

## Review Article

# A Survey of the Use of Iterative Reconstruction Algorithms in Electron Microscopy

**C. O. S. Sorzano,<sup>1,2</sup> J. Vargas,<sup>1</sup> J. Otón,<sup>1</sup> J. M. de la Rosa-Trevín,<sup>1</sup> J. L. Vilas,<sup>1</sup> M. Kazemi,<sup>1</sup> R. Melero,<sup>1</sup> L. del Caño,<sup>1</sup> J. Cuenca,<sup>1</sup> P. Conesa,<sup>1</sup> J. Gómez-Blanco,<sup>1</sup> R. Marabini,<sup>3</sup> and J. M. Carazo<sup>1</sup>**

<sup>1</sup>Centro Nacional de Biotecnología, CSIC, c/Darwin 3, Cantoblanco, 28049 Madrid, Spain

<sup>2</sup>Universidad CEU San Pablo, Campus Urbanización Montepríncipe, Boadilla del Monte, 28668 Madrid, Spain

<sup>3</sup>Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain

Correspondence should be addressed to C. O. S. Sorzano; [coss@cnb.csic.es](mailto:coss@cnb.csic.es)

Received 16 September 2016; Accepted 9 March 2017; Published 17 September 2017

Academic Editor: Achim Langenbacher

Copyright © 2017 C. O. S. Sorzano et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

One of the key steps in Electron Microscopy is the tomographic reconstruction of a three-dimensional (3D) map of the specimen being studied from a set of two-dimensional (2D) projections acquired at the microscope. This tomographic reconstruction may be performed with different reconstruction algorithms that can be grouped into several large families: direct Fourier inversion methods, back-projection methods, Radon methods, or iterative algorithms. In this review, we focus on the latter family of algorithms, explaining the mathematical rationale behind the different algorithms in this family as they have been introduced in the field of Electron Microscopy. We cover their use in Single Particle Analysis (SPA) as well as in Electron Tomography (ET).

## 1. Introduction

Electron Microscopy has been established as one of the key players in Structural Biology with the goal of elucidating the three-dimensional structure of macromolecular complexes in order to better understand their function and molecular mechanisms [1–3]. One of the most important steps in the image processing pipeline is the 3D reconstruction of a map compatible with the projections acquired at the microscope [4]. In practice, projections of the macromolecule are contaminated by a huge amount of noise (typical Signal-to-Noise Ratios in the order of 0.01; i.e., there is 100 times more noise power than signal power, Hosseinizadeh et al. [5]), and the 3D reconstruction emerges, in a simplified way, as the 3D “average” of thousands of projections, each one looking at the molecule from a different point of view [4]. In the 3D space, the signal coming from the macromolecules is reinforced by the averaging process. However, random noise tends to be canceled by this averaging. Currently, the 3D reconstruction step is no longer seen as a limiting step (except

for its execution time) in Single Particle Analysis due to the large number of particles involved in the reconstruction (between tens and hundreds of thousands), and direct Fourier inversion methods are currently the standard de facto [6–8]. These latter methods are especially well suited to handle a large number of projections thanks to their computational speed and their accuracy when the angular coverage of the set of projections fully fills the 3D Fourier space, which is currently normally the case in SPA (a word of caution should be expressed in those cases in which subsequent rounds of 3D classification significantly reduce the number of images per 3D class). However, in the past, there was an intense work of research in selecting the best reconstruction algorithm between four different families of reconstruction algorithms: (1) direct Fourier inversion [6–8], (2) back-projection algorithms [9–18], (3) iterative algorithms [19–43], and (4) Radon inversion methods [44–46].

The situation in Electron Tomography (ET) is, in general, specially involved, due to the combined effect of a smaller number of projection images, the existence of a missing

wedge, and the fact that images are very large, posing an important challenge to the computational resources (especially in terms of memory; for a comparison between Single Particle Analysis and Electron Tomography, see Jonic et al. [47]). The research of 3D reconstruction algorithms trying to reduce reconstruction artifacts and the execution time has been rather intense, especially in the last decade [48–61]. Similarly, pure reconstruction algorithms in ET may be combined with a priori information incorporating the fact that the reconstruction is sparse in some domain or exploiting the discrete nature of the object being imaged [62–76].

In this review, we focus on the family of iterative reconstruction algorithms, also known as series expansion methods. The classical algorithms (ART, Block ART, and SIRT) have been followed by a number of more modern algorithms like Conjugate Gradient, Subgradient Descent, Projected Subgradient, Superiorization, ADMM, and so forth, allowing all kinds of regularizations with a special emphasis on sparsity promoting regularizers. We start by showing how the 3D reconstruction problem can be posed as a problem of solving a linear system of equations. Then, we put the more classical methods in a common algebraic framework. Finally, we introduce the rationale behind the more modern methods.

## 2. 3D Reconstruction as an Equation System Problem

Images collected by the electron microscope can be understood as the parallel projection of the Coulomb potential of the macromolecule being imaged. The relationship between the 3D model,  $V(\mathbf{r}) : \mathbb{R}^3 \times 1 \rightarrow \mathbb{R}$ , and a projection image,  $I(\mathbf{s}) : \mathbb{R}^2 \times 1 \rightarrow \mathbb{R}$ , is given by [77]

$$I_{\bar{A}}(\mathbf{s}) = \int_{-\infty}^{\infty} V(\bar{A}^{-1}\tilde{H}^T\mathbf{s}) dt, \quad (1)$$

where  $\mathbf{r} = (x, y, z, 1)$  and  $\mathbf{s} = (x, y, 1)$  are the homogeneous coordinates of a 3D point and a 2D point, respectively;  $\tilde{H}^T$  is

$$\tilde{H}^T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & t \\ 0 & 0 & 1 \end{pmatrix}, \quad (2)$$

with  $t$  being the integration variable, and

$$\bar{A} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} R & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \quad (3)$$

is the matrix that specifies the point of view from where the projection has been taken ( $R$  is a rotation matrix normally specified by 3 Euler angles and  $\mathbf{t}$  is an in-plane displacement). Every projection has its own matrix  $\bar{A}$  reflecting the different projection directions and individual shifts of each image.

In practice, this ideal image is never observed but it is corrupted by random noise (whose nature is related to

the structure of the ice surrounding the macromolecule, the random arrival of electrons, etc.) and the projected image is not known at any position  $\mathbf{s}$  but at a discrete set of positions ( $\mathbf{s}_i$ , normally the centers of the pixels of the acquired image). In this way, we may rewrite the above equation as

$$I_{\bar{A},i} = I_{\bar{A}}(\mathbf{s}_i) = \int_{-\infty}^{\infty} V(\bar{A}^{-1}\tilde{H}^T\mathbf{s}_i) dt + n_i. \quad (4)$$

Note that this image formation model uses additive noise, since the main source of noise is not the low number of electron counts at each pixel but the structure of the amorphous ice embedding the molecule. This noise has a Gaussian distribution as was shown in [78].

Let us assume now that we express the volume as a linear combination of functions,  $b(\mathbf{r})$ , shifted to a set of known positions  $\mathbf{r}_j$  (e.g., usually these positions are regularly distributed on a grid [30]):

$$V(\mathbf{r}) = \sum_j x_j b(\mathbf{r} - \mathbf{r}_j). \quad (5)$$

The basis functions may be voxels or any other function with some interesting property from the tomographic point of view (for instance, Kaiser-Bessel modified functions, also known as blobs, [32, 36] are known to reduce the noise in the 3D reconstruction).

The goal of the tomographic problem is to determine the basis coefficients,  $x_j$ , such that the experimental projection is actually the line integral of the volume  $V$ . Substituting (5) into (4) and disregarding the noise term (it will be recovered later), we have

$$\begin{aligned} I_{\bar{A},i} &= \int_{-\infty}^{\infty} V(\bar{A}^{-1}\tilde{H}^T\mathbf{s}_i) dt \\ &= \int_{-\infty}^{\infty} \left( \sum_j x_j b(\bar{A}^{-1}\tilde{H}^T\mathbf{s}_i - \mathbf{r}_j) \right) dt \\ &= \sum_j x_j \left( \int_{-\infty}^{\infty} b(\bar{A}^{-1}\tilde{H}^T\mathbf{s}_i - \mathbf{r}_j) dt \right) = \sum_j x_j a_{\bar{A},ij}, \end{aligned} \quad (6)$$

where  $a_{\bar{A},ij}$  is the line integral of the basis function located at  $\mathbf{r}_j$  at the location  $\mathbf{s}_i$  along the direction determined by the geometrical transformation  $\bar{A}$ , that is, the contribution of the  $j$ -th basis function of the volume to the  $i$ -th pixel in the image. We may write the image pixels in a single vector using lexicographical order and we may rewrite the above equation in a matrix form:

$$\mathbf{I}_{\bar{A}} = A_{\bar{A}}\mathbf{x}. \quad (7)$$

If we have images of size  $P \times P$  pixels and assuming that the volume is reconstructed in a cube of size  $B \times B \times B$ , then  $\mathbf{I}_{\bar{A}} \in \mathbb{R}^{P^2}$  is a vector in a space of dimension  $P^2$ ,  $\mathbf{x} \in \mathbb{R}^{B^3}$  is a vector in a space of dimension  $B^3$ , and  $A_{\bar{A}}$  is a matrix of size  $P^2 \times B^3$ . If we look again at the pixel level, we may write

$$I_{\bar{A},i} = \mathbf{a}_{\bar{A},i}^T \mathbf{x} = \langle \mathbf{a}_{\bar{A},i}, \mathbf{x} \rangle, \quad (8)$$

where  $\mathbf{a}_{\bar{A}_i}^T$  is the  $i$ -th row of the matrix  $A_{\bar{A}}$ . This is the equation of a hyperplane in the  $\mathbb{R}^{B^3}$  space, meaning that every single pixel of the projection image provides a hyperplane constraint for locating the volume coefficients  $\mathbf{x}$ . The true  $\mathbf{x}$ , in the ideal noiseless case, must be at the intersection of all the hyperplanes defined by all pixels (the noise observed in the measurements simply randomly shifts each hyperplane along the direction perpendicular to the hyperplane normal). In this way, a single projection provides  $P^2$  equations, but we have  $B^3$  unknowns so that there is an infinite number of structures compatible with a single projection. In Electron Microscopy (EM), we collect thousands of projection images (let us say  $M$ ) and we may stack all the measurements in a big vector. The corresponding equation system would then be

$$\begin{pmatrix} \mathbf{I}_{\bar{A}_1} \\ \mathbf{I}_{\bar{A}_2} \\ \vdots \\ \mathbf{I}_{\bar{A}_M} \end{pmatrix} = \begin{pmatrix} A_{\bar{A}_1} \\ A_{\bar{A}_2} \\ \vdots \\ A_{\bar{A}_M} \end{pmatrix} \mathbf{x} \quad (9)$$

which in general can be written as the linear equation system

$$\mathbf{b} = \mathbf{A}\mathbf{x}, \quad (10)$$

where  $\mathbf{b} \in \mathbb{R}^{MP^2}$  is a vector collecting all the experimental images and  $A$  is a matrix of size  $MP^2 \times B^3$  collecting all the projection matrices. Note that this matrix is very sparse since, for common basis functions like voxels, blobs, and so forth, just a few coefficients in the volume contribute to each pixel (only those coefficients along the integration line passing through that pixel). Just to give some typical numbers in EM, let us say we have  $M = 50,000$  images of size  $P \times P = 256 \times 256$ . We would have  $MP^2 = 3.3$  billion equations and  $B^3 = P^3 = 16.8$  million unknowns (if voxels are used as basis function). The equation system is inconsistent because of the measurement noise and some kind of least squares solution must be sought [79]. We will do so in the next section.

We have presented the linear equation system only in connection to the data collection geometry. However, the Contrast Transfer Function (CTF) of the microscope (i.e., how the microscope blurs the ideal images) can easily be incorporated. The convolution can be represented by a matrix multiplication using a Toeplitz matrix, such that instead of the  $A_{\bar{A}}$  matrices that contain purely geometrical information we may use the matrices  $CA_{\bar{A}}$ , where  $C$  is the CTF Toeplitz matrix (each experimental image may have its own CTF matrix).

