ITERATIVE METHODS IN LARGE FIELD ELECTRON MICROSCOPE TOMOGRAPHY*

XIAOHUA WAN†, SÉBASTIEN PHAN‡, ALBERT LAWRENCE‡, FA ZHANG†,
RENMIN HAN§, ZHIYONG LIU¶, AND MARK ELLISMAN‡

Abstract. Electron tomography (ET) is a powerful technology allowing the three-dimensional (3D) imaging of cellular ultrastructure. These structures are reconstructed from a set of micrographs taken at different sample orientations, the final volume being the solution of a general inverse problem. Two different approaches are used in this context: iterative methods and filtered backprojection. Iterative methods are known to provide high-resolution 3D reconstructions for ET under noisy and incomplete data conditions. However, all previous implementations have been restricted to the straight-line projection models. This is not accurate since electron trajectories in electron microscopes do not follow the straight-line optics assumed for X-rays, and biological samples may warp as a result of being exposed to an electron beam. Compensation for curvilinear trajectories, nonlinear electron optics, and sample warping constitutes a major advance in large-field ET and has made possible resolution down to the molecular level in reconstructions of whole cells. At present these advances are limited to filtered backprojection and have been implemented in the software package TxBR. As the next step in this development, we have modified the ASART method in conjunction with a 3D model. By employing alignment based on general curvilinear trajectories we have been able to show that further improvements can be achieved with iterative methods. We also discuss a unified treatment of the alignment and reconstruction problems within the framework of iterative methods, and the relationship between formulas employed in the update step and cross-validation methods.

Key words. electron tomography, three-dimensional reconstruction, iterative methods, nonlinear projection model, ASART, TxBR

AMS subject classifications. 65F10, 15A29, 53C65

DOI. 10.1137/120881464

1. Background. Electron tomography (ET) enables the detailed imaging of complex biological assemblies within the context of a cell [1]. In combination with other techniques, it can provide three-dimensional (3D) reconstructions of protein assemblies, correlate 3D structures with functional investigations at the light microscope level, and provide structural information which extends the findings of genomics and molecular biology. Three-dimensional reconstructions of those structures are typically generated from a set of projection images acquired at different sample orientations, for example, by tilting the specimen around one [2] or more axes [3]. Large-field

high-resolution ET allows one to visualize and understand global structures, such as the cell nucleus, extended neural processes, or even whole cells on scales approaching molecular resolution.

Realistic physical details are essential for the task of modeling cellular processes over many spatial scales. While the electron microscope resolution can extend to a fraction of a nanometer, a typical 3D reconstruction may just cover $1/10^{15}$ of the volume of an optical microscope reconstruction. In order to bridge the gap between those two approaches, the available spatial range of an ET reconstruction has been expanded by various techniques. Large sensor arrays and wide-field camera assemblies have increased the field dimensions by a factor of ten over the past decade, and new techniques for serial tomography and montaging make possible the assembly of many 3D reconstructions. The number of tomographic volumes necessary to cover an average cell down to the protein assembly level is of the order 10^4 or more, and given the imaging and algorithm requirements, the computational problem associated with bridging the gap is well in the exascale range. Single-step tomographic reconstruction can be made parallel to a very high degree, and the associated algorithms can be mapped to the simplified processors comprising, for example, a graphics processor unit (GPU). In particular, single-step reconstruction algorithms employing backprojection as the key step are computationally simple, generally parallelizable, and can be programmed on a GPU board with a consequent large speedup [4, 5]. Although there are no intrinsic advantages in the use of current GPU architectures, these are very cost effective in the implementation of such reconstructions. This is an important consideration because we expect that many more orders of magnitude improvement in computational capabilities will still be required in the coming decade. These wide-field reconstructions are now feasible because, on one hand, developments in camera technologies make very large detectors available (for instance, containing of $8k \times 8k$ pixel sensors or larger [6]), and on the other hand, reconstruction techniques such as montaging and serial sectioning have been improved [7]. We discuss a particular class of algorithms for parallel reconstruction below.

Three major obstacles hamper obtaining high-quality reconstruction results from large-field ET. First, by taking a larger field of view, distortions in the reconstructions become more pronounced and require special attention. Major distortions are also the consequence of the intrinsic nature of electron microscope optics: electron trajectories are helical since an electron beam is focused by means of magnetic fields [8, 9]. This has been noted many times in the literature, but ignored in most cases, because helicity has little effect on the typical small images taken before the rise of interest in wide-field tomography. A second class of distortions originates from the sample warping that results from sample mass loss during the data acquisition step as electron collisions degrade the specimen; this is especially true when the number of micrographs taken is large. Distortions also result from spherical and other optical aberrations, which are the unavoidable consequence of electron optics. Second, owing to technical and material limitations, a single-tilt series in ET always has a relatively small number of views and an incomplete span angle. This generally leads to undesirable artifacts in the final reconstruction, such as hourglass patterns and filtration artifacts in the plane containing the optical axis and orthogonal to the tilt axis, and poor resolution in the plane containing the optical and tilt axes. Additionally electron micrographs can be quite noisy, especially in low dose conditions, when preserving the sample integrity is of importance [10]. Iterative methods avoid the problem of filtration and are better suited to suppression of discretization artifacts and noise. Finally, the need for high

resolution requires, in the case of ET, larger projection images, which yields larger reconstructed files and requires more intensive use of computational resources and considerable processing time [5].

The three main problems we address in this study are (i) incorporating a projection model in an iterative method, which makes it consistent with the classical ray optics of imaging in the electron microscope, (ii) adopting our method to multiple-tilt reconstructions in order to reduce the effect of a missing wedge as well as taking advantage of averaging phenomena, and (iii) improving the parallelism by using various parallelization schemes in tandem. The latter will achieve quicker turnaround for the user. Although improving the computational efficiency of iterative methods in order to make them more competitive with single-step methods for large data sets is very desirable, research in this area is still in progress.

1.1. Imaging in the electron microscope. Because electrons move in curvilinear trajectories under observational conditions in the electron microscope [9], we are led to consider the general situation of integrals along curved paths. The problem is to reconstruct an object with a 3D density distribution from image intensities, where the density is related to a local scattering cross-section. The image intensity at each point of each image represents the exponential of a line integral along a specific trajectory through the object.

Transforms by means of line integrals along curvilinear trajectories have been studied as generalized X-ray transforms [11]. The generalized ray transform is defined for a family of curves in $\mathbb{R}^3$, or electron trajectories through the object (for example, a family of helices):

$$(1.1) \qquad \Gamma = \{\gamma_{\mathbf{x},\omega}(t) | t_0 \leq t \leq t_1\},$$

where $\mathbf{x} = (x_1, x_2)$ denotes a point in the image plane and ω corresponds to a physical rotation of the object. The points t_0 and t_1 correspond to the entry and exit points of the trajectory through the object, respectively. For simplicity we consider a slab containing the object, with zero density outside the object. Direct determination of the effective trajectory may be rather complicated because a typical trajectory through a slab containing the object is of the form [9]

$$(1.2) \quad \gamma(t) = a\,(1 - bt)\,(X_0 \cos{(ct)} + Y_0 \sin{(ct)}, Y_0 \cos{(ct)} - X_0 \sin{(ct)}, Z_0 + t)\,,$$

where (X_0, Y_0, Z_0) represents the entry point of the trajectory in the object, and the unknown quantities a, b, c may vary over data acquisition runs. Note that this equation accounts for the apparent spiral inward, as discussed in [9] and other publications. This picture is complicated by sample warping as the data acquisition proceeds, and by lens distortions in the optics, which require a 3D warping and a 2D warping function, respectively, in order to determine the effective trajectories. Difficulties in writing the full equations for the trajectories tend to disappear if we shift to the projection picture, where the effective trajectory is determined to sufficient accuracy by the zeros of two algebraic functions. Procedures for determining the effective trajectories and associated projection maps have been discussed in previous publications [12, 13] and are addressed below.

