INSTITUTE OF PHYSICS PUBLISHING INVERSE PROBLEMS

Inverse Problems **19** (2003) 933–949 PII: S0266-5611(03)54271-X

Three-dimensional reconstruction by Chahine's method from electron microscopic projections corrupted by instrumental aberrations

J P Zubelli[1,3,5], **R Marabini**[2,3], **C O S Sorzano**[2] **and G T Herman**[4]

[1] IMPA, Est. D. Castorina 110, RJ 22460-320, Brazil
[2] CNB, Campus Universidad Autónoma s/n 28049 Cantoblanco, Madrid, Spain
[3] EPS, Universidad Autónoma de Madrid, 28049 Cantoblanco, Madrid, Spain
[4] Computer Science Program, Graduate Center—CUNY, 365 Fifth Avenue, New York, NY 10016-4309, USA

E-mail: zubelli@impa.br, roberto@cnb.uam.es, coss@cnb.uam.es and GHerman@gc.cuny.edu

Received 1 October 2002, in final form 23 June 2003
Published 22 July 2003
Online at stacks.iop.org/IP/19/933

Abstract

Three-dimensional reconstruction of nano-scale objects (such as biological macromolecules) can be accomplished using data recorded with a transmission electron microscope. An image obtained by a transmission electron microscope can be conceived of as an 'ideal' projection subjected to a contrast transfer function, which attenuates most frequencies, reverses the phase of others and even eliminates some of them. Such instrumental aberrations make the problem of reconstruction from such data difficult. We reformulate the problem so that Chahine's method becomes applicable to it. We demonstrate the performance of our approach with numerical evidence using both simulated and actual electron microscopy data.

1. Introduction

Many important cellular functions are carried out by protein complexes that act as molecular machines. Three-dimensional electron microscopy (3DEM) is a powerful technique for imaging these complex biological macromolecules in order to understand their function. Unfortunately, the 3DEM reconstruction process cannot be made truly quantitative unless the aberrations introduced by the electron microscope are taken into account. We follow a model that is well accepted within the electron microscopy community [9, 29, 30, 32, 33], incorporates the aberrations and assumes a linear relation between the data and the measurements as described below. This paper presents a way of explicitly incorporating these aberrations into a version of the algorithm due to Chahine [3] for solving certain types of linear equation system.

[5] Author to whom any correspondence should be addressed.

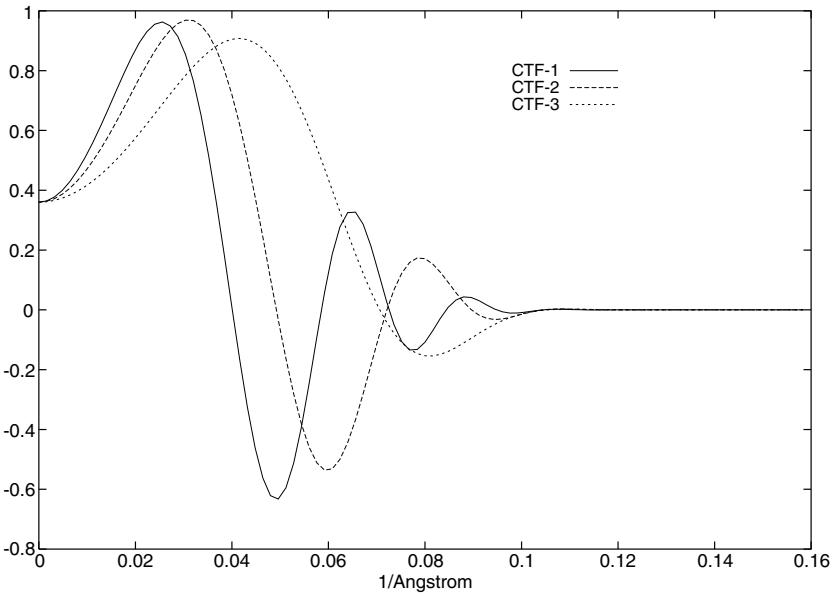

Figure 1. Radial profiles of the CTFs used in the experiments.

The microscope aberrations are described (in Fourier space) by the so-called *contrast transfer function* (CTF). The CTF filters both the high and the low frequencies, introduces zones of alternate contrast and eliminates all information at certain frequencies; see, for example, [33]. Figure 1 shows the radial profiles of three CTFs used in our simulations.

The model we consider [33] is the following: the measured data for the ith projection y_i consist of the line integrals parallel to a chosen direction of the original compactly supported volume density $\rho : \mathbb{R}^3 \to \mathbb{R}_{\geqslant 0}$ followed by the convolution with the point spread function (PSF) h_i. The PSF h_i is the inverse Fourier transform of the CTF H_i. In practical applications, such projections y_i would also be affected by noise.