Interestingly, posing the tomographic problem as a linear equation system can also be done in Fourier space. Thanks to the Central Slice Theorem, the relationship between the 3D Fourier transform of the macromolecule model and the 2D Fourier transform of the projection can be expressed as [77]

$$\hat{I}_{\bar{A}}(\tilde{\mathbf{S}}) = e^{-i(\tilde{\mathbf{S}}, H_0 \mathbf{R}t)} \hat{V}(\bar{A}^T \tilde{H}_0^T \tilde{\mathbf{S}}), \quad (11)$$

where  $\bar{A}^T = \begin{pmatrix} R & 0 \\ 0^T & 1 \end{pmatrix}$ ,  $\tilde{H}_0$  is the matrix defined at (2) with  $t = 0$ , and  $H_0$  is the same matrix except the last row (the one

adding its homogeneous nature).  $\tilde{\mathbf{S}}$  is the 2D frequency coordinate and  $\hat{I}_{\bar{A}}$  and  $\hat{V}$  are the 2D and 3D Fourier transforms of the projection image and the macromolecular model, respectively. Equation (11) means that, to evaluate the Fourier coefficient at the 2D frequency coordinate  $\tilde{\mathbf{S}}$ , we need to evaluate the Fourier transform of the volume at the coordinate  $\bar{A}^T \tilde{H}_0^T \tilde{\mathbf{S}}$  and then multiply by the corresponding phase term,  $e^{-i(\tilde{\mathbf{S}}, H_0 \mathbf{R}t)}$ , to account for the shift in the images. In practice, we do not have the Fourier transform of the volume at any possible location. We will have to interpolate its value from the 3D Fourier coefficients in its surroundings. Let us concentrate at the frequency  $\tilde{\mathbf{S}}_j$  and let us call the corresponding 2D Fourier coefficient  $\hat{I}_j$ . Let us refer to the  $j$ -th 3D Fourier coefficient by  $\hat{V}_j$ . Then, we may interpolate the 2D Fourier coefficient as

$$\hat{I}_i = \sum_j \hat{V}_j \hat{a}_{\bar{A},ij}, \quad (12)$$

where the terms  $\hat{a}_{\bar{A},ij}$  comprise the phase shift as well as the interpolation weights. This equation is formally identical to (6), meaning that the 3D reconstruction problem in Fourier space can also be expressed as a linear system of equations, and we have to be careful to perform the complex interpolation. Actually, this is the position taken in Scheres [80] and Chen and Förster [59]. Including the CTF in this framework is even easier than in the real space case, since it suffices to multiply the  $\hat{a}_{\bar{A},ij}$  terms by the  $c_i$  CTF coefficients.

### 3. (Weighted) Least Squares Solution

Given the inconsistent equation system  $\mathbf{A}\mathbf{x} = \mathbf{b}$ , we may try to look for the  $\mathbf{x}$  that minimizes its distance to all the hyperplanes given by the experimentally measured pixel values. This is achieved by minimizing the norm of the residuals vectors  $\mathbf{A}\mathbf{x} - \mathbf{b}$ . In tomographic terms,  $\mathbf{A}\mathbf{x}$  may be interpreted as the expected experimental projections given a macromolecular model  $\mathbf{x}$ , while  $\mathbf{b}$  are the experimentally measured projections. The idea is to find the model  $\mathbf{x}$  such that its reprojections are as similar as possible to the experimentally acquired images. The residual of the comparison between the experimental images and the reprojections should be just the noise present in the experimental images:

$$\mathbf{x} = \arg \min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2, \quad (13)$$

where the norm is calculated using the standard inner product:

$$\|\mathbf{u}\|^2 = \langle \mathbf{u}, \mathbf{u} \rangle = \mathbf{u}^T \mathbf{u}. \quad (14)$$

This is a linear least squares problem. Note that the solution is not unique (any vector  $\mathbf{x}_0$  in the kernel of the matrix  $A$  will yield the same error norm, because  $A(\mathbf{x} + \mathbf{x}_0) = \mathbf{A}\mathbf{x} + \mathbf{0} = \mathbf{A}\mathbf{x}$ ). In practice,  $\mathbf{x}_0$  refers to any structure whose Fourier coefficients are in areas not measured by the experimental images. Developing the norm in the optimization problem,

we have

$$\begin{aligned} \mathbf{x} &= \arg \min_{\mathbf{x}} (\mathbf{A}\mathbf{x} - \mathbf{b})^T (\mathbf{A}\mathbf{x} - \mathbf{b}) \\ &= \arg \min_{\mathbf{x}} \|\mathbf{b}\|^2 + \mathbf{x}^T \mathbf{A}^T \mathbf{A} \mathbf{x} - 2\mathbf{x}^T \mathbf{A}^T \mathbf{b}. \end{aligned} \quad (15)$$

If we now differentiate and equate to 0, we have

$$\frac{\partial \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2}{\partial \mathbf{x}} = 2\mathbf{A}^T \mathbf{A} \mathbf{x} - 2\mathbf{A}^T \mathbf{b} = 0 \quad (16)$$

or equivalently

$$\mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{A}^T \mathbf{b}. \quad (17)$$

This latter equation is known as the normal equation of the least squares problem. It can be shown that any solution of the normal equations is also a solution of the least squares problem, and, conversely, any solution of the least squares problem must also be a solution of the normal equations [81, Chapter 6]. Except for degenerate cases,  $\mathbf{A}^T \mathbf{A}$  is a positive definite matrix, implying that the matrix is invertible, and consequently the normal equations have a unique solution. For complex numbers, the norm is defined as

$$\|\mathbf{u}\|^2 = \mathbf{u}^H \mathbf{u} \quad (18)$$

with  $H$  being the conjugate transpose, so that the normal equations in Fourier space would be

$$\mathbf{A}^H \mathbf{A} \mathbf{x} = \mathbf{A}^H \mathbf{b}. \quad (19)$$

We may now introduce weights in the minimization if there are some measurements that are more reliable than others (in real space, this is more difficult, but in Fourier space, low frequency components are overrepresented with respect to high frequency ones and they are consequently downweighted, Chen and Förster [59]). Given a diagonal, positive definite matrix  $W$  of size  $MP^2 \times MP^2$ , we may define the  $W$  inner product as

$$\langle \mathbf{u}, \mathbf{v} \rangle_W = \mathbf{u}^H W \mathbf{v}. \quad (20)$$

The normal equations become in this case

$$\mathbf{A}^H W \mathbf{A} \mathbf{x} = \mathbf{A}^H W \mathbf{b}. \quad (21)$$

The most straightforward approach to solving this equation system is the use of the Moore-Penrose pseudoinverse:

$$\mathbf{x} = (\mathbf{A}^H W \mathbf{A})^{-1} \mathbf{A}^H W \mathbf{b} = \mathbf{A}^\dagger \mathbf{b}. \quad (22)$$

The matrix  $\mathbf{A}^H W \mathbf{A}$  is of size  $B^3 \times B^3$  and its direct inversion is normally out of the reach of any current computer. Numerical algorithms that avoid the need to invert this matrix are normally used and they are explained in the next section.

## 4. Iterative Solutions of a Linear Equation System

Note that all equation systems posed so far (see (6), (12), (17), (19), and (21)) are of the generic form:

$$\mathcal{A} \mathbf{x} = \mathcal{B}. \quad (23)$$

In this section,  $\mathcal{A}$  and  $\mathcal{B}$  are the generic system matrix and data terms, independently of their actual expressions in terms of the microscopic projection images, CTFs, Fourier or real space versions, and so forth, since these details have already been carefully presented in the sections above.

A very simple, but general, approach to produce an iterative algorithm solving the equation system in (23) is to decompose matrix  $\mathcal{A}$  as the difference of two other matrices,  $\mathcal{M}$  and  $\mathcal{N}$ , as in

$$\mathcal{A} = \mathcal{M} - \mathcal{N}. \quad (24)$$

Then,

$$\begin{aligned} (\mathcal{M} - \mathcal{N}) \mathbf{x} &= \mathcal{B} \\ \mathcal{M} \mathbf{x}^{(k+1)} &= \mathcal{N} \mathbf{x}^{(k)} + \mathcal{B} \\ \mathbf{x}^{(k+1)} &= \mathcal{M}^{-1} \mathcal{N} \mathbf{x}^{(k)} + \mathcal{M}^{-1} \mathcal{B} \\ &= \mathbf{x}^{(k)} + \mathcal{M}^{-1} (\mathcal{B} - \mathcal{A} \mathbf{x}^{(k)}). \end{aligned} \quad (25)$$

This iterative scheme has the interesting property that if the residuals are  $\mathcal{B} - \mathcal{A} \mathbf{x}^{(k)} = \mathbf{0}$  (i.e., we have successfully found a solution of the equation system), then there is no update of the current solution.

(i) *Jacobi*.  $\mathcal{A}$  is decomposed in its diagonal matrix  $\mathcal{D}$  and two lower  $\mathcal{L}$  and upper  $\mathcal{U}$  triangular matrices:

$$\begin{aligned} \mathcal{A} &= \mathcal{D} - (\mathcal{L} + \mathcal{U}) \implies \\ \mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} + \mathcal{D}^{-1} (\mathcal{B} - \mathcal{A} \mathbf{x}^{(k)}). \end{aligned} \quad (26)$$

(ii) *Gauss-Seidel*.  $\mathcal{A}$  is decomposed in a similar way to the Jacobi decomposition above, but the  $\mathcal{L}$  matrix is used differently:

$$\begin{aligned} \mathcal{A} &= (\mathcal{L} + \mathcal{D}) - \mathcal{U} \implies \\ \mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} + (\mathcal{L} + \mathcal{D})^{-1} (\mathcal{B} - \mathcal{A} \mathbf{x}^{(k)}). \end{aligned} \quad (27)$$

(iii) *Richardson*.  $\mathcal{A}$  is decomposed using the identity matrix as reference:

$$\begin{aligned} \mathcal{A} &= I - (I - \mathcal{A}) \implies \\ \mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} + (I - \mathcal{A}) \mathbf{x}^{(k)}. \end{aligned} \quad (28)$$

The reasons why these numerical schemes succeed in solving the equation system  $\mathcal{A} \mathbf{x} = \mathcal{B}$  can be nicely illustrated

in Richardson's scheme. We may rewrite the iterative algorithm as

$$\mathbf{x}^{(k+1)} = \mathcal{C} + (I - \mathcal{A}) \mathbf{x}^{(k)}. \quad (29)$$

We start with  $\mathbf{x}^{(0)} = \mathbf{0}$ . Then, the estimates of  $\mathbf{x}$  would be

$$\begin{aligned} \mathbf{x}^{(1)} &= \mathcal{C} \\ \mathbf{x}^{(2)} &= \mathcal{C} + (I - \mathcal{A}) \mathcal{C} \\ \mathbf{x}^{(3)} &= \mathcal{C} + (I - \mathcal{A}) \mathcal{C} + (I - \mathcal{A})^2 \mathcal{C} \\ &\vdots \\ \mathbf{x}^{(k)} &= \left( \sum_{m=0}^{k-1} (I - \mathcal{A})^m \right) \mathbf{b}. \end{aligned} \quad (30)$$

If  $\mathcal{A}$  is an invertible matrix, then  $\sum_{m=0}^{\infty} (I - \mathcal{A})^m$  converges to  $\mathcal{A}^{-1}$  as long as for all eigenvalues of the matrix  $\mathcal{A}$ ,  $\lambda_i$ , it is fulfilled that  $0 < |1 - \lambda_i| < 1$ , which for real eigenvalues translates into  $0 < \lambda_i < 2$  [82, Section 3.5.1]. Actually, this family of iterative algorithms is normally modified to introduce a relaxation factor  $\mu_k$  (which may be different for each iteration):

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mu_k \mathcal{M}^{-1} (\mathcal{C} - \mathcal{A} \mathbf{x}^{(k)}). \quad (31)$$

This relaxation factor helps to increase the radius of convergence of the algorithm by modifying the eigenvalues of the matrix involved in the convergence analysis (we have seen the analysis for Richardson's iteration but different matrices are needed for the rest of the schemes). There are works analyzing the convergence of the reconstruction algorithm in terms of the properties of the sequence of numbers  $\mu_k$  [83].