Returning to the physics associated with the forward problem, for a given electron path the image intensity at rotation ω is given by

$$(1.3) \qquad I = I_0 e^{-\int_{t_0}^{t_1} u[\gamma_{\mathbf{x},\omega}(t)]dt},$$

where $I(\mathbf{x})$ represents the electrons impinging the image plane, and I_0 represents the initial beam intensity. In the experimental setting the values for ω and (x_1, x_2) are discrete, but in the following we discuss the continuous case for simplicity.

1.2. Generalized radon transform. By taking the log of the image intensity, we obtain the generalized ray transform [11] as follows:

$$(1.4) \qquad R_\Gamma u(\mathbf{x}, \omega) \equiv v(\mathbf{x}, \omega) = \int_{t_0}^{t_1} u[\gamma_{\mathbf{x}, \omega}(t)] dt,$$

where the integral is taken along the curve $\gamma_{\mathbf{x}, \omega}$, and u and v are the object density and image intensity, respectively. This formalism is in agreement with the variable beam model situation.

Determination of the curves $\gamma_{\mathbf{x}, \omega}(t)$ is equivalent to the image alignment problem in the projection picture [12]. This problem is intrinsically three-dimensional. This is because inverting the transform R_Γ via reconstructions requires the construction of projection maps which "invert" the trajectories through the sample. In particular, we require a set of projection maps P_ω so that

$$(1.5) \qquad P_\omega(\gamma_{\mathbf{x}, \omega}^1(t), \gamma_{\mathbf{x}, \omega}^2(t), \gamma_{\mathbf{x}, \omega}^3(t)) = (x_1, x_2).$$

In practice, the departures from orthogonal projections are small, and the P_ω can be represented to sufficient accuracy as a pair of polynomial maps of small degree. A precise knowledge of the projection maps between points X in the volume and micrograph pixels $\mathbf{x}$ is critical to achieving high quality reconstructions.

1.3. Reconstruction. The ET problem consists in providing a solution to the inverse problem of (1.4), that is, finding operator R_Γ^{-1}:

$$(1.6) \qquad u(X) = R_\Gamma^{-1} v(\mathbf{x}),$$

where $X = (X_1, X_2, X_3)$ is a point in the volume. A general treatment of the reconstruction problem would pose the forward ray transform and its dual as Fourier integral operators and the inversion as the problem of inverting these integral operators. Although exact inversion formulas are known in specific cases, in the most general case inversion is given with an error term [14]. For example, in the case of curvilinear Radon transforms in dimensions $n - 1$, one may define Fourier integral operators $F = R_\Gamma^* K R_\Gamma$, where K is a 1D convolution operator. By proper choice of K and R_Γ^* we have $F = I + T$, where I is the identity, and T is an operator with more desirable properties, and which is small if the Radon transform is a small perturbation of the classical Radon transform. More generally K may be a classical pseudodifferential operator. The details of this approach vary according to the assumptions we make on the specific forward transform and measures that define the transform and its adjoint [15], so this should be regarded as describing the theory only in a formal sense. In the original work [14], this analysis was confined to strict Radon transforms, but more recently this approach has been applied to the case of a ray transform in 3D space which models electron trajectories through a sample in an electron microscope [16]. In this paper, the authors propose a filtered backprojection (FBP) algorithm $\mathcal{L}u = R_\Gamma^* \mathcal{D} R_\Gamma u$, where $\mathcal{D}$ is a differential operator acting on the data $R_\Gamma u$. Generally, in the case of curvilinear trajectories, singularities may be added to the reconstruction (or masked). In [16] the operator D is chosen to de-emphasize some of the singularities which may be added to the reconstruction. More details on the mathematical

foundations of reconstruction from curvilinear trajectories, the necessary microlocal analysis, and the problem of singularities in FBP may be found in [17, 18, 16]. In any case, filtered backprojection may be used as a starting point in iterative methods, and good control of reconstruction artifacts is useful for speeding convergence.

For computer-based implementations, the reconstruction problem is discretized, and considered as a finite-dimensional linear problem. Specification of the linear map between object and images requires the solution of the alignment problem, i.e., the determination of the coefficients of the polynomials P_ω. The alignment problem, i.e., the determination of the coefficients of the polynomials P_ω, is carried along with the discretization. We discuss this problem below.

1.3.1. Filtered backprojection. FBP is a popular method in the field of 3D reconstruction of ET, due to its algorithmic simplicity and computational efficiency. The first step of the reconstruction is to apply a filter on the aligned projection images. The density of an object point is then evaluated from its corresponding values in the filtered images. This correspondence, as described in (1.5), is established through a projection determined by the electron paths. We can express the operator R_Γ^{-1} as the composition of the backprojection with a generalized filtration:

$$(1.7) \qquad\qquad R_\Gamma^{-1} = R_\Gamma^* \circ \Psi^{*-1},$$

where R_Γ^* is the backprojection operator

$$(1.8) \qquad\qquad R_\Gamma^* u(X) = \int_{P_\omega(X)=\mathbf{x}} v(\mathbf{x}, \omega)\, \mathrm{d}\omega.$$

In the case where the projection maps are rotations around a fixed axis followed by projection onto the first two coordinates, the operator Ψ^{*-1} is often approximated by the standard R-weighted filter associated with the ray transform in two dimensions [19, pp. 81–99].

TxBR adopts an extension of the standard orthogonal backprojection process [7, 12]. As mentioned above, the electron trajectories determine projection maps, where the points along a given trajectory are all mapped into the same point in the image. These projection maps can be approximated by polynomial functions. The density values of the points in the filtered images are pulled back to the point in the object and averaged into the density value there. Formally

$$(1.9) \qquad\qquad u(X, Y, Z) = \frac{1}{M} \sum_{\omega=1}^{M} v_\omega(P_\omega(X, Y, Z)),$$

where u is the reconstructed value at (X, Y, Z), M is the number of electron microscope (EM) projections, and v_ω is the aligned filtered EM view corresponding to P_ω. The backprojection process itself is a smoothing operation, that is, an implicit low-pass filtering that makes reconstructed volumes strongly blurred. A high-pass filter is applied to the projection images in order to compensate for the point spread function of the backprojection process. The major disadvantage of FBP, however, is that the results may be strongly affected by the conditions found in ET, such as the limited tilt angle, low signal-to-noise ratio (SNR), and low contrast [20]. An additional problem is that with general polynomial projections, the point spread function is three-dimensional itself, and also a function of position, so removing it becomes a six-dimensional problem. Nevertheless in most applications removing the point

spread function by a 1D filter gives acceptable results [12]; thus, FBP is computationally efficient, which has made it the method of choice for large-scale reconstructions. Nevertheless, artifacts due to discretization and limited angle range remain and thus may become a problem for reconstructions from high-resolution images.