The problem is to reconstruct ρ from the collection of projections y_i, where $i = 1, \ldots, I$. Here, by 'to reconstruct' we mean to find a vector $c \in \mathbb{R}^J$ whose components are the coefficients of a sufficiently good approximation ρ_c to ρ in a finite dimensional space generated by suitable basis functions. One example is the representation of ρ_c using a voxel (volume element) basis, in which case the coefficients should be near to the averages of ρ over the corresponding voxels. Following [17] we do not use voxels as the basis functions, but spherical symmetric functions which are not only spatially limited but can also be chosen to be smooth. They are the generalized Kaiser–Bessel window functions, also known as *blobs*. The property of blobs that makes them very promising for the formulation of many algorithms is that they are smooth both in the real and in the Fourier domain. These blobs are placed in space so that they are centred on the points of the so-called body-centred cubic grid (BCC, obtained by interlacing two cubic grids in such a way that the points of either grid are central to eight points forming a cube in the other grid). For a detailed description of blobs and of their placement in space see [16, 20–23].

From now on, we consider all the objects as defined on discrete and finite domains. Hence, our convolutions will be discrete ones (defined using wraparounds) and our Fourier transforms

will be discrete Fourier transforms. We write each (noiseless) projection y_i as

$$y_i = h_i * \mathcal{P}_i c, \tag{1}$$

where $\mathcal{P}_i$ is the so-called *projection operator* along the ith direction. It maps the volume coefficients into the two-dimensional array p_i whose entries correspond to the line integrals of the density ρ_c along the set of lines in the ith direction. We will use J to denote the number of basis functions (and hence the dimensionality of the vector c) and N to denote the number of parallel lines in one projection (and hence the dimensionality of the vectors h_i and y_i); consequently, $\mathcal{P}_i$ is an $N \times J$ matrix of non-negative entries (a fact of which we make use later). In a typical 3DEM experiment for single-particle reconstruction [10], we have a total of 10^3–10^5 projections and each vector y_i has of the order of 64^2–315^2 components. Furthermore, the signal to noise ratio of each measured y_i is typically less than unity, and can be as small as one-third.

We denote by $\mathcal{H}_i$ the two-dimensional convolution operators (in our approach $N \times N$ matrices), defined for $z \in \mathbb{R}^N$ by

$$\mathcal{H}_i z = h_i * z, \qquad i = 1, \ldots, I. \tag{2}$$

For convenience we introduce the auxiliary block-matrices $\mathcal{H}$, and $\mathcal{P}$ and the vector y, consisting of

$$\mathcal{H} \stackrel{\text{def}}{=} \operatorname{diag}[\mathcal{H}_1, \ldots, \mathcal{H}_I], \qquad \mathcal{P} \stackrel{\text{def}}{=} \begin{bmatrix} \mathcal{P}_1 \\ \vdots \\ \mathcal{P}_I \end{bmatrix}, \qquad \text{and} \qquad y \stackrel{\text{def}}{=} \begin{bmatrix} y_1 \\ \vdots \\ y_I \end{bmatrix}. \tag{3}$$

In the following we also make use of the (real) matrix $\mathcal{B} = \mathcal{P}^{\mathrm{T}}$, which is known in the reconstruction literature as the (discrete) *back-projection* operator. The sum of all components of a given vector z will be denoted by $\overline{z}$.

In an ideal situation, which is far from reality in microscopy, the CTF would be identically one and there would be no noise. In this case, our problem would reduce to data

$$p \stackrel{\text{def}}{=} \mathcal{P} c. \tag{4}$$

This problem has attracted a significant amount of attention; see [14, 24, 25] and references therein. The presence of the CTF, however, introduces further difficulties. In the microscopy literature it was recently considered by several authors: Frank and Penczek [11] applied Wiener filtering in the three-dimensional space to the reconstructed volume; Zhu *et al* [34] incorporated a three-dimensional PSF into the data model and used a regularized steepest-descent technique; reconstruction algorithm; Stark *et al* [28] applied inverse CTF filtering to the reconstructed volume; Skoglund *et al* [27] incorporated a two-dimensional CTF particular to each projection to the projection model in a maximum-entropy reconstruction algorithm; Grigorieff [13] provided a Fourier reconstruction algorithm in which the CTF for each projection is considered in a Wiener-like fashion; Ludtke *et al* [18] proposed a CTF correction applied to the individual projections with a weighting function in the Fourier space computed from a set of images sharing a common CTF; Ludtke *et al* [19] added a Wiener filter to the weighting function defined in [18]. The presence of these multiple approaches is indicative of the fact that there is no agreed standard technique in 3DEM of single particles to correct for CTF effects.

Many of the proposed approaches (and, in particular, the three approaches that we experimentally compare in this paper) fall into the following category: given the measured data y, 'solve' the system

$$\mathcal{D}\mathcal{P}c = \mathcal{E}y \tag{5}$$

for some 'suitable' matrices $\mathcal{D}$ and $\mathcal{E}$. In this paper we consider three choices of $\mathcal{D}$ and $\mathcal{E}$; the first two are standard approaches and the third has not previously been explicitly studied in the context of reconstruction from microscopy projections corrupted by CTF.

(i) Choose both $\mathcal{D}$ and $\mathcal{E}$ to be the identity matrix. This simply ignores the CTF and so results in attempting to solve (1). Even with this most simple-minded approach one can sometimes get biologically useful reconstructions [1].

