ORIENTATION DETERMINATION OF CRYO-EM IMAGES USING LEAST UNSQUARED DEVIATIONS

LANHUI WANG*, AMIT SINGER†, AND ZAIWEN WEN‡

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 [41] 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 resulted problems are solved by 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.

Key words. Angular reconstitution, cryo-electron microscopy, single particle reconstruction, common lines, least unsquared deviations, semidefinite relaxation, alternating direction method of multipliers, iteratively reweighted least squares

AMS subject classification. 92E10, 68U10, 94A08, 92C55, 90C22, 90C25

1. Introduction. In single particle analysis, cryo-electron microscopy (Cryo-EM) is used to attain a resolution sufficient to interpret fine details in three-dimensional (3D) macromolecular structures [12, 13, 50, 53]. 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. In addition, the unknown orientational information of the imaged particles need to be estimated for 3D reconstruction. An ab-initio estimation of the orientations of images using the random-conical tilt technique [33] or common-lines based approaches [40, 41, 48] are often applied after multivariate statistical analysis [18, 49] and classification techniques [29, 42, 47] that are used to sort and partition the large set of images by their viewing directions, producing "class averages" of enhanced signal-to-noise ratio (SNR). Using the ab-initio estimation of the orientations, a preliminary 3D map is reconstructed from the images by a 3D reconstruction algorithm. The initial model is then iteratively refined [28] in order to obtain a higher-resolution 3D reconstruction.

The Fourier projection-slice theorem (see, e.g., [26]) plays a fundamental role in the 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

*The Program in Applied and Computational Mathematics (PACM), Princeton University, Fine Hall, Washington Road, Princeton, NJ 08544-1000, USA, lanhuiw@math.princeton.edu, Corresponding author. Tel.: +1 609 258 5785; fax: +1 609 258 1735.

†Department of Mathematics and PACM, Princeton University, Fine Hall, Washington Road, Princeton, NJ 08544-1000, USA, amits@math.princeton.edu

‡Department of Mathematics, MOE-LSC and Institute of Natural Sciences, Shanghai Jiao Tong University, Pao Yue-Kong Library, 800 Dongchuan Rd, Shanghai, China, zw2109@sjtu.edu.cn

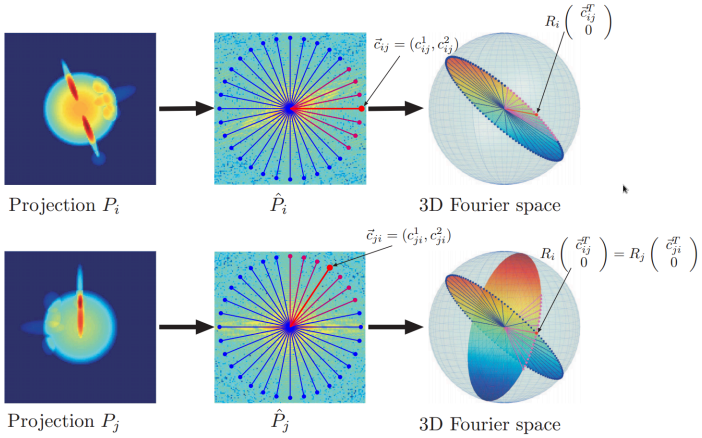


Fig. 1.1: Fourier projection-slice theorem. In the middle, $\hat{P}_i$ is a polar Fourier transform of projection P_i on the left. The red line $\vec{c}_{ij}$ represents the direction of a common-line between $\hat{P}_i$ and $\hat{P}_j$ on $\hat{P}_i$. On the right, the two transformed images $\hat{P}_i$ and $\hat{P}_j$ intersect with each other at the common-line after rotations R_i and R_j, yielding the equation (3.3).

(Figure 1.1). Thus, any two projections imaged from non-parallel 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 [48], which was also developed independently by Vainshtein and Goncharov [46]. 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 square sense. The main problem with such techniques is that they are sensitive to false detection of common lines that leads to the accumulation of errors. Penczek et.al. [31] 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. [22] and Singer et al. [40] applied Bayesian approaches to use common-lines information from different groups of projections. Recently, Singer and Shkolnisky [41] 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 signal-to-noise ratio (SNR) of the image 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, that have no structure. The standard common-lines based methods, including those using least squares (LS) [10, 41], 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 [27, 44], rather than the sum of squared residuals of the LS formulation. Convex relaxations of least unsquared deviations (LUD) have been recently proposed for other applications, such as robust principal component analysis [19] and robust synchronization of orthogonal transformations [54]. Under certain noise models for the distribution of the outliers (e.g., the haystack model of [19]), 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 to consider in this paper the application of LUD to the problem of orientation estimation from common-lines in single particle reconstruction.

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. This artificial clustering can be explained by the fact that images that share the same viewing direction also share more than one common line. In order to mitigate this spurious clustering of 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), which has been proved to converge to the global minimizer in many cases [16]. 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 resulted by our new methods are more accurate and require fewer refinement iterations compared to least squares 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.

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 non-uniform fast Fourier transform (NUFFT) [9, 11, 15]. The transformed images have resolution n_r in the radial direction and resolution n_θ 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_θ is the number of angularly equally-spaced Fourier rays computed for each image (Figure 1.1). For simplicity, we let n_θ be an even number. The transformed images are denoted as $\left(\vec{l}_0^{\,k}, \vec{l}_1^{\,k}, \ldots, \vec{l}_{n_\theta-1}^{\,k}\right)$, where $\vec{l}_m^{\,k} = \left(l_{m,1}^k, l_{m,2}^k, \ldots, l_{m,n_r}^k\right)$ 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 DC term is shared by all lines independently of the image, and is therefore excluded for comparison. To determine the common line between two images P_i and P_j, the similarity between all n_θ radial lines $\vec{l}_0^{\,i}, \vec{l}_1^{\,i}, \ldots, \vec{l}_{n_\theta-1}^{\,i}$ from the first image with all n_θ radial lines $\vec{l}_0^{\,j}, \vec{l}_1^{\,j}, \ldots, \vec{l}_{n_\theta-1}^{\,j}$ from the second image are measured (overall n_θ^2 comparisons), and the pair of radial lines $\vec{l}_{m_{i,j}}^{\,i}$ and $\vec{l}_{m_{j,i}}^{\,j}$ 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_θ). 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_θ possibilities (see also [48] and [30], 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 [38, 48]).

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

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

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 [48], 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 a given discretization error tolerance) as "correctly detected" common lines, or "inliers" and to the remaining common lines as "falsely detected", or "outliers".

3. Weighted LS and least unsquared deviation (LUD). We define the directions of detected common-lines between the transformed image i and transformed image j as unit vectors (Figure 1.1)

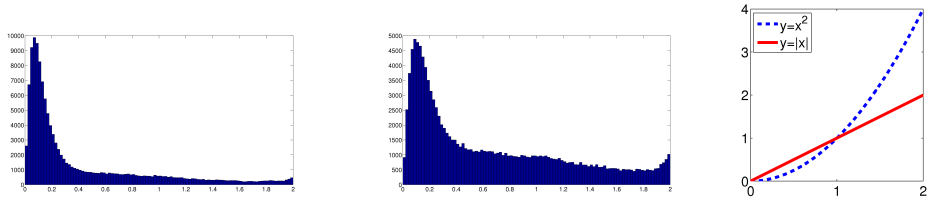
$$\vec{c}_{ij} = \left(c^1_{ij}, c^2_{ij}\right) = \left(\cos\left(2\pi m_{ij}/n_\theta\right), \sin\left(2\pi m_{ij}/n_\theta\right)\right), \tag{3.1}$$

$$\vec{c}_{ji} = \left(c^1_{ji}, c^2_{ji}\right) = \left(\cos\left(2\pi m_{ji}/n_\theta\right), \sin\left(2\pi m_{ji}/n_\theta\right)\right), \tag{3.2}$$

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

$$R_i \begin{pmatrix} \vec{c}^{\,T}_{ij} \\ 0 \end{pmatrix} = R_j \begin{pmatrix} \vec{c}^{\,T}_{ji} \\ 0 \end{pmatrix} \text{ for } 1 \le i < j \le K. \tag{3.3}$$

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



(a) SNR=1/32 (b) SNR=1/64 (c) x^2 vs $|x|$

Fig. 3.1: Left and Middle: The histogram plots of errors in the detected common-lines $\vec{c}_{ij}$ for all i and j, i.e., $\left\| R_i \left(\vec{c}_{ij}, 0\right)^T - R_j \left(\vec{c}_{ji}, 0\right)^T \right\|$ where R_i is a true rotation matrix for all i. The fat tail in (b) indicates the detected common-lines contain a large amount of outliers. Right: elucidating the difference between the squared distance and the absolute deviation.

matrix does not contribute in (3.3) due to the zero third entries in the common-line vectors in $\mathbb{R}^3$). The weighted LS approach for solving this system can be formulated as the minimization problem

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

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

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

When the weight $w_{ij} = 1$ for each pair $i \neq j$, (3.5) is equivalent to the LS problem that was considered in [31], and more recently in [41] using convex relaxation of the non-convex constraint set. The solution to the LS problem may not be optimal due to the typically large proportion of outliers (Figure 3.1).