Evaluating polynomials like $P_\omega(X, Y, Z)$ at each point of a large regular grid can be computationally very costly. However, a significant speed improvement is achieved with the use of a simple recursive scheme [4]. Polynomial functions can be computed using only additions when moving from neighbor to neighbor along the X direction. The variables Y and Z are held constant. The problem is here reduced to a single variable recursion problem. In the case of a polynomial function P of order n (of a single variable), knowledge of the n first finite differences at one node P_m^i is sufficient to be able to calculate their values on the next node $i + 1$: $P_m^{i+1} = P_m^i + P_{m+1}^{i+1}$. The table of values is built from the bottom. Finite differences of order n (and higher) are a constant (or zero) over the entire grid, making this scheme possible. Evaluating P_m^{i+1} from $m = n-1$ to 0 yields to the polynomial evaluation on node $n + 1$. This scheme may also be used in iterative methods where the forward function is polynomial. In effect, the algorithm for specifying the voxel for the backprojection is reduced to a set of additions, the number of which is linearly dependent on the degree of the polynomial. Because the forward projection can be made independently along each X-line in the object, this particular algorithm yields a high degree of parallelism [4], which can also be exploited in cost-effective GPU implementations of both FBP and portions of the update step of iterative algorithms.

1.3.2. Alignment. In order to perform the backprojection, the images must be aligned. This, in essence, is a 3D problem, as the projection parameters are necessary for backprojection. Alignment is performed via a generalized bundle adjustment, in which the coordinates of identified markers (the XYZ model) are computed simultaneously with the coefficients of the polynomials describing the projections (the projection model). The collective coordinates of the particles and projections are described as a combined model in [12]. The number of unknown parameters in this combined model is the sum of the number of 3D coordinates necessary to specify the positions of the markers plus the number of coefficients needed to specify a polynomial projection map for each tilt. Because the number of knowns is twice the number of tilt observations, this model can be made overspecified by the tracking of a sufficient number of markers in the images. We let

$$(1.10) \qquad \boldsymbol{T}_{\omega\varrho} = \left\{ (x_{\omega\varrho,1}, x_{\omega\varrho,2}) \, | \omega \epsilon \left\{ \omega_1, \omega_2, \ldots, \omega_N \right\}, \, P_\omega \mathbf{X}_\varrho = \mathbf{x}_{\omega\varrho} \right\},$$

where $\boldsymbol{T}_{\omega\varrho}$ denotes the track of the features through the image series $\mathbf{X}_\varrho$. The features are labeled by ϱ, and $\mathbf{x}_{\omega\varrho}$ denotes the image of the corresponding feature in track ω. The sum of the squared errors of the reprojections gives us the required error term for the optimization:

$$(1.11) \qquad E = \sum_{\omega, \varrho} \left\| P_\omega(\boldsymbol{X}_\varrho) - \boldsymbol{x}_{\omega\varrho} \right\|^2.$$

Calculation of the unknowns is accomplished by a conjugate gradient optimization. For an $8k$-by-$8k$ image, about 100 markers suffice for good statistical calculation of cubic projection maps. A good determination via conjugate gradient techniques takes only a few minutes.

Generally $\{\mathbf{X}_\varrho\}$ is a set of point features, for example, small gold beads deposited onto the surface of the object. Determination of the positions of the point features or

markers in the image may be done manually or semiautomatically before the bundle adjustment procedure. Initial estimates of the 3D positions of the markers in the object can be given through triangulation, as the object rotations are already known approximately. Note that either the XYZ model can be altered by warping the object or the projection maps may be composed with a warp to give a new projection model. This gives rise to a phenomenon similar to a gauge ambiguity [21, 22, 23], which has been noted in other contexts [12], and can be exploited to "flatten" serial sections of a biological sample in order to stack them in a larger reconstruction. Methods to eliminate this ambiguity have been discussed by Phan et al. [7].

Although the trajectories can be straightened by warping the sample space, and the projections made orthogonal, in many cases, such as the helical trajectories above, the warping map and the rotations do not commute, and thus the warps are different for each view [24]. This complicates the reconstruction problem [18] as well as the stitching together of multiple adjacent reconstructions in order to cover larger volumes [7]. Other mathematical issues can be gleaned from the literature. The problems of determining in a precise manner the numbers of objects and their surface areas and volumes in a given reconstruction within a typical region, i.e., the stereological problem, appear in the interpretation of tomographic reconstructions [25, pp. 1–18], [26]. In particular we would like the warps to be volume preserving. The beam convergence associated with the factor $a\,(1 - bt)$ (1.2) complicates the stereological problem, as might sample mass loss.

1.3.3. Iterative methods. Iterative methods constitute one of the main alternatives to FBP in 3D reconstruction of ET, owing to their good performance in handing incomplete, noisy data. Because the general inversion problem for curvilinear trajectories is unsolved, in the sense of having specific inversion formulas [18], iterative methods can supply a useful alternative to FBP. In general, iterative methods are real-space reconstruction algorithms that formulate the 3D reconstruction problem as a large system of linear equations. Thus, algebraic reconstruction techniques constitute a general approach to inverting linear systems discretizing the geometric optics models of the image formation process. Iterative methods also avoid the filtering problem.

In the following, we adopt a discrete representation of the space; volume reconstructions are made of voxels, with u_j representing the unknown density value and N the total number of voxels. Note that we are dropping the notation for ω above because we have written the image series as a single vector. Similarly we drop the XYZ notation for points in the object. The projection constraints are then expressed with the following linear system:

$$(1.12) \qquad v_i \approx \sum_{j=1}^{N} w_{ij} u_j, \quad 1 \le i \le M,$$

where v_i denotes the ith measured projection value and $M = B \times S$ is the dimension of v, B being the number of projection angles and S the number of projection values per view.

The matrix $W = (w_{ij})_{M,N}$ is a sparse matrix where w_{ij} can be computed according to the contribution of the jth point in the volume to the ith projection, once the projection maps are specified. With algebraic reconstruction techniques, the solution of (1.12) for the volume density $(u_j)_{1 \le j \le N}$ is calculated iteratively from an initial guess. Note that the coefficients w_{ij} must be recomputed at each iteration, as the matrix W is generally too large to be stored efficiently.

Various reconstruction methods will handle various difficulties with different levels of success, and it is often argued for the reasons mentioned above that iterative methods provide higher quality reconstructions [20]. In particular, ASART, an adaptive simultaneous algebraic reconstruction technique, has been shown to be very efficient when dealing with noisy and incomplete data conditions [27]. However, most iterative-based reconstruction packages [28, 29, 30, 20] adopt straight-line correspondences between the 2D micrographs and the 3D reconstruction, ignoring the complex 3D distortions. The distortions are very serious, especially in large-field electron microscopy [13, 24].

