} \rangle + \varepsilon^2 \right) \quad \text{s.t. } \mathcal{A}(G) = \mathbf{b}, \text{ (optional: } \|G\|_2 \leq \alpha K) \quad \text{(7.10)}$$

on the $(k + 1)$th iteration, where $w_{ij}^0 = 1$, and

$$w_{ij}^k = 1 / \sqrt{2 - 2 \langle G_{ij}^k, S_{ij} \rangle + \varepsilon^2} \quad \text{for all } k > 0.$$

In other words, in each iteration, more emphasis is given to detected common-lines that are better explained by the current estimate $G^k$ of the Gram matrix. The inclusion of the regularization parameter $\varepsilon$ ensures that no single detected common-line can gain undue influence when solving problem (7.10), which is equivalent to

$$G^{k+1} = \arg\max_{G \succcurlyeq 0} \langle W^k \circ S, G \rangle \quad \text{s.t. } \mathcal{A}(G) = \mathbf{b} \text{ (optional: } \|G\|_2 \leq \alpha K). \quad \text{(7.11)}$$

We repeat the process until the residual sequence $\{r^k\}$ has converged or the maximum number of iterations has been reached.

## 7.2. Convergence analysis

We shall verify below that the value of the cost function is nonincreasing, and that every cluster point of the sequence of IRLS is a stationary point of (7.5)–(7.7) in the following

lemma and theorem, for the problem without the spectral norm constraint on $G$. The arguments can be generalized to the case with the spectral norm constraint. The proof of Theorem 7.2 follows the method of proof for Theorem 3 in the paper [30] by Mohan and Fazel.

**Lemma 7.1—**The value of the cost function sequence is monotonically nonincreasing; i.e.,

$$F(G^{k+1}, \varepsilon) \leq F(G^k, \varepsilon), \quad (7.12)$$

where $\{G^k\}$ is the sequence generated by the IRLS procedure of Algorithm 1.

### Algorithm 1

(**the IRLS procedure**). Solve optimization problem (7.1)–(7.3) (with the spectral norm constraint on $G$ (7.4) if the input parameter $\alpha$ satisfies $\frac{2}{3} \leq \alpha < 1$), and then recover the orientations by rounding.

---

**Input:** a $2K \times 2K$ common-line matrix $S$, a regularization parameter $\varepsilon$, a parameter $\alpha$, and the total number of iterations $N_{\text{iter}}$.

$w_{ij}^0 = 1$ for all $i, j = 1, \ldots, K$;

$G^0 = 0$;

**for** $k = 1 \to N_{\text{iter}}$, step size = 1 **do**

update $W$ by setting $w_{ij} = w_{ij}^{k-1}$;

if $\frac{2}{3} \leq \alpha < 1$, obtain $G^k$ by solving the problem (6.1)–(6.3) using ADMM; otherwise, obtain $G^k$ by solving (4.6)–(4.8) using SDPLR (with initial guess $G^{k-1}$);

$r_{ij}^k = \sqrt{2 - 2\sum_{p,q=1,2} G_{ij}^{pq} S_{ij}^{pq} + \varepsilon^2}$;

$w_{ij}^k = 1/r_{ij}^k$;

the residual $r^k = \sum_{i,j=1}^K r_{ij}^k$;

**end for**

obtain estimated orientations from $G^{N_{\text{iter}}}$ using the randomized rounding procedure in section 4.4.

**Output:** estimated orientations $\hat{R_1}, \ldots, \hat{R_K}$.

---

**<u>Proof:</u>** Since $G^k$ is the solution of (7.11), using the Karush–Kuhn–Tucker (KKT) conditions, there exist $\mathbf{y}^k \in \mathbb{R}^{2K}$ and $X^k \in \mathbb{R}^{2K \times 2K}$ such that

$$-\mathscr{A}^*(\mathbf{y}^k) + X^k + W^{k-1} \circ S = 0, \quad \mathscr{A}(G^k) - \mathbf{b} = 0, \quad (7.13)$$

$$G^k \succcurlyeq 0, \ X^k \succcurlyeq 0, \left\langle G^k, X^k \right\rangle = 0. \quad (7.14)$$

Hence we have

$$0 = -(\mathbf{y}^k)^T \left(\mathscr{A}(G^k) - \mathbf{b}\right) + (\mathbf{y}^{(k+1)})^T \left(\mathscr{A}(G^{k+1}) - \mathbf{b}\right)$$
$$= (\mathbf{y}^{k+1} - \mathbf{y}^k)^T \left(\mathscr{A}(G^k) - \mathbf{b}\right) + \left\langle \mathscr{A}^*(\mathbf{y}^{k+1}), G^{k+1} - G^k \right\rangle$$
$$= \left\langle X^{k+1} + W^k \circ S, G^{k+1} - G^k \right\rangle$$
$$= \left\langle X^{k+1}, G^{k+1} \right\rangle - \left\langle X^{k+1}, G^k \right\rangle + \left\langle W^k \circ S, G^{k+1} - G^k \right\rangle \qquad (7.15)$$
$$\leq \left\langle W^k \circ S, G^{k+1} - G^k \right\rangle$$
$$= \tfrac{1}{2} \sum_{i \neq j} \left( -w_{ij}^k \left( 2 - 2 \left\langle G_{ij}^{k+1}, S_{ij} \right\rangle + \varepsilon^2 \right) + w_{ij}^k \left( 2 - 2 \left\langle G_{ij}^k, S_{ij} \right\rangle + \varepsilon^2 \right) \right)$$

$$= \frac{1}{2} \sum_{i \neq j} \left( -\frac{2 - 2\left\langle G_{ij}^{k+1}, S_{ij} \right\rangle + \varepsilon^2}{\sqrt{2 - 2\left\langle G_{ij}^k, S_{ij} \right\rangle + \varepsilon^2}} + \sqrt{2 - 2\left\langle G_{ij}^k, S_{ij} \right\rangle + \varepsilon^2} \right), \quad (7.16)$$

where the first and third equalities use (7.13), and the inequality (7.15) uses (7.14). From (7.16) we obtain

$$F(G^k, \varepsilon)^2 = \left( \sum_{i \neq j} \sqrt{2 - 2\left\langle G_{ij}^k, S_{ij} \right\rangle + \varepsilon^2} \right)^2$$
$$\geq \left( \sum_{i \neq j} \sqrt{2 - 2\left\langle G_{ij}^k, S_{ij} \right\rangle + \varepsilon^2} \right) \left( \sum_{i \neq j} \frac{2 - 2\left\langle G_{ij}^{k+1}, S_{ij} \right\rangle + \varepsilon^2}{\sqrt{2 - 2\left\langle G_{ij}^k, S_{ij} \right\rangle + \varepsilon^2}} \right) \quad (7.17)$$
$$\geq \left( \sum_{i \neq j} \sqrt{2 - 2\left\langle G_{ij}^{k+1}, S_{ij} \right\rangle + \varepsilon^2} \right)^2 = F(G^{k+1}, \varepsilon)^2,$$

where the last inequality uses the Cauchy–Schwarz inequality and the equality holds if and only if

$$\frac{\sqrt{2 - 2\left\langle G_{ij}^{k+1}, S_{ij} \right\rangle + \varepsilon^2}}{\sqrt{2 - 2\left\langle G_{ij}^k, S_{ij} \right\rangle + \varepsilon^2}} = c \quad \text{for all } i \neq j, \quad (7.18)$$

where $c$ is a constant. Thus (7.12) is confirmed.

**Theorem 7.2**—The sequence of iterates $\{G^k\}$ of IRLS is bounded, and every cluster point of the sequence is a stationary point of (7.5)–(7.7).

**Proof:** Since trace$(G^k) = 2K$ and $G^k \succcurlyeq 0$, the sequence $\{G^k\}$ is bounded. It follows that $W^k$ and trace$((W^k \circ S)G^{k+1})$ are bounded. Using the strong duality of SDP, we conclude that $\mathbf{b}^T \mathbf{y}^{k+1} = \text{trace}((W^k \circ S)G^{k+1})$ is bounded. In addition, from the KKT conditions (7.13)–(7.14) we obtain $-\mathscr{A}^*(\mathbf{y}^{k+1}) + W^k \circ S \succcurlyeq 0$. Using the definition of $\mathscr{A}^*$ and $S$, the property of semidefinite matrices, and the fact that $W^k$ is bounded, it can be verified that $\mathbf{y}^k$ is bounded. Using (7.13) again, we obtain

$$\|X^k\| = \|\mathscr{A}^*(\mathbf{y}^k) - W^{k-1} \circ S\| \leq \|\mathscr{A}^*(\mathbf{y}^k)\| + \|W^{k-1} \circ S\|,$$

which implies that $X^k$ is bounded.