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

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

or equivalently,

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

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 (Figure 3.1c). We remark that it is also possible to consider the weighted version of (3.6), namely

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

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 non-convex and therefore extremely difficult to solve if one requires the matrices R_i to be rotations, that is, when adding the constraints

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

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

The relaxation in [41] that uses semidefinite programming (SDP) can be modified in a straightforward manner in order to deal with non-unity 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 SDP employed here is slightly different than the one presented in [41] and is closer in spirit to the rounding procedure of Goemans and Williamson for the MAX-CUT problem [14]. Finally, in the Appendix we prove exact recovery of the rotations by the semidefinite relaxation 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:

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

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,

$$G = R^T R. \tag{4.3}$$

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

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

where G_{ij} is the 2×2 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

$$G_{ii} = I_2,\ i = 1, 2, \cdots, K, \tag{4.4}$$

where I_2 is the 2×2 identity matrix.

4.2. SDR for weighted LS. We first define two $2K \times 2K$ matrices $S = (S_{ij})_{i,j=1,\cdots,K}$ and $W = (W_{ij})_{i,j=1,\cdots,K}$, where the 2×2 sub-blocks S_{ij} and W_{ij} are given by

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

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 $(W \circ S)\, G$:

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

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

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

$$\text{s.t.} \quad G_{ii} = I_2,\ i = 1, 2, \cdots, K, \tag{4.7}$$

$$G \succcurlyeq 0 \tag{4.8}$$

The non-convex rank-3 constraint on the Gram matrix G is missing from this semidefinite relaxation (SDR) [21]. 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 $\tilde{G} = (\tilde{G}_{ij})_{i,j=1,\cdots,K}$, where each $\tilde{G}_{ij}$ is a 3×3 block defined as $\tilde{G}_{ij} = R_i^T R_j$. Then, a natural SDR for (3.7) is given by

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

The constraints missing in this SDP formulation are the non-convex 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 the solution to (3.7), then the set of conjugated rotation matrices $\{JR_iJ\}$ is also the 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 $\tilde{G}^J = (\tilde{G}^J_{ij})_{i,j=1,\cdots,K}$ with the 3×3 sub-blocks given by $\tilde{G}^J_{ij} = JR_i^T JJR_j J = JR_i^T R_j J$. It can be verified that $\frac{1}{2}(\tilde{G}+\tilde{G}^J)$ is also a solution to (4.9). Using the fact that

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

the problem (4.9) is reduced to

$$\min_{G\succcurlyeq 0} \sum_{i\neq j} \left\| \vec{c}^T_{ij} - G_{ij}\vec{c}^T_{ji} \right\|, \text{ s.t. } G_{ii} = I_2. \tag{4.10}$$

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, $P = Q\ \mathrm{sign}(\mathrm{diag}(R))$, where sign stands for the entry-wise sign function and $\mathrm{diag}(R)$ 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 [24]. We project the solution G onto the subspace spanned by the three columns of the matrix GP [1].

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

$$(GP)^T = \begin{pmatrix} A_1 & A_2 & \cdots & A_K \end{pmatrix}$$

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 $R_i^{[1,2]} = \begin{pmatrix} R_i^1 & R_i^2 \end{pmatrix}$ 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 $A_i = U_i\Sigma_i V_i^T$ is the singular value decomposition of A_i. We note that except for the orthogonality constraint (4.7), the semidefinite program (4.6)–(4.8) is identical to the Goemans–Williamson SDP for finding the maximum cut in a weighted graph [14], where the SDR and the randomized rounding procedure [21, 43] for maximum cut problem is proved to have a 0.87 performance guarantee. 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 [41].

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

[1] The 3 dimensional 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 [41].

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

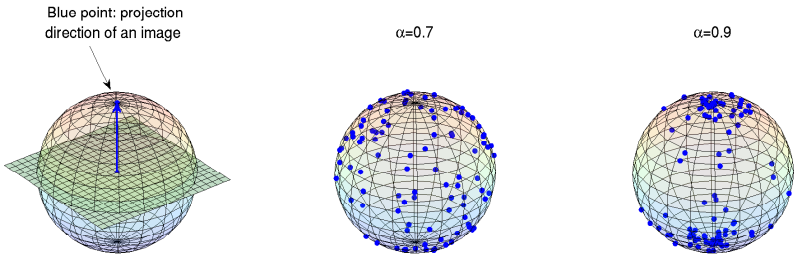


Fig. 5.1: The dependency of the spectral norm of G (denoted as αK here) on the distribution of orientations of the images. Here $K = 100$. The larger α is, the more clustered the orientations are.

by either solving (4.6)-(4.8) or (4.10) are highly clustered (Figure 5.1). This empirical behavior of the solutions can be explained by the fact that images whose viewing directions are parallel share many common lines. In other words, when the viewing directions of R_i and R_j are nearby, the fidelity term $\left\| R_i \left(\vec{c}_{ij}, 0\right)^T - R_j \left(\vec{c}_{ji}, 0\right)^T \right\|$ (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.

In order to prevent the viewing directions from clustering, 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):

$$G \preccurlyeq \alpha K I_{2K}, \tag{5.1}$$

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

$$\|G\|_2 \leq \alpha K, \tag{5.2}$$

where $\|G\|_2$ is the spectral norm of the matrix G, and the parameter $\alpha \in [\frac{2}{3}, 1)$ controls the spread of the viewing directions. If the true image orientations are uniformly sampled from the rotation group $\mathbf{SO(3)}$, 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 $\mathrm{Tr}(G) = \mathrm{Tr}(R^T R) = \mathrm{Tr}(RR^T) = \mathrm{Tr}(KI_2) = 2K$. 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 non-trivial 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, we fix the parameter α to some number satisfying $\frac{2}{3} \leq \alpha < 1$, and perhaps even try a few possible values for α and choose the best value by examining the resulting reconstructions.

6. The Alternating Direction Method of Multipliers (ADMM) for SDRs with Spectral Norm Constraint. The application of ADMM to SDP problems was considered in [55]. Here we generalize the application of ADMM to the optimization problems considered in 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 update the Lagrange multipliers. We apply ADMM to the dual problems since the linear constraints (6.2) satisfy $\mathcal{A}\mathcal{A}^* = I$ 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 the 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 after 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 $\alpha K I - G \succcurlyeq 0$, but $\alpha K I - G$ 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, we devise here a version of 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

$$\min_{G \succcurlyeq 0} \quad -\langle C, G\rangle \tag{6.1}$$

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

$$\|G\|_2 \le \alpha K \tag{6.3}$$

where

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

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

G_{ij}^{pq} denotes the (p,q) th element in the 2×2 sub-block 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 $\mathbf{y} = \begin{pmatrix} y_i^1 \\ y_i^2 \\ y_i^3 \end{pmatrix}_{i=1,2,\ldots,K}$, 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 $i = 1, 2, \ldots, K$

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

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

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

By rearranging terms in (6.5), we obtain

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

Using the fact that the dual norm of the spectral norm is the nuclear norm (Proposition 2.1 in [34]), we can obtain from (6.6) the dual problem

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

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

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

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

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

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

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:

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

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

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

and the Lagrange multiplier G is updated by

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

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

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

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

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

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

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

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

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

$$\min_{\mathbf{z}} \frac{\alpha K}{\mu} \|\mathbf{z}\|_1 + \frac{1}{2} \|\mathbf{z} - \boldsymbol{\lambda}\|_2^2, \tag{6.15}$$