SIRT is one of the most popular reconstruction algorithms used in Electron Tomography. As we show below, SIRT is the result of the Jacobi algorithm applied to the normal equations with a particular weighting scheme. Let us consider the normal equations of the Weighted Least Squares problem in real space:

$$A^T W A \mathbf{x} = A^T W \mathbf{b}. \quad (32)$$

The Jacobi update with relaxation parameter is

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mu_k \mathcal{D}^{-1} (\mathcal{C} - \mathcal{A} \mathbf{x}^{(k)}). \quad (33)$$

Let us decompose matrix  $A$  in its different rows (remember that the  $i$ -th row indicates how each basis function in the volume contributes to the  $i$ -th pixel),

$$A = \begin{pmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \vdots \\ \mathbf{a}_{MP^2}^T \end{pmatrix}, \quad (34)$$

and columns (the  $j$ -th column indicates how the  $j$ -th basis function affects all the pixels in the measurements),

$$A = (\boldsymbol{\alpha}_1 \quad \boldsymbol{\alpha}_2 \quad \cdots \quad \boldsymbol{\alpha}_{B^3}). \quad (35)$$

Then,

$$\begin{aligned} \mathcal{A} &= A^T W A = \sum_{i=1}^{MP^2} w_i \mathbf{a}_i \mathbf{a}_i^T \\ \mathcal{C} &= A^T W \mathbf{b} = \sum_{i=1}^{MP^2} w_i b_i \mathbf{a}_i \end{aligned} \quad (36)$$

$$\mathcal{A} \mathbf{x}^{(k)} = \sum_{i=1}^{MP^2} w_i \mathbf{a}_i \mathbf{a}_i^T \mathbf{x}^{(k)} = \sum_{i=1}^{MP^2} w_i \left( \sum_{m=1}^{B^3} a_{im} x_m^{(k)} \right) \mathbf{a}_i.$$

Let us now concentrate on a given basis function  $j$ :

$$\begin{aligned} \mathcal{A}_{jj} &= \sum_{i=1}^{MP^2} w_i a_{ij}^2 = \|\boldsymbol{\alpha}_j\|_W^2 \\ \mathcal{C}_j &= \sum_{i=1}^{MP^2} w_i b_i a_{ij} = \langle \boldsymbol{\alpha}_j, \mathbf{b} \rangle_W \\ (\mathcal{A} \mathbf{x}^{(k)})_j &= \sum_{i=1}^{MP^2} w_i \left( \sum_{m=1}^{B^3} a_{im} x_m^{(k)} \right) a_{ij} = \langle \boldsymbol{\alpha}_j, \mathcal{A} \mathbf{x}^{(k)} \rangle_W \\ x_j^{(k+1)} &= x_j^{(k)} + \mu_k \mathcal{A}_{jj}^{-1} (\mathcal{C}_j - (\mathcal{A} \mathbf{x}^{(k)})_j) \\ &= x_j^{(k)} + \mu_k \frac{\langle \boldsymbol{\alpha}_j, \mathbf{b} - \mathcal{A} \mathbf{x}^{(k)} \rangle_W}{\|\boldsymbol{\alpha}_j\|_W^2}. \end{aligned} \quad (37)$$

Interestingly, if the current image residual  $(\mathbf{b} - \mathcal{A} \mathbf{x}^{(k)})$  is  $W$ -orthogonal to  $\boldsymbol{\alpha}_j$ , that is, changing the coefficient of the  $j$ -th basis function does not affect the residual, then, as expected, the coefficient of the  $j$ -th basis function is not changed.

In EM, we are used to formulate SIRT as

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mu_k \sum_{i=1}^{MP^2} w_i \frac{b_i - \langle \mathbf{a}_i, \mathbf{x}^{(k)} \rangle}{\|\mathbf{a}_i\|^2} \mathbf{a}_i. \quad (38)$$

In the following, let us show that both formulations (see (37) and (38)) are not equivalent. Actually SIRT is not a single algorithm but a full family of reconstruction algorithms [84]; each one provides a different insight into the reconstruction process. If we look at a particular basis function  $j$  in (38), we have

$$\begin{aligned} x_j^{(k+1)} &= x_j^{(k)} + \mu_k \sum_{i=1}^{MP^2} w_i \frac{b_i - \langle \mathbf{a}_i, \mathbf{x}^{(k)} \rangle}{\|\mathbf{a}_i\|^2} a_{ij} \\ &= x_j^{(k)} + \mu_k \sum_{i=1}^{MP^2} \frac{w_i}{\|\mathbf{a}_i\|^2} (b_i - \langle \mathbf{a}_i, \mathbf{x}^{(k)} \rangle) a_{ij}. \end{aligned} \quad (39)$$

Let us now work on (37):

$$\begin{aligned} x_j^{(k+1)} &= x_j^{(k)} + \mu_k \mathcal{A}_{jj}^{-1} \left( \mathcal{B}_j - \left( \mathcal{A} \mathbf{x}^{(k)} \right)_j \right) = x_j^{(k)} + \mu_k \\ &\cdot \frac{1}{\|\boldsymbol{\alpha}_j\|_W^2} \left( \sum_{i=1}^{MP^2} w_i b_i a_{ij} - \sum_{i=1}^{MP^2} w_i \left( \sum_{m=1}^{B^3} a_{im} x_m^{(k)} \right) a_{ij} \right) \quad (40) \\ &= x_j^{(k)} + \mu_k \frac{1}{\|\boldsymbol{\alpha}_j\|_W^2} \sum_{i=1}^{MP^2} w_i \left( b_i - \langle \mathbf{a}_i, \mathbf{x}^{(k)} \rangle \right) a_{ij}. \end{aligned}$$

The algorithms in (39) and (40) are clearly different and, however, both are called SIRT and both belong to the SIRT family of algorithms. Any algorithm of the form

$$x_j^{(k+1)} = x_j^{(k)} + \mu_k \frac{1}{\gamma_j} \sum_{i=1}^{MP^2} \frac{w_i}{\rho_i} \left( b_i - \langle \mathbf{a}_i, \mathbf{x}^{(k)} \rangle \right) a_{ij} \quad (41)$$

for suitable  $\gamma_j$  and  $\rho_i$  numbers is also considered a SIRT algorithm [84]. A particular class for which convergence has been proven [84] is for

$$\begin{aligned} \gamma_j &= \sum_{i=1}^{MP^2} w_i |a_{ij}|^\alpha \\ \rho_i &= \sum_{j=1}^{B^3} |a_{ij}|^{2-\alpha} \quad (42) \end{aligned}$$

$$0 < \alpha \leq 2, \quad 0 < \mu_k < 2.$$

The Jacobi iteration resulting in (40) corresponds to  $\alpha = 2$ ; the typical EM SIRT (see (39)) corresponds to  $\alpha = 0$  and is the one implemented in Xmipp [85, 86] and TomoJ [87]. However, ASTRA toolbox [88], also used in EM, has a SIRT algorithm with  $\alpha = 1$ .

The case where  $\alpha = 0$  (see (38)) makes an interesting connection to the theory of feasibility problems [89–91]. Consider the hyperplane defined by all volumes compatible with the  $i$ -th experimental measurement:

$$H_i = \left\{ \mathbf{x} \in \mathbb{R}^{B^3} \mid \langle \mathbf{a}_i, \mathbf{x} \rangle = b_i \right\}. \quad (43)$$

Solving the tomographic problem amounts to finding a volume  $\mathbf{x}$  such that it is compatible with all measurements (note that, in this formulation, we are disregarding the effect of noise which makes such an intersection be empty):

$$\mathbf{x} \in H_1 \cap H_2 \cap \dots \cap H_{MP^2}. \quad (44)$$

Given a volume  $\mathbf{x}^{(k)}$ , we may orthogonally project it onto the hyperplane given by the  $i$ -th measurement by

$$\text{Proj}_{H_i} \left\{ \mathbf{x}^{(k)} \right\} = \mathbf{x}^{(k)} + \frac{b_i - \langle \mathbf{a}_i, \mathbf{x}^{(k)} \rangle}{\|\mathbf{a}_i\|^2} \mathbf{a}_i. \quad (45)$$

In this way, we may rewrite the SIRT iteration with  $\alpha = 0$  as

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mu_k \sum_{i=1}^{MP^2} w_i \left( \text{Proj}_{H_i} \left\{ \mathbf{x}^{(k)} \right\} - \mathbf{x}^{(k)} \right). \quad (46)$$

That is, at every iteration, we update the volume with a weighted sum of the orthogonal projections of the current solution onto the set of hyperplanes defined by the experimental measurements. The relaxation factor  $\mu_k$  (normally chosen between 0 and 1, although there are convergence theorems for values between 0 and 2) could be understood as how much we go from our current position  $\mathbf{x}^{(k)}$  to the desired position  $\text{Proj}_{H_i} \left\{ \mathbf{x}^{(k)} \right\}$ . In a noiseless case, we are rather certain about the update and may set  $\mu_k = 1$ . In a noisy environment, we may be more conservative and use a smaller relaxation factor reflecting our distrust in the experimental measurements.

Actually, we may update our estimate of the current solution after a single hyperplane projection; that is, we do not have to wait to “see” all measurements at the same time but we may update the volume just after seeing each pixel value:

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mu_k w_{i(k)} \left( \text{Proj}_{H_{i(k)}} \left\{ \mathbf{x}^{(k)} \right\} - \mathbf{x}^{(k)} \right). \quad (47)$$

The index  $i(k)$  will go over all the experimental measurements. This scheme is known as ART (algebraic reconstruction technique) and the difference between SIRT and ART is the same as the difference between a Jacobi update (SIRT) in the solution of a linear equation system and a Gauss-Seidel update (ART). In the ART case, we update the volume as soon as we have new information available, while in the SIRT case, we update the volume with a consensus of all the information available. This property of ART can be exploited to use ART in a streaming mode as data is acquired and stop data acquisition as soon as the reconstruction achieves a specific criterion (this could minimize beam damage, for instance) [92]. ART converges much faster than SIRT, although it tends to produce little bit noisier reconstructions. This problem can be alleviated through the choice of a relaxation factor that decreases over time. A trade-off between updating after seeing every pixel (ART) and updating after seeing all pixels (SIRT) is given by Block-ART or Simultaneous ART (SART). The volume is updated after seeing a small set of pixels  $S_k$  (normally all those in the same experimental image):

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mu_k \sum_{i \in S_k} w_i \left( \text{Proj}_{H_i} \left\{ \mathbf{x}^{(k)} \right\} - \mathbf{x}^{(k)} \right). \quad (48)$$

Kunz and Frangakis [73] studied how the order in which these blocks were chosen affected the reconstruction’s quality.

Additionally, we do not need to be restricted to orthogonal projections, and oblique projections can be undertaken. Given a symmetric, positive definite matrix  $G$ , the oblique projection onto the hyperplane  $H_i$  is defined as [93]

$$\text{Proj}_{H_i}^G \left\{ \mathbf{x}^{(k)} \right\} = \mathbf{x}^{(k)} + \frac{b_i - \langle \mathbf{a}_i, \mathbf{x}^{(k)} \rangle}{\|\mathbf{a}_i\|_{G^{-1}}^2} G^{-1} \mathbf{a}_i. \quad (49)$$

The study of these alternatives has given rise to a whole family of iterative algorithms (Cimmino, Component Averaging (CAV), Block-iterative CAV (BiCAV), Block-Simplified SART, iterative algorithms with Bregman projections, Block-iterative Underrelaxed Entropy Projections, Averaging Strings, etc.) [93–98]. A different variant of ART is a

multiplicative ART (MART), where the iterative step is multiplicative instead of additive. Although some of these algorithms have been tested in Electron Microscopy [40, 99], none of these more advanced variants have made their way into a massive adoption. For extensive reviews of the algebraic reconstruction techniques, the reader is referred to Gordon and Herman [100], Gordon [101], and Herman [102].

All SIRT algorithms can be written in a compact matrix form:

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mu_k CA^T R (\mathbf{b} - A\mathbf{x}^{(k)}), \quad (50)$$

with  $A$  being the system matrix ( $A$  projects the current estimate of the solution onto the image projection space, while  $A^T$  back-projects the residual into the volume space) and  $R$  and  $C$  being suitable diagonal matrices ( $C$  acting on the basis function space  $c_{jj} = 1/\gamma_j$ ; and  $R$  acting on the experimental measurements space  $r_{ii} = w_i/\rho_i$ ).

Stated in this matrix form, we may easily find another SIRT algorithm that is very popular in image processing: the Landweber iteration. A possible solution to the Weighted Least Squares problem,

$$\begin{aligned} \mathbf{x} &= \arg \min_{\mathbf{x}} E(\mathbf{x}) = \frac{1}{2} \|\mathbf{Ax} - \mathbf{b}\|_W^2 \\ &= \arg \min_{\mathbf{x}} \frac{1}{2} (\mathbf{Ax} - \mathbf{b})^T W (\mathbf{Ax} - \mathbf{b}) \\ &= \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{b}\|_W^2 + \frac{1}{2} \mathbf{x}^T A^T W A \mathbf{x} - \mathbf{x}^T A^T W \mathbf{b}, \end{aligned} \quad (51)$$

is to use gradient descent iterations:

$$\begin{aligned} \mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} - \mu_k \nabla E(\mathbf{x}^{(k)}) \\ &= \mathbf{x}^{(k)} - \mu_k (A^T W A \mathbf{x} - A^T W \mathbf{b}) \\ &= \mathbf{x}^{(k)} + \mu_k A^T W (\mathbf{b} - A\mathbf{x}). \end{aligned} \quad (52)$$

This latter iteration is called Landweber iteration and it can be easily recognized that it fits into the generic SIRT form of (50).

Another interesting variant of this family of traditional algorithms is the possibility to use unmatched projectors, for example, a relatively complicated forward operator,  $A$ , including the projection geometry and multiple defoci effects and a simple backward operator,  $B$ , considering only the projection geometry [103]:

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mu_k CB^T R (\mathbf{b} - A\mathbf{x}). \quad (53)$$

The convergence of this kind of algorithms depends on the eigenvalues of the matrix  $I - CB^T RA$  and the specific details can be found in [103]. In EM, this strategy has been implemented in Xmipp (the forward projection considers geometry and CTF, but the backward projection only considers geometry). Although it was shown to successfully recover the underlying 3D structure, this method has not been further pursued because of the general problem of most iterative

algorithms: their computational speed compared to direct Fourier inversion algorithms.