We now show that every cluster point of $\{G^k\}$ is a stationary point of (7.5)–(7.7). Suppose to the contrary, and let $\bar{G}$ be a cluster point of $\{G^k\}$ that is not a stationary point. By the definition of cluster point, there exists a subsequence $\{G^{k_i}, W^{k_i}, X^{k_i}, \mathbf{y}^{k_i}\}$ of $\{G^k, W^k, X^k, \mathbf{y}^k\}$ converging to $(\bar{G}, \bar{W}, \bar{X}, \bar{y})$. By passing to a further subsequence if necessary, we can assume that $\{G^{k_i+1}, W^{k_i+1}, X^{k_i+1}, \mathbf{y}^{k_i+1}\}$ is also convergent, and we denote its limit by $(\hat{G}, \hat{W}, \hat{X}, \hat{\mathbf{y}})$. $G^{k_i+1}$ is defined as (7.10) or (7.11) and satisfies the KKT conditions (7.13)–(7.14). Passing to limits, we see that

$$-\mathscr{A}^*(\hat{\mathbf{y}})+\hat{X}+\overline{W}\circ S=0, \qquad \mathscr{A}(\hat{G})-\mathbf{b}=0,$$
$$\hat{G}\geqslant 0, \qquad \hat{X}\geqslant 0, \qquad \langle \hat{G}, \hat{X}\rangle=0.$$

Thus we conclude that $\hat{G}$ is a maximizer of the following convex optimization problem:

$$\max_{G\geqslant 0}\langle \overline{W}\circ S, G\rangle \qquad \text{s.t. } \mathscr{A}(\hat{G})=\mathbf{b}.$$

Next, by assumption, $\bar{G}$ is not a stationary point of (7.5)–(7.7). This implies that $\bar{G}$ is not a maximizer of the problem above, and thus $\langle \overline{W}\circ S, \hat{G}\rangle > \langle \overline{W}\circ S, \bar{G}\rangle$. From this last relation and (7.17)–(7.18) it follows that

$$F(\hat{G}, \varepsilon)<F(\overline{G}, \varepsilon). \quad (7.19)$$

Otherwise if $F(\hat{G}, \varepsilon) = F(\bar{G}, \varepsilon)$, then $\langle \hat{G}_{ij}, S_{ij}\rangle = \langle \bar{G}_{ij}, S_{ij}\rangle$ due to (7.17)–(7.18), and thus we would obtain $\langle \overline{W}\circ S, \hat{G}\rangle = \langle \overline{W}\circ S, \bar{G}\rangle$, which is a contradiction.

On the other hand, it follows from Lemma 7.1 that the sequence $\{F(G^k, \varepsilon)\}$ converges. Thus we have that

$$\lim F(G^k, \varepsilon)= \lim F(G^{k_i}, \varepsilon)=F(\overline{G}, \varepsilon)= \lim F(G^{k_i+1}, \varepsilon)=F(\hat{G}, \varepsilon),$$

which contradicts (7.19). Hence, every cluster point of the sequence is a stationary point of (7.5)–(7.7).

In addition, using Hölder's inequality, the analysis can be generalized to the reweighted approach to solve

$$\min_{G\geqslant 0}\sum_{i\neq j}(2-2\langle G_{ij}, S_{ij}\rangle)^{\frac{p}{2}} \qquad \text{s.t. } \mathscr{A}(G)=\mathbf{b} \text{ (optional:} \|G\|_2 \leq \alpha K), \quad (7.20)$$

where $0 < p < 1$. The convergence analysis of IRLS ($p = 1$) in [6, 23] can be extended to different applications with $p < 1$. The problem (7.20) is an SDR of the problem

$$\min_{R_1,\dots,R_K\in\mathbf{SO}(3)}\sum_{i\neq j}\|R_i\,(\vec{c}_{ij},0)^T-R_j\,(\vec{c}_{ji},0)^T\|^p. \quad (7.21)$$

The smaller $p$ is, the more penalty the outliers in the detected common-lines receive.

## 8. Numerical results

All numerical experiments were performed on a machine with 2 Intel(R) Xeon(R) X5570 CPUs, each with 4 cores, running at 2.93 GHz. In all the experiments, the polar Fourier transform of images for common-line detection had radial resolution $n_r = 100$ and angular resolution $n_\theta = 360$. The number of iterations was set to be $N_{\text{iter}} = 10$ in all IRLS procedures. Reconstruction from the images with estimated orientations used the Fourier-based 3D reconstruction package FIRM[4] [59]. The reconstructed volumes are shown in Figures 5, 6, and 9 using the visualization system Chimera [38].

To evaluate the accuracy or the resolution of the reconstructions, we used the 3D Fourier shell correlation (FSC) [43]. FSC measures the normalized cross correlation coefficient between two 3D volumes over corresponding spherical shells in Fourier space, i.e.,

$$\text{FSC}(i)=\frac{\sum_{\mathbf{j}\in Shell_i}\mathscr{F}(\mathbf{V}_1)(\mathbf{j})\cdot\overline{\mathscr{F}(\mathbf{V}_2)(\mathbf{j})}}{\sqrt{\sum_{\mathbf{j}\in Shell_i}|\mathscr{F}(\mathbf{V}_1)(\mathbf{j})|^2\cdot\sum_{\mathbf{j}\in Shell_i}|\mathscr{F}(\mathbf{V}_2)(\mathbf{j})|^2}}, \quad (8.1)$$

where $\mathcal{F}(\mathbf{V}_1)$ and $\mathcal{F}(\mathbf{V}_2)$ are the Fourier transforms of volumes $\mathbf{V}_1$ and $\mathbf{V}_2$, respectively; the spatial frequency $i$ ranges from 1 to $N/2 - 1$ times the unit frequency $1/(N\cdot\text{pixel size})$; $N$ is the size of a volume; and $Shell_i := \{\mathbf{j} : 0.5 + (i-1) + \varepsilon \quad \|\mathbf{j}\| < 0.5 + i + \varepsilon\}$, where $\varepsilon = 1e-4$. In this form, the FSC takes two 3D volumes and converts them into a 1D array. In section 8.2, we used the FSC 0.143 cutoff criterion [44, 2] to determine the resolutions of the ab initio models and the refined models.

### 8.1. Experiments on simulated images

We simulated 500 centered images of size $129 \times 129$ pixels with pixel size 2.4Å of the 50S ribosomal subunit (the top image in Figure 5), where the orientations of the images were sampled from the uniform distribution over $\mathbf{SO}(3)$. White Gaussian noise was added to the clean images to generate noisy images with SNR = 1/16, 1/32, and 1/64 (Figure 3). Common-line pairs that were detected with an error smaller than $10°$ were considered to be correct. The common-line detection rates were 64%, 44%, and 23% for images with SNR = 1/16, 1/32, and 1/64, respectively (Figure 2).

To measure the accuracy of the estimated orientations, we defined the mean squared error (MSE) of the estimated rotation matrices $\hat{R_1}, \dots, \hat{R_K}$ to be

$$\text{MSE}=\frac{1}{K}\sum_{i=1}^{K}\|R_i-\hat{O}\hat{R}_i\|^2, \quad (8.2)$$

where $\hat{O}$ is the optimal solution to the registration problem between the two sets of rotations $\{R_1, \dots, R_K\}$ and $\{\hat{R}_1, \dots, \hat{R}_K\}$ in the sense of minimizing the MSE. As shown in [48], there

---

[4]The FIRM package is available at https://web.math.princeton.edu/~lanhuiw/software.html.

is a simple procedure for obtaining both $\hat{O}$ and the MSE from the singular value decomposition of the matrix $\frac{1}{K}\sum_{i=1}^{K}\hat{R}_i R_i^T$.

We applied the LS approach using SDP and ADMM, and the LUD approach using ADMM and IRLS to estimate the images' orientations, then computed the MSEs of the estimated rotation matrices, and lastly reconstructed the volume (Figures 5 and 6). In order to measure the accuracy of the reconstructed volumes, we measured each volume's FSC (8.1) (Figure 7) against the clean 50S ribosomal subunit volume; that is, in our measurement $\mathbf{V}_1$ was the reconstructed volume, and $\mathbf{V}_2$ was the "ground truth" volume.

When SNR= 1/16 and 1/32, the common-line detection rate was relatively high (64% and 44%), and the algorithms without the spectral norm constraint on $G$ were enough to make a good estimation. The LUD approach using ADMM and IRLS outweighed the LS approach in terms of accuracy measured by MSE and FSC (Figures 5–7). Note that the LS approach using SDP failed when SNR = 1/32, while the LUD approach using either ADMM or IRLS succeeded. When SNR = 1/64, the common-line detection rate was relatively low (23%), and most of the detected common-lines were outliers (Figure 2). As a consequence, we observed that the algorithms without spectral norm constraint $\|G\|_2 \quad aK$ did not work. In particular, the viewing directions of images estimated by the IRLS procedures without $\|G\|$ $aK$ converged to two clusters around two antipodal directions, yielding no 3D reconstruction. The LUD approach using ADMM failed in this case; however, the IRLS procedure with an appropriate regularization on the spectral norm (i.e., $a = 0.67$ since the true rotations were uniformly sampled over $SO(3)$) gave the best reconstruction.