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

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

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

$$\min_{X} \left\|X - H^k\right\|_F^2, \text{ s.t. } X \succcurlyeq 0,$$

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

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

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

It follows from the update rule (6.14) that

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

6.2. The relaxed LUD problem. Consider the LUD problem after SDR:

$$\min_{G \succcurlyeq 0} \sum_{i<j} \left\| \vec{c}_{ij}^T - G_{ij} \vec{c}_{ji}^T \right\| \text{ s.t. } \mathcal{A}(G) = \mathbf{b}, \tag{6.16}$$

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

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

The dual problem of problem (6.17) is

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

By rearranging terms in (6.18), we obtain

$$\begin{aligned}\max_{\boldsymbol{\theta}_{ij},\mathbf{y},X\succcurlyeq 0}\min_{\mathbf{x}_{ij},\|G\|_2\leq\alpha K} & -\langle Q\left(\boldsymbol{\theta}\right)+X+\mathcal{A}^{*}\left(\mathbf{y}\right),G\rangle+\mathbf{y}^{T}\mathbf{b}\\ & +\sum_{i<j}\left(\|\mathbf{x}_{ij}\|-\langle\boldsymbol{\theta}_{ij},\mathbf{x}_{ij}\rangle+\left\langle\boldsymbol{\theta}_{ij},\vec{c}_{ij}^{T}\right\rangle\right),\end{aligned} \tag{6.19}$$

where $\boldsymbol{\theta}=\left(\boldsymbol{\theta}_{ij}\right)_{i,j=1,\ldots,K}$, $\boldsymbol{\theta}_{ij}=\left(\theta_{ij}^{1},\theta_{ij}^{2}\right)^{T}$, $\vec{c}_{ij}=\left(c_{ij}^{1},c_{ij}^{2}\right)$,

$$Q\left(\boldsymbol{\theta}\right)=\frac{1}{2}\begin{pmatrix}Q^{11}\left(\boldsymbol{\theta}\right) & Q^{12}\left(\boldsymbol{\theta}\right)\\ Q^{21}\left(\boldsymbol{\theta}\right) & Q^{22}\left(\boldsymbol{\theta}\right)\end{pmatrix}\text{ and }Q^{pq}\left(\boldsymbol{\theta}\right)=\begin{pmatrix}0 & \theta_{12}^{p}c_{21}^{q} & \cdots & \theta_{1K}^{p}c_{K1}^{q}\\ c_{21}^{q}\theta_{12}^{p} & 0 & \cdots & \theta_{2K}^{p}c_{K2}^{q}\\ \vdots & \vdots & \ddots & \vdots\\ c_{K1}^{q}\theta_{1K}^{p} & c_{K2}^{q}\theta_{2K}^{p} & \cdots & 0\end{pmatrix}$$

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

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

In fact, (6.20) is obtained using the inequality

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

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

$$\min_{\boldsymbol{\theta}_{ij},\mathbf{y},X\succcurlyeq 0}-\mathbf{y}^{T}\mathbf{b}-\sum_{i<j}\left\langle\boldsymbol{\theta}_{ij},\vec{c}_{ij}^{T}\right\rangle+\alpha K\|Z\|_{*} \tag{6.22}$$

$$\text{s.t.}\qquad Z=Q\left(\boldsymbol{\theta}\right)+X+\mathcal{A}^{*}\left(\mathbf{y}\right),\qquad\text{and }\|\boldsymbol{\theta}_{ij}\|\leq 1. \tag{6.23}$$

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

$$\begin{aligned}\mathcal{L}\left(\mathbf{y},\boldsymbol{\theta},Z,X,G\right)=&-\mathbf{y}^{T}\mathbf{b}+\alpha K\|Z\|_{*}-\sum_{i<j}\left\langle\boldsymbol{\theta}_{ij},\vec{c}_{ij}^{T}\right\rangle+\langle G,Q\left(\boldsymbol{\theta}\right)+X+\mathcal{A}^{*}\left(\mathbf{y}\right)-Z\rangle\\ &+\frac{\mu}{2}\|Q\left(\boldsymbol{\theta}\right)+X+\mathcal{A}^{*}\left(\mathbf{y}\right)-Z\|_{F}^{2},\end{aligned} \tag{6.24}$$

for $\|\boldsymbol{\theta}_{ij}\|\leq 1$, 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:

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

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

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

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

and the Lagrange multiplier G is updated by

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

where $\gamma \in \left(0, \frac{1+\sqrt{5}}{2}\right)$ is an approprately chosen step length. The methods to solve 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 eqivalent problem

$$\min_{\boldsymbol{\theta}_{ij}} -\left\langle \boldsymbol{\theta}_{ij}, \vec{c}_{ij}^T \right\rangle + \frac{\mu}{2}\left\|\boldsymbol{\theta}_{ij}\vec{c}_{ji} + \Phi_{ij}\right\|_F^2, \text{ s.t. } \|\boldsymbol{\theta}_{ij}\| \le 1,$$

where $\Phi = X^k + \mathcal{A}^*\left(\mathbf{y}^{k+1}\right) - Z^k + \frac{1}{\mu}G^k$, $\Phi = \begin{pmatrix} \Phi^{11} & \Phi^{12} \\ \Phi^{21} & \Phi^{22} \end{pmatrix}$ and $\Phi_{ij} = \begin{pmatrix} \Phi_{ij}^{11} & \Phi_{ij}^{12} \\ \Phi_{ij}^{21} & \Phi_{ij}^{22} \end{pmatrix}$. Problem (6.29) is further simplified as

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

whose solution is

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

The practical issues related to how to take advantage of low-rank assumption of G in the eigenvalue decomposition performed at each iteration, strategies for adjusting the penalty parameter μ, the use of a step size γ for updating the primal variable X and termination rules using the in-feasibility measures are discussed in details in [55]. The convergence analysis on ADMM using more than two blocks of variables can be found in [16]. However, there is one condition of Assumption A (page 5) in [16] that cannot be satisfied for our problem: the condition that the feasible set should be polyhedral, whereas the SDP cone in our problem is not a polyhedral. To generalize the convergence analysis in [16] to our problem, we will need to show that the local error bounds (page 8 - 9 in [16]) hold for the SDP cone. Currently we do not have a rigorous convergence proof for ADMM for our problem.

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 semidefinite relaxation:

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

$$\text{s.t.} \quad G_{ii} = I_2,\ i = 1, 2, \cdots, K, \tag{7.2}$$

$$G \succcurlyeq 0, \tag{7.3}$$

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

where α 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 an variant of the IRLS procedure [5, 6, 19], 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.

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 α satisfies $\frac{2}{3} \leq \alpha < 1$), and then recover the orientations by rounding.

Require: a $2K \times 2K$ common-line matrix S, a regularization parameter ϵ, a parameter α and the total number of iterations N_{iter}

$w_{ij}^0 = 1\ \forall i, j = 1, \cdots, K$;

$G^0 = 0$;

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});

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

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

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

end for

obtain estimated orientations $\hat{R}_1, \ldots, \hat{R}_K$ from $G^{N_{\text{iter}}}$ using the randomized rounding procedure in section 4.4.

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

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

$$\text{s.t.} \quad G_{ii} = I_2,\ i = 1, 2, \cdots, K, \tag{7.6}$$

$$G \succcurlyeq 0, \tag{7.7}$$

$$\|G\|_2 \leq \alpha K \text{ (optional)}. \tag{7.8}$$

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

$\arg\min F(G)$, then we shall verify that

$$|F(G_\epsilon^*) - F(G^*)| \leq 4K^2\epsilon. \tag{7.9}$$

Using the fact that

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

we obtain

$$\begin{aligned}&(F(G^*,\epsilon) - F(G_\epsilon^*,\epsilon)) + (F(G_\epsilon^*) - F(G^*))\\ =&(F(G^*,\epsilon) - F(G^*)) - (F(G_\epsilon^*,\epsilon) - F(G_\epsilon^*)) \leq 4K^2\epsilon.\end{aligned}$$

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

In each iteration, the IRLS procedure solves the problem

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

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

$$w_{ij}^k = 1/\sqrt{2 - 2\left\langle G_{ij}^k, S_{ij}\right\rangle + \epsilon^2},\ \forall k > 0.$$

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 ϵ ensures that no single detected common-line can gain undue influence when solving

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

We repeat the process until the residual sequence $\{r^k\}$ has converged, or the maximum number of iterations has been reached. We shall verify that the value of the cost function is non-increasing, 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 [25] by Mohan et. al..