In this work, we have applied ASART to the generalized ray transform case in a manner similar to TxBR, by using nonlinear correspondences between the projection series and the reconstruction volume. Compared to the original TxBR FBP scheme, iterative techniques are desirable because they remove the problem of calculating the filter expression. The rest of the paper is organized as follows: In section 2, we introduce a new reconstruction method applying ASART to the case of the generalized X-ray transform and also introduce a highly parallel scheme for our reconstruction method, which entails evaluation of polynomial functions on a regular grid; in section 3, experimental results are shown; in section 4 we discuss the results and conclude the paper.

2. Methods.

2.1. Computational methods for iterative reconstruction. In this work, we used an iterative approach recently developed [27] and called ASART, short for adaptive simultaneous algebraic reconstruction technique. In the following sections, we outline the main characteristics and key features of ASART: an appropriate choice for the initial guess, a modified multilevel scheme for data access (MMAS), and an adaptive adjustment for the relaxation parameters (AAR). We have implemented our method in C language.

2.1.1. Initial value. Two main choices are possible to start the iteration process for the initial value of the volume density u_j.

The first possibility is to use a simple backprojection technique (BPT): the initial guess for the reconstruction is given by the weighted average of the projections of all the possible rays passing through the pixel [31], that is,

$$(2.1) \qquad u_j^{(0)} = \frac{\sum_{i=1}^{M} w_{ij} v_i}{\sum_{i=1}^{M} w_{ij}}.$$

The w_{ij} are a discretization of the projection maps, in particular the ones estimated in section 2.1 through the bundle adjustment technique. As a rule, the value of w_{ij} is 1 when the ith curvilinear ray crosses the jth pixel; otherwise w_{ij} is equal to 0. With algebraic reconstruction techniques, the solution of (2.1) for the volume density is calculated iteratively from an initial guess $u_j^{(0)}$.

The second possibility is to use an FBP solution such as TxBR in (1.9) as an input of ASART. This may improve the convergence rate, as simple backprojection is subject to a high degree of smoothing which obliterates the details, and in effect places the starting point further from an actual solution.

2.1.2. Modified multilevel scheme for data access. ASART adopts a view-by-view strategy, for which every pixel of each single view is involved at each iteration

cycle. A good choice of the access order of these views can greatly facilitate the convergence rate, especially when they are ordered so as to minimize their correlation [32]. For the case of a complete tilt series (from 0° to 180°), this was accomplished in MAS, a multilevel accessing scheme, for which a view is always next to one of its orthogonal counterparts.

MAS is only suitable for a complete set of projections distributed uniformly over a 180° range. As mentioned before, a tilt series in ET is incomplete due to the missing wedge problem; therefore pairs of views 90° apart are not always available. For this work, we use a modified multilevel accessing scheme (MMAS) to order the projection views within ASART, where the analogue of B is the total span range θ of the tilt angles. The details of MMAS are shown in [27].

Before the ASART iteration process is carried out, all the views are reordered according to MMAS. This allows for faster convergence especially in the first iterations when the volume corrections are large; this scheme prevents the error from accumulating between two consecutive iterations, as may happen in a sequential scheme.

2.1.3. Adaptive adjustment for relaxation parameters. After rearranging the projection views order data according to MMAS, the iterative process updates via the following scheme:

$$(2.2) \qquad u_j^{(k+1)} = u_j^k + \sum_{s=1}^{S} c_{ij} \left(v_i - \sum_{h=1}^{N} w_{ih} u_h^{(k)} \right),$$

where k is the iteration index, c_{ij} is a relaxation parameter, S is the number of pixels per view, and $i = bS + s$ denotes the ith pixel of the system (B is the number of all views and $b = (k \bmod B)$ is the view index). $u_j^{(k+1)}$ is the next iterative value by updating $u_j^{(k)}$.

In previous iterative approaches, the relaxation parameters c_{ij} usually remain constant and are only related to the projection matrix W. In ASART, we use an adaptive scheme for the relaxation, the parameters c_{ij} being taken dependent on the volume density during the iteration process. This data-driven adjustment procedure is regulated by the following expression of c_{ij}:

$$(2.3) \qquad c_{ij} = \frac{\lambda w_{ij} u_j^{(k)}}{\sum_{s=1}^{S} w_{ij} \sum_{h=1}^{N} w_{ih} u_h^{(k)}},$$

where λ is the fixed value (in general, $0 < \lambda < 2$).

With this approach, the readjustment rate of the volume density is directly dependent on the density value in the previous iteration. We note that the true contribution of the jth pixel to the ith ray integral is $w_{ij}x_j$ rather than w_{ij}, and that w_{ij} only represents a geometrical contribution of the jth pixel to the ith ray integral. Pixels with a large contribution $w_{ij}x_j$ to the ith ray integral will have a large adjustment step in our adaptive adjustment.

Before reconstructing the volume u_j, we should first define the projection map coefficients of P_b by means of a bundle adjustment procedure discussed above. Then according to the projection map P_b, ASART is formulated as follows:

$$(2.4) \qquad \begin{cases} u_j^{(0)} = \frac{\sum_{i=1}^{M} w_{ij} v_i(P_b(j))}{\sum_{i=1}^{M} w_{ij}}, \\ u_j^{(k+1)} = u_j^k + \sum_{s=1}^{S} \frac{\lambda w_{ij} u_j^{(k)}}{\sum_{s=1}^{S} w_{ij} \sum_{h=1}^{N} w_{ih} u_h^{(k)}} \left(v_i(P_b(j)) - \sum_{h=1}^{N} w_{ih} u_h^{(k)} \right). \end{cases}$$

2.2. Updating scheme for multiple-tilt reconstructions. Owing to technical limitations of electron microscopes, i.e., the penetration depth of electrons in the beam, the tilt range cannot cover the widest achievable angular range (up to $\pm 90°$) [1]. The lack of specimen views at high tilt angles will cause artifacts in the tomograms. This is because the limited tilt range in ET results in a region empty of information in the Fourier space of the 3D reconstruction (Fourier slice theorem) and is called the missing wedge. In real space, the missing wedge produces artifacts such as blurring of the spatial features in the beam direction. This will make some features look elongated in that direction and result in a significant loss of resolution. In order to reduce these artifacts caused by the missing wedge, double-tilt or multiple-tilt axis acquisition geometry is used for 3D electron tomographic reconstruction. Typically for stained samples, which are less degraded by radiation, another tilt series is taken with the specimen rotated by 90° (the double-axis tilting geometry). The use of double-tilt axis acquisition geometry significantly reduces the missing information [33, 3] and gives a better distribution of information in Fourier space.

There are two main choices for handling a double-tilt reconstruction. The first is that each tilt series can be reconstructed and the resulting volumes combined, possibly with some warping applied. The second choice is to align both tilt series together and perform a reconstruction on the combined series. Fortunately, the iterative methods described here are independent of the particular rotation scheme employed. In this paper, we give the results for an updating scheme for multiple-tilt reconstructions which is also based on the second choice. On the other hand, if, for example, we desire to perform cross-validation to reduce artifacts, we may apply the first choice. That is, we must assess the comparative quality of the separate double-tilt 3D reconstructions and combine the reconstructions accordingly.