## 8.2. Experiments on a real dataset

A set of micrographs of E. coli 50S ribosomal subunits was provided by Dr. M. van Heel. These micrographs were acquired by a Philips CM20 at defocus values between 1.37 and 2.06$\mu$m, and they were scanned at 3.36 Å/pixel. The particles (particularly E. coli 50S ribosomal subunits) were picked using the automated particle picking algorithm in EMAN Boxer [24]. Then, using the IMAGIC software package [52, 58], the 27,121 particle images of size $90 \times 90$ pixels were phase-flipped to remove the phase-reversals in the contrast transfer function (CTF), bandpass filtered at 1/150 and 1/8.4 Å, normalized by their variances, and then translationally aligned with the rotationally averaged total sum. The particle images were randomly divided into two disjoint groups of equal number of images. The following steps were performed on each group separately.

The images were rotationally aligned and averaged to produce class averages of better quality, following the procedure detailed in [66]. For each group, the images were denoised and compressed using Fourier–Bessel-based steerable principal component analysis (FBSPCA) [65]. Then, triple products of Fourier–Bessel expansion coefficients obtained in FBSPCA were used to compute rotationally invariant features of the images, i.e, the bispectrums [42, 21, 27]. For each image, an initial set of neighboring images was computed using the normalized cross-correlation of the bispectrums, which was later refined using the method described in [49] to produce new sets of neighbors. Finally, we averaged each image with its ten nearest neighbors after alignment.[5] Three examples of averaged images are shown in Figure 8.

---

[5]We observed that, using either the LS or LUD method, the initial model reconstructed from 200 averaged images, which are averaged over 50 nearest neighbors, has the highest resolution, 11.1Å, that we can achieve. However, averaging too many neighbors may result in oversmoothing.

One thousand class averages were randomly selected from each group. The LS and LUD approaches with and without the spectral norm constraint were applied. Two reconstructed volumes were obtained from the two groups of images. The two resulting volumes were aligned and averaged to obtain the ab initio model (Figure 9(a)). We observed that the LUD approach gives much more reasonable ab initio models compared to the LS approach. In addition, the FSC of the two volumes was computed to estimate the resolution of the ab initio model (Figure 10). Among all the ab initio models, the one obtained by LS is at the lowest resolution, 17.2Å, while the one obtained by LUD through IRLS procedure has the highest resolution, 16.1Å. Notice that the FSC measures the variance error but not the bias error of the ab initio model. We also notice that the viewing directions of images estimated by the IRLS procedures without the spectral norm constraint converged to two clusters around two antipodal directions, resulting in no 3D reconstruction. Moreover, for this dataset, adding the spectral norm constraint on $G$ with $\alpha = 0.85$ did not improve the accuracy of the result, although this helped with regularizing the convergence in the IRLS procedure.

The two resulting volumes were then iteratively refined using 10,000 raw images in each group. In each refinement iteration, 2,000 template images were generated by projecting the 3D model from the previous iteration, then the orientations of the raw images were estimated using reference-template matching, and finally a new 3D model was reconstructed from the 10,000 raw images with highest correlation with the reference images. Each refinement iteration took about four hours. Therefore, a good ab initio model should be able to accelerate the refinement process by reducing the total number of refinement iterations. The FSC plots in Figure 10(a)–10(e) show the convergence of the refinement process using different ab initio models. We observed that all the refined models are at the resolution of 11.1Å. However, the worst ab initio model obtained by LS needed 7 iterations (about 28 hours) for convergence (Figure 10(a)), while the best ab initio model obtained by LUD needed 3 iterations (about 12 hours) for convergence (Figure 10(b) and (d)). Figure 10(f) uses FSC plots to compare the refined models. We observed that the refined models in Figure 10(b)–10(e) were consistent with each other, while the refined model obtained by LS in Figure 10(a) was slightly different from the others.

The average CPU times for computing the ab initio models in these two subsections are shown in Table 1. It is not surprising to see that the LS approach was the fastest and that adding the spectral norm constraint slowed down the ADMM and IRLS procedures. The reason for this slowing is that a large portion of the CPU time in ADMM and IRLS is due to the projections onto the semidefinite cone. These steps are expected to be accelerated by the recent advance in eigenspace computation [63]. However, when using the LUD approach for the real dataset, the time saved in the refinement was about 16 hours, which is much more than the time cost for computing the ab initio models (about 0.5–1 hour when ADMM was used).

## 9. Discussion

To estimate image orientations, we introduced a robust self-consistency error and used the ADMM or IRLS procedure to solve the associated LUD problem after SDR. Numerical experiments demonstrate that the solution is less sensitive to outliers in the detected common-lines than is the LS method approach. In addition, when the common-line detection rate is low, the spectral norm constraint on the Gram matrix $G$ can help to tighten the semidefinite relaxation, and thus improves the accuracy of the estimated rotations in some cases. Moreover, the numerical experiments using the real dataset (section 8.2) demonstrate that the ab initio models resulting from the LUD-based methods are more accurate than initial models that result from least squares–based methods. In particular, our initial models

require fewer time-consuming refinement iterations. We note that it is also possible to consider other self-consistency errors involving the unsquared deviations raised to some power $p$ (e.g., the cases $p = 1, 2$ correspond to LUD and LS, respectively). We observed that the accuracy of the estimated orientations can be improved by using $p < 1$, provided that the initial guess is "sufficiently good." The LUD approach and the spectral norm constraint on $G$ can be generalized to the synchronization approach to estimate the images' orientations [46].

Both LS and LUD-based algorithms have weaknesses and strengths in applications. The LS method is much faster than the LUD method, while the LUD method is more robust to outliers in the detected common-lines than is the LS method. Moreover, adding a spectral norm constraint can tighten the convex relaxation and thus improve the accuracy of the result, while it slows down the whole procedure. Based on the observations in our numerical experiments, we recommend using the LS method, the LUD method, and both methods with spectral norm constraints, respectively, for images with SNR from high to low. However, it is hard to quantify the actual value of SNR, due to the unknown 3D structure, and to choose a method accordingly. We suggest trying the LS method without any spectral norm constraint first, then examine the top four or five eigenvalues of the Gram matrix $G$ to see whether $G$ has three dominate eigenvalues. If the spectrum of $G$ is reasonable, we can examine the ab initio model that is reconstructed using the estimated rotations. If one of the two examinations fails, the LUD and the spectral norm constraint can be considered to improve the result. One can also experiment with different orientation determination algorithms and choose the result with the highest likelihood score. Once an ab initio estimation of orientations is given, it is possible to evaluate the likelihood score of the estimation in one of two ways: (1) determine the "common-lines" by the estimated rotations, compute the Fourier coefficients in the common-lines, and then compute the sum of squared distances between the implied common-lines in the Fourier space; or (2) reconstruct a 3D volume using the estimated rotations and compute the sum of deviations between the original input 2D images and the corresponding 2D projections of the 3D volume (i.e., the reprojection error).

In [61], the LUD approach is shown to be more robust than the LS approach for the synchronization problem over the rotation group $SO(d)$. Given some relative rotations $R_i^T R_j$, the synchronization problem is to estimate the rotations $R_i \in SO(d)$, $i = 1, \ldots, K$, up to a global rotation. It is verified that under a specific model of the measurement noise and the measurement graph for $R_i^T R_j$, the rotations can be exactly and stably recovered using LUD, exhibiting a phase transition behavior in terms of the proportion of noisy measurements. The problem of orientation determination using common-lines between cryo-EM images is similar to the synchronization problem. The difference is that the pairwise information given by the relative rotation $R_i^T R_j$ is full, while that given by the common-lines $\vec{c}_{ji}^T \vec{c}_{ij}$ is partial. Moreover, the measurement noise of each detected common-line $\vec{c_{ij}}$ depends on images $i$ and $j$, and thus it cannot be simply modeled, which brings the difficulties in verifying the conditions for the exact and stable orientation determination that we observed.

## Acknowledgments

# References

1. Arun KS, Huang TS, Blostein SD. Least-squares fitting of two 3-D point sets. IEEE Trans Pattern Anal Mach Intell. 1987; 9:698–700. [PubMed: 21869429]

2. Bai X, Fernandez IS, McMullan G, Scheres SHW. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. eLife Sciences. 2013; 2:e00461.

3. Boyd, S.; Vandenberghe, L. Convex Optimization. Cambridge University Press; New York: 2004.

4. Burer S, Monteiro RDC. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. Math Program. 2003; 95:329–357.

5. Candès EJ, Wakin MB, Boyd SP. Enhancing sparsity by reweighted $\ell_1$ minimization. J Fourier Anal Appl. 2008; 14:877–905.

6. Daubechies I, DeVore R, Fornasier M, Güntürk CS. Iteratively reweighted least squares minimization for sparse recovery. Comm Pure Appl Math. 2010; 63:1–38.

7. de Klerk, E. Applied Optimization. Springer; New York: 2002. Aspects of Semidefinite Programming: Interior Point Algorithms and Selected Applications.

8. Donoho DL, Johnstone IM. Adapting to unknown smoothness via wavelet shrinkage. J Amer Statist Assoc. 1995; 90:1200–1224.