LEMMA 7.1. *The value of the cost function sequence is monotonically non-increasing, i.e.,*

$$F(G^{k+1},\epsilon) \leq F(G^k,\epsilon). \tag{7.12}$$

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

Proof. Since G^k is the solution of (7.11), there exists $\mathbf{y}^k \in \mathbb{R}^{2K}$ and $X^k \in \mathbb{R}^{2K\times 2K}$ such that

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

$$G^k \succcurlyeq 0,\ X^k \succcurlyeq 0,\ \left\langle G^k, X^k\right\rangle = 0. \tag{7.14}$$

Hence we have

$$
\begin{aligned}
0 &= -(\mathbf{y}^k)^T(\mathcal{A}(G^k)-\mathbf{b}) + (\mathbf{y}^{k+1})^T(\mathcal{A}(G^{k+1})-\mathbf{b}) \\
&= (\mathbf{y}^{k+1}-\mathbf{y}^k)^T(\mathcal{A}(G^k)-\mathbf{b}) + \left\langle \mathcal{A}^*(\mathbf{y}^{k+1}), G^{k+1}-G^k \right\rangle \\
&= \left\langle X^{k+1} + W^k \circ S, G^{k+1}-G^k \right\rangle \\
&\leq \left\langle W^k \circ S, G^{k+1}-G^k \right\rangle \qquad (7.15) \\
&= \frac{1}{2}\sum_{i\neq j} \left(-\beta_{ij}^k \left(2 - 2\left\langle G_{ij}^{k+1}, S_{ij}\right\rangle + \epsilon^2\right) + \beta_{ij}^k \left(2 - 2\left\langle G_{ij}^{k}, S_{ij}\right\rangle + \epsilon^2\right)\right) \\
&= \frac{1}{2}\sum_{i\neq j} \left(-\frac{2 - 2\left\langle G_{ij}^{k+1}, S_{ij}\right\rangle + \epsilon^2}{\sqrt{2 - 2\left\langle G_{ij}^{k}, S_{ij}\right\rangle + \epsilon^2}} + \sqrt{2 - 2\left\langle G_{ij}^{k}, S_{ij}\right\rangle + \epsilon^2}\right), \qquad (7.16)
\end{aligned}
$$

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

$$
\begin{aligned}
F(G^k,\epsilon)^2 &= \left(\sum_{i\neq j} \sqrt{2 - 2\left\langle G_{ij}^{k}, S_{ij}\right\rangle + \epsilon^2}\right)^2 \\
&\geq \left(\sum_{i\neq j} \sqrt{2 - 2\left\langle G_{ij}^{k}, S_{ij}\right\rangle + \epsilon^2}\right)\left(\sum_{i\neq j} \frac{2 - 2\left\langle G_{ij}^{k+1}, S_{ij}\right\rangle + \epsilon^2}{\sqrt{2 - 2\left\langle G_{ij}^{k}, S_{ij}\right\rangle + \epsilon^2}}\right) \\
&\geq \left(\sum_{i\neq j} \sqrt{2 - 2\left\langle G_{ij}^{k+1}, S_{ij}\right\rangle + \epsilon^2}\right)^2 = F(G^{k+1},\epsilon)^2, \qquad (7.17)
\end{aligned}
$$

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

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

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

THEOREM 7.2. *The sequence of iterates $\left\{G^k\right\}$ 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 $\left\{G^k\right\}$ 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 $\mathbf{b}^T\mathbf{y}^{k+1}$ =trace$((W^k \circ S)G^{k+1})$ is bounded. In addition, from the KKT conditions (7.13) - (7.14) we obtain $-\mathcal{A}^*(\mathbf{y}^{k+1}) + W^k \circ S \succcurlyeq 0$. Using the definition of $\mathcal{A}^*$ and S, the property of semi-definite 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

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

which implies that X^k is bounded.

We now show that every cluster point of $\left\{G^k\right\}$ is a stationary point of (7.5) - (7.7). Suppose to the contrary and let $\bar{G}$ be a cluster point of $\left\{G^k\right\}$ that is not a stationary point. By the definition of cluster point, there exists a subsequence

$\left\{G^{k_i}, W^{k_i}, X^{k_i}, \mathbf{y}^{k_i}\right\}$ of $\left\{G^k, W^k, X^k, \mathbf{y}^k\right\}$ converging to $\left(\bar{G}, \bar{W}, \bar{X}, \bar{y}\right)$. By passing to a further subsequence if necessary, we can assume that $\left\{G^{k_i+1}, W^{k_i+1}, X^{k_i+1}, \mathbf{y}^{k_i+1}\right\}$ is also convergent and we denote its limit by $\left(\hat{G}, \hat{W}, \hat{X}, \hat{\mathbf{y}}\right)$. 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

$$-\mathcal{A}^*(\hat{\mathbf{y}}) + \hat{X} + \bar{W} \circ S = 0, \mathcal{A}(\hat{G}) - \mathbf{b} = 0,$$
$$\hat{G} \succcurlyeq 0, \hat{X} \succcurlyeq 0, \left\langle \hat{G}, \hat{X} \right\rangle = 0.$$

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

$$\max_{G \succcurlyeq 0} \left\langle \bar{W} \circ S, G \right\rangle \text{ s.t. } \mathcal{A}(\hat{G}) = \mathbf{b}.$$

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 $\left\langle \bar{W} \circ S, \hat{G} \right\rangle > \left\langle \bar{W} \circ S, \bar{G} \right\rangle$. From this last relation and (7.17) - (7.18) it follows that

$$F(\hat{G}, \epsilon) < F(\bar{G}, \epsilon). \tag{7.19}$$

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

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

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

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

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

where $0 < p < 1$. Convergence analysis of IRLS for different applications with $p < 1$ can be found in [6, 19]. The problem (7.20) is a SDR of the problem

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

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) CPUs X5570, 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. The reconstruction from the images with estimated orientations used the Fourier based 3D reconstruction

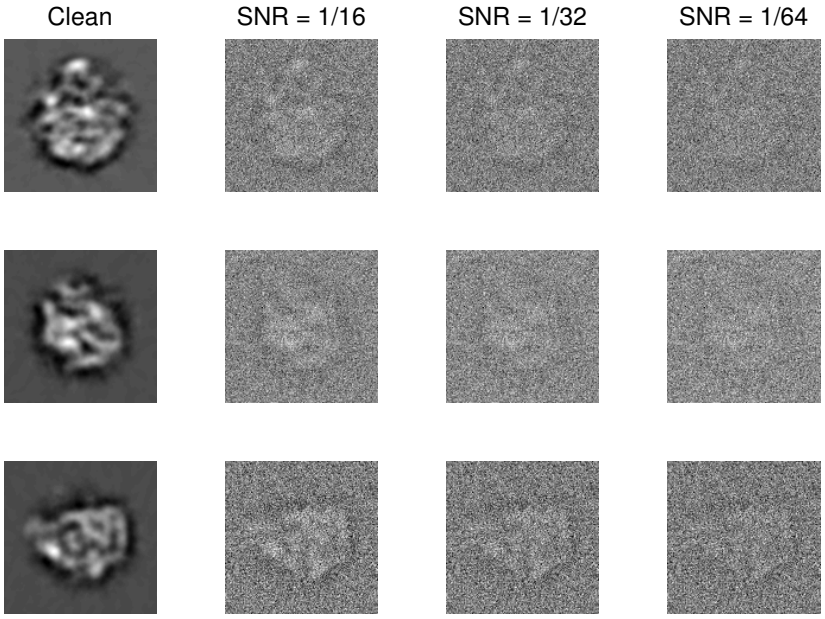


Fig. 8.1: The first column shows three clean images of size 129×129 pixels generated from a 50S ribosomal subunit volume with different orientations. The other three columns show three noisy images corresponding to those in the first column with SNR= 1/16, 1/32 and 1/64, respectively.

package FIRM[3] [52]. The reconstructed volumes are shown in Figure 8.2 and 8.5 using the visualization system Chimera [32].