For the results shown in this paper, we consider all the double-tilt images and calculate the total difference between the value of the image data and that of computed projection. Next we use the total difference to update the value of volume. This updating scheme can also be applied to multiple tile reconstructions, as follows:

$$u_j^{(k+1)} = u_j^k + \sum_{s=1}^{S} c_{ij}(v_i(P_{b1}(j) : P_{b2}(j) : \cdots : P_{bn}(j)) - \sum_{h=1}^{N}(w_{ih1} : w_{ih2} : \cdots : w_{ihn})u_h^{(k)}),$$

(2.5)

where there are n-tilt series images. Using this updating scheme, we can handle multiple-tilt reconstruction as a single unified tilt reconstruction.

2.3. Parallel strategy for iterative reconstructions. Three-dimensional reconstruction in ET entails large computational costs and resources which are a function of the computational complexity of the reconstruction algorithms and the size and number of the projection images involved. This is especially true for wide-field tomography. Traditionally, high-performance computing [34] has been used to address such computational requirements by means of parallel computing on supercomputers [35], large computer clusters [30], and multicore computers [36].

Our parallelization strategy is based on two separate schemes.

First, our method, as described in the previous section, permits the decomposition of the reconstruction problem into independent slabs along the Z-axis and makes the process well suited for parallelization. We have a natural choice for coarse-grained parallel computation, in which the reconstruction along each Z-slice is calculated on a different processor. Thus, we can implement a parallel strategy where the subrecon-

struction along each Z-slice is calculated at the same time. This strategy makes use of a message passing interface (MPI), standard in parallel programming. We can also apply a single program multiple data (SPMD) approach to perform the parallelization of the reconstruction on each Z-slice.

The 3D volume is decomposed into several slabs with equal heights along the Z-axis. These slabs are assigned and reconstructed on an individual node on a cluster. The number of slabs equals to the number of nodes. In this work, the generic iterative reconstructing process is described as follows:

(1) Initialization: compute the projection map P_ω and make an initial value for $u_j^{(0)}$ by BPT or FBP.

(2) Reprojection: estimate the computed projection data v' based on the current approximation $u_j^{(k)}$.

(3) Backprojection: backproject the discrepancy Δv between the experimental v and calculated projections v', and refine the current approximation u by incorporating the weighted backprojection Δu.

(4) Barrier: synchronize the computation. There are two synchronous barriers during the reconstruction process. One is after the initialization. The other is required during the reconstruction process per iteration because the updated value $u_j^{(k+1)}$ is used in the next iteration.

(5) Reduction: combine the subreconstructions. After the process of reprojection, the calculated projections based on subreconstruction of Z-slices should be stacked together.

The second scheme, as described in section 1.3.1, evaluating polynomials by using the projection map P_ω for each point, can be simplified by a recursion trick based on higher order differences. Because this scheme can be used in each X-line, we can compute the polynomials of each X-line in parallel. In the reprojection and backprojection procedures, we adopt this scheme to eliminate multiplications and accelerate the process of calculating these polynomials. Our implementation uses both the coarse-grain and fine-grain schemes in tandem.

3. Results. In this section, we report results of the reconstruction technique described in the previous sections applied to a biological specimen involving a synaptic terminal of a mouse brain. The sample thickness is about 350nm; micrographs were taken in a 300kV FEI Titan TEM with a 37k magnification. The tilt series in this example is composed of 121 micrographs, each micrograph being 512×512. The size of reconstructions is $651 \times 684 \times 101$. The specimen is tilted from $-60°$ to $+60°$ in one-degree increments. Alignment of the tilt series as presented in section 1.3.2 was performed by tracking 10nm size gold markers. In the alignment process, we used 174 gold markers to get the projection map.

We have considered three different reconstruction cases: First is an FBP reconstruction obtained with TxBR for a quadratic projection map model. Two are implemented using ASART, respectively, with an affine and a quadratic projection model ($n = 1$ and $n = 2$ in (1.10)). The quadratic maps used both for ASART and FBP were identical, and designed to flatten the reconstruction of the biological specimen section [7]. Here, ASART adopts the FBP described in section 2.1.1 to calculate the initial value. We have implemented the two ASART methods for 90 iterations. Furthermore, we adopt the parallel strategy for iterative process and recursion scheme to accelerate the process of the reconstruction. In this experiment, on a cluster with 8 nodes the running time per iteration is only 2.5 minutes on average.

Figure 3.1 shows X-Z, X-Y, and X-Z slices of the reconstruction using the three

(a) Result of FBP quadratic map.

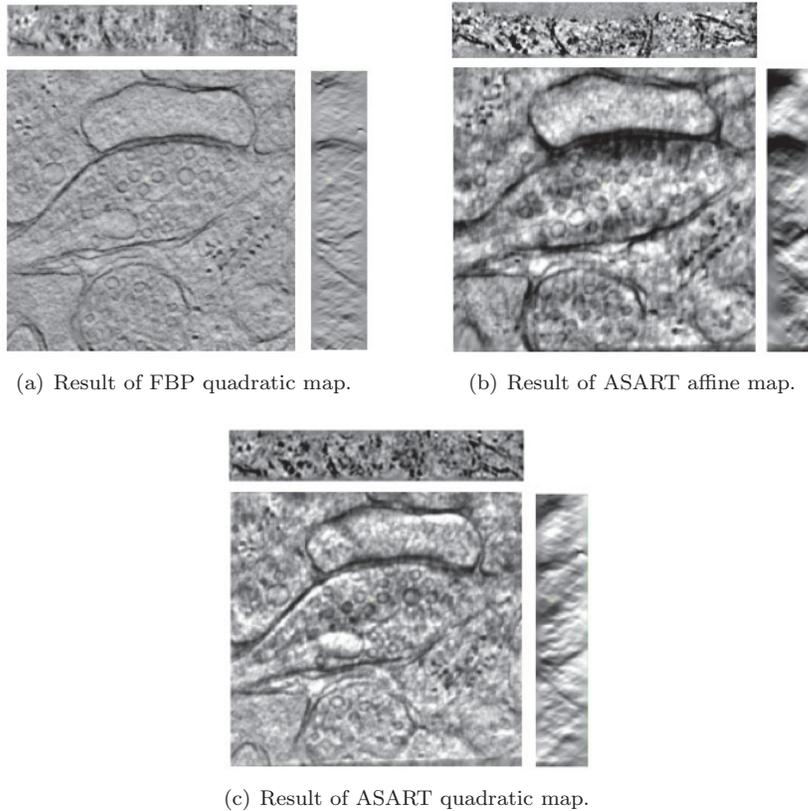(b) Result of ASART affine map.

(c) Result of ASART quadratic map.