In recent years, one of the most popular algorithms for solving a linear equation system is Conjugate Gradient [59, 104–106]. The success of this algorithm is demonstrated in its ability to converge to the local solution in  $B^3$  steps (as many steps as the dimension of the solution being sought). Note that this number is finite and much smaller than the number of steps of ART or SIRT, whose convergence theorems are on the limit as the number of iterations goes to infinity. Given the equation system

$$\mathcal{A}\mathbf{x} = \mathcal{B}, \quad (54)$$

the trick is to use a set of directions  $\{\mathbf{p}_k\}$  which is orthogonal with respect to the inner product induced by  $\mathcal{A}$  ( $k \neq k' \Rightarrow \langle \mathbf{p}_k, \mathbf{p}_{k'} \rangle_{\mathcal{A}} = 0$ ; these directions are said to be conjugate with respect to  $\mathcal{A}$ ). Then the solution sought can be written in the form

$$\mathbf{x}^{(k)} = \sum_{k'=0}^k \frac{\langle \mathbf{p}_{k'}, \mathcal{B} \rangle}{\|\mathbf{p}_{k'}\|_{\mathcal{A}}^2} \mathbf{p}_{k'}. \quad (55)$$

The set of conjugate directions is also iteratively constructed as

$$\mathbf{p}_{k+1} = (\mathcal{B} - \mathcal{A}\mathbf{x}^{(k)}) - \sum_{k'=0}^k \frac{\langle \mathbf{p}_{k'}, \mathcal{B} - \mathcal{A}\mathbf{x}^{(k)} \rangle}{\|\mathbf{p}_{k'}\|_{\mathcal{A}}^2} \mathbf{p}_{k'}. \quad (56)$$

Actually, these summations can be nicely reorganized so that in real implementation we only need to keep a vector for the current solution, a vector with the residuals, and a vector with the current conjugate direction [104].

In the context of EM, the Conjugate Gradient was used by Chen and Förster [59] in Fourier space. However, there are a number of variants to the basic Conjugate Gradient algorithm (Conjugate Residuals, Biconjugate Gradient, Stabilized Biconjugate Gradient, Lanczos method, Generalized Minimal Residuals (GMINRES), Bi-Lanczos, Conjugate Gradient Squared, Quasi Minimal Residuals, etc.; most of them belong to a family of algorithms called Krylov subspace algorithms [107, 108]), none of which have been tested in EM.

In practice, none of these iterative algorithms are run to convergence. Instead, the algorithms are typically run for a fixed number of iterations (typically  $M$  or  $2M$  in the case of Block ART, 20 for CG, and 100–150 in the case of SIRT). How deep these algorithms have gone in the objective function landscape depends on the conditioning number of the equation system:

$$\kappa(\mathcal{A}) = \frac{\sigma_{\max}}{\sigma_{\min}}, \quad (57)$$

where  $\sigma_{\max}$  and  $\sigma_{\min}$  are the maximum and minimum singular values of matrix  $\mathcal{A}$ . If this ratio is close to 1, iterative algorithms will quickly converge to the minimum of the error function. If the ratio is much larger than 1, then the problem is said to be ill-posed (small perturbations in the  $\mathcal{B}$  vector may translate into large variations in the solution vector  $\mathbf{x}$ ) and

the convergence speed of iterative algorithms is slow. In case of solving normal equations, we have

$$\kappa(\mathcal{A}) = \kappa(A^T A) = \kappa^2(A). \quad (58)$$

That is, the ill-posed character of matrix  $A$  is worsened. For this reason, in the theory of linear equation solving, it is customary to use a preconditioning matrix  $P$  so that we do not solve the problem

$$\mathcal{A}\mathbf{x} = \mathcal{C} \quad (59)$$

but

$$P^{-1}\mathcal{A}\mathbf{x} = P^{-1}\mathcal{C}. \quad (60)$$

The preconditioning  $P$  matrix is chosen in such a way that

$$\kappa(P^{-1}\mathcal{A}) < \kappa(\mathcal{A}). \quad (61)$$

Although several preconditioners have been tested in different kinds of tomographic setups (computerized tomography [109], optical diffusion tomography [110], positron emission tomography [111], electrical capacitance tomography [112], acoustic waveform tomography [113], etc.), in EM, there has not been any attempt to use any preconditioning, although this is an issue the community is aware of [114].

## 5. Constrained 3D Reconstruction

The feasibility problem of (44) can be complemented with some additional constraints representing a priori knowledge about the reconstruction. We may do so by imposing the fact that the volume also belongs to some convex set (a set is convex if, for any two volumes in this set,  $\mathbf{x}_1$  and  $\mathbf{x}_2$ , the linear combination  $(1 - \lambda)\mathbf{x}_1 + \lambda\mathbf{x}_2$ , with  $0 \leq \lambda \leq 1$ , also belongs to that set; the set of all nonnegative volumes is convex as well as the set of all volumes defined within a mask, the set of all volumes with a given symmetry, the set of all volumes bandlimited to a given frequency, etc. [64]). Given a collection of convex sets representing our a priori knowledge about the reconstructed particle,  $C_1, C_2, \dots, C_K$ , we may try to find a feasible volume at the intersection

$$\mathbf{x} \in H_1 \cap H_2 \cap \dots \cap H_{MP^2} \cap C_1 \cap C_2 \cap \dots \cap C_K. \quad (62)$$

This problem has been extensively explored in the EM community under the name Projection Onto Convex Sets (POCS) and the interested reader is referred to Carazo and Carrascosa [115, 116], Carazo [117], García et al. [31], Sorzano et al. [64], and Deng et al. [118]. The idea is to alternate between projections onto the subspaces defined by the experimental measurements ( $H_i$ ) and projections onto the convex sets (the projector onto the set of nonnegative volumes simply sets all negative values to 0; the projector onto the set of volumes defined within a mask applies that mask to the current solution; the projector onto the set of symmetric volumes symmetrizes the current solution; the projector onto the set of bandlimited volumes applies a low-pass filter to the current solution; etc.). In the signal processing community,

POCS algorithms have been generalized by algorithms using the so-called Proximity operators. Despite being out of the scope of this review, because these algorithms have not been introduced in EM, the interested reader may follow the review of Combettes and Pesquet [119] and the references therein.

Landweber iterations can also be set in a constrained setup. Let us assume that we have the a priori knowledge that  $\mathbf{x}$  is in a convex set  $C$ . Then, the solution of the constrained problem

$$\mathbf{x} = \arg \min_{\mathbf{x} \in C} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_W^2 \quad (63)$$

can be found with the iterative algorithm [119]:

$$\mathbf{x}^{(k+1)} = \text{Proj}_C \{ \mathbf{x}^{(k)} + \mu_k A^T W (\mathbf{b} - \mathbf{A}\mathbf{x}) \}. \quad (64)$$

As seen in the two examples above, one of the most interesting ideas of this constrained optimization is the possibility to alternate between the standard tomographic update ( $\mathbf{x}^{(k)} + \mu_k A^T W (\mathbf{b} - \mathbf{A}\mathbf{x})$ ) and the projection onto convex sets (POCS,  $\text{Proj}_C$ ).

These ideas of constrained optimization can be further extended to nondifferentiable, convex functions (the  $l_1$  norm of the residual is a function of this kind). Let us assume that we are minimizing a nondifferentiable, convex function  $\phi(\mathbf{x})$ :

$$\mathbf{x} = \arg \min \phi(\mathbf{x}). \quad (65)$$

At a differentiable point of  $\phi$ , a gradient descent iteration is perfectly suitable to find the minimum of the objective function:

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \mu_k \nabla \phi(\mathbf{x}^{(k)}). \quad (66)$$

The problem at a nondifferentiable point (which occurs normally at the frontiers of the intersection of convex sets) is that the gradient is not well defined. We may define instead the subgradient. A vector  $\mathbf{g}_k \in \mathbb{R}^{B^3}$  is a subgradient of  $\phi$  at  $\mathbf{x}^{(k)}$  if for any  $\mathbf{x}$  we have

$$\phi(\mathbf{x}) - \phi(\mathbf{x}^{(k)}) \geq \langle \mathbf{g}_k, \mathbf{x} - \mathbf{x}^{(k)} \rangle. \quad (67)$$

Intuitively, the gradient defines at differentiable points a unique hyperplane that is tangent to  $\phi$  and, since  $\phi$  is convex, this tangent plane is always below  $\phi$ . At nondifferentiable points, there are an infinite number of hyperplanes touching at  $(\mathbf{x}^{(k)}, \phi(\mathbf{x}^{(k)}))$  and below  $\phi$ . The normal vectors to these hyperplanes constitute the set of subgradients. The subgradient method iteration is then, at these nondifferentiable points [120, 121],

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \mu_k \mathbf{g}_k. \quad (68)$$

We may easily add convex constraints to this method obtaining constrained subgradient minimization (known as Projected Subgradient minimization) [122, 123]:

$$\mathbf{x}^{(k+1)} = \text{Proj}_C \{ \mathbf{x}^{(k)} - \mu_k \mathbf{g}_k \}. \quad (69)$$



Superiorization has been proposed as an alternative to Projected Subgradients [43, 124–126]. The idea of Superiorization is to steer the reconstruction iterations towards the constraints. One of the main differences between Projected Subgradient and Superiorization is that if there are several convex constraints, Projected Subgradient requires the projection onto the intersection of all of them, the feasible set, while Superiorization requires the sequential projection onto each of the constraints. Note that the intersection may be empty. For instance, the set of all volumes defined within a finitely supported mask and bandlimited is empty. Also, the projector onto the intersection set may be more complicated than the projector onto each one of the individual constraints. This limits the applicability of Projected Subgradient, while this is not a problem for Superiorization. Although these algorithms are available in the tomography community and are well characterized in other domains, none of them have made their way into EM.

A different approach to constrained optimization which has been actively explored in EM is by defining new equations that must be simultaneously solved along with the equation system coming from the measurements [64]. For instance, a mask can be easily expressed in terms of basis functions. Let us assume that there is no density coming from the molecule at a given location  $\mathbf{r}_0$ . Then, we may particularize the series expansion at  $\mathbf{r}_0$  (see (5)) and add the linear equation

$$V(\tilde{\mathbf{r}}_0) = \sum_j x_j b(\tilde{\mathbf{r}}_0 - \tilde{\mathbf{r}}_j) = 0. \quad (70)$$

In this way, we handled a priori information coming from masks, symmetry, total molecular mass, and nonnegativity. Adding this a priori information was certainly relevant when the number of projections was small. However, as the number of projections increased, the improvement by the a priori information was less noticeable, showing that most of the information was already present in the experimental dataset. Now, with increasing pursuit of high-resolution results, there is again room for exploring these extra sources of information available, although for the moment this line of research has not been resumed.

Iteratively steering algorithms tend to promote reconstructions with certain characteristics. For instance, [65, 71, 72, 127, 128] focused on the problem of reconstructing discrete valued objects (the reconstructed volume could only have a few, normally two (background and foreground), values) or objects with very few active voxels. They use projectors similar to those used in convex sets; for instance, bgART [71] defined

$$\text{Proj} \{x_j^{(k)}\} = \begin{cases} m_k & x_j^{(k)} < m_k + Ks_k \\ x_j^{(k)} & x_j^{(k)} \geq m_k + Ks_k, \end{cases} \quad (71)$$

where  $m_k$  is the estimated mean of the background gray values at iteration  $k$ ,  $s_k$  is their estimated standard deviation, and  $K$  is a user-selected multiple (normally a number between 3 and 6). The main difference between these iterative steering algorithms and those projecting onto convex sets is that the

former project onto a nonconvex set; additionally, the set onto which the reconstruction is projected changes from iteration to iteration. The convergence of this kind of algorithms was studied in [129]. Reference [75] presented an algorithm for Electron Tomography in which the steering is driven by a nonlinear diffusion denoising algorithm. After each reconstruction step, the reconstructed volume is denoised by applying a step of a nonlinear diffusion algorithm (the projector). For these steering algorithms to converge, the reconstruction algorithm must be “perturbation resilient” [124] and the perturbation must be “small enough” as not to destroy the work of the reconstruction algorithm.