9. Dutt A, Rokhlin V. Fast Fourier transforms for nonequispaced data. SIAM J Sci Comput. 1993; 14:1368–1393.

10. Farrow NA, Ottensmeyer FP. A posteriori determination of relative projection directions of arbitrarily oriented macromolecules. J Opt Soc Amer A. 1992; 9:1749–1760.

11. Fessler JA, Sutton BP. Nonuniform fast Fourier transforms using min-max interpolation. IEEE Trans Signal Process. 2003; 51:560–574.

12. Frank, J. Three Dimensional Electron Microscopy of Macromolecular Assemblies. Academic Press; New York: 1996.

13. Frank J. Cryo-electron microscopy as an investigative tool: The ribosome as an example. BioEssays. 2001; 23:725–732. [PubMed: 11494321]

14. Frank J. Single-particle imaging of macromolecules by cryo-electron microscopy. Annu Rev Biophys Biomol Struct. 2002; 31:303–319. [PubMed: 11988472]

15. Gilbert P. Iterative methods for the three-dimensional reconstruction of an object from projections. J Theoret Biol. 1972; 36:105–117. [PubMed: 5070894]

16. Goemans MX, Williamson DP. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. J ACM. 1995; 42:1115–1145.

17. Gordon R, Bender R, Herman GT. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. J Theoret Biol. 1970; 29:471–481. [PubMed: 5492997]

18. Greengard L, Lee JY. Accelerating the nonuniform fast Fourier transform. SIAM Rev. 2004; 46:443–454.

19. Harauz G, van Heel M. Exact filters for general geometry three dimensional reconstruction. Optik. 1986; 73:146–156.

20. Hong, M.; Luo, Z-Q. On the Linear Convergence of the Alternating Direction Method of Multipliers. 2012. http://arxiv.org/abs/1208.3922, preprints

21. Kondor, RI. A Complete Set of Rotationally and Translationally Invariant Features for Images. 2007. http://arxiv.org/abs/cs/0701127

22. Lebart, L.; Morineau, A.; Warwick, KM. Wiley Ser Prob Math Statist. Wiley-Interscience; New York: 1984. Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for Large Matrices.

23. Lerman, G.; McCoy, M.; Tropp, JA.; Zhang, T. Robust Computation of Linear Models, or How to Find a Needle in a Haystack. 2012. http://arxiv.org/abs/1202.4044

24. Ludtke SJ, Baldwin PR, Chiu W. EMAN: Semiautomated software for high-resolution single-particle reconstructions. J Struct Biol. 1999; 128:82–97. [PubMed: 10600563]

25. Luo Z, Ma W, So A, Ye Y, Zhang S. Semidefinite relaxation of quadratic optimization problems. IEEE Signal Process Mag. 2010; 27:20–34.

26. Mallick, SP.; Agarwal, S.; Kriegman, DJ.; Belongie, SJ.; Carragher, B.; Potter, CS. Computer Vision and Pattern Recognition, Proceedings of the 2006 IEEE Computer Society Conference. Vol. 2. IEEE Press; Piscataway, NJ: 2006. Structure and view estimation for tomographic reconstruction: A Bayesian approach; p. 2253-2260.

27. Marabini R, Carazo JM. On a new computationally fast image invariant based on bispectral projections. Pattern Recognition Lett. 1996; 17:959–967.

28. Marabini R, Herman GT, Carazo JM. 3D reconstruction in electron microscopy using ART with smooth spherically symmetric volume elements (blobs). Ultramicroscopy. 1998; 72:53–65. [PubMed: 9561797]

29. Mezzadri F. How to generate random matrices from the classical compact groups. Notices Amer Math Soc. 2007; 54:592–604.

30. Mohan K, Fazel M. Iterative reweighted algorithms for matrix rank minimization. J Mach Learning Res. 2012; 13:3441–3473.

31. Natterer, F. Classics in Appl Math. Vol. 32. SIAM; Philadelphia: 2001. The Mathematics of Computerized Tomography.

32. Nyquist H. Least orthogonal absolute deviations. Comput Statist Data Anal. 1988; 6:361–367.

33. Penczek P, Grassucci R, Frank J. The ribosome at improved resolution: New techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. Ultramicroscopy. 1994; 53:251–270. [PubMed: 8160308]

34. Penczek P, Radermacher M, Frank J. Three-dimensional reconstruction of single particles embedded in ice. Ultramicroscopy. 1992; 40:33–53. [PubMed: 1580010]

35. Penczek P, Renka R, Schomberg H. Gridding-based direct Fourier inversion of the three-dimensional ray transform. J Opt Soc Amer A. 2004; 21:499–509.

36. Penczek PA, Grassucci RA, Frank J. The ribosome at improved resolution: New techniques for merging and orientation refinement in 3d cryo-electron microscopy of biological particles. Ultramicroscopy. 1994; 53:251–270. [PubMed: 8160308]

37. Penczek PA, Zhu J, Frank J. A common-lines based method for determining orientations for N > 3 particle projections simultaneously. Ultramicroscopy. 1996; 63:205–218. [PubMed: 8921628]

38. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera—A visualization system for exploratory research and analysis. J Comput Chem. 2004; 25:1605–1612. [PubMed: 15264254]

39. Radermacher, M. Electron Tomography. Springer; New York: 2006. Weighted back-projection methods; p. 245-273.

40. Radermacher M, Wagenknecht T, Verschoor A, Frank J. A new 3-D reconstruction scheme applied to the 50S ribosomal subunit of E. coli. Ultramicroscopy. 1986; 141:RP1–2.

41. Recht B, Fazel M, Parrilo PA. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM Rev. 2010; 52:471–501.

42. Sadler BM, Giannakis GB. Shift- and rotation-invariant object reconstruction using the bispectrum. J Opt Soc Amer A. 1992; 9:57–69.

43. Saxton WO, Baumeister W. The correlation averaging of a regularly arranged bacterial cell envelope protein. J Microscopy. 1982; 127:127–138.

44. Scheres SHW, Chen S. Prevention of overfitting in cryo-EM structure determination. Nature Methods. 2012; 9:853–854. [PubMed: 22842542]

45. Serysheva II, Orlova EV, Chiu W, Sherman MB, Hamilton SL, van Heel M. Electron cryomicroscopy and angular reconstitution used to visualize the skeletal muscle calcium release channel. Nature Struct Biol. 1995; 2:18–24. [PubMed: 7719847]

46. Shkolnisky Y, Singer A. Viewing direction estimation in cryo-EM using synchronization. SIAM J Imaging Sci. 2012; 5:1088–1110.

47. Singer A, Coifman RR, Sigworth FJ, Chester DW, Shkolnisky Y. Detecting consistent common lines in cryo-EM by voting. J Struct Biol. 2010; 169:312–322. [PubMed: 19925867]

48. Singer A, Shkolnisky Y. Three-dimensional structure determination from common lines in cryo-EM by eigenvectors and semidefinite programming. SIAM J Imaging Sci. 2011; 4:543–572. [PubMed: 22536457]

49. Singer A, Zhao Z, Shkolnisky Y, Hadani R. Viewing angle classification of cryo-electron microscopy images using eigenvectors. SIAM J Imaging Sci. 2011; 4:723–759. [PubMed: 22506089]

50. So A, Zhang J, Ye Y. On approximating complex quadratic optimization problems via semidefinite programming relaxations. Math Program. 2007; 110:93–110.

51. Späth H, Watson GA. On orthogonal linear approximation. Numer Math. 1987; 51:531–543.

52. Stark H, Rodnina MV, Wieden H, Zemlin F, Wintermeyer W, van Heel M. Ribosome interactions of aminoacyl-tRNA and elongation factor Tu in the codon-recognition complex. Nature Struct Mol Biol. 2002; 9:849–854.

53. Vainshtein, B.; Goncharov, A. Determination of the spatial orientation of arbitrarily arranged identical particles of an unknown structure from their projections. Proceedings of the 11th International Congress on Electron Microscopy; Tokyo: Japanese Society for Electron Microscopy; 1986. p. 459-460.

54. van Heel M. Multivariate statistical classification of noisy images (randomly oriented biological macromolecules). Ultramicroscopy. 1984; 13:165–183. [PubMed: 6382731]

55. van Heel M. Angular reconstitution: A posteriori assignment of projection directions for 3D reconstruction. Ultramicroscopy. 1987; 21:111–123. [PubMed: 12425301]

56. van Heel M, Frank J. Use of multivariates statistics in analysing the images of biological macromolecules. Ultramicroscopy. 1981; 6:187–194. [PubMed: 7268930]

57. van Heel M, Gowen B, Matadeen R, Orlova EV, Finn R, Pape T, Cohen D, Stark H, Schmidt R, Schatz M, Patwardhan A. Single-particle electron cryo-microscopy: Towards atomic resolution. Quart Rev Biophys. 2000; 33:307–369.

58. van Heel M, Harauz G, Orlova EV, Schmidt R, Schatz M. A new generation of the imagic image processing system. J Struct Biol. 1996; 116:17–24. [PubMed: 8742718]

59. Vonesch, C.; Wang, L.; Shkolnisky, Y.; Singer, A. Fast wavelet-based single-particle reconstruction in Cryo-EM. Biomedical Imaging: From Nano to Macro; Proceedings of the 2011 IEEE International Symposium; Piscataway, NJ: IEEE Press; 2011. p. 1950-1953.