(ii) Choose $\mathcal{D}$ to be the identity matrix and choose $\mathcal{E}$ so that negative values in the CTFs (see figure 1) are turned into positive values. (This is referred to as *phase flipping* [9, p 45].) Such an $\mathcal{E}$ can in practice be obtained by taking the discrete Fourier transform of the two-dimensional array y_i (this is a matrix multiplication), reversing the sign for all points in Fourier space where the CTF has a negative value, and taking the inverse discrete Fourier transform.

(iii) Choose $\mathcal{D} = \mathcal{B}\mathcal{H}^{\mathrm{T}}\mathcal{H}$ and $\mathcal{E} = \mathcal{B}\mathcal{H}^{\mathrm{T}}$, which results in (5) being the normal equations corresponding to (1), namely

$$\mathcal{B}\mathcal{H}^{\mathrm{T}}\mathcal{H}\mathcal{P}c = d \stackrel{\mathrm{def}}{=} \mathcal{B}\mathcal{H}^{\mathrm{T}}y. \tag{6}$$

We remark that the operations of projection and back-projection are relatively fast, and the convolution can be efficiently implemented by the fast Fourier transform (FFT), so in practice we do not need to store the (prohibitively large) matrix

$$\mathcal{A} \stackrel{\mathrm{def}}{=} \mathcal{B}\mathcal{H}^{\mathrm{T}}\mathcal{H}\mathcal{P}. \tag{7}$$

In what follows, we will report on comparison experiments using the same data y, but with the three different pairs of $\mathcal{D}$ and $\mathcal{E}$ listed above. An additional choice needs to be made: what algorithm should be applied to solve (5)?

For the first two choices of $\mathcal{D}$ and $\mathcal{E}$, we decided to apply one of the standard approaches in this context, namely ART (short for algebraic reconstruction techniques [14, 15, 20, 21]). Some specifics of this algorithm are provided below in section 5; here, we remark only that for the first two choices of $\mathcal{D}$ and $\mathcal{E}$ our ART method is very efficient due to the sparse nature of the matrix $\mathcal{P}$. For the third choice of $\mathcal{D}$ and $\mathcal{E}$, ART cannot make good use of the sparsity of $\mathcal{P}$ and so loses its computational efficiency. Therefore, we looked for an alternative algorithm for the third choice of $\mathcal{D}$ and $\mathcal{E}$.

Due to its size and the non-negativity of ρ, one approach to solving (6) is to try an algorithm due to Chahine [3] that has been used in many applications including atmospheric science [2, 3] and medical imaging [4, 5]. This method, due to its convergence properties [6] and simplicity, has proved to be competitive when applied to positron emission tomography. We discuss further the choice of Chahine's method in section 2. It turns out, as we shall see below, that not all the necessary conditions for the published convergence results of Chahine's method are satisfied by the formulation of our problem given in (6).

The first goal of the present work is to present an answer to the following question: based on the knowledge of the CTFs and the data y, how do we adapt problem (6) so that Chahine's algorithm becomes applicable to it? The second goal is to show, by numerical examples, that such a method indeed produces good results, at least when compared to no CTF correction or to a simple phase flipping.

The plan for the paper is the following. In section 2 we describe Chahine's method and give some background information on 3DEM. In section 3 we present our answer to the question stated in the previous paragraph. In section 4 we describe some of the implementation details. In section 5 we discuss the numerical results. We close in section 6 with some conclusions and suggestions for further research.

2. Previous work and background

We first describe Chahine's method for solving systems of linear equations of the form

$$Mx = b, \tag{8}$$

where $x, b \in \mathbb{R}^J$ and M is a square $J \times J$ matrix satisfying some further hypotheses that will be spelled out in the following. The solution method goes as follows.

Consider an initial guess (say a constant vector of ones) $x^{(0)}$ in the positive orthant $\mathbb{R}^J_{>0}$ and iterate the following.

For $k = 0, 1, 2, \ldots$ do

$$e^{(k)} = Mx^{(k)}$$
$$x_j^{(k+1)} = x_j^{(k)} \cdot b_j / e_j^{(k)}, \qquad j = 1, \ldots, J.$$

Chahine's method has been used in the medical imaging literature where it is sometimes referred to as the image space reconstruction algorithm (ISRA) [4, 5]. In fact, it was independently proposed by Gold and Scofield in 1960 for the solution of certain linear integral equations [12]; see [6] and references therein. A fairly strong result concerning the method is given by the following.

Theorem 1 (De Pierro [5]). *If M is a non-negative definite matrix with positive diagonal elements and if the entries of b are all positive, then the Chahine iteration generates a sequence $(x^{(k)})_{k=0,1,\ldots}$ that possesses a convergent subsequence. Furthermore, if M is nonsingular, then the sequence converges.*

It turns out [6] that the limit of any convergent subsequence is a solution of the symmetric linear complementarity problem

$$Mx - b \geqslant 0, \qquad x \geqslant 0, \qquad x^{\mathrm{T}}(Mx - b) = 0. \tag{9}$$

The latter are the Kuhn–Tucker conditions associated with minimizing the quadratic functional

$$F(x) \stackrel{\text{def}}{=} \tfrac{1}{2} x^{\mathrm{T}} M x - x^{\mathrm{T}} b \tag{10}$$

subject to $x \geqslant 0$. The conditions in (9) are necessary and sufficient for x to be a minimizer of (10) subject to $x \geqslant 0$.