To evaluate the accuracy or the resolution of the reconstructions, we used the 3D Fourier Shell Correlation (FSC) [36]. FSC measures the normalized cross-correlation coefficient between two 3D volumes over corresponding spherical shells in Fourier space, i.e.,

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

where $\mathcal{F}(\mathbf{V}_1)$ and $\mathcal{F}(\mathbf{V}_2)$ are the Fourier transforms of volume $\mathbf{V}_1$ and volume $\mathbf{V}_2$ 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 $Shell_i := \{\mathbf{j} : 0.5 + (i-1) + \epsilon \leq \|\mathbf{j}\| < 0.5 + i + \epsilon\}$ where ϵ =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 [2, 37] 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 129×129 pixels with pixel size 2.4Å of the 50S ribosomal subunit (the top volume in Figure 8.2), 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 respectively (Figure 8.1).

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

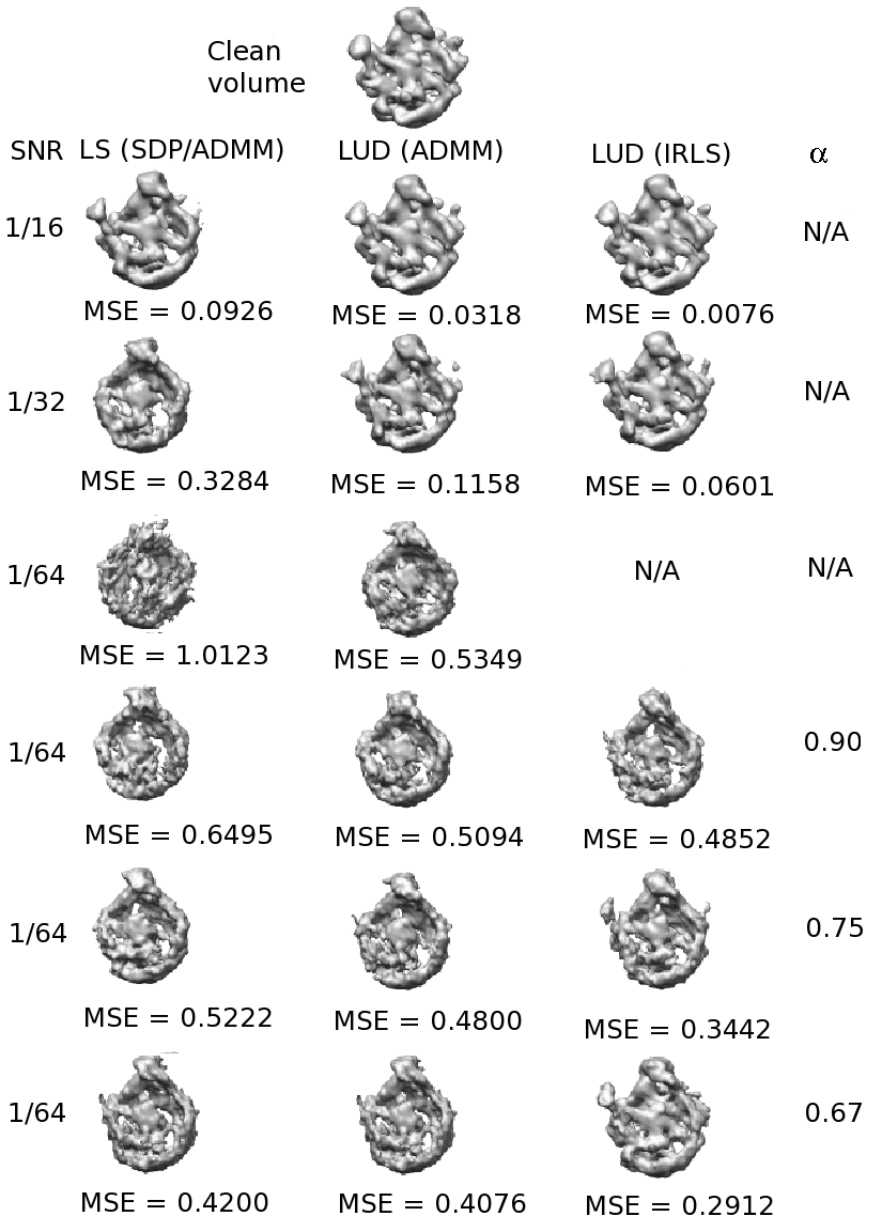


Fig. 8.2: The clean volume (top), the reconstructed volumes and the MSEs (8.2) of the estimated rotations. From 2nd to 4th row, no spectral norm constraint was used (i.e., $\alpha = \mathrm{N/A}$) for all algorithms. The last 4 rows are all results of very noisy images with SNR = 1/64, where the result using the IRLS procedure without α is not available due to the highly clustered estimated projection directions, and the result from the IRLS procedure with $\alpha = 0.67$ for the spectral norm constraint is best.

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 3.1).

To measure the accuracy of the estimated orientations, we defined the mean squared error (MSE) of the estimated rotation matrices $\hat{R}_1, \ldots, \hat{R}_K$ as

$$\text{MSE} = \frac{1}{K} \sum_{i=1}^{K} \left\| R_i - \hat{O}\hat{R}_i \right\|^2, \tag{8.2}$$

where $\hat{O}$ is the optimal solution to the registration problem between the two sets of rotations $\{R_1, \ldots, R_K\}$ and $\left\{\hat{R}_1, \ldots, \hat{R}_K\right\}$ in the sense of minimizing the MSE. As shown in [41], there is a simple procedure to obtain 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 (Figure 8.2). In order to measure the accuracy of the reconstructed volumes, we measured each volume's FSC (8.1) (Figure 8.3) 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%), 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 (Figure 8.2-8.3). 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 small (23%), and most of the detected common-lines were outliers (Figure 3.1), the algorithms without spectral norm constraint $\|G\|_2 \leq \alpha K$ did not work. Especially, the viewing directions of images estimated by the IRLS procedures without $\|G\|_2 \leq \alpha K$ 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., $\alpha = 0.67$ since the true rotations were uniformly sampled over $SO(3)$) gave the best reconstruction.

8.2. Experiments on a real dataset. A set of micro-graphs of E. coli 50S ribosomal subunits was provided by Dr. M. van Heel. These micro-graphs were acquired by a Philips CM20 at defocus values between 1.37 and 2.06 μm, 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 [20]. Then using the IMAGIC software package ([45, 51]), the 27,121 particle images of size 90×90 pixels were phase-flipped to remove the phase-reversals in the 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 2 disjoint groups of equal number of images. The following steps were performed to each group separately.

The images were rotationally aligned and averaged to produce class averages of better quality, following the procedure detailed in [58]. For each group, the images were denoised and compressed using Fourier-Bessel based principal component analysis (FBsPCA) [57]. Then, triple products of Fourier-Bessel expansion coefficients

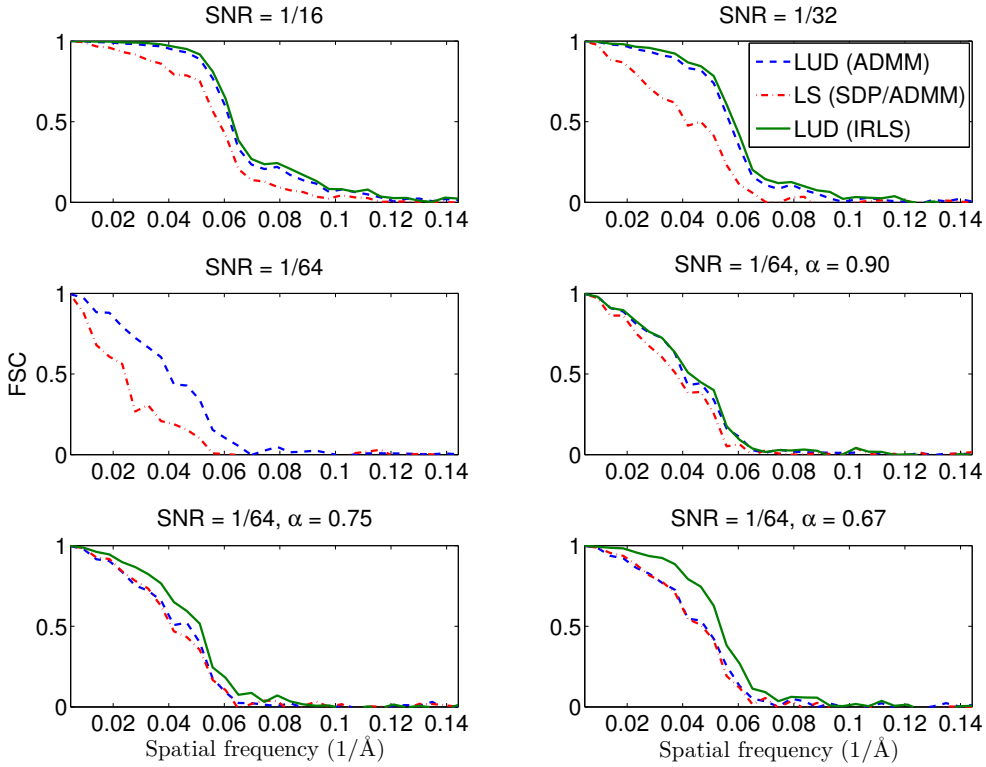


Fig. 8.3: FSCs (8.1) of the reconstructed volumes against the clean volume in Figure 8.2. The plots of the correlations show that the LUD approach using ADMM and IRLS (denoted as the blue dashed lines and green solid lines) outweighed the LS approach using SDP or ADMM (denoted as the red dot-dashed lines). Note that all the last four sub-figures are results for images with SNR = 1/64, where the last three sub-figures are results using different α for the spectral norm constraint. In the third sub-figure (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.