FIG. 3.1. *Results (the top is an X-Z section, the left is an X-Y section, and the right is a Y-Z section) from three methods:* (a) *FBP combined with a quadratic projection map;* (b) *ASART combined with an affine projection map after* 50 *iterations;* (c) *ASART combined with a quadratic projection map after* 50 *iterations.*

methods mentioned above. The results of the TxBR FBP are shown in Figure 3.1(a). Figures 3.1(b) and 3.1(c) display the results of ASART after 50 iterations. The two results are produced by different projection maps: an affine and a quadratic map, respectively. Comparing the results in Figures 3.1(a) and 3.1(c), we can see that structures and overall shapes between the ASART and FBP reconstructions implemented with a quadratic projection model are similar, thus validating our approach. The contrast in the ASART reconstructions seems to be enhanced in comparison to FBP. We can see more details in the result of ASART than that of FBP. From Figure 3.1(b), as it is clear in the reconstruction using the affine projection map (compatible with straight line projections), the sample appears to be strongly warped as a result of a long beam exposure. The difference in shape of the object in the two cases, where quadratic and affine projection maps are employed, is due to the inherent gauge ambiguity described in the alignment section above. In particular, the reconstructed shape of the object can be changed by applying a warping map to the positions of the markers in the object or alternatively changing the projection maps. Note the curvatures of this sample in the X-Z and Y-Z sections.

For evaluation of the quality of the reconstructed images, we applied three criteria for the comparison of the results using the different measures: maximum absolute
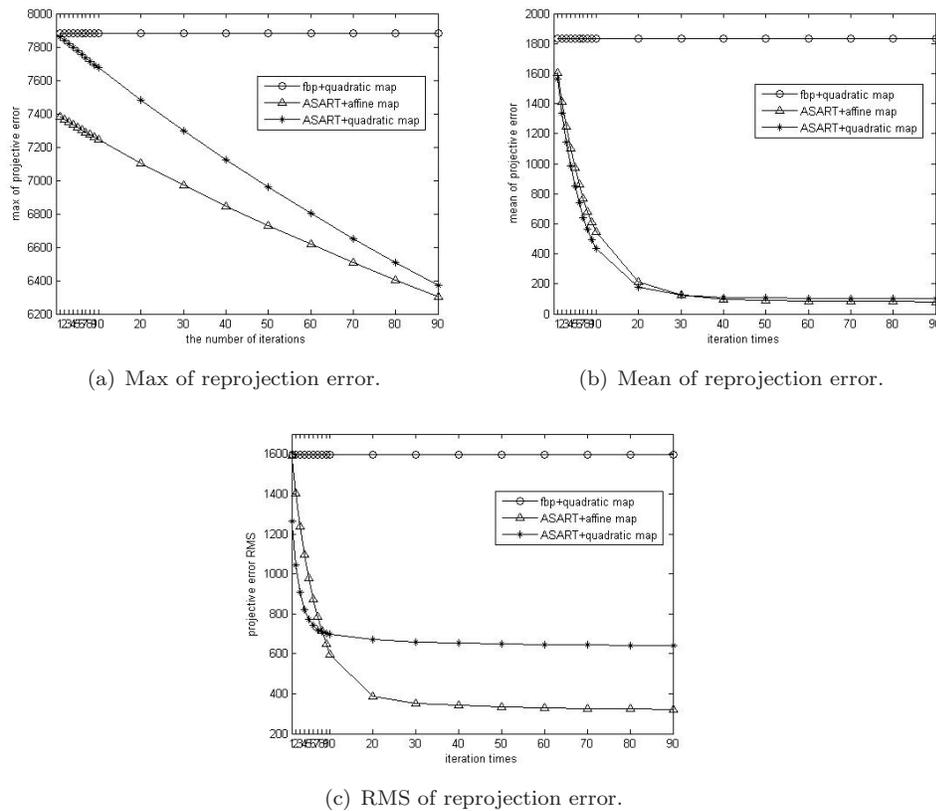
(a) Max of reprojection error.

(b) Mean of reprojection error.

(c) RMS of reprojection error.

Fig. 3.2. *Three criteria for the comparison of the error using single-tilt series.*

pixel difference, mean of the reprojection error, and root mean square (RMS) of the reprojection error. These criteria are based on the discrepancy between the experimental images and the images calculated by reprojection. For example, the RMS of the reprojection error is calculated as follows:

$$(3.1) \qquad \varepsilon = \left[\frac{1}{M} \sum_{i=1}^{M} (v_i - v_i')^2 \right]^{\frac{1}{2}},$$

where v_i is the experimental projection images and v_i' the calculated projection images. In general, a smaller reprojection error indicates a better reconstruction. The curves of these measures versus the number of iterations are presented in Figure 3.2. Because the result of FBP is not related to the number of iterations, the values of these measures are constant. According to Figures 3.2(b) and 3.2(c), we can see that the values of the averages for ASART using two projection maps decrease rapidly before 20 iterations and afterward begin to change more slowly. It appears from these measures that the iteration processes become more stable after 20 iterations. As shown in Figure 3.2(a), contrary to the averages, the maxima of reprojection error decrease nearly linearly throughout the iterations shown. We will discuss the cause in the next section.

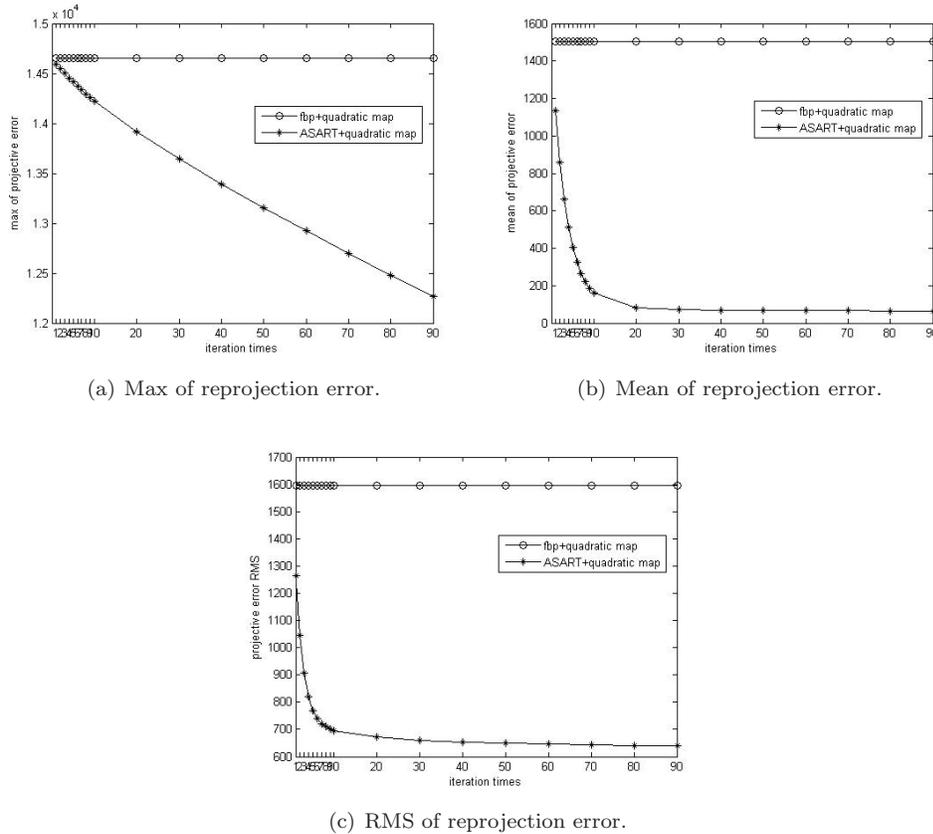To evaluate our updating scheme for multiple-tilt reconstructions, we calculated

(a) Max of reprojection error.