60. Wang L, Sigworth FJ. Cryo-EM and single particles. Physiology. 2006; 21:13–18. [PubMed: 16443818]

61. Wang L, Singer A. Exact and stable recovery of rotations for robust synchronization. Inform and Inference. 2013 to appear. 10.1093/imaiai/iat005

62. Wen Z, Goldfarb D, Yin W. Alternating direction augmented Lagrangian methods for semidefinite programming. Math Program Comput. 2010; 2:203–230.

63. Wen, Z.; Yang, C.; Liu, X.; Zhang, Y. Trace-penalty minimization for large-scale eigenspace computation. Optimization Online. 2013. http://www.optimization-online.org/DBHTML/2013/03/3799.html

64. Zhang, T.; Lerman, G. A Novel M-Estimator for Robust PCA. Arxiv preprint. 2012. http://arxiv.org/abs/1112.4863

65. Zhao Z, Singer A. Fourier-Bessel rotational invariant eigenimages. J Opt Soc Amer A. 2013; 30:871–877.

66. Zhao, Z.; Singer, A. Rotationally Invariant Image Representation for Viewing Angle Classification. 2013. submittedavailable at http://arxiv.org/abs/1309.7643

67. Zhu Y, Carragher B, Glaeser RM, Fellmann D, Bajaj R, Bern M, Mouche F, Haas FD, Hall RJ, Kriegman DJ, Ludtke C, Mallick SP, Penczek PA, Roseman AM, Sigworth FJ, Potter CS. Automatic particle selection: Results of a comparative study. J Struct Biol. 2004; 145:3–14. [PubMed: 15065668]

68. Zhu Z, So AMC, Ye Y. Universal rigidity and edge sparsification for sensor network localization. SIAM J Optim. 2010; 20:3059–3081.

## Appendix. Exact recovery of the Gram matrix G from correct common-lines

Here we prove that if the detected common-lines $\vec{c_{ji}}$ (defined in (3.1)) are all correct and at least three images have linearly independent projection directions (i.e., the viewing

directions of the three images are not on the same great circle on the sphere shown in Figure 4), then the Gram matrix $G$ obtained by solving the LS problem (4.6)–(4.8) or the LUD problem (4.10) is uniquely that defined in (4.3). To verify the uniqueness of the solution $G$, it is enough to show rank$(G)= 3$ due to the SDP solution uniqueness theorem (see [7, pp. 36–39], [68]). Without loss of generality, we consider the SDP for the LS approach when applied on three images (i.e., $K = 3$ and $w_{ij} = 1$ in the problem (4.6)–(4.8)):

$$\max_{G_{6\times 6}\geqslant 0} \sum_{i,j=1,2,3} \langle G_{ij}, \overrightarrow{c}_{ji}^T\overrightarrow{c}_{ij}\rangle \quad \text{s.t. } G_{ii}=I_2.$$

Since the solution $G$ is positive semidefinite, we can decompose $G$ as

$$G= \begin{pmatrix} \mathbf{u}_1^{1T} \\ \mathbf{u}_1^{2T} \\ \mathbf{u}_2^{1T} \\ \mathbf{u}_2^{2T} \\ \mathbf{u}_3^{1T} \\ \mathbf{u}_3^{2T} \end{pmatrix} \begin{pmatrix} \mathbf{u}_1^1 & \mathbf{u}_1^2 & \mathbf{u}_2^1 & \mathbf{u}_2^2 & \mathbf{u}_3^1 & \mathbf{u}_3^2 \end{pmatrix},$$

where $\mathbf{u}_i^p$, $p = 1, 2$ and $i = 1, 2, 3$, are column vectors. We will show rank$(G)= 3$; i.e., any four vectors among $\{ \mathbf{u}_1^1, \mathbf{u}_1^2, \mathbf{u}_2^1, \mathbf{u}_2^2, \mathbf{u}_3^1, \mathbf{u}_3^2\}$ span a space with dimensionality at most 3.

Define arrays $\mathbf{u}_i$ as

$$\mathbf{u}_i=(\mathbf{u}_i^1, \mathbf{u}_i^2);$$

then the inner product

$$\begin{aligned} \langle G_{ij}, \overrightarrow{c}_{ji}^T\overrightarrow{c}_{ij}\rangle &= \langle \mathbf{u}_i^T\mathbf{u}_j, \overrightarrow{c}_{ji}^T\overrightarrow{c}_{ij}\rangle \\ &= \langle \overrightarrow{c}_{ji}\mathbf{u}_i, \overrightarrow{c}_{ij}\mathbf{u}_j\rangle \\ &= \langle c_{ji}^1\mathbf{u}_i^1+c_{ji}^2\mathbf{u}_i^2, c_{ij}^1\mathbf{u}_j^1+c_{ij}^2\mathbf{u}_j^2\rangle \\ &\leq 1, \end{aligned}$$

where the last inequality follows from the Cauchy–Schwarz inequality and the facts that all $\overrightarrow{c_{ij}}$ are unit vectors, $\mathbf{u}_i^1$ and $\mathbf{u}_i^2$ are unit vectors and orthogonal to each other due to the constraint $G_{ii} = I_2$, and thus all $c_{ij}^1\mathbf{u}_j^1+c_{ij}^2\mathbf{u}_j^2$ are unit vectors on the Fourier slices of the images. The equality holds if and only if

$$c_{ji}^1\mathbf{u}_i^1+c_{ji}^2\mathbf{u}_i^2=c_{ij}^1\mathbf{u}_j^1+c_{ij}^2\mathbf{u}_j^2. \quad \text{(A.1)}$$

Thus when the maximum is achieved, due to (A.1) and the fact that the projection directions of the images are linearly independent, $\dim(\text{span}\{\mathbf{u}_i^1, \mathbf{u}_i^2\} \cap \text{span}\{\mathbf{u}_j^1, \mathbf{u}_j^2\})=1$, and thus $\dim(\text{span}\{\mathbf{u}_i^1, \mathbf{u}_i^2, \mathbf{u}_j^1, \mathbf{u}_j^2\})=3$. Therefore, without loss of generality, we only have to show

that $\dim(\mathrm{span}\{\mathbf{u}_1^1, \mathbf{u}_2^1, \mathbf{u}_3^1, \mathbf{u}_3^2\}) \leq 3$. Using (A.1), assume that

$\mathrm{span}\{\mathbf{u}_1^1, \mathbf{u}_1^2\} \cap \mathrm{span}\{\mathbf{u}_3^1, \mathbf{u}_3^2\} = \mathrm{span}(\mathbf{v}_1)$ and $\mathrm{span}\{\mathbf{u}_2^1, \mathbf{u}_2^2\} \cap \mathrm{span}\{\mathbf{u}_3^1, \mathbf{u}_3^2\} = \mathrm{span}\{\mathbf{v}_2\}$, where $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent vectors (otherwise all three projection directions are linearly dependent, and thus the three Fourier slices of the images intersect at the same line). Therefore we have $\mathrm{span}\{\mathbf{v}_1, \mathbf{v}_2\} = \mathrm{span}\{\mathbf{u}_3^1, \mathbf{u}_3^2\}, \mathrm{span}\{\mathbf{v}_1, \mathbf{u}_1^1\} \subseteq \mathrm{span}\{\mathbf{u}_1^1, \mathbf{u}_1^2\}$, and $\mathrm{span}\{\mathbf{v}_2, \mathbf{u}_2^1\} \subseteq \mathrm{span}\{\mathbf{u}_2^1, \mathbf{u}_2^2\}$. Thus