## 6. Sparse 3D Reconstructions

Sparse representations has been one of the most active research fields in the area of image and signal processing in the last 10–15 years [130–132]. The idea is that natural images and objects have a representation in some appropriate space in which very few nonnull coefficients are needed. This space may be fixed (e.g., wavelet transform or discrete cosine transform and DCT) or computed ad hoc for a particular problem (dictionary based algorithms). Knowing that our object has a sparse representation helps the algorithm to concentrate the energy in a few coefficients, preventing the energy dispersion normally caused by noise. In this problem setup, vector norms different from the Euclidean are normally employed. In general, the  $l_p$  norm is used:

$$\|\mathbf{x}\|_p = \left( \sum_j |x_j|^p \right)^{1/p}. \quad (72)$$

Euclidean norm is obtained for  $p = 2$ , Manhattan norm is obtained for  $p = 1$ , and for  $p = 0$  the norm of a vector is simply the count of the number of nonzero coefficients in the vector (technically,  $l_0$  is not a norm because it does not fulfill the condition  $\|\lambda \mathbf{x}\|_0 = |\lambda| \|\mathbf{x}\|_0$ ). The goal of sparsity is to find a representation of  $\mathbf{x}$  such that

$$\begin{aligned} \mathbf{x} &= \arg \min \|\mathcal{D}\mathbf{x}\|_0 \\ \text{s.t.} \quad & D\mathbf{x} = \mathbf{x}, \end{aligned} \quad (73)$$

where  $D$  is the dictionary of elements available to represent  $\mathbf{x}$  (the wavelet, DCT, or ad hoc dictionary) and  $\mathcal{D}$  is the representation of  $\mathbf{x}$  in that dictionary. The aim is to have as few coefficients different from 0 as possible.

As stated above, sparse representations are mostly interesting in  $l_0$ -norms. However, having this norm in the objective function requires combinatorial optimization techniques, known to be NP-hard in computational complexity. Interestingly,  $l_p$  norms with  $0 < p < 2$  tend to promote sparse representations ( $\mathcal{D}$  has relatively few nonzero coefficients) and efficient algorithms have been developed in the recent years for these minimizations.

The first class of algorithms we will review involve an  $l_0$ -norm and are those solving any of these two related problems:

$$\begin{aligned} \mathbf{x} = \arg \min \quad & \|\mathbf{x}\|_0 \\ \text{s.t.} \quad & \|\boldsymbol{\mathcal{E}} - \mathcal{A}D\mathbf{x}\|_2^2 \leq \epsilon \end{aligned} \quad (74)$$

$$\begin{aligned} \text{or } \mathbf{x} = \arg \min \quad & \|\boldsymbol{\mathcal{E}} - \mathcal{A}D\mathbf{x}\|_2^2 \\ \text{s.t.} \quad & \|\mathbf{x}\|_0 \leq N. \end{aligned} \quad (75)$$

Note that our ‘‘tomographic dictionary’’ ( $\mathcal{A}D = A^T A D$ ) includes the standard dictionary representation,  $D$ , as well as information about the projection structure of the tomographic setup  $\mathcal{A} = A^T A$ . In the following, let us refer to this ‘‘tomographic dictionary’’ by  $\mathcal{D}$  and to its  $j$ -th column by  $\mathbf{d}_j$  (in the dictionary jargon, this would be the  $j$ -th atom and we will assume that there are  $J \ll B^3$  atoms in the dictionary). Many different algorithms exist to solve this problem [130]; none of them have been tested in EM. Among them, Matching Pursuit [133] and Orthogonal Matching Pursuit [134] are two of the most popular algorithms (actually the latter was tested in a tomography setting by researchers working on EM, although it was not applied to EM data [135]). For its simplicity, let us show the iterations in Matching Pursuit to illustrate the flavor of these algorithms and how they differ from the iterative algorithms presented so far.

Let  $\mathbf{x}^{(k)}$  represent our current reconstruction. In the first iteration, it will be  $\mathbf{x}^{(0)} = \mathbf{0}$ . We now choose the atom that is maximally aligned with the residual of our equation system,  $\mathbf{r}^{(k)}$  (note that  $\mathbf{r}^{(0)} = \boldsymbol{\mathcal{E}}$ ), calculate its coefficient, and update the residual. This process is iterated until the desired number of atoms is reached. The following algorithm is executed from  $k = 0$  to  $k = N - 1$ .

(i) *Step 1.* Seek the atom maximally aligned with this residual

$$j = \arg \max_{j'} \left| \langle \mathbf{r}^{(k)}, \mathbf{d}_{j'} \rangle \right|. \quad (76)$$

(ii) *Step 2.* Update the current solution with the projection of the residual onto this atom

$$\mathbf{x}_{j'}^{(k+1)} = \begin{cases} \mathbf{x}_{j'}^{(k)} & j' \neq j \\ \frac{\langle \mathbf{r}^{(k)}, \mathbf{d}_j \rangle}{\|\mathbf{d}_j\|^2} \mathbf{d}_j & j' = j. \end{cases} \quad (77)$$

(iii) *Step 3.* Update the residual

$$\mathbf{r}^{(k+1)} = \mathbf{r}^{(k)} - \mathbf{x}_j^{(k+1)} \mathbf{d}_j. \quad (78)$$

This algorithm is greedy and finds a suboptimal solution of the reconstruction problem (see (75)) [133]. However, it is very easy to implement. A significant improvement is provided by Orthogonal Matching Pursuit in which Step 2 is modified to update all coefficients (not just one) by orthogonally projecting  $\boldsymbol{\mathcal{E}}$  onto the subspace spanned by all the atoms employed so far.

The second class of algorithms substitute the  $l_0$ -norm by an  $l_1$ -norm:

$$\begin{aligned} \mathbf{x} = \arg \min \quad & \|\mathbf{x}\|_1 \\ \text{s.t.} \quad & \|\boldsymbol{\mathcal{E}} - \mathcal{A}D\mathbf{x}\|_2^2 \leq \epsilon \end{aligned} \quad (79)$$

$$\begin{aligned} \text{or } \mathbf{x} = \arg \min \quad & \|\boldsymbol{\mathcal{E}} - \mathcal{A}D\mathbf{x}\|_2^2 \\ \text{s.t.} \quad & \|\mathbf{x}\|_1 \leq \delta. \end{aligned} \quad (80)$$

This allows more efficient optimization algorithms to be employed. Classical algorithms are Least Absolute Shrinkage and Selection Operator (Lasso also known as Basis Pursuit), Iterative Reweighted Least Squares (IRLS), Iterative Shrinkage algorithms, Least Angle Regression (LARS), and any of their variants [130]. For their simplicity, Iterative Shrinkage algorithms have found their way into Electron Microscopy [136, 137] (for a theoretical background, you may also see [138, 139]). However, the EM versions are aimed at solving a regularized problem,

$$\mathbf{x} = \arg \min \|\boldsymbol{\mathcal{E}} - \mathcal{A}D\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1, \quad (81)$$

rather than the constrained problems above. In their most simplified formulation, these algorithms alternate between a standard reconstruction update (let us, for instance, take a SIRT step) and a soft-thresholding step in some suitable, sparse space (e.g., the wavelet space). The rationale is relatively simple and resembles the line of thought of the iterative steering algorithms. After applying a standard reconstruction step, the solution is steered towards a sparse solution by setting to 0 all small coefficients in some space known to promote sparsity (as the wavelet space). A basic iterative step would be

$$\begin{aligned} \mathbf{x}^{(k+1/2)} &= \mathbf{x}^{(k)} + \mu_k C A^T R (\mathbf{b} - A \mathbf{x}^{(k)}) \\ \mathbf{x}^{(k+1)} &= \sum_{j=1}^J S_\lambda (\langle \mathbf{x}^{(k+1/2)}, \mathbf{d}_j \rangle) \mathbf{d}_j, \end{aligned} \quad (82)$$

where  $S_\lambda(x)$  is the soft-thresholding function:

$$S_\lambda(x) = \begin{cases} x + \frac{\lambda}{2} & x < -\frac{\lambda}{2} \\ 0 & |x| \leq \frac{\lambda}{2} \\ x - \frac{\lambda}{2} & x > \frac{\lambda}{2}. \end{cases} \quad (83)$$

In the above iteration, we have assumed an orthonormal dictionary or transformation, but equivalent formulas can be found for nonorthonormal dictionaries.

The use of Lagrangian augmented objective functions of the form

$$\mathbf{x} = \arg \min_{\mathbf{x}} F(\mathbf{b} - A\mathbf{x}) + \lambda G(\mathbf{x}) \quad (84)$$

is a widespread technique in data analysis. The first term,  $F$ , is called the data fidelity term, while the second term,

$G$ , is a regularizer. Many different schemes respond to this structure. For instance, a typical regularized Weighted Least Squares problem would be given by Tikhonov quadratic regularization (also known as ridge regression) involving some “preprocessing” matrix  $L$ ,

$$\mathbf{x} = \arg \min_{\mathbf{x}} \|\mathbf{b} - A\mathbf{x}\|_W^2 + \lambda \|\mathbf{Lx}\|_2^2, \quad (85)$$

whose normal equations would be

$$(A^T W A + \lambda L^T L) \mathbf{x} = A^T W \mathbf{b}. \quad (86)$$

Assuming Gaussian errors for the measurements and a Gaussian prior for the volume coefficients, we would have  $L = \Sigma_{\mathbf{x}}^{-1/2}$ .

$$(A^T W A + \lambda \Sigma_{\mathbf{x}}^{-1}) \mathbf{x} = A^T W \mathbf{b}. \quad (87)$$

If we think that  $\mathbf{x}$  and  $\mathbf{b}$  are in Fourier space, these would be the normal equations associated with the same Bayesian problem that Relion is solving [7].  $W$  contains the CTF information,  $A$  the data collection geometry (including the probability of each projection having a given projection direction and shift), and  $\Sigma_{\mathbf{x}}$  our prior about the energy of the coefficients in Fourier space. A difference between Relion and these normal equations is that Relion reestimates the prior after each iteration.

We might have gone one step further to the generalized Tikhonov regularization,

$$\mathbf{x} = \arg \min_{\mathbf{x}} \|\mathbf{b} - A\mathbf{x}\|_W^2 + \lambda \|L(\mathbf{x} - \mathbf{x}_0)\|_Q^2, \quad (88)$$

whose normal equations are

$$(A^T W A + \lambda L^T Q L) \mathbf{x} = A^T W \mathbf{b} + Q L \mathbf{x}_0. \quad (89)$$

If we assume independence between the preprocessed coefficients, most maximum a posteriori (MAP) algorithms could be written using

$$G(\mathbf{Lx}) = \sum_j \phi((\mathbf{Lx})_j), \quad (90)$$

where  $\phi(x)$  is the a priori, negative log likelihood of observing a value of  $x$  ( $\phi(x) = -\log p_x(x)$ ). Assume that a Gaussian distribution of the  $\mathbf{x}$  coefficients results in  $L = \Sigma_{\mathbf{x}}^{-1/2}$  and  $\phi(x) = x^2 + C$  ( $C$  is a constant that does not affect the MAP optimization). Assume that a centered Laplace distribution of standard deviation  $\sigma$  ( $p_x(x) = (1/\sqrt{2}\sigma)e^{-2|x|/\sqrt{2}\sigma}$ ) results in  $L = (\sqrt{2}/\sigma)I$  and  $\phi(x) = |x| + C$  with  $G$  being an  $l_1$ -norm (instead of an  $l_2$ -norm as in the case of the Gaussian prior). In an EM setup, Moriya et al. [74] assumed a Median Root Prior which favors locally monotonic reconstructions.

If we take  $L$  to be a volume derivative operator and take  $G$  to be the  $l_1$ -norm of this derivative (that is, a Laplace prior on the derivative coefficients),

$$G(\mathbf{x}) = \|\mathbf{Lx}\|_1, \quad (91)$$

then we have a total variation regularization (also named as TVL1). The problem with the  $l_1$ -norm is that it cannot be differentiated and sometimes it is substituted by

$$G(\mathbf{x}) = \int_{\mathbb{R}^3} \sqrt{\left(\frac{\partial V(\mathbf{r})}{\partial x}\right)^2 + \left(\frac{\partial V(\mathbf{r})}{\partial y}\right)^2 + \left(\frac{\partial V(\mathbf{r})}{\partial z}\right)^2} d\mathbf{r}, \quad (92)$$

where  $V(\mathbf{r})$  is the volume reconstructed using the  $\mathbf{x}$  coefficients (see (5)). This was explored for EM by Zhu et al. [62], Aganj et al. [63], Li et al. [66], Goris et al. [57], and Zhuge et al. [76].

The approach of Albarqouni et al. [140] also falls into this regularization category. The function  $G$  is, in their case, a function borrowed from robust statistics, the Huber function:

$$\phi_{\tau}(x) = \begin{cases} 0.5x^2 & |x| \leq \tau \\ \tau|x| - 0.5\tau^2 & |x| > \tau. \end{cases} \quad (93)$$

This function is half-way between a quadratic and a linear penalization ( $\tau$  controls the switch between these two behaviors). For low values of the derivatives, the function behaves as a quadratic term and for high values it behaves as a linear term. The idea is not to let large derivatives dominate the optimization process.