(b) Mean of reprojection error.

(c) RMS of reprojection error.

FIG. 3.3. *Three criteria for the comparison of the error using double-tilt series.*

a reconstruction from a double-tilt series. In the second calculation, the results are reconstructed from two series of projection images including an X-axis tilt series used in the first calculation and the other Y-axis tilt series. In order to evaluate the performance of our method for double-tilt reconstruction, we have implemented the reconstructions using FBP and ASART. Here, both methods adopt the same quadratic projection map. The three measures of reprojection error of the reconstructed results are shown in Figure 3.3. As shown in Figure 3.3, we can see curves similar to those in Figure 3.2. Comparing the value of the mean reprojection error for one iteration in Figures 3.2(b) and 3.3(b), it is clear that this value using double-tilt series is much smaller than that from single-tilt series. The same holds for the RMS measure of reprojection error. Thus, an updating scheme for multiple-tilt reconstructions can accelerate the convergence of the iterative process.

4. Discussion and conclusions. Electron tomography is an essential imaging technique for determining the 3D structures of complex cellular samples. Large-field ET extends the findings of systems biology because the 3D structure of the whole cells and large subcellular organelles such as the cell nucleus can be visualized and understood globally in the contexts of normal metabolic function and disease. We review in this paper the main difficulties in obtaining high-resolution reconstruction results by means of large-field ET, as well as obstacles to the reconstruction of a high-

quality 3D representation of a biological sample, and how reconstruction quality can be affected by various departures from the ideal instrument. In particular, a review of the current literature indicates that the inversion problem associated with the helical trajectories of the electrons through the sample does not have a closed form solution. Furthermore the dose-sensitivity of biological samples in the microscope limits the number of view angles. This exacerbates the effect of discretization artifacts, especially for single-step methods, such as FBP.

The nonlinear alignment scheme described above can compensate for significant deviations caused by the curvilinearity of electron trajectories. On the other hand, iterative methods are known to be effective in minimizing the deleterious effects of limited angular range, discretization, and statistical noise. In this work, we have used an iterative method, ASART, together with a nonlinear alignment scheme specially developed to handle distortion effects in the microscope due to curvilinearity of the electron trajectories, lens aberrations, and sample warping.

We have shown through a concrete example that this new method is effective in compensating for the nonlinearities in the effective projection maps and yields high-quality reconstructions in both single- and double-tilt series, and that the behavior of the convergence is similar in both cases. We also note that the contrast in the ASART reconstructions seems more pronounced than in the FBP. Additionally, the use of an iterative method to invert this generalized X-ray problem has major technical advantages, as opposed to the FBP case, for which the most accurate filter would be spatially variable, and thus would pose the computational problem of working with a six-dimensional object.

From a computational point of view, incorporating beam curvilinearity and sample warping effects in the reconstruction procedure has the main consequence in the evaluation of the projection matrix W, which needs to be computed twice for each iteration. The matrix W is relatively sparse, and so may be amenable to handling by more efficient alternative procedures. Further improvements may also be available in the alignment model and the updating step of the iteration.

One of the shortcomings of our present alignment models is that the projection parameters are calculated semi-independently for each tilt. Specifically the model for the marker position is shared by all tilts, but the projection maps are independent. This potentially has the effect of introducing unwanted spatial noise in the projection maps. Furthermore, marker position in the image data is shared in this limited way among the tilt projections, so the use of data is inefficient. For statistical purposes, the degree of overdetermination must be relatively large in semi-independent alignment models. This defect can be eliminated by introducing an angle parameter into the polynomial coefficients of the individual nonlinear tilt projections. The polynomial coefficients then become a function of angle, and the projection maps for each tilt can then be made dependent on the joint information in all the particle tracks. By using only the first few Fourier components for each angle-dependent polynomial coefficient, we reduce the number of marker positions necessary to calculate polynomial projection coefficients.

The graphs of the mean error and RMS error (see Figures 3.2(b) and 3.2(c)) indicate that two different processes may contribute to these measures: a process which converges rapidly, and one that converges slowly. This observation indicates further investigation is needed. When these two graphs are compared with the maximum of the absolute difference, which appears to decrease at a nearly constant rate up to 90 iterations, we see that this measure behaves quite differently. One possibility is

the presence of artifact in the reconstruction. Three processes contribute to artifact: discretization, missing information at high tilt angles, and differential sampling of regions of high density gradient. The discretization artifact shows up strongly in ray patterns around dark features, the missing data artifact in white patches around dark features, and the differential sampling artifact in the different appearance of the X-Z and Y-Z sections shown in the reconstructions above. In particular, much of the artifacts appear to contribute high gradients in directions perpendicular to the tilt axis. These geometric differences can be used for artifact reduction.

We propose that nonlinear and cross-validation methods can be used for artifact reduction and may also speed convergence. We have obtained some promising results in artifact reduction using several nonlinear filtering techniques. A short list includes two cross-validation methods: a method using differential weighting of back-projected densities from separate reconstructions arising from double-tilt series, where the weights come from comparing the local gradients of the separate reconstructions; and a method using a technique which weights the contribution from two series according to local spatial variance. We have also investigated a technique which normalizes the pointwise variance, and a technique which computes the reconstruction via replacement of the addition in the backprojection by a "sum" operator associated with Maslov dequantization. In particular, Maslov dequantization may be used to suppress the extremes which are associated with reconstruction artifacts. Many of these techniques can be adopted in the update step of the iteration by modification of the update formula and are described in an online wiki [37].

Beyond more sophisticated update schemes, a second strategy for increasing computational speed is exploiting parallelism. Although we have not employed this strategy in our efforts described in this paper, we have shown that a fine-grain scheme will work and have implemented this scheme on a cluster. The next step is to use machines or clusters of machines with GPU boards, for cost-effect speedup.

Finally, we should mention that further investigation is necessary to specify a stopping criterion good for all purposes. At present, stopping at the turning point of the graph of the mean square error (at about 20 iterations) appears to give good results overall in comparison to FBP, but artifacts may still confound small features in the reconstructions.

REFERENCES

[1] J. FRANK, *Electron Tomography: Methods for Three-Dimensional Visualization of Structures in the Cell*, 2nd ed., Springer, New York, 2006.

[2] R. MARABINI, E. RIETZEL, R. SCHROEDER, G.T. HERMAN, AND J.M. CARAZO, *Three-dimensional reconstruction from reduced sets of very noisy images acquired following a single-axis tilt schema: Application of a new three-dimensional reconstruction algorithm and objective comparison with weighted back-projection*, J. Structural Biol., 120 (1997), pp. 363–371.

[3] D.N. MASTRONARDE, *Dual-axis tomography: An approach with alignment methods that preserve resolution*, J. Structural Biol., 120 (1997), pp. 343–352.