obtained in FBsPCA were used to compute rotational invariant features of the images, i.e, the bispectrums [17, 23, 35]. 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 [42] to produce new sets of neighbors. Finally, for each image, we averaged it with its 10 nearest neighbors after alignment. Three examples of averaged images are shown in Figure 8.4.

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 8.5a). 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

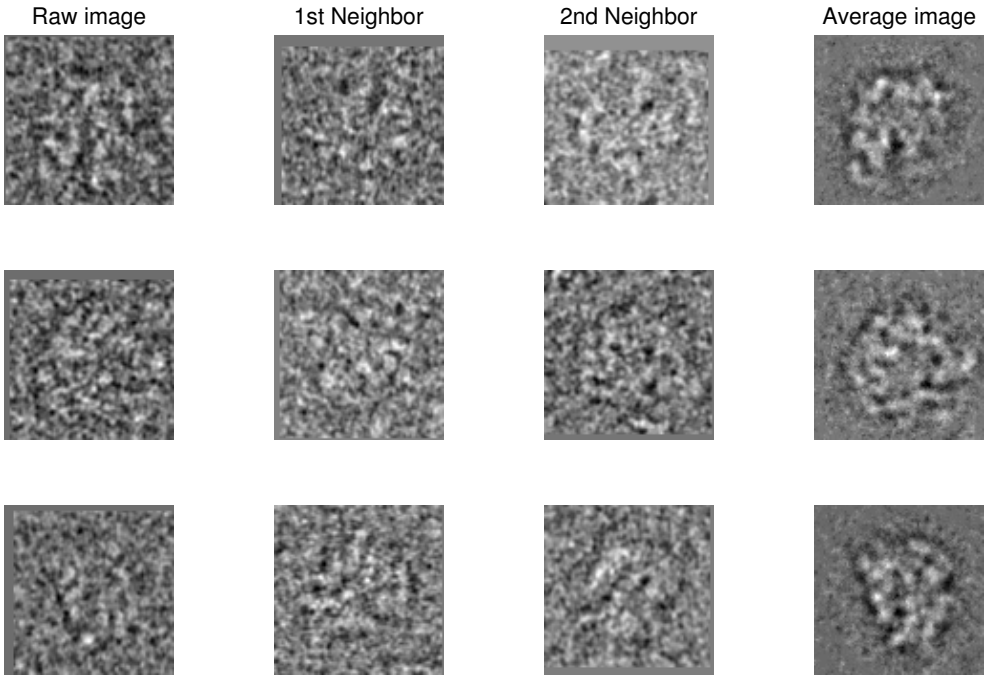


Fig. 8.4: Noise reduction by image averaging. Three raw ribosomal images are shown in the first column. Their closest two neighbours (i.e., raw images having similar orientations after alignments) are shown in the second and third columns. The average images shown in the last column were obtained by averaging over 10 neighbours of each raw image.

was computed to estimate the resolution of the ab-initio model (Figure 8.6). 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 is 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 4 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 8.6a - 8.6e show the convergence of the refinement process using different ab-initio models. We observed that all the refined models are at the resolution 11.1Å. However, the worst ab-initio

model obtained by LS needed 7 iterations (about 28 hours) for convergence (Figure 8.6a), while the best ab-initio model obtained by LUD needed 3 iterations (about 12 hours) for convergence (Figure 8.6b and Figure 8.6d). Figure 8.6f uses FSC plots to compare the refined models. We observed that the refined models in Figure 8.6b - 8.6e were consistent to each other, while the refined model obtained by LS in Figure 8.6a was slightly different from others.

The average cost time for computing the ab-initio models in these two subsections are shown in Table (8.1). 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 is that a large portion of the cost 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 on eigenspace computation [56]. However, when using the LUD approach for the real data set, 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).

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

Table 8.1: The average cost time using different algorithms on 500 and 1000 images in the two experimental subsections. The notation $\alpha = \text{N/A}$ means no spectral norm constraint $\|G\|_2 \leq \alpha K$ is used.

9. Discussion. To estimate image orientations, we introduced a robust self consistency error and used ADMM or the 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 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 data set (Section 8.2) demonstrate that the ab-initio models resulted by the LUD based methods are more accurate than initial models that are resulted by least squares based methods. In particular, our initial models requires 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 in [39].

In [54], 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 image 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 we observed.

10. Acknowledgements. The authors would like to thank Zhizhen Zhao for producing class averages from the experimental ribosomal images. The work of L. Wang and A. Singer was partially supported by Award Number FA9550-12-1-0317 from AFOSR, by Award Number R01GM090200 from the NIGMS, by the Alfred P. Sloan Foundation, and by the Simons Foundation. The work of Z. Wen was partially supported by NSFC grant 11101274.

References.

[1] K. S. Arun, T. S. Huang, and S. D. Blostein. Least-Squares Fitting of Two 3-D Point Sets. *IEEE Trans. Pattern Anal. Mach. Intell.*, 9(5):698–700, May 1987.

[2] X. Bai, I. S. Fernandez, G. McMullan, and S. HW Scheres. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *eLife Sciences*, 2, 2013.

[3] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.

[4] S. Burer and R. D. C. Monteiro. A Nonlinear Programming Algorithm for Solving Semidefinite Programs via Low-rank Factorization. *Mathematical Programming (series B)*, 95:2003, 2001.

[5] Emmanuel J. Candès, Michael B. Wakin, and Stephen P. Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *Journal of Fourier Analysis and Applications*, 14:877–905, 2008.

[6] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk. Iteratively reweighted least squares minimization for sparse recovery. *Communications on Pure and Applied Mathematics*, 63(1):1–38, 2010.

[7] E. de Klerk. *Aspects of Semidefinite Programming: Interior Point Algorithms and Selected Applications*. Applied Optimization. Springer, 2002.

[8] D. L. Donoho and I. M. Johnstone. Adapting to Unknown Smoothness via Wavelet Shrinkage. *Journal of the American Statistical Association*, 90(432):1200+, December 1995.

[9] A. Dutt and V. Rokhlin. Fast Fourier Transforms for Nonequispaced Data. *SIAM Journal on Scientific Computing*, 14(6):1368–1393, 1993.

[10] N. A. Farrow and F. P. Ottensmeyer. A posteriori determination of relative projection directions of arbitrarily oriented macromolecules. *J. Opt. Soc. Am. A*, 9(10):1749–1760, Oct 1992.

[11] J. A. Fessler and B. P. Sutton. Nonuniform fast Fourier transforms using min-max interpolation. *IEEE Transactions on Signal Processing*, 51(2):560 – 574, 2003.

[12] J. Frank. *Three Dimensional Electron Microscopy of Macromolecular Assemblies*. Academic Press, Inc., 1996.

[13] J. Frank. Cryo-electron microscopy as an investigative tool: the ribosome as an example. *BioEssays*, 23(8):725–732, 2001.

[14] M. X. Goemans and D.P. Williamson. Improved Approximation Algorithms

for Maximum Cut and Satisfiability Problems Using Semidefinite Programming. *Journal of the ACM*, 42:1115–1145, 1995.

[15] L. Greengard and J. Lee. Accelerating the Nonuniform Fast Fourier Transform. *SIAM Review*, 46(3):443–454, 2004.

[16] M. Hong and Z.-Q. Luo. On the Linear Convergence of the Alternating Direction Method of Multipliers. *ArXiv e-prints*, August 2012.