$\dim(\mathrm{span}\{\mathbf{u}_1^1, \mathbf{u}_2^1, \mathbf{u}_3^1, \mathbf{u}_3^2\}) = \dim(\mathrm{span}\{\mathbf{u}_1^1, \mathbf{u}_2^1, \mathbf{v}_1, \mathbf{v}_2\}) \leq \dim(\mathrm{span}\{\mathbf{u}_1^1, \mathbf{u}_1^2, \mathbf{u}_2^1, \mathbf{u}_2^2\}) \leq 3$.
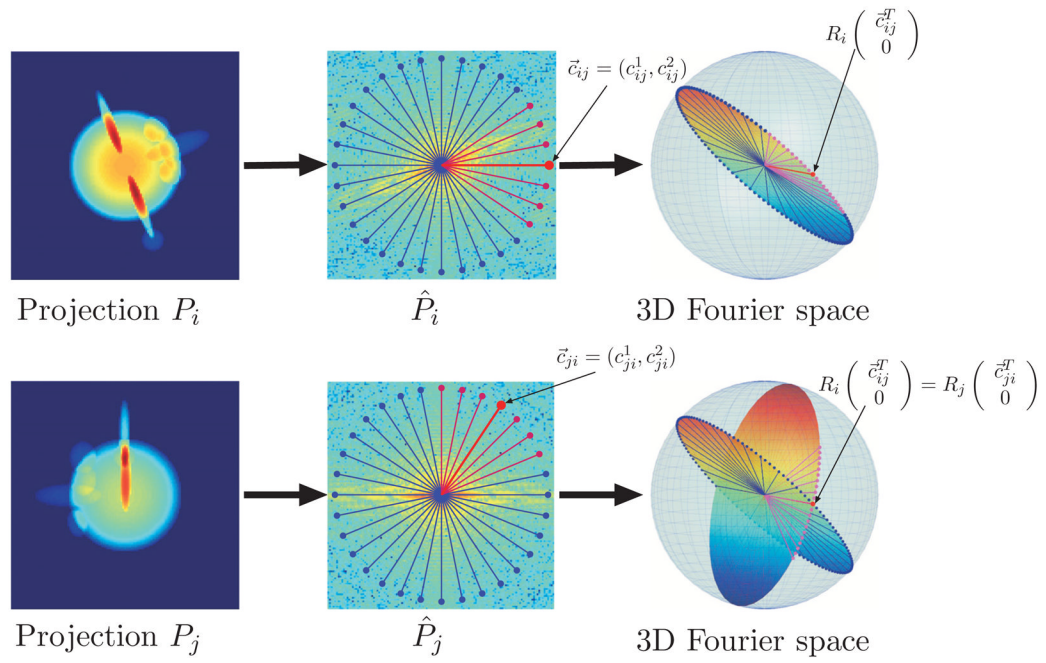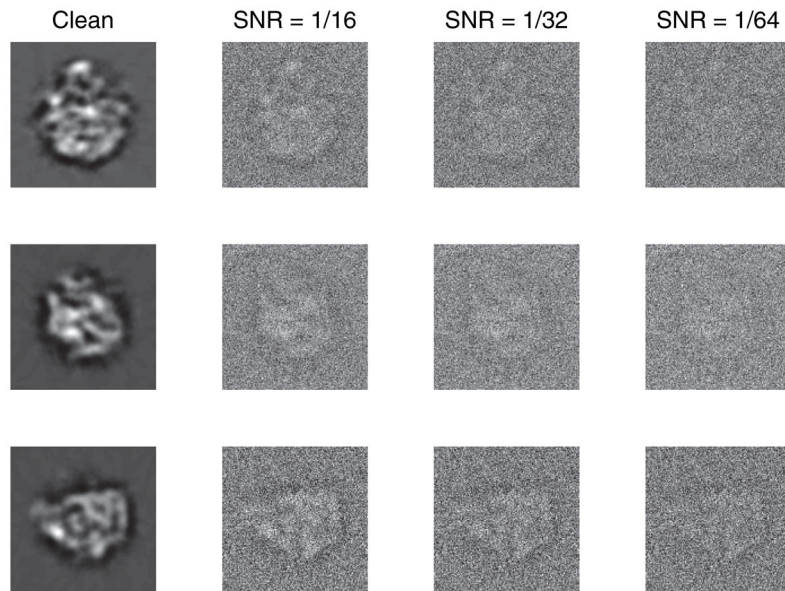
**Figure 1.**
Fourier projection-slice theorem. In the middle, $\hat{P_i}$ is the 2D Fourier transform of projection $P_i$ (left) computed on a polar grid. $\vec{c_{ij}}$ is a unit vector representing the direction of the common-line (indicated in red) between $\hat{P_i}$ and $\hat{P_j}$ on $\hat{P_i}$. On the right, the two transformed images $\hat{P_i}$ and $\hat{P_j}$ intersect with each other at the common-line after rotations $R_i$ and $R_j$, yielding (3.3), where $R_i$ and $R_j$ represent the orientations of the projections $P_i$ and $P_j$, respectively.
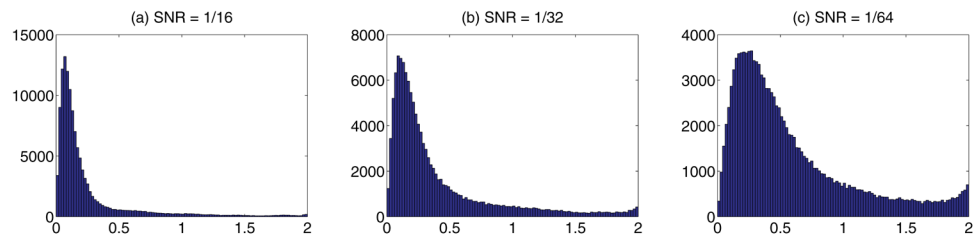
**Figure 2.**
The histogram plots of errors in the detected common-lines $\vec{c_{ij}}$ for all $i$ and $j$, i.e., $\|R_i(\vec{c_{ij}}, 0)^T - R_j(\vec{c_{ji}}, 0)^T\|$, where $R_i$ is the true rotation matrix for all $i$. The fat tail in (c) indicates that the detected common-lines contain a large number of outliers. The cryo-EM images used for common-line detection are shown in Figure 3.