We can see that the above regularized problems include either some a priori knowledge on the volume ( $\mathbf{x}$ ) or a property of that volume ( $\mathbf{Lx}$ , normally its derivative). However, we could include both through a new algorithm (alternating-direction method of multipliers, ADMM) recently introduced for EM [141]:

$$\begin{aligned} \mathbf{x} = \arg \min \quad & \|\mathbf{b} - A\mathbf{x}\|_2^2 + \lambda_1 G_1(\mathbf{x}) + \lambda_2 G_2(\mathbf{u}) \\ \text{s.t.} \quad & \mathbf{u} = \mathbf{Lx}. \end{aligned} \quad (94)$$

The actual problem being solved is an augmented Lagrangian:

$$\begin{aligned} \mathbf{x} = \arg \min_{\mathbf{x}, \mathbf{u}, \boldsymbol{\alpha}} \mathcal{L}_{\mu}(\mathbf{x}, \mathbf{u}, \boldsymbol{\alpha}) \\ = \|\mathbf{b} - A\mathbf{x}\|_2^2 + \lambda_1 G_1(\mathbf{x}) + \lambda_2 G_2(\mathbf{u}) + \boldsymbol{\alpha}^T (\mathbf{Lx} - \mathbf{u}) \\ + \frac{\mu}{2} \|\mathbf{Lx} - \mathbf{u}\|^2, \end{aligned} \quad (95)$$

where  $\boldsymbol{\alpha}$  is the set of Lagrangian multipliers. The ADMM proceeds iteratively as

$$\begin{aligned} \mathbf{x}^{(k+1)} &= \arg \min_{\mathbf{x}} \mathcal{L}_{\mu}(\mathbf{x}, \mathbf{u}^{(k)}, \boldsymbol{\alpha}^{(k)}) \\ \mathbf{u}^{(k+1)} &= \arg \min_{\mathbf{u}} \mathcal{L}_{\mu}(\mathbf{x}^{(k+1)}, \mathbf{u}, \boldsymbol{\alpha}^{(k)}) \\ \boldsymbol{\alpha}^{(k+1)} &= \boldsymbol{\alpha}^{(k)} + \mu (\mathbf{Lx}^{(k+1)} - \mathbf{u}^{(k+1)}). \end{aligned} \quad (96)$$

Related to these sparse reconstruction problems is the one of compressed sensing. The idea is to perform a 2D or 3D reconstruction problem starting, not from a full image, but

from a “few” incoherent points from the projection images [142–144]. The trick is that the a priori knowledge that the solution is sparse allows reconstructing the full image or volume from fewer points than Nyquist theorem requires (there is a lower limit on the number of measurements required). This theory has not been explored in biological EM applications, but it has been studied in material science EM [145–150], especially when Scanning TEM (STEM) is used [151]. For biological samples, Energy Filtered TEM (EFTEM) is a clear candidate to benefit from a compressed sensing acquisition [149].

## 7. Conclusions

The field of iterative reconstruction algorithms has been very much studied, particularly in its application to Electron Microscopy data, as we have shown in this review. The 3D reconstruction problem is no longer seen as a bottleneck in Single Particle Analysis (this is a technique in which many single particles, assumed to come from an homogeneous population but at different angular orientations, are combined into a single 3D map; Jonic et al. [47]), and the few attempts to complement the data with a priori knowledge are not in widespread use, probably due to their higher computational cost. However, as we approach resolutions of 2–3 Å, it may be worthy to retake this line of research as a way to increase the reconstruction resolution. In Electron Tomography, the situation is different due to the reconstruction artifacts induced by the maximum tilt angle limitation, the presence of gold beads, and the low number of projections. Additionally, the particular data collection geometry (normally, single tilt axis) favors the adoption of a pure 2D reconstruction approach that reduces the reconstruction problem in one dimension, implying a great reduction in the computational cost. In this field, iterative algorithms capable of incorporating a priori information are still a very active field of research. Traditional iterative algorithms as ART or SIRT still dominate the “market.” However, very powerful reconstruction algorithms with more modern approaches incorporating convex, nonconvex, and sparsity constraints are continuously appearing and most likely, in the near future, one of these algorithms will eventually become the standard de facto.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

The authors would like to acknowledge economical support from Comunidad de Madrid through Grant CAM (S2010/BMD-2305), the NSF through Grant I114901, the Spanish Ministry of Economy and Competitiveness through Grants AIC-A-2011-0638, BIO2010-16566, and BIO2013-44647-R, Instituto de Salud Carlos III (PT13/0001/0009), and Fundación General CSIC (Programa ComFuturo). This work was funded by Instruct, part of the European Strategy Forum on Research Infrastructures (ESFRI), and supported by national member subscriptions.

## References

- [1] M. Eisenstein, “The field that came in from the cold,” *Nature Methods*, vol. 13, no. 1, pp. 19–22, 2015.
- [2] R. M. Glaeser, “How good can cryo-EM become?” *Nature Methods*, vol. 13, no. 1, pp. 28–32, 2015.
- [3] E. Nogales, “The development of cryo-EM into a mainstream structural biology technique,” *Nature Methods*, vol. 13, no. 1, pp. 24–27, 2015.
- [4] J. M. Carazo, C. O. S. Sorzano, J. Otón, R. Marabini, and J. Vargas, “Three-dimensional reconstruction methods in Single Particle Analysis from transmission electron microscopy data,” *Archives of Biochemistry and Biophysics*, vol. 581, pp. 39–48, 2015.
- [5] A. Hosseinizadeh, A. Dashti, P. Schwander, R. Fung, and A. Ourmazd, “Single-particle structure determination by X-ray free-electron lasers: Possibilities and challenges,” *Structural Dynamics*, vol. 2, no. 4, Article ID 041601, 2015.
- [6] P. A. Penczek, R. Renka, and H. Schomberg, “Gridding-based direct Fourier inversion of the three-dimensional ray transform,” *Journal of the Optical Society of America A. Optics, Image Science, and Vision*, vol. 21, no. 4, pp. 499–509, 2004.
- [7] S. H. W. Scheres, “RELION: implementation of a Bayesian approach to cryo-EM structure determination,” *Journal of Structural Biology*, vol. 180, no. 3, pp. 519–530, 2012.
- [8] V. Abrishami, J. R. Bilbao-Castro, J. Vargas, R. Marabini, J. M. Carazo, and C. O. S. Sorzano, “A fast iterative convolution weighting approach for gridding-based direct Fourier three-dimensional reconstruction with correction for the contrast transfer function,” *Ultramicroscopy*, vol. 157, pp. 79–87, 2015.
- [9] B. P. Medoff, W. R. Brody, M. Nassi, and A. Macovski, “Iterative convolution backprojection algorithms for image reconstruction from limited data,” *Journal of the Optical Society of America*, vol. 73, no. 11, pp. 1493–1500, 1983.
- [10] G. Harauz and M. van Heel, “Exact filters for general geometry three dimensional reconstruction,” *Optik*, vol. 73, pp. 146–156, 1986.
- [11] M. Radermacher, “Weighted back-projection methods,” in *Electron Tomograph*, J. Frank, Ed., pp. 91–115, Plenum, 1992.
- [12] S. Horbelt, M. Liebling, and M. Unser, “Discretization of the radon transform and of its inverse by spline convolutions,” *IEEE Transactions on Medical Imaging*, vol. 21, no. 4, pp. 363–376, 2002.
- [13] A. Lawrence, J. C. Bouwer, G. Perkins, and M. H. Ellisman, “Transform-based backprojection for volume reconstruction of large format electron microscope tilt series,” *Journal of Structural Biology*, vol. 154, no. 2, pp. 144–167, 2006.
- [14] I. G. Kazantsev, J. Klukowska, G. T. Herman, and L. Cernetic, “Fully three-dimensional defocus-gradient corrected backprojection in cryoelectron microscopy,” *Ultramicroscopy*, vol. 110, no. 9, pp. 1128–1142, 2010.
- [15] F. Vázquez, E. M. Garzón, and J. J. Fernández, “A matrix approach to tomographic reconstruction and its implementation on GPUs,” *Journal of Structural Biology*, vol. 170, no. 1, pp. 146–151, 2010.
- [16] D. M. Pelt and K. J. Batenburg, “Fast tomographic reconstruction from limited data using artificial neural networks,” *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5238–5251, 2013.
- [17] D. M. Pelt and K. J. Batenburg, “Improving filtered back-projection reconstruction by data-dependent filtering,” *IEEE*

- Transactions on Image Processing*, vol. 23, no. 11, pp. 4750–4762, 2014.
- [18] E. Bladt, D. M. Pelt, S. Bals, and K. J. Batenburg, “Electron tomography based on highly limited data using a neural network reconstruction technique,” *Ultramicroscopy*, vol. 158, pp. 81–88, 2015.
- [19] R. Gordon, R. Bender, and G. T. Herman, “Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography,” *Journal of Theoretical Biology*, vol. 29, no. 3, pp. 471–481, 1970.
- [20] S. H. Bellman, R. Bender, R. Gordon, and J. E. Rowe Jr., “ART is science being a defense of algebraic reconstruction techniques for three-dimensional electron microscopy,” *Journal of Theoretical Biology*, vol. 32, no. 1, pp. 205–IN2, 1971.
- [21] R. A. Crowther and A. Klug, “ART and science or conditions for three-dimensional reconstruction from electron microscope images,” *Journal of Theoretical Biology*, vol. 32, no. 1, pp. 199–203, 1971.
- [22] P. Gilbert, “Iterative methods for the three-dimensional reconstruction of an object from projections,” *Journal of Theoretical Biology*, vol. 36, no. 1, pp. 105–117, 1972.
- [23] G. T. Herman, A. Lent, and S. W. Rowland, “ART: mathematics and applications. a report on the mathematical foundations and on the applicability to real data of the algebraic reconstruction techniques,” *Journal of Theoretical Biology*, vol. 42, pp. 1–3, 1973.
- [24] A. V. Lakshminarayanan and A. Lent, “Methods of least squares and SIRT in reconstruction,” *Journal of Theoretical Biology*, vol. 76, no. 3, pp. 267–295, 1979.
- [25] A. H. Andersen and A. C. Kak, “Simultaneous Algebraic Reconstruction Technique (SART): a superior implementation of the art algorithm,” *Ultrasonic Imaging*, vol. 6, no. 1, pp. 81–94, 1984.
- [26] R. M. Lewitt, “Multidimensional digital image representations using generalized kaiser-bessel window functions,” *Journal of the Optical Society of America A: Optics and Image Science, and Vision*, vol. 7, no. 10, pp. 1834–1846, 1990.
- [27] P. Penczek, M. Radermacher, and J. Frank, “Three-dimensional reconstruction of single particles embedded in ice,” *Ultramicroscopy*, vol. 40, no. 1, pp. 33–53, 1992.
- [28] A. J. Koster, M. B. Braunfeld, J. C. Fung et al., “Towards automatic three-dimensional imaging of large biological structures using intermediate voltage electron microscopy,” *Microscopy Society of America Bulletin*, vol. 23, no. 2, pp. 176–188, 1993.
- [29] H. Guan and R. Gordon, “A projection access order for speedy convergence of ART (algebraic reconstruction technique): a multilevel scheme for computed tomography,” *Physics in Medicine & Biology*, vol. 39, pp. 2005–2022, 1994.
- [30] S. Matej and R. M. Lewitt, “Efficient 3D Grids for Image Reconstruction Using Spherically-Symmetric Volume Elements,” *IEEE Transactions on Nuclear Science*, vol. 42, no. 4, pp. 1361–1370, 1995.
- [31] I. García, J. Roca, J. Sanjurjo, J. M. Carazo, and E. L. Zapata, “Implementation and experimental evaluation of the constrained ART algorithm on a multicomputer system,” *Signal Processing*, vol. 51, no. 1, pp. 69–76, 1996.
- [32] S. Matej and R. M. Lewitt, “Practical considerations for 3-D image reconstruction using spherically symmetric volume elements,” *IEEE Transactions on Medical Imaging*, vol. 15, no. 1, pp. 68–78, 1996.
- [33] I. Garca, P. M. Ortigosa, L. G. Casado, G. T. Herman, and S. Matej, “Multidimensional optimization in image reconstruction from projections,” in *Developments in Global Optimization*, I. M. Bomze, T. Csendes, R. Horst, and P. M. Pardalos, Eds., Nonconvex Optimization and Applications, pp. 289–300, Kluwer Academic Pub., 1997.
- [34] 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 backprojection,” *Journal of Structural Biology*, vol. 120, no. 3, pp. 363–371, 1997.
- [35] G. T. Herman, “Algebraic reconstruction techniques in medical imaging,” in *Medical Imaging, Systems Techniques and Applications: Computational Techniques*, T. C. Leondes, Ed., pp. 1–42, Gordon and Breach Science Publishers, Amsterdam, Netherlands, 1998.
- [36] R. Marabini, G. T. Herman, and J. M. Carazo, “3D reconstruction in electron microscopy using ART with smooth spherically symmetric volume elements (blobs),” *Ultramicroscopy*, vol. 72, no. 1-2, pp. 53–65, 1998.
- [37] R. Marabini, G. T. Herman, and J. M. Carazo, “Fully Three-Dimensional Reconstruction in Electron Microscopy,” in *Computational Radiology and Imaging*, vol. 110 of *The IMA Volumes in Mathematics and its Applications*, pp. 251–281, Springer, New York, NY, USA, 1999.
- [38] C. O. S. Sorzano, R. Marabini, N. Boisset et al., “The effect of overabundant projection directions on 3D reconstruction algorithms,” *Journal of Structural Biology*, vol. 133, no. 2-3, pp. 108–118, 2001.
- [39] C. O. S. Sorzano, R. Marabini, G. T. Herman, and J. M. Carazo, “Multiobjective algorithm parameter optimization using multivariate statistics in three-dimensional electron microscopy reconstruction,” *Pattern Recognition*, vol. 38, no. 12, pp. 2587–2601, 2005.
- [40] J. R. Bilbao-Castro, J. M. Carazo, I. García, and J. J. Fernández, “Parallelization of reconstruction algorithms in three-dimensional electron microscopy,” *Applied Mathematical Modelling*, vol. 30, no. 8, pp. 688–701, 2006.
- [41] J. Tong, I. Arslan, and P. Midgley, “A novel dual-axis iterative algorithm for electron tomography,” *Journal of Structural Biology*, vol. 153, no. 1, pp. 55–63, 2006.
- [42] J. J. Fernández, “High performance computing in structural determination by electron cryomicroscopy,” *Journal of Structural Biology*, vol. 164, no. 1, pp. 1–6, 2008.
- [43] G. T. Herman, E. Garduño, R. Davidi, and Y. Censor, “Superiorization: An optimization heuristic for medical physics,” *Medical Physics*, vol. 39, no. 9, pp. 5532–5546, 2012.
- [44] S. Lanzavecchia and P. L. Bellon, “Fast computation of 3D radon transform via a direct Fourier method,” *Bioinformatics*, vol. 14, no. 2, pp. 212–216, 1998.
- [45] S. Lanzavecchia, P. L. Bellon, and M. Radermacher, “Fast and accurate three-dimensional reconstruction from projections with random orientations via radon transforms,” *Journal of Structural Biology*, vol. 128, no. 2, pp. 152–164, 1999.
- [46] S. Lanzavecchia, F. Cantele, M. Radermacher, and P. Luigi Bellon, “Symmetry embedding in the reconstruction of macromolecular assemblies via the discrete Radon transform,” *Journal of Structural Biology*, vol. 137, no. 3, pp. 259–272, 2002.
- [47] S. Jonic, C. O. Sorzano, and N. Boisset, “Comparison of single-particle analysis and electron tomography approaches: an overview,” *Journal of Microscopy*, vol. 232, no. 3, pp. 562–579, 2008.