Other results are known [2, 5–8]. In particular, Eggermont [7] strengthened theorem 1 showing convergence even without the condition that M is nonsingular. In any case, the known results seem to rely on the hypotheses that M is non-negative definite with positive diagonal elements and that all the entries of b are positive. The simplicity of the method and its demonstrated effectiveness for a number of large-scale problems plagued by low signal-to-noise ratio makes it an attractive candidate for our three-dimensional electron microscopy problem. Indeed, the number of operations necessary to compute the matrix $\mathcal{A}$ applied to a vector c representing the volume's density is surprisingly small. All that needs to be done is to compute projections of the current estimate (for blobs this can be efficiently performed by 'footprint' techniques [23]), an FFT, a product by the squared CTF, an inverse FFT and a back-projection (that can also be done using footprints). To complete an iteration, this is followed by one multiplication and one division per component of the unknown c. Furthermore, for some bases (for example, the pixel basis) the volume has a *non-negative* coefficient expansion. This means that the assumption that $c \geqslant 0$ is naturally satisfied. For other bases, such as blobs, this corresponds to a regularization of the problem since it imposes smoothness on the approximate ρ. For the applicability of Chahine's method we need to discuss whether or not the above mentioned hypotheses are satisfied.

We start with some technical assumptions.

(i) The PSFs are even functions (or equivalently the CTFs are real functions). This implies that the adjoint to the convolution operator $\mathcal{H}$ is itself, and so we may omit the super-index T in $\mathcal{H}^{\mathrm{T}}$.
(ii) Every voxel (or blob) in the decomposition of the image is intersected by at least one of the projecting rays.
(iii) For every i, $\overline{h_i} \geqslant 0$ (recall that $\overline{h_i}$ denotes the sum of all components of h_i). This assumption is harmless because if $\overline{h_i} < 0$, then we can replace our data y_i by $-y_i$ and proceed with h_i replaced by $-h_i$.

It is easy to check that the matrix $\mathcal{A}$ in (7) is non-negative definite since the h_i are real and even and $\mathcal{B} = \mathcal{P}^{\mathrm{T}}$. The diagonal entries of $\mathcal{A}$ are positive, except for degenerate cases. Indeed, each entry $\mathcal{A}_{jj}$ corresponds to the ℓ^2 norm of the jth column of $\mathcal{HP}$. Such an ℓ^2 norm would be zero only if, for all i, the zero set of the ith CTF contained all the nonzero points of the Fourier transform of $\mathcal{P}_i e_j$, where e_j is the vector of zeros except for a one in the jth position. Though possible, this is unlikely. Furthermore, if the number of projections is large, the CTFs are 'mildly' behaved and a good angular coverage is available, then one would expect $\mathcal{A}$ to be invertible (although we cannot say much *a priori* about its conditioning).

However, in the case of 3DEM in the presence of the CTFs, a major difficulty arises: the PSFs h_i change sign. Furthermore, their Fourier transforms have zeros and decay at infinity. The change in sign of the PSFs may cause the components of the vector d to change sign. Because of the decay of the CTFs, the matrix $\mathcal{A}$ may be singular or nearly singular (and thus highly ill conditioned). Indeed, our first attempts to apply Chahine's method directly to the 3DEM problem with realistic CTFs did not lead to acceptable results.

3. Methodology

We now discuss a technique to circumvent the above problems in the case of 3DEM or in any other field where a similar data formation model holds. Our idea is to pre-process the data in a way that is equivalent to replacing the PSFs by some *virtual* PSFs. This pre-processing consists of adding a suitable multiple of a constant-component vector to each projection. For notational convenience take

$$\mathbf{1} \stackrel{\text{def}}{=} [1, 1, \ldots, 1]^{\mathrm{T}} \in \mathbb{R}^N.$$

We propose the following idea: add the vector $v_i \stackrel{\text{def}}{=} \lambda_i \mathbf{1}$ to y_i before and after it is convolved with h_i, where λ_i will be suitably chosen. We set v to be the vector

$$v = \begin{bmatrix} l v_1 \\ \vdots \\ v_I \end{bmatrix}.$$

We shall show that the above mentioned pre-processing corresponds to solving a modification of problem (6) with a different set of PSFs and the right-hand side (RHS) vector

$$\tilde{d} = \mathcal{B}(\mathcal{H}(y + v) + v). \tag{11}$$

Note that for any vector $z \in \mathbb{R}^N$,

$$z * \mathbf{1} = \overline{z}\mathbf{1}. \tag{12}$$

The following claims are fundamental to the following.

Claim 1. If we choose the numbers λ_i, $i = 1, \ldots, I$, appropriately, then all components of the vector $\tilde{d}$ are positive.