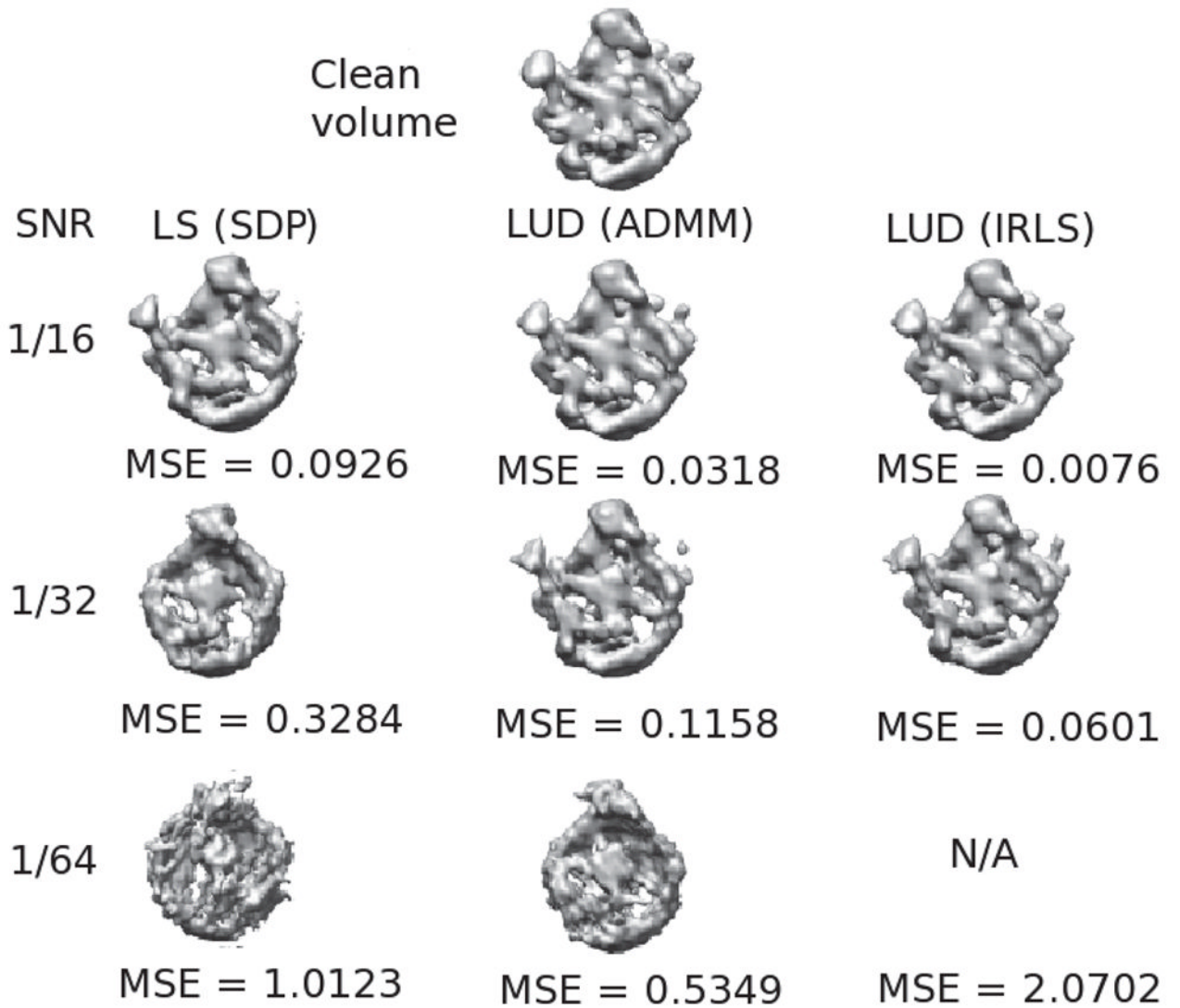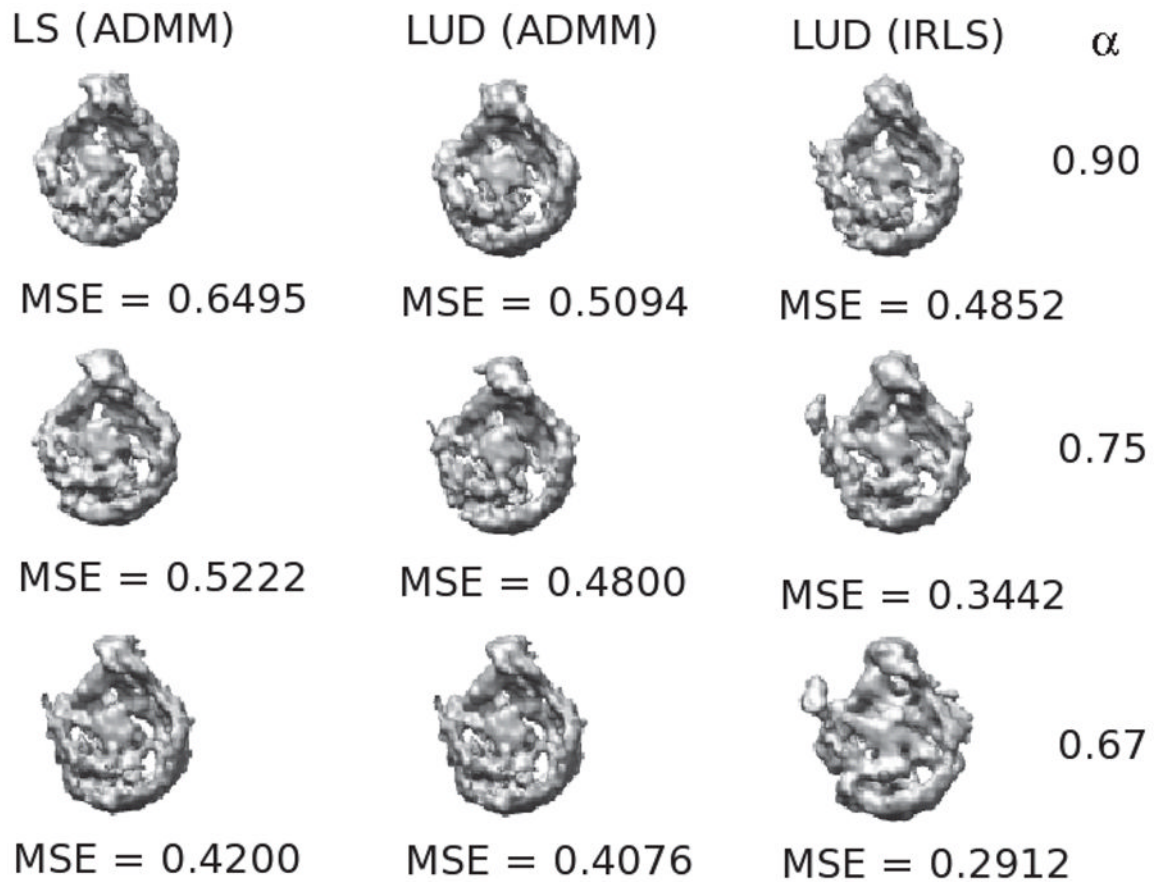
**Figure 3.**
The first column shows three clean images of size 129×129 pixels generated from a 50S ribosomal subunit volume with different orientations. The other three columns show three noisy images corresponding to those in the first column with SNR = 1/16, 1/32, and 1/64, respectively.
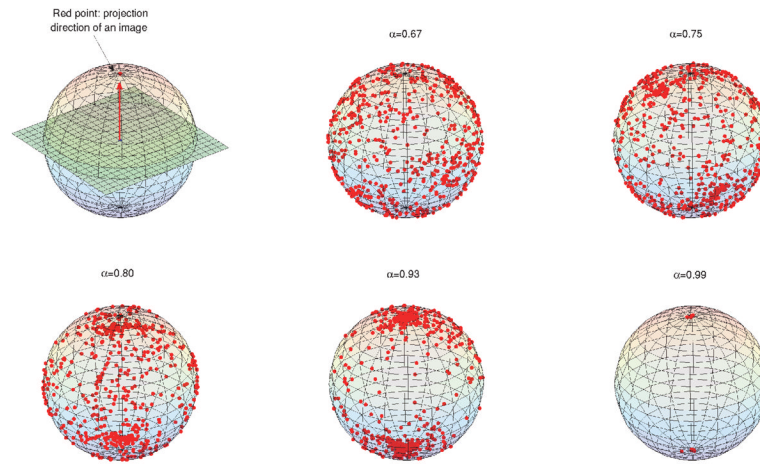
**Figure 4.**
The dependency of the spectral norm of $G$ (denoted as $aK$ here) on the distribution of orientations of the images. Each red point denotes the viewing direction of a projection. Five different distributions of viewing directions of $K = 500$ projections are plotted on spheres. The larger the spectral norm $aK$ is, the more clustered the viewing directions are.
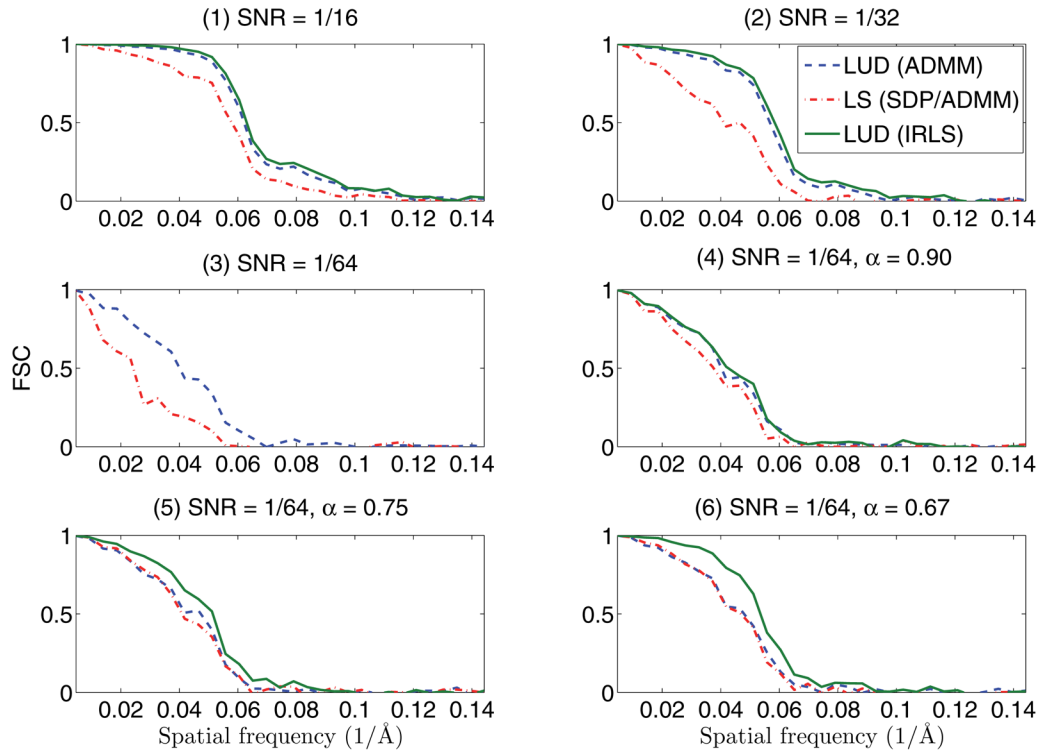
**Figure 5.**
The clean volume (top), the reconstructed volumes, and the MSEs (8.2) of the estimated rotations. No spectral norm constraint was used (i.e., $\alpha$ = N/A) for all algorithms. The result using the IRLS procedure without $\alpha$ is not available due to the highly clustered estimated projection directions. All the volumes have size $129 \times 129 \times 129$ voxels.