- [48] R. Jonges, P. N. M. Boon, J. Van Marle, A. J. J. Dietrich, and C. A. Grimbergen, "CART: A controlled algebraic reconstruction technique for electron microscope tomography of embedded, sectioned specimen," *Ultramicroscopy*, vol. 76, no. 4, pp. 203–219, 1999.
- [49] J.-J. Fernández, A. F. Lawrence, J. Roca, I. García, M. H. Ellisman, and J.-M. Carazo, "High-performance electron tomography of complex biological specimens," *Journal of Structural Biology*, vol. 138, no. 1-2, pp. 6–20, 2002.
- [50] D. Castaño Díez, H. Mueller, and A. S. Frangakis, "Implementation and performance evaluation of reconstruction algorithms on graphics processors," *Journal of Structural Biology*, vol. 157, no. 1, pp. 288–295, 2007.
- [51] J. R. Bilbao-Castro, R. Marabini, C. O. S. Sorzano, I. García, J. M. Carazo, and J. J. Fernández, "Exploiting desktop supercomputing for three-dimensional electron microscopy reconstructions using ART with blobs," *Journal of Structural Biology*, vol. 165, no. 1, pp. 19–26, 2009.
- [52] J. I. Agulleiro, E. M. Garzón, I. García, and J. J. Fernández, "Vectorization with SIMD extensions speeds up reconstruction in electron tomography," *Journal of Structural Biology*, vol. 170, no. 3, pp. 570–575, 2010.
- [53] J. I. Agulleiro and J. J. Fernández, "Fast tomographic reconstruction on multicore computers," *Bioinformatics*, vol. 27, no. 4, Article ID btq692, pp. 582–583, 2011.
- [54] W. J. Palenstijn, K. J. Batenburg, and J. Sijbers, "Performance improvements for iterative electron tomography reconstruction using graphics processing units (GPUs)," *Journal of Structural Biology*, vol. 176, no. 2, pp. 250–253, 2011.
- [55] J. I. Agulleiro, F. Vázquez, E. M. Garzón, and J. J. Fernández, "Hybrid computing: CPU+GPU co-processing and its application to tomographic reconstruction," *Ultramicroscopy*, vol. 115, pp. 109–114, 2012.
- [56] K. J. Batenburg and L. Plantagie, "Fast approximation of algebraic reconstruction methods for tomography," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3648–3658, 2012.
- [57] B. Goris, T. Roelandts, K. J. Batenburg, H. Heidari Mezerji, and S. Bals, "Advanced reconstruction algorithms for electron tomography: From comparison to combination," *Ultramicroscopy*, vol. 127, pp. 40–47, 2013.
- [58] J.-I. Agulleiro and J.-J. Fernández, "Tomo3D 2.0 - exploitation of advanced vector eXtensions (AVX) for 3D reconstruction," *Journal of Structural Biology*, vol. 189, no. 2, pp. 147–152, 2015.
- [59] Y. Chen and F. Förster, "Iterative reconstruction of cryo-electron tomograms using nonuniform fast Fourier transforms," *Journal of Structural Biology*, vol. 185, no. 3, pp. 309–316, 2014.
- [60] N. C. Dvornek, F. J. Sigworth, and H. D. Tagare, "SubspaceEM: A fast maximum-a-posteriori algorithm for cryo-EM single particle reconstruction," *Journal of Structural Biology*, vol. 190, no. 2, pp. 200–214, 2015.
- [61] B. Turoňová, L. Marsalek, T. Davidovič, and P. Slusallek, "Progressive stochastic reconstruction technique (PSRT) for cryo electron tomography," *Journal of Structural Biology*, vol. 189, no. 3, pp. 195–206, 2015.
- [62] J. Zhu, P. A. Penczek, R. Schröder, and J. Frank, "Three-dimensional reconstruction with contrast transfer function correction from energy-filtered cryoelectron micrographs: Procedure and application to the 70S Escherichia coli ribosome," *Journal of Structural Biology*, vol. 118, no. 3, pp. 197–219, 1997.
- [63] I. Aganj, A. Bartesaghi, M. Borgnia, H. Y. Liao, G. Sapiro, and S. Subramaniam, "Regularization for inverting the radon transform with wedge consideration," in *Proceedings of the 2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro; ISBI'07*, pp. 217–220, USA, April 2007.
- [64] C. O. S. Sorzano, J. A. Velázquez-Muriel, R. Marabini, G. T. Herman, and J. M. Carazo, "Volumetric restrictions in single particle 3DEM reconstruction," *Pattern Recognition*, vol. 41, no. 2, pp. 616–626, 2008.
- [65] K. J. Batenburg and J. Sijbers, "DART: a practical reconstruction algorithm for discrete tomography," *IEEE Transactions on Image Processing*, vol. 20, no. 9, pp. 2542–2553, 2011.
- [66] M. Li, G. Xu, C. O. S. Sorzano, F. Sun, and C. L. Bajaj, "Single-particle reconstruction using L2-gradient flow," *Journal of Structural Biology*, vol. 176, no. 3, pp. 259–267, 2011.
- [67] A. Gopinath, G. Xu, D. Röss, O. Öktem, S. Subramaniam, and C. Bajaj, "Shape-based regularization of electron tomographic reconstruction," *IEEE Transactions on Medical Imaging*, vol. 31, no. 12, pp. 2241–2252, 2012.
- [68] A. Kucukelbir, F. J. Sigworth, and H. D. Tagare, "A Bayesian adaptive basis algorithm for single particle reconstruction," *Journal of Structural Biology*, vol. 179, no. 1, pp. 56–67, 2012.
- [69] C. V. Sindelar and N. Grigorieff, "Optimal noise reduction in 3D reconstructions of single particles using a volume-normalized filter," *Journal of Structural Biology*, vol. 180, no. 1, pp. 26–38, 2012.
- [70] W. van Aarle, K. J. Batenburg, and J. Sijbers, "Automatic parameter estimation for the discrete algebraic reconstruction technique (DART)," *IEEE Transactions on Image Processing*, vol. 21, no. 11, pp. 4608–4621, 2012.
- [71] C. Messaoudi, N. Aschman, M. Cunha, T. Oikawa, C. O. Sanchez Sorzano, and S. Marco, "Three-dimensional chemical mapping by EFTEM-TomoJ including improvement of SNR by PCA and ART reconstruction of volume by noise suppression," *Microscopy and Microanalysis*, vol. 19, no. 6, pp. 1669–1677, 2013.
- [72] A. Dabrovolski, K. J. Batenburg, and J. Sijbers, "A multiresolution approach to discrete tomography using DART," *PLoS ONE*, vol. 9, no. 9, Article ID e106090, 2014.
- [73] M. Kunz and A. S. Frangakis, "Super-sampling SART with ordered subsets," *Journal of Structural Biology*, vol. 188, no. 2, pp. 107–115, 2014.
- [74] T. Moriya, E. Acar, R. H. Cheng, and U. Ruotsalainen, "A Bayesian approach for suppression of limited angular sampling artifacts in single particle 3D reconstruction," *Journal of Structural Biology*, vol. 191, no. 3, article no. 6755, pp. 318–331, 2015.
- [75] Y. Chen, Y. Zhang, K. Zhang et al., "FIRT: Filtered iterative reconstruction technique with information restoration," *Journal of Structural Biology*, vol. 195, no. 1, pp. 49–61, 2016.
- [76] X. Zhuge, W. J. Palenstijn, and K. J. Batenburg, "TVR—DART: A more robust algorithm for discrete tomography from limited projection data with automated gray value estimation," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 455–468, 2016.
- [77] C. O. S. Sorzano, R. Marabini, J. Vargas et al., "Interchanging Geometry Conventions in 3DEM: Mathematical Context for the Development of Standards," in *Computational Methods for Three-Dimensional Microscopy Reconstruction*, Applied and Numerical Harmonic Analysis, pp. 7–42, Springer, New York, NY, USA, 2014.
- [78] C. O. S. Sorzano, L. G. De La Fraga, R. Clackdoyle, and J. M. Carazo, "Normalizing projection images: A study of image normalizing procedures for single particle three-dimensional electron microscopy," *Ultramicroscopy*, vol. 101, no. 2-4, pp. 129–138, 2004.