[4] A. LAWRENCE, S. PHAN, AND R. SINGH, *Parallel processing and large-field electron microscope tomography*, in Proceedings of the World Congress on Computer Science and Information Engineering, 2009, pp. 339–343.

[5] X. Wan, F. Zhang, Q. Chu, and Z. Liu, *High-performance blob-based iterative three-dimensional reconstruction in electron tomography using multi-GPUs*, BMC Bioinform., 13 (2012), pp. 61–72.

[6] S.T. Peltier, J.C. Bouwer, L. Jin, K. Khodjasaryan, S. Geist, N.H. Xuong, and M.H. Ellisman, *Design of a new $8k \times 8k$ lens coupled detector for wide-field, high-resolution transmission electron microscopy*, Microscopy and Microanalysis, 11 (2005), pp. 610–611.

[7] S. Phan, A. Lawrence, T. Molita, J. Lanman, M. Berlanga, M. Terada, A. Kulungowski, and J. Obayashi, *TxBR montage reconstruction for large field electron tomography*, J. Structural Biol., 180 (2012), pp. 154–164.

[8] K. Tsuna and Y. Harada, *Minimisation of radial and spiral distortion in electron microscopy through the use of a triple pole-piece lens*, J. Phys. E Sci. Instruments, 14 (1981), pp. 313–320.

[9] L. Reimer, *Transmission Electron Microscopy: Physics of Image Formation and Microanalysis*, 4th ed., Springer, New York, 1997.

[10] J.J. Fernández, *Computational methods for electron tomography*, Micron, 43 (2012), pp. 1010–1030.

[11] V. Sharafutdinof, *Ray Transforms on Riemannian Manifolds*, Lecture Notes, University of Washington, Seattle, 1999.

[12] A. Lawrence, J.C. Bouwer, G. Perkins, and M.H. Ellisman, *Transform-based backprojection for volume reconstruction of large format electron microscope tilt series*, J. Structural Biol., 154 (2006), pp. 144–167.

[13] S. Phan and A. Lawrence, *Tomography of large format electron microscope tilt series: Image alignment and volume reconstruction*, in Proceedings of the Congress on Image and Signal Processing (CISP08), pp. 176–182.

[14] G. Beylkin, *The inversion problem and applications of the generalized radon transform*, Comm. Pure Appl. Math., 37 (1984), pp. 579–599.

[15] E.T. Quinto, *Radon transforms, differential equations and microlocal analysis*, Contemp. Math., 278 (2001), pp. 57–68.

[16] E.T. Quinto and H. Rullgard, *Local Singularity reconstruction from integrals over curves in $\mathbb{R}^3$*, Inverse Problems and Imaging, 7 (2013), pp. 585–609.

[17] A. Greenleaf and G. Uhlmann, *Non-local inversion formulas for the X-ray transform*, Inverse Problems and Imaging 7 (2013), pp. 585–609.

[18] V. Palamodov, *A uniform reconstruction formula in integral geometry*, Inverse Problems, 28 (2012), pp. 65014–65028.

[19] F. Natterer and F. Wübbeling, *Mathematical Methods in Image Reconstruction*, SIAM, Philadelphia, 2001.

[20] J.J. Fernández, A. Lawrence, J. Roca, I. García, M.H. Ellisman, and J.M. Carazo, *High performance electron tomography of complex biological specimens*, J. Structural Biol., 138 (2002), pp. 6–20.

[21] O. Faugeras, Q. Luong, and T. Papadopoulo, *The Geometry of Multiple Images: The Laws That Govern the Formation of Multiple Images of a Scene and Some of Their Applications*, MIT Press, Cambridge, MA, 2001.

[22] A. Heyden and K. Astrom, *Euclidean reconstruction from image sequences with varying and unknown focal length and principal point*, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1997, pp. 438–443.

[23] K. Denecker, J.V. Overloop, and F. Sommen, *The general quadratic Radon transform*, Inverse Problems, 14 (1998), pp. 615–633.

[24] S. Phan, J. Bouwer, J. Lanman, M. Terada, and A. Lawrence, *Non-linear bundle adjustment for electron tomography*, in Proceedings of the World Congress on Computer Science and Information Engineering, 2009, pp. 604–612.

[25] R.V. Ambartzumian, *Factorization Calculus and Geometric Probability*, Cambridge University Press, Cambridge, UK, 1900.

[26] D. Vanhecke, D. Studer, and M. Ochs, *Stereology meets electron tomography: Towards quantitative 3D electron microscopy*, J. Structural Biol., 159 (2007), pp. 443–450.

[27] X. Wan, F. Zhang, Q. Chu, K. Zhang, F. Sun, B. Yuan, and Z. Liu, *Three-dimensional reconstruction using an adaptive simultaneous algebraic reconstruction technique in electron tomography*, J. Structural Biol., 175 (2011), pp. 277–287.

[28] H. Chen, W.K. Clyborne, J.W. Sedat, and D.A. Agard, *PRIISM: An integrated system for display and analysis of 3-d microscope images*, SPIE: Biomedical Image Processing and Three-Dimensional Microscopy, 1660 (1992), pp. 784–790.

[29] J. Bilbao-castro, R. Marabini, C. Sorzano, I. García, J.M. Carazo, and J.J. Fernández, *Exploiting desktop supercomputing for three-dimensional electron microscopy reconstructions using ART with blobs*, J. Structural Biol., 165 (2009), pp. 19–26.

[30] X. WAN, F. ZHANG, AND Z. LIU, *Modified simultaneous algebraic reconstruction technique and its parallelization in cryo-electron tomography*, in Proceedings of the International Conference on Parallel and Distributed Systems (ICPADS 2009), 2009, pp. 384–390.

[31] G.T. HERMAN, *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography*, Academic Press, New York, 1980.

[32] H. GUAN AND R. GORDON, *A projection access order for speedy convergence of ART (algebraic reconstruction technique): A multilevel scheme for computed tomography*, Phys. Med. Biol., 39 (1994), pp. 2005–2022.

[33] P. PENCZEK, M. MARKO, K. BUTTLE, AND J. FRANK, *Double-tilt electron tomography*, Ultramicroscopy, 60 (1995), pp. 393–410.

[34] J.J. FERNÁNDEZ, *High performance computing in structural determination by electron cryomicroscopy*, J. Structural Biol., 164 (2008), pp. 1–6.

[35] J.J. FERNÁNDEZ, J.M. GARAZO, AND I. GARCÍA, *Three-dimensional reconstruction of cellular structures by electron microscope tomography and parallel computing*, J. Parallel Distrib. Comput., 64 (2004), pp. 285–300.

[36] J.I. AGULLEIRO AND J.J. FERNÁNDEZ, *Fast tomographic reconstruction on multicore computers*, Bioinform., 27 (2011), pp. 582–583.

[37] A. LAWRENCE, *Recent Developments in Electron Microscope Tomography*, Center for Research in Biological Systems University of California, San Diego, CA, 2011, pp. 49–58; http://www2.nbcr.net/wiki/images/a/a6/Presentation-on-contour.pdf.