Proof. Fix $i \in \{1, \ldots, I\}$. Because of (12), $h_i * v_i = \overline{h}_i v_i$. So,

$$h_i * (y_i + v_i) + v_i = h_i * y_i + (\overline{h}_i + 1)v_i.$$

Now we choose λ_i so that

$$\lambda_i(\overline{h}_i + 1) > -\min_l\{0, (h_i * y_i)_l\}. \tag{13}$$

Since the back-projection matrix has only non-negative entries, the resulting vector $\tilde{d}$ has only non-negative entries. Furthermore, none of the entries in $\tilde{d}$ can be zero since by the choice of the λ_i all the entries of $\tilde{y} \stackrel{\text{def}}{=} (\mathcal{H}(y + v) + v)$ are positive. Therefore, the only way an entry of $\mathcal{B}\tilde{y}$ can be zero is if the corresponding row of $\mathcal{B}$ vanishes. But this contradicts the second of our technical assumptions in section 2 since it would correspond to a column of zeros in $\mathcal{P}$, in other words a voxel (or blob) which is not intersected by any of the line integrals that produce the totality of projections. □

Our next goal is to show that the pre-processing in claim 1 leads to a set of virtual PSFs of the form

$$\tilde{h}_i = h_i + \alpha_i v_i, \tag{14}$$

where the constants α_i will be chosen appropriately. Such a virtual PSF may not be realizable in a real microscope, but it leads to a set of equations that is amenable to Chahine's algorithm, as we shall see below. Associated with these virtual PSFs we define the operators

$$\tilde{\mathcal{H}}_i z \stackrel{\text{def}}{=} \tilde{h}_i * z \qquad \text{and} \qquad \tilde{\mathcal{H}} \stackrel{\text{def}}{=} \text{diag}[\tilde{\mathcal{H}}_1, \ldots, \tilde{\mathcal{H}}_I]. \tag{15}$$

Claim 2. Let $c \in \mathbb{R}^J_{\geqslant 0} \setminus \{0\}$ be such that $\mathcal{A}c = d$ and suppose that $y = \mathcal{H}\mathcal{P}c$. Take $\tilde{d}$ as in (11) with λ_i satisfying (13), for $i = 1, \ldots, I$. Then, there exists $\alpha = (\alpha_1, \ldots, \alpha_I) \in \mathbb{R}^I$ such that c is a solution of $\tilde{\mathcal{A}}c = \tilde{d}$ where $\tilde{\mathcal{A}} = \mathcal{B}\tilde{\mathcal{H}}\tilde{\mathcal{H}}\mathcal{P}$, and $\tilde{\mathcal{H}}$ is associated with the PSFs $\tilde{h}_i$ as defined in (14) and (15).

Proof. For $i = 1, \ldots, I$, let $p_i = \mathcal{P}_i c$, and consequently $\overline{p}_i > 0$. We will show how to choose the α_i so that $\mathcal{B}\tilde{\mathcal{H}}\tilde{\mathcal{H}}\mathcal{P}c = \tilde{d}$. By (11) it is sufficient to prove that, for $i = 1, \ldots, I$,

$$\tilde{\mathcal{H}}_i\tilde{\mathcal{H}}_i\mathcal{P}_i c = \mathcal{H}_i(y_i + v_i) + v_i; \tag{16}$$

or, using equations (14) and (15) and the assumption that $y = \mathcal{H}\mathcal{P}c$, that

$$(h_i + \alpha_i\lambda_i\mathbf{1}) * (h_i + \alpha_i\lambda_i\mathbf{1}) * p_i = h_i * (h_i * p_i + \lambda_i\mathbf{1}) + \lambda_i\mathbf{1}. \tag{17}$$

Using that for $\mathbf{1} \in \mathbb{R}^N$ we have $\mathbf{1} * \mathbf{1} = N\mathbf{1}$ and equation (12), it is easy to check that equation (17) is a consequence of

$$(\overline{p}_i N)(\alpha_i\lambda_i)^2 + 2\overline{h}_i\overline{p}_i(\alpha_i\lambda_i) - \lambda_i(\overline{h}_i + 1) = 0. \tag{18}$$

We must have by (13), and our third technical assumption that $\overline{h}_i \geqslant 0$, that $\lambda_i > 0$. Consequently, we can find the following real-valued solutions α_i of equation (18). If $\overline{h}_i = 0$, then we can take $\alpha_i = \pm 1/\sqrt{\lambda_i\overline{p}_i N}$. If $\overline{h}_i \neq 0$, then using the fact that $\overline{y}_i = \overline{h}_i\overline{p}_i$, we can take

$$\alpha_i = \frac{\overline{h}_i}{\lambda_i N}\left(\pm\sqrt{1 + \frac{N(\overline{h}_i + 1)\lambda_i}{\overline{h}_i\overline{y}_i}} - 1\right). \tag{19}$$

This concludes the proof. □

Remark 1. For each i, if $\overline{h}_i \neq 0$, then α_i can be chosen to be a non-negative number that depends only on $\overline{h}_i$, $\overline{y}_i$ and N. If $\overline{h}_i = 0$, then we can take $\alpha_i = \pm 1/\sqrt{\lambda_i \overline{p}_i N}$.

Remark 2. Technical assumption (iii) of section 2 gives us that for each $i = 1, \ldots, I$ the DC component of the PSF h_i is non-negative. For reasons of numerical stability it is sound to choose the positive value of α_i in our computations, so as to avoid loss of significant digits in the computations. In this case a more numerically sound way of computing α_i is by the expression

$$\alpha_i = \frac{[(\overline{h}_i + 1)/\overline{y}_i]}{1 + \left(1 + \frac{\lambda_i [(\overline{h}_i+1)/\overline{y}_i]}{\overline{h}_i/N}\right)^{1/2}}. \tag{20}$$

From now on, we take α_i to be this positive root of equation (18).