- [79] R. Gordon, "Artifacts in reconstructions made from a few projections," in *Proc 1st Intl Joint Conf on Pattern Recognition*, vol. 30, pp. 275–285, 1973.
- [80] S. H. W. Scheres, "A bayesian view on cryo-EM structure determination," *Journal of Molecular Biology*, vol. 415, no. 2, pp. 406–418, 2012.
- [81] D. Lay, *Linear Algebra and Its Applications*, Pearson, 2006.
- [82] W. Hackbusch, *Iterative Solution of Large Sparse Systems of Equations*, vol. 95, Springer, 2016.
- [83] G. T. Herman, *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography*, Academic Press, New York, NY, USA, 1980.
- [84] A. van der Sluis and H. A. van der Vorst, "Sirt-and cg-type methods for the iterative solution of sparse linear least-squares problems," *Linear Algebra and its Applications*, vol. 130, pp. 257–303, 1990.
- [85] C. O. S. Sorzano, R. Marabini, J. Velázquez-Muriel et al., "XMIPP: a new generation of an open-source image processing package for electron microscopy," *Journal of Structural Biology*, vol. 148, no. 2, pp. 194–204, 2004.
- [86] J. M. De la Rosa-Trevín, J. Otón, R. Marabini et al., "Xmipp 3.0: an improved software suite for image processing in electron microscopy," *Journal of Structural Biology*, vol. 184, no. 2, pp. 321–328, 2013.
- [87] C. Messaoudi, T. Boudier, C. O. S. Sorzano, and S. Marco, "TomoJ: tomography software for three-dimensional reconstruction in transmission electron microscopy," *BMC Bioinformatics*, vol. 8, article 288, 2007.
- [88] W. van Aarle, W. J. Palenstijn, J. De Beenhouwer et al., "The ASTRA Toolbox: A platform for advanced algorithm development in electron tomography," *Ultramicroscopy*, vol. 157, pp. 35–47, 2015.
- [89] P. L. Combettes, "The convex feasibility problem in image recovery," *Advances in Imaging and Electron Physics*, vol. 95, pp. 155–270, 1996.
- [90] C. Byrne, "Iterative oblique projection onto convex sets and the split feasibility problem," *Inverse Problems. An International Journal on the Theory and Practice of Inverse Problems, Inverse Methods and Computerized Inversion of Data*, vol. 18, no. 2, pp. 441–453, 2002.
- [91] Y. Censor, T. Elfving, N. Kopf, and T. Bortfeld, "The multiple-sets split feasibility problem and its applications for inverse problems," *Inverse Problems. An International Journal on the Theory and Practice of Inverse Problems, Inverse Methods and Computerized Inversion of Data*, vol. 21, no. 6, pp. 2071–2084, 2005.
- [92] G. Alvarez and R. Gordon, "Ct brush and cancerzap!: two video games for computed tomography dose minimization," *Theoretical Biology & Medical Modelling*, vol. 12, no. 7, 2015.
- [93] Y. Censor and T. Elfving, "Block-iterative algorithms with diagonally scaled oblique projections for the linear feasibility problem," *SIAM Journal on Matrix Analysis and Applications*, vol. 24, no. 1, pp. 40–58, 2002.
- [94] Y. Censor, T. Elfving, and G. T. Herman, "Averaging strings of sequential iterations for convex feasibility problems," in *Inherently parallel algorithms in feasibility and optimization and their applications (Haifa, 2000)*, vol. 8 of *Stud. Comput. Math.*, pp. 101–113, Amsterdam, The Netherlands, 2001.
- [95] Y. Censor, D. Gordon, and R. Gordon, "Component averaging: an efficient iterative parallel algorithm for large and sparse unstructured problems," *Parallel Computing*, vol. 27, no. 6, pp. 777–808, 2001.
- [96] Y. Censor, D. Gordon, and R. Gordon, "BICAV: A block-iterative parallel algorithm for sparse systems with pixel-related weighting," *IEEE Transactions on Medical Imaging*, vol. 20, no. 10, pp. 1050–1060, 2001.
- [97] Y. Censor and G. T. Herman, "Block-iterative algorithms with underrelaxed Bregman projections," *SIAM Journal on Optimization*, vol. 13, no. 1, pp. 283–297, 2002.
- [98] Y. Censor, T. Elfving, G. T. Herman, and T. Nikazad, "On diagonally relaxed orthogonal projection methods," *SIAM Journal on Scientific Computing*, vol. 30, no. 1, pp. 473–504, 2007/08.
- [99] J. R. Bilbao-Castro, J. M. Carazo, I. García, and J. J. Fernández, "Parallel iterative reconstruction methods for structure determination of biological specimens by electron microscopy," in *Proceedings of the IEEE International Congress on Image Processing*, vol. 1, pp. 565–568, Barcelona, Spain, 2003.
- [100] R. Gordon and G. T. Herman, "Three-Dimensional Reconstruction from Projections: A Review of Algorithms," in *International Review of Cytology*, vol. 38 of *International Review of Cytology*, pp. 111–151, Elsevier, 1974.
- [101] R. Gordon, "A tutorial on art (algebraic reconstruction techniques)," *IEEE Transactions on Nuclear Science*, vol. 21, no. 3, pp. 78–93, 1974.
- [102] G. T. Herman, *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*, Springer, 2009.
- [103] G. L. Zeng and G. T. Gullberg, "Unmatched projector/back-projector pairs in an iterative reconstruction algorithm," *IEEE Transactions on Medical Imaging*, vol. 19, no. 5, pp. 548–555, 2000.
- [104] J. R. Shewchuk, 1994, An introduction to the conjugate gradient method without the agonizing pain.
- [105] W. Zhu, Y. Wang, Y. Yao et al., "Iterative total least-squares image reconstruction algorithm for optical tomography by the conjugate gradient method," *Journal of the Optical Society of America A*, vol. 14, pp. 799–807, 1997.
- [106] E. L. Piccolomini and F. Zama, "The conjugate gradient regularization method in computed tomography problems," *Applied Mathematics and Computation*, vol. 102, no. 1, pp. 87–99, 1999.
- [107] G. H. Golub and H. A. van der Vorst, "Closer to the solution: iterative linear solvers," in *The state of the art in numerical analysis*, pp. 63–92, 2001.
- [108] T. Sakuma, S. Schneider, and Y. Yasuda, "Fast solution methods," in *Computational Acoustics of Noise Propagation in Fluids - Finite and Boundary Element Methods*, pp. 333–366, Springer, New York, NY, USA, 2008.
- [109] C. Chirm and S.-C. Huang, "A general class of preconditioners for statistical iterative reconstruction of emission computed tomography," *IEEE Transactions on Medical Imaging*, vol. 16, no. 1, pp. 1–10, 1997.
- [110] A. Jin, B. Yazici, A. Ale, and V. Ntziachristos, "Preconditioning of the fluorescence diffuse optical tomography sensing matrix based on compressive sensing," *Optics Letters*, vol. 37, no. 20, pp. 4326–4328, 2012.
- [111] J. A. Fessler and S. D. Booth, "Conjugate-gradient preconditioning methods for shift-variant pet image reconstruction," *IEEE Transactions on Image Processing*, vol. 8, no. 5, pp. 688–699, 1999.
- [112] G. Lu, L. Peng, B. Zhang, and Y. Liao, "Preconditioned Landweber iteration algorithm for electrical capacitance tomography," *Flow Measurement and Instrumentation*, vol. 16, no. 2-3, pp. 163–167, 2005.

- [113] R. Kamei, R. G. Pratt, and T. Tsuji, "On acoustic wave-form tomography of wide-angle OBS data-strategies for pre-conditioning and inversion," *Geophysical Journal International*, vol. 194, no. 2, pp. 1250–1280, 2013.
- [114] C. Vonesch, L. Wang, Y. Shkolnisky, and A. Singer, "Fast wavelet-based single-particle reconstruction in Cryo-EM," in *Proceedings of the 2011 8th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI'11*, pp. 1950–1953, usa, April 2011.
- [115] J. M. Carazo and J. L. Carrascosa, "Information recovery in missing angular data cases: an approach by the convex projections method in three dimensions," *Journal of Microscopy*, vol. 145, no. 1, pp. 23–43, 1987.
- [116] J. M. Carazo and J. L. Carrascosa, "Restoration of direct Fourier three-dimensional reconstructions of crystalline specimens by the method of convex projections," *Journal of Microscopy*, vol. 145, pp. 159–177, 1987.
- [117] J. M. Carazo, "The fidelity of 3D reconstructions from incomplete data and the use of restoration methods," in *Electron Tomography. Three-Dimensional Imaging with the Transmission Electron Microscope*, J. Frank, Ed., pp. 117–166, Plenum Press, New York, NY, USA, 1992.
- [118] Y. Deng, Y. Chen, Y. Zhang, S. Wang, F. Zhang, and F. Sun, "ICON: 3D reconstruction with 'missing-information' restoration in biological electron tomography," *Journal of Structural Biology*, vol. 195, no. 1, pp. 100–112, 2016.
- [119] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-point algorithms for inverse problems in science and engineering*, vol. 49 of *Springer Optim. Appl.*, pp. 185–212, Springer, New York, NY, USA, 2011.
- [120] M. Held, P. Wolfe, and H. P. Crowder, "Validation of subgradient optimization," *Mathematical Programming*, vol. 6, pp. 62–88, 1974.
- [121] P. Wolfe, "A method of conjugate subgradients for minimizing nondifferentiable functions," in *Nondifferentiable optimization*, pp. 145–173, 1975.
- [122] A. Beck and M. Teboulle, "Mirror descent and nonlinear projected subgradient methods for convex optimization," *Operations Research Letters*, vol. 31, no. 3, pp. 167–175, 2003.
- [123] P.-E. Maingé, "Strong convergence of projected subgradient methods for nonsmooth and nonstrictly convex minimization," *Set-Valued Analysis. An International Journal Devoted to the Theory of Multifunctions and its Applications*, vol. 16, no. 7-8, pp. 899–912, 2008.
- [124] Y. Censor, R. Davidi, and G. T. Herman, "Perturbation resilience and superiorization of iterative algorithms," *Inverse Problems. An International Journal on the Theory and Practice of Inverse Problems, Inverse Methods and Computerized Inversion of Data*, vol. 26, no. 6, Article ID 065008, 065008, 12 pages, 2010.
- [125] Y. Censor, R. Davidi, G. T. Herman, R. W. Schulte, and L. Tetrushvili, "Projected subgradient minimization versus superiorization," *Journal of Optimization Theory and Applications*, vol. 160, no. 3, pp. 730–747, 2014.
- [126] Y. Censor and A. J. Zaslavski, "Strict fejer monotonicity by superiorization of feasibility—seeking projection methods," *Journal of Optimization Theory and Applications*, vol. 165, no. 1, pp. 172–187, 2015.
- [127] Y. Censor and S. Matej, "Binary steering of nonbinary iterative algorithms," in *Discrete tomography*, pp. 285–297, 1999.
- [128] K. J. Batenburg and J. Sijbers, "Dart: A fast heuristic algebraic reconstruction algorithm for discrete tomography," in *Proceedings of the 14th IEEE International Conference on Image Processing, ICIP 2007*, pp. IV133–IV136, September 2007.
- [129] Y. Censor, I. Pantelimon, and C. Popa, "Family constraining of iterative algorithms," *Numerical Algorithms*, vol. 66, no. 2, pp. 323–338, 2014.
- [130] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Review*, vol. 51, no. 1, pp. 34–81, 2009.
- [131] J.-L. Starck, F. Murtagh, and J. M. Fadili, *Sparse image and signal processing*, Cambridge University Press, Cambridge, UK, 2010.
- [132] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 948–958, 2010.
- [133] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.
- [134] Y. C. Pati, R. Rezaifar, and P. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Proceedings of the 27th Asilomar Conference on Signals, Systems and Computers*, pp. 40–44, Pacific Grove, Calif, USA, 1993.
- [135] H. Y. Liao and G. Sapiro, "Sparse representations for limited data tomography," in *Proceedings of the 2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI*, pp. 1375–1378, May 2008.
- [136] M. W. Kim, J. Choi, L. Yu, K. E. Lee, S.-S. Han, and J. C. Ye, "Cryo-electron microscopy single particle reconstruction of virus particles using compressed sensing theory," in *Proceedings of the Computational Imaging V*, January 2007.
- [137] M. Li, Z. Fan, H. Ji, and Z. Shen, "Wavelet frame based algorithm for 3D reconstruction in electron microscopy," *SIAM Journal on Scientific Computing*, vol. 36, no. 1, pp. B45–B69, 2014.
- [138] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.
- [139] J. Frikel, "Sparse regularization in limited angle tomography," *Applied and Computational Harmonic Analysis. Time-Frequency and Time-Scale Analysis, Wavelets, Numerical Algorithms, and Applications*, vol. 34, no. 1, pp. 117–141, 2013.
- [140] S. Albarqouni, T. Lasser, W. Alkhaldi, A. Al-Amoudi, and N. Navab, "Gradient projection for regularized cryo-electron tomographic reconstruction," *Lecture Notes in Computational Vision and Biomechanics*, vol. 22, pp. 43–51, 2015.
- [141] M. Nilchian, *High performance reconstruction framework for straight ray tomography [Ph.D. thesis]*, École Polytechnique Fédérale de Lausanne, 2015.
- [142] D. L. Donoho, "Compressed sensing," *Institute of Electrical and Electronics Engineers. Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [143] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling: A sensing/sampling paradigm that goes against the common knowledge in data acquisition," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.
- [144] J. Romberg, "Imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 14–20, 2008.
- [145] Z. Saghi, D. J. Holland, R. Leary et al., "Three-dimensional morphology of iron oxide nanoparticles with reactive concave surfaces. A compressed sensing-electron tomography (CS-ET) approach," *Nano Letters*, vol. 11, no. 11, pp. 4666–4673, 2011.



- [146] P. Binev, W. Dahmen, R. DeVore et al., “Compressed Sensing and Electron Microscopy,” in *Modeling Nanoscale Imaging in Electron Microscopy*, pp. 73–126, Springer, Boston, MA, USA, 2012.
- [147] R. Leary, Z. Saghi, P. A. Midgley, and D. J. Holland, “Compressed sensing electron tomography,” *Ultramicroscopy*, vol. 131, pp. 70–91, 2013.
- [148] A. Stevens, H. Yang, L. Carin, I. Arslan, and N. D. Browning, “The potential for bayesian compressive sensing to significantly reduce electron dose in high-resolution STEM images,” *Microscopy*, vol. 63, no. 1, pp. 41–51, 2014.
- [149] A. Al-Afeef, W. P. Cockshott, I. MacLaren, and S. McVitie, “Electron tomography image reconstruction using data-driven adaptive compressed sensing,” *Scanning*, vol. 38, no. 3, pp. 251–276, 2016.
- [150] Z. Saghi, G. Divitini, B. Winter et al., “Compressed sensing electron tomography of needle-shaped biological specimens - Potential for improved reconstruction fidelity with reduced dose,” *Ultramicroscopy*, vol. 160, pp. 230–238, 2016.
- [151] M. D. Guay, W. Czaja, M. A. Aronova, and R. D. Leapman, “Compressed sensing electron tomography for determining biological structure,” *Scientific Reports*, vol. 6, Article ID 27614, 2016.