[17] R. I. Kondor. A complete set of rotationally and translationally invariant features for images. *CoRR*, abs/cs/0701127, 2007.

[18] L. Lebart, A. Morineau, and K. M. Warwick. *Multivariate descriptive statistical analysis: correspondence analysis and related techniques for large matrices*. Wiley series in probability and mathematical statistics: Applied probability and statistics. Wiley, 1984.

[19] G. Lerman, M. McCoy, J. A. Tropp, and T. Zhang. Robust computation of linear models, or How to find a needle in a haystack. `arXiv:1202.4044v1` [cs.IT], 2012.

[20] S.J. Ludtke, P. R. Baldwin, and W. Chiu. EMAN: Semiautomated Software for High-Resolution Single-Particle Reconstructions. *Journal of Structural Biology*, 128(1):82 – 97, 1999.

[21] Z. Luo, W. Ma, A. So, Y. Ye, and S. Zhang. Semidefinite Relaxation of Quadratic Optimization Problems. *IEEE Signal Processing Magazine*, 27(3):20–34, may 2010.

[22] S.P. Mallick, S. Agarwal, D.J. Kriegman, S.J. Belongie, B. Carragher, and C.S. Potter. Structure and view estimation for tomographic reconstruction: A bayesian approach. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 2253–2260, 2006.

[23] R. Marabini and J. M. Carazo. On a new computationally fast image invariant based on bispectral projections. *Pattern Recogn. Lett.*, 17(9):959–967, 1996.

[24] F. Mezzadri. How to generate random matrices from the classical compact groups. *Notices of the AMS*, 54:592–604, 2007.

[25] K. Mohan and M. Fazel. Iterative reweighted algorithms for matrix rank minimization. *Journal of Machine Learning Research*, 13:3441–3473, 2012.

[26] F. Natterer. *The Mathematics of Computerized Tomography*. Classics in Appl. Math. 32. SIAM, Philadelphia, 2001.

[27] H. Nyquist. Least orthogonal absolute deviations. *Computational Statistics & Data Analysis*, 6(4):361–367, June 1988.

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

[29] P. Penczek, M. Radermacher, and J. Frank. Three-dimensional reconstruction of single particles embedded in ice. *Ultramicroscopy*, 40(1):33–53, 1992.

[30] P. A. Penczek, R. A. Grassucci, and J. Frank. The ribosome at improved resolution: New techniques for merging and orientation refinement in 3d cryo-electron microscopy of biological particles. *Ultramicroscopy*, 53(3):251 – 270, 1994.

[31] P. A. Penczek, J. Zhu, and J. Frank. A common-lines based method for determining orientations for N $>$ 3 particle projections simultaneously. *Ultramicroscopy*, 63(3-4):205 – 218, 1996.

[32] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25:1605–1612, 2004.

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

[34] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization. *SIAM Rev.*, 52(3):471–501, August 2010.

[35] B. M. Sadler and G. B. Giannakis. Shift- and rotation-invariant object reconstruction using the bispectrum. *J. Opt. Soc. Am. A*, 9(1):57–69, Jan 1992.

[36] W. O. Saxton and W. Baumeister. The correlation averaging of a regularly arranged bacterial cell envelope protein. *Journal of Microscopy*, 127(2):127–138, 1982.

[37] S. HW Scheres and S. Chen. Prevention of overfitting in cryo-EM structure determination. *Nat Meth*, 9:853–854, 2012.

[38] I. I. Serysheva, E. V. Orlova, W. Chiu, M. B. Sherman, S. L. Hamilton, and M. van Heel. Electron cryomicroscopy and angular reconstitution used to visualize the skeletal muscle calcium release channel. *Nat Struct Mol Biol*, 2:18–24, 1995.

[39] Y. Shkolnisky and A. Singer. Viewing direction estimation in cryo-EM using synchronization. *SIAM Journal on Imaging Sciences*, 5(3):1088–1110, 2012.

[40] A. Singer, R. R. Coifman, F. J. Sigworth, D. W. Chester, and Y. Shkolnisky. Detecting consistent common lines in cryo-EM by voting. *Journal of Structural Biology*, 169(3):312–322, 2010.

[41] A. Singer and Y. Shkolnisky. Three-Dimensional Structure Determination from Common Lines in Cryo-EM by Eigenvectors and Semidefinite Programming. *SIAM Journal on Imaging Sciences*, 4(2):543–572, 2011.

[42] A. Singer, Z. Zhao, Y. Shkolnisky, and R. Hadani. Viewing Angle Classification of Cryo-Electron Microscopy Images Using Eigenvectors. *SIAM Journal on Imaging Sciences*, 4(2):723–759, 2011.

[43] A. So, J. Zhang, and Y. Ye. On approximating complex quadratic optimization problems via semidefinite programming relaxations. *Math. Program.*, 110(1):93–110, March 2007.

[44] H. Späth and G. A. Watson. On orthogonal linear approximation. *Numer. Math.*, 51(5):531–543, October 1987.

[45] H. Stark, M. V. Rodnina, H. Wieden, F. Zemlin, W. Wintermeyer, and M. van Heel. Ribosome interactions of aminoacyl-tRNA and elongation factor Tu in the codon-recognition complex. *Nat Struct Mol Biol*, 9:849–854, 2002.

[46] B. Vainshtein and A. Goncharov. Determination of the spatial orientation of arbitrarily arranged identical particles of an unknown structure from their projections. In *Proc. llth Intern. Congr. on Elec. Mirco.*, pages 459–460, 1986.

[47] M. van Heel. Multivariate statistical classification of noisy images (randomly oriented biological macromolecules). *Ultramicroscopy*, 13(1-2):165 – 183, 1984.

[48] M. van Heel. Angular reconstitution: A posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy*, 21(2):111 – 123, 1987.

[49] M. van Heel and J. Frank. Use of multivariates statistics in analysing the images of biological macromolecules. *Ultramicroscopy*, 6(1):187 – 194, 1981.

[50] M. van Heel, B. Gowen, R. Matadeen, E. V. Orlova, R. Finn, T. Pape, D. Cohen, H. Stark, R. Schmidt, M. Schatz, and A. Patwardhan. Single-particle electron cryo-microscopy: towards atomic resolution. *Quarterly Reviews of Biophysics*, 33(04):307–369, 2000.

[51] M. van Heel, G. Harauz, E. V. Orlova, R. Schmidt, and M. Schatz. A new generation of the imagic image processing system. *Journal of Structural Biology*, 116(1):17 – 24, 1996.

[52] C. Vonesch, Lanhui Wang, Y. Shkolnisky, and A. Singer. Fast wavelet-based single-particle reconstruction in Cryo-EM. In *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, pages 1950 –1953, 2011.

[53] L. Wang and F. J. Sigworth. Cryo-EM and single particles. *Physiology (Bethesda)*, 21:13–18, 2006.

[54] L. Wang and A. Singer. Exact and stable recovery of rotations for robust synchronization, 2012. submitted. Also availabe at `http://arxiv.org/abs/1211.2441`.

[55] Z. Wen, D. Goldfarb, and W. Yin. Alternating direction augmented Lagrangian methods for semidefinite programming. *Mathematical Programming Computation*, 2:203–230, 2010.

[56] Z. Wen, C. Yang, X. Liu, and Y. Zhang. Trace-Penalty Minimization for Large-scale Eigenspace Computation. *Optimization Online*, 2013.

[57] Z. Zhao and A. Singer. Fourier-Bessel rotational invariant eigenimages, 2012. Submitted. Also available at `http://arxiv.org/abs/1211.1968`.

[58] Z. Zhao and A. Singer. Rotationally Invariant Image Representation for Viewing Angle Classification, 2013. In preparation.

[59] Z. Zhu, A. So, and Y. Ye. Universal Rigidity and Edge Sparsification for Sensor-Network Localization. *SIAM Journal on Optimization*, 20(6):3059–3081, 2010.

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 5.1), then the Gram matrix G obtained by solving the LS problem (4.6)-(4.8) or the LUD problem (4.10) is uniquely the one 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 (page 36-39 in [7], [59]). 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)):

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

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 $\left\{\mathbf{u}_1^1, \mathbf{u}_1^2, \mathbf{u}_2^1, \mathbf{u}_2^2, \mathbf{u}_3^1, \mathbf{u}_3^2\right\}$ span a space with dimensionality at most 3.