It would be desirable for any solution to our modified equations $\tilde{\mathcal{A}}c = \tilde{d}$ to satisfy the original equations $\mathcal{A}c = d$. This, however, may not be the case. We have instead the following weaker result.

Claim 3. Let $\tilde{c}$ be a solution of $\tilde{\mathcal{A}}\tilde{c} = \tilde{d}$, where α_i and $\tilde{h}_i$ are computed according to (20) and (14), for $i = 1, \ldots, I$, and $\tilde{d}$ is computed by (11). Then, $\tilde{c}$ satisfies

$$\mathcal{A}\tilde{c} = d + \sum_{i=1}^{I} r_i \mathcal{P}_i^{\mathrm{T}} \mathbf{1}, \tag{21}$$

where

$$r_i = (1 + \overline{h}_i)\lambda_i - \overline{\mathcal{P}_i \tilde{c}}(N(\alpha_i \lambda_i)^2 + 2\overline{h}_i(\alpha_i \lambda_i)). \tag{22}$$

Proof. If $\tilde{c}$ satisfies $\tilde{\mathcal{A}}\tilde{c} = \tilde{d}$, then by (6) we get that

$$\sum_{i=1}^{I} \mathcal{P}_i^{\mathrm{T}}((h_i + \alpha_i\lambda_i \mathbf{1}) * (h_i + \alpha_i \lambda_i \mathbf{1}) * \mathcal{P}_i \tilde{c}) = d + \sum_{i=1}^{I} \mathcal{P}_i^{\mathrm{T}}(h_i * v_i + v_i). \tag{23}$$

Expanding the product $(h_i + \alpha_i\lambda_i \mathbf{1}) * (h_i + \alpha_i \lambda_i \mathbf{1})$, and using that $\mathbf{1} * \mathbf{1} = N\mathbf{1}$, we have that (21) and (22) follow. □

In our numerical experiments, we found that the norm of $\sum_{i=1}^{I} r_i \mathcal{P}_i^{\mathrm{T}} \mathbf{1}$ was small as compared to the norm of d, and so in practice the solutions of $\tilde{\mathcal{A}}\tilde{c} = \tilde{d}$ yielded 'near solutions' of $\mathcal{A}c = d$.

4. Practical implementation

In view of claims 1–3 we transform our problem of seeking a solution to equation (6) into searching for a solution to

$$\tilde{\mathcal{A}}c = \tilde{d}, \tag{24}$$

where now $\tilde{d}$ has positive components and $\tilde{\mathcal{A}}$ is a non-negative definite matrix with positive diagonal entries. Indeed, $\tilde{d}$ has positive components because of claim 1. Furthermore, $\tilde{\mathcal{A}}$ is a non-negative definite matrix since $\tilde{\mathcal{H}}$ is symmetric and thus we can write $\tilde{\mathcal{A}} = (\tilde{\mathcal{H}}\mathcal{P})^{\mathrm{T}}\tilde{\mathcal{H}}\mathcal{P}$. Finally, the diagonal entries of $\tilde{\mathcal{A}}$ are positive because each entry $\mathcal{A}_{jj}$ corresponds to the ℓ^2 norm of the jth column of $\tilde{\mathcal{H}}\mathcal{P}$. Such an ℓ^2 norm is not zero. To see this claim, notice that it is the sum for $i = 1, \ldots, I$ of the ℓ^2 norm of $h_i * \mathcal{P}_i e_j$. By Plancherel's theorem this is a (positive) multiple of the ℓ^2 norm of $\tilde{\mathsf{H}}_i \mathcal{F}(\mathcal{P}_i e_j)$, where $\tilde{\mathsf{H}}_i$ is the Fourier transform of the

virtual PSF $\tilde{h}_i$. The claim follows now because the DC components of all the *virtual* PSFs are nonzero.

Thus all the hypotheses in theorem 1 hold and Chahine's method is applicable to our reformulated problem.

The above remarks lead to the following practical implementation of Chahine's method in 3DEM.

Given the PSF h_i, and the data y_i, $i = 1, \ldots, I$:

(i) for each $i = 1, \ldots, I$, compute $\overline{h}_i$ and $\overline{y}_i$, and make sure (by changing (h_i, y_i) into $(-h_i, -y_i)$, if necessary) that they are both non-negative (see remark 3 below);
(ii) choose λ_i so that equation (13) holds for $i = 1, \ldots, I$;
(iii) for $i = 1, \ldots, I$, determine $\tilde{h}_i$ by computing α_i according to claim 2 and equation (20) (equivalently, add to the DC value of the ith CTF the value of $N\alpha_i$) and
(iv) apply Chahine's method to solve the equation

$$\tilde{\mathcal{A}}c = \tilde{d},$$

where $\tilde{d}$ is given by (11).