LS (ADMM)          LUD (ADMM)          LUD (IRLS)          α

0.90

MSE = 0.6495       MSE = 0.5094        MSE = 0.4852

0.75

MSE = 0.5222       MSE = 0.4800        MSE = 0.3442

0.67

MSE = 0.4200       MSE = 0.4076        MSE = 0.2912

**Figure 6.**
The reconstructed volumes from images with SNR = 1/64 and the MSEs (8.2) of the estimated rotations using spectral norm constraints (i.e., $\alpha = 0.90$, 0.75, and 0.67) for all algorithms. All the volumes have size $129 \times 129 \times 129$ voxels. The result from the IRLS procedure with $\alpha = 0.67$ for the spectral norm constraint is best.
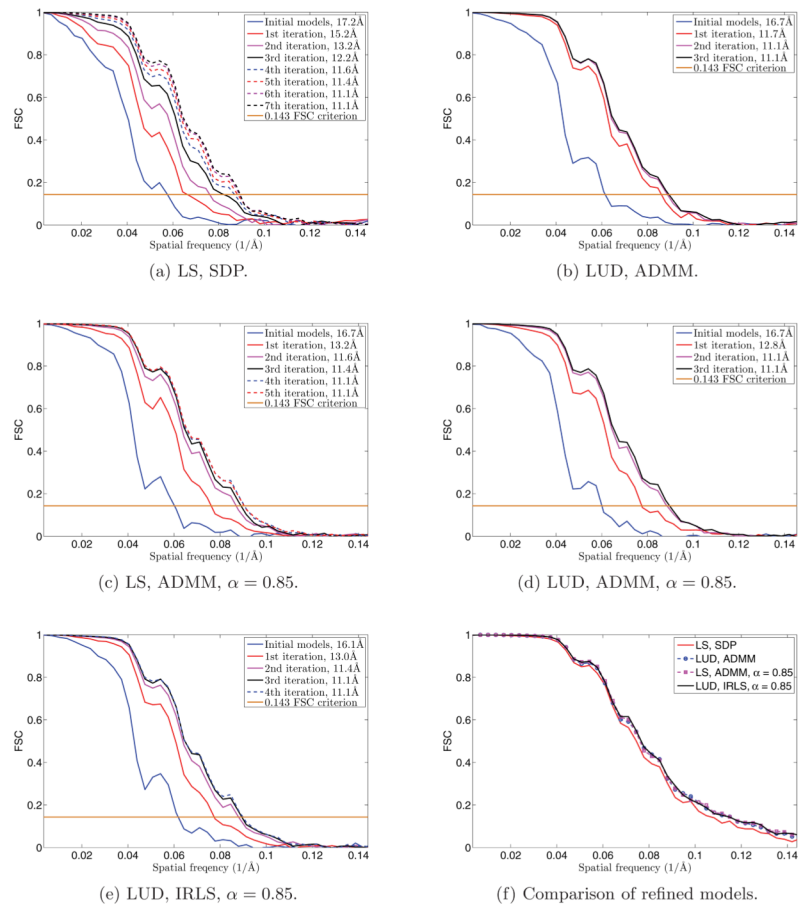
**Figure 7.**
FSCs (8.1) of the reconstructed volumes against the clean volume in Figures 5 and 6. The plots of the correlations show that the LUD approach using ADMM and IRLS (denoted as the blue dashed lines and green solid lines) outweighed the LS approach using SDP or ADMM (denoted as the red dot-dashed lines). Note that all of the last four subfigures are results for images with SNR = 1/64, where the last three subfigures are results using different $a$ for the spectral norm constraint. In the third subfigure (left to right, top to bottom), there is no green solid line for the LUD approach using IRLS, since the IRLS procedure without the spectral norm constraint converges to a solution where the estimated projection directions are highly clustered and no 3D reconstruction can be computed.
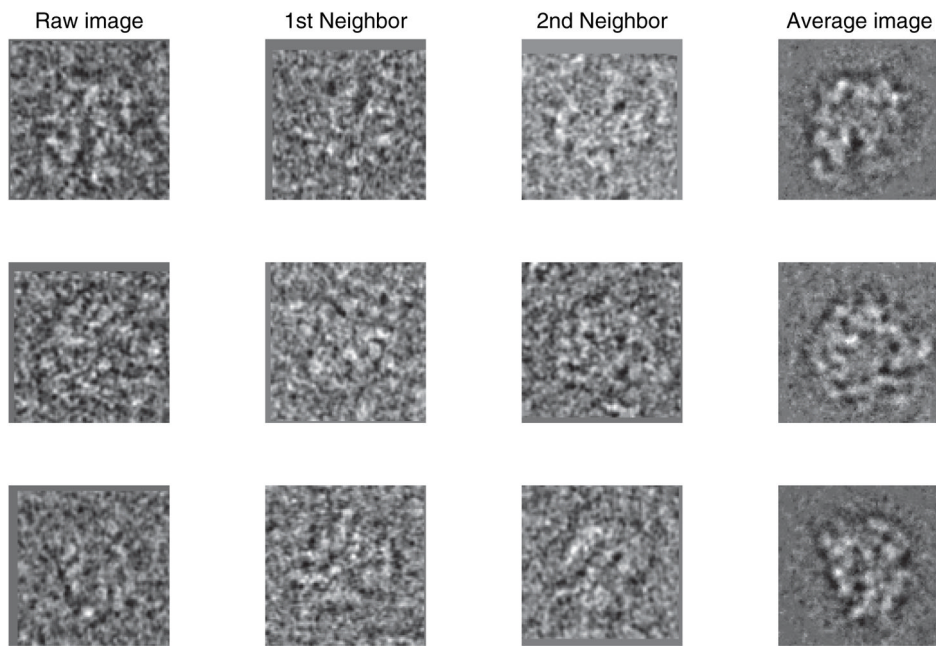
**Figure 8.**
Noise reduction by image averaging. Three raw ribosomal images are shown in the first column. Their closest two neighbors (i.e., raw images having similar orientations after alignments) are shown in the second and third columns. The average images shown in the last column were obtained by averaging over 10 neighbors of each raw image.

(a) Initial models.
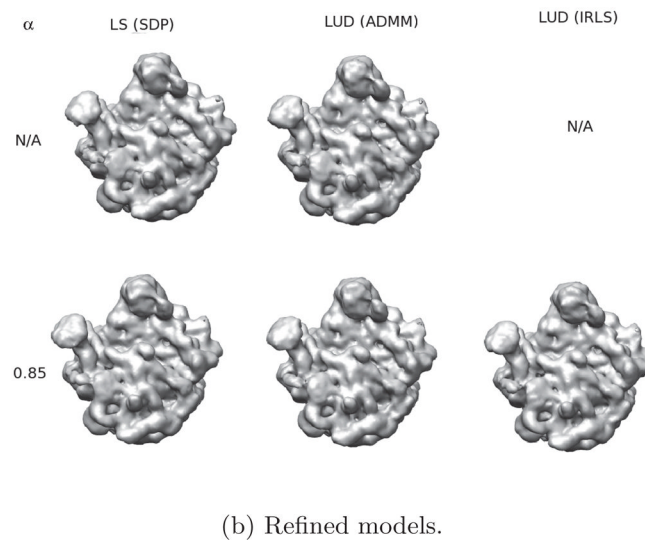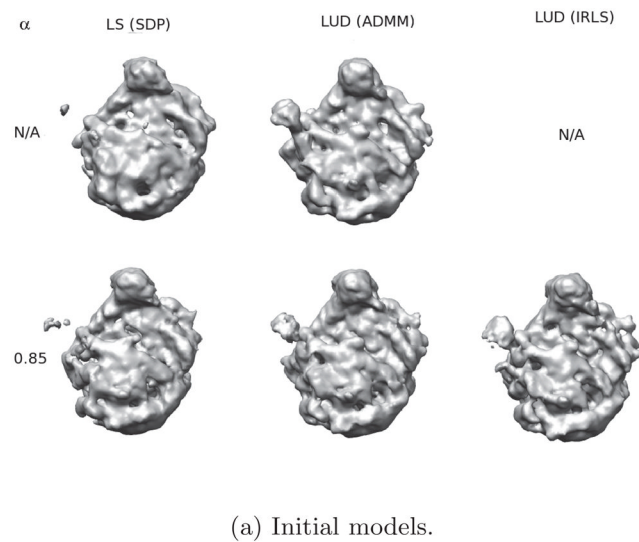


(b) Refined models.

**Figure 9.**
Initial models and refined models. (a) The ab initio models estimated by merging two independent reconstructions, each obtained from 1000 class averages. The resolutions of the models are 17.2Å, 16.7Å, 16.7Å, 16.7Å, and 16.1Å (from top to bottom, left to right) using the FSC 0.143 resolution cutoff (Figure 10). The model using the IRLS procedure without the spectral norm constraint (i.e., $\alpha$ = N/A) is not available, since the estimated projection directions are highly clustered. (b) The refined models corresponding to the ab initio models in (a). The resolutions of the models are all 11.1Å.

(a) LS, SDP.

(b) LUD, ADMM.

(c) LS, ADMM, $\alpha = 0.85$.

(d) LUD, ADMM, $\alpha = 0.85$.

(e) LUD, IRLS, $\alpha = 0.85$.

(f) Comparison of refined models.

**Figure 10.**
Convergence of the refinement process. In panels (a)–(e), the FSC plots show the convergence of the refinement iterations. The ab initio models (Figure 9(a) used in (a)–(e) were obtained by solving the LS/LUD problems using SDP/ADMM/IRLS. The numbers of refinement iterations performed in (a)–(e) are 7, 3, 5, 3, and 4, respectively. Panel (f) shows FSC plots of the refined models in (a), (b), (c), and (e) against the refined model in (d), which are measurements of similarities between the refined models in Figure 9(b).

**Table 1**

The average cost time using different algorithms on 500 and 1000 images in the two experimental subsections. The notation $\alpha = $ N/A means that no spectral norm constraint $\|G\|_2 \quad \alpha K$ is used.

| $K$ | $\alpha = $ N/A | | | | $\frac{2}{3} \leq \alpha \leq 1$ | | | |
|---|---|---|---|---|---|---|---|---|
| | LS (SDP) | LUD | | LS (ADMM) | | | LUD | |
| | | ADMM | IRLS | | | | ADMM | IRLS |
| 500 | 7s | 266s | 469s | 78s | | | 454s | 3353s |
| 1000 | 31s | 1864s | 3913s | 619s | | | 1928s | 20918s |