Define arrays $\mathbf{u}_i$ as

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

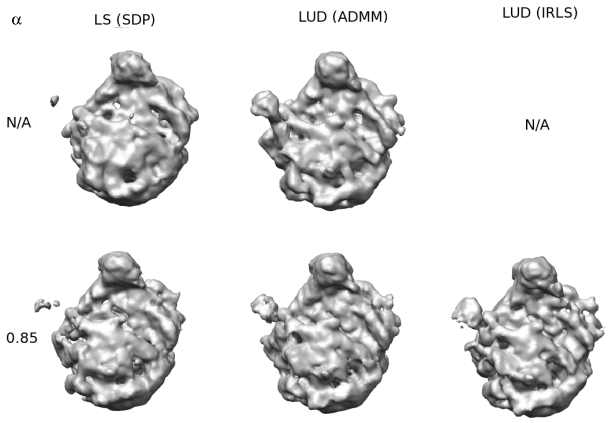
then the inner product

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

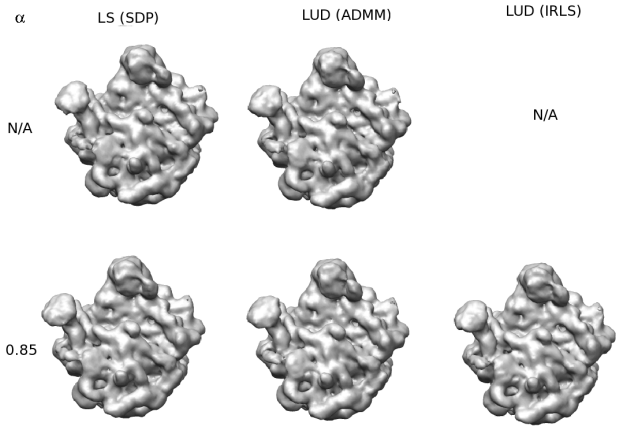
where the last inequality follows the Cauchy-Schwarz inequality and the facts that all $\vec{c}_{ij}$ are unit vectors, $\mathbf{u}_i^1$ and $\mathbf{u}_i^2$ are unit vectors and orthogonal to each other due to the constraint $G_{ii} = I_2$, and thus all $c_{ij}^1 \mathbf{u}_j^1 + c_{ij}^2 \mathbf{u}_j^2$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

$$c_{ji}^1 \mathbf{u}_i^1 + c_{ji}^2 \mathbf{u}_i^2 = c_{ij}^1 \mathbf{u}_j^1 + c_{ij}^2 \mathbf{u}_j^2. \tag{A.1}$$

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 $\dim(\text{span}\{\mathbf{u}_1^1, \mathbf{u}_2^1, \mathbf{u}_3^1, \mathbf{u}_3^2\}) \leq 3$. Using (A.1), assume that $\text{span}\{\mathbf{u}_1^1, \mathbf{u}_1^2\} \cap \text{span}\{\mathbf{u}_3^1, \mathbf{u}_3^2\} = \text{span}\{\mathbf{v}_1\}$ and $\text{span}\{\mathbf{u}_2^1, \mathbf{u}_2^2\} \cap \text{span}\{\mathbf{u}_3^1, \mathbf{u}_3^2\} = \text{span}\{\mathbf{v}_2\}$, where $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent vectors (otherwise all three projection directions are linearly dependent and thus the 3 Fourier slices of the images intersect at the same line). Therefore we have $\text{span}\{\mathbf{v}_1, \mathbf{v}_2\} = \text{span}\{\mathbf{u}_3^1, \mathbf{u}_3^2\}$, $\text{span}\{\mathbf{v}_1, \mathbf{u}_1^1\} \subseteq \text{span}\{\mathbf{u}_1^1, \mathbf{u}_1^2\}$ and $\text{span}\{\mathbf{v}_2, \mathbf{u}_2^1\} \subseteq \text{span}\{\mathbf{u}_2^1, \mathbf{u}_2^2\}$. Thus $\dim(\text{span}\{\mathbf{u}_1^1, \mathbf{u}_2^1, \mathbf{u}_3^1, \mathbf{u}_3^2\}) = \dim(\text{span}\{\mathbf{u}_1^1, \mathbf{u}_2^1, \mathbf{v}_1, \mathbf{v}_2\}) \leq \dim(\text{span}\{\mathbf{u}_1^1, \mathbf{u}_1^2, \mathbf{u}_2^1, \mathbf{u}_2^2\}) \leq 3$.

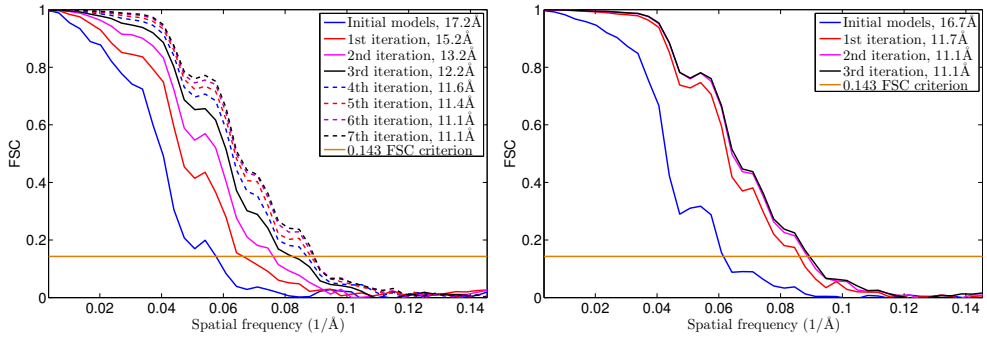


(a) Initial models.



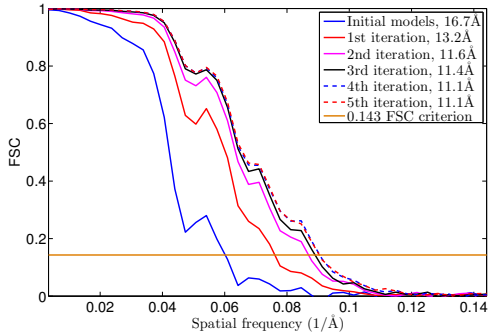
(b) Refined models.

Fig. 8.5: 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 8.6). The model using the IRLS procedure without the spectral norm constraint (i.e., $\alpha = \mathrm{N/A}$) 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.

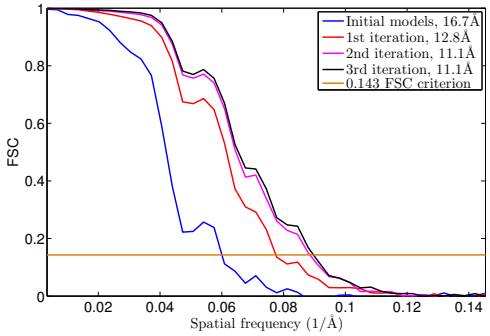


(a) LS, SDP

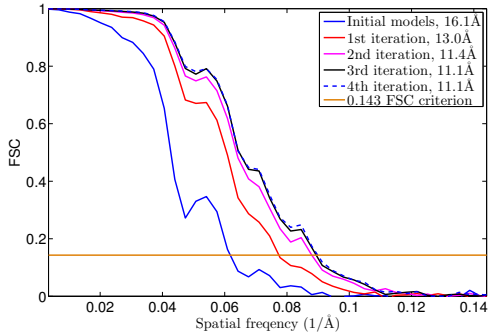
(b) LUD, ADMM



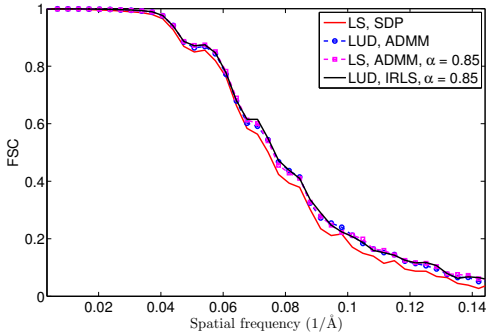
(c) LS, ADMM, $\alpha = 0.85$



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



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



(f) Comparison of refined models.

Fig. 8.6: Convergence of the refinement process. In sub-figure (a) - (e), the FSC plots show the convergence of the refinement iterations. The ab-initio models (Figure 8.5a 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. The sub-figure (f) are 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 8.5b.