Remark 3. The effect of noise and mismatch on the estimation of the angles and on the estimation of the CTF can lead to further corruption of the data y_i, $i = 1 \ldots, I$. One problem that could arise is that the averages $\overline{y_i}$ and $\overline{h}_i$ may have different signs, which could lead to complex roots in equation (20). In such a case, the matrix $\tilde{\mathcal{A}}$ may not satisfy the hypotheses of theorem 1 and Chahine's method would no longer be applicable. This cannot happen in claim 2 itself, because one of its hypotheses is that $y = \mathcal{H}\mathcal{P}c$, which implies that the data are perfect except for the CTF aberration. One practical way of dealing with this is by adding a constant volume to the sought value c. To wit, we choose a $\mu \in \mathbb{R}$ so that the data satisfy the condition that $\operatorname{sgn}(\overline{y_i} + \mu\overline{\mathcal{H}_i\mathcal{P}_i\mathbf{1}}) = \operatorname{sgn}(\overline{h}_i)$, for $i \leqslant i \leqslant I$.

5. Numerical results

In this section we describe some of the numerical experiments that we used to validate the technique described in section 4. We compared our results with two implementations of the ART; see [14]. They are both based on the *block-ART* variant of ART, which can be described as follows.

Starting from a J-dimensional zero vector as an estimate of c, we update this estimate iteratively. In an iterative step we make use of data from one projection only; we repeatedly cycle through all the projections in the complete algorithm. The update of the estimate is done in a computationally efficient manner [23] so as to reduce the discrepancy between the measured data provided by the projection used in that iterative step and the matching *pseudo-projection data* that are obtained from the current estimate of c. A more detailed description of this implementation, including convergence properties, is given in [15].

In all the implementations (both ART and Chahine's method) we used blobs as the basis functions. Following [20], we selected blobs with the following parameters: order $m = 2$, shape $\alpha = 13.363\,3059$, radius $a = 3.394\,212$ and grid sample separation 2.0. The initial estimate of c is constant valued (unit valued for Chahine's method and zero valued for ART).

In one implementation of ART, we apply a simple CTF phase flipping, and in the other we do not take the CTF into account at all. By CTF phase flipping, we mean that the algorithm

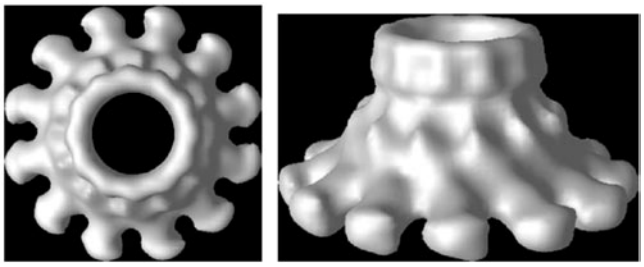

Figure 2. Surface rendering of the three-dimensional ϕ-29 phantom.

makes use of the sign of the CTF so that the processed measurements (i.e., the input to the algorithm) are obtained by the inverse Fourier transform of the Fourier transform of the original measurements multiplied by -1 whenever the CTF is negative. All of our reconstructions were implemented using the XMIPP software available at http://www.cnb.uam.es/~bioinfo/.

In the following, we show the evolution of the objective function

$$F(c) = \tfrac{1}{2}c^{\mathrm{T}}\tilde{\mathcal{A}}c - c^{\mathrm{T}}\tilde{d}. \tag{25}$$

Notice that for the special case $\tilde{\mathcal{H}} = \mathcal{H}$ and $d = \tilde{d}$, a simple computation gives that

$$\tfrac{1}{2}\|\mathcal{H}\mathcal{P}c - y\|^2 = F(c) + \tfrac{1}{2}\|y\|^2.$$

$F(c)$ is a way of monitoring the evolution of the iterates. It is known that $F(c^{(k)})$ is a non-increasing sequence [5]. We will also show the evolution of the squared relative residual error

$$\eta^{(k)} \stackrel{\text{def}}{=} \frac{\|e^{(k)} - \tilde{d}\|^2}{\|\tilde{d}\|^2}, \tag{26}$$

where $e^{(k)} = \tilde{\mathcal{A}}c^{(k)}$.

In the synthetic data examples, we used a total of $I = 1000$ projections with uniformly distributed random angles and $N = 128^2$ pixels each. The CTFs used were randomly chosen from the three examples displayed in figure 1. The number of blob coefficients was $J = 34\,327$.

In the real data experiments, we used a total of $I = 570$ projections, with $N = 128^2$ pixels each, and the number of blobs $J = 130\,869$.

The grey-level scale of the displayed results of the reconstructions was adjusted so that the minimum value was mapped to zero and the maximum level was mapped to 255. In the plots of $F(c^{(k)})$ and $\eta^{(k)}$ we do not display the values for $k = 0$, since they would change the scale of the graphs and reduce the distinguishability of the remaining ordinates.

The experiments were performed using the XMIPP software running under the Red Hat Linux 7.2 operating system, on an Intel Xeon CPU of 1.70 GHz. In the synthetic data experiments, the typical (system) time for one iterative step was of the order of 10 min. In the real data experiments, the typical (system) time for one iterative step was of the order of 25 min. In typical 3DEM experiments, the specimen preparation and data gathering exceed such times by several orders of magnitude.

5.1. Synthetic data: the bacteriophage ϕ-29 connector

We chose as a phantom for our simulations a rather realistic one. It was produced from the structure of the bacteriophage ϕ-29 connector. The atomic coordinates of this protein can be found in the 'protein data base' (http://www.pdb.org) under the PDB-ID 1FOU. In our model,

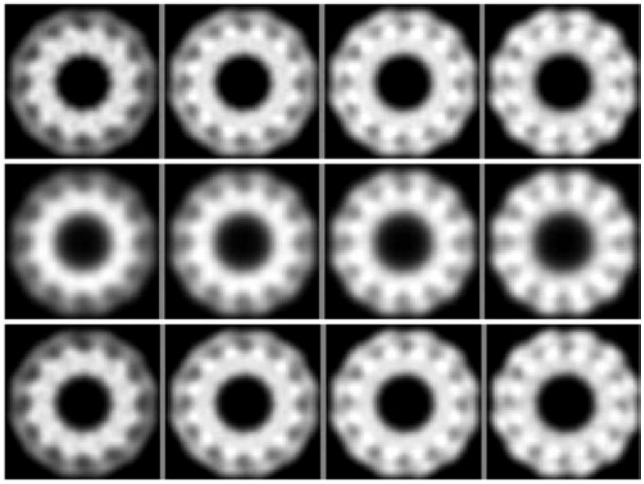

Figure 3. Central slices of the original ϕ-29 phantom (top) and the reconstruction after six iterations (middle) and after 49 iterations (bottom), using Chahine's method.

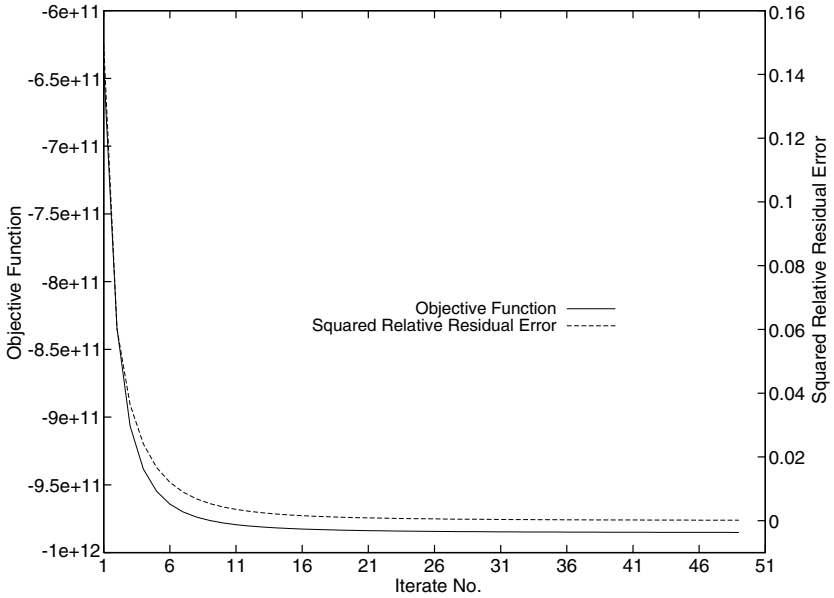

Figure 4. Evolution of the objective function $F(c^{(k)})$ of (25) and of the squared relative residual error $\eta^{(k)}$ of (26), with $k = 1, 2, \ldots$ for the ϕ-29 example by Chahine's method in the no-noise and no-CTF case.

the atoms are replaced by small blobs. Such blobs, however, have different parameters from the ones used in the reconstructions; more precisely the blob radius was 12, the shape 0.1 and the order 2. The surface of the resulting phantom is displayed in figure 2.

We first present the reconstruction by Chahine's method of the original phantom in the (unrealistic) noiseless and no-CTF case. As can be seen in figures 3 and 4, the reconstruction and the original phantom are in excellent agreement. In figure 4 we display the values

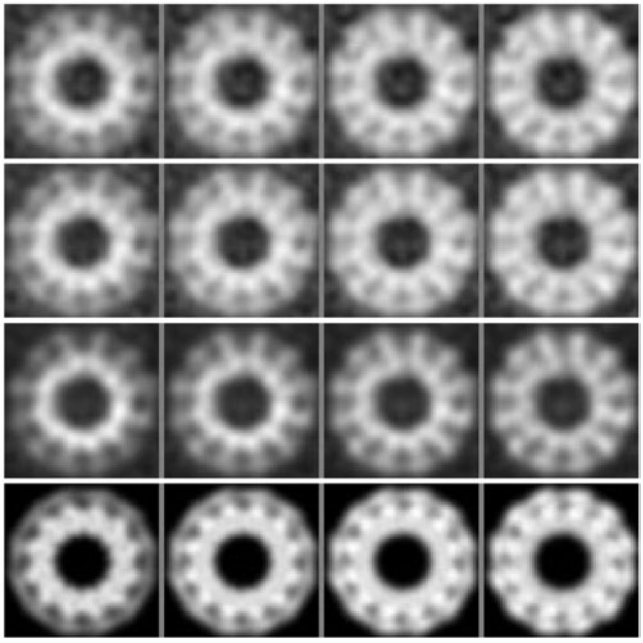

Figure 5. Central slices of the ϕ-29 phantom and its reconstructions from noisy data using ART, ART with phase flipping and the sixth iteration of Chahine's method. From top to bottom: ART; ART with phase flipping; Chahine's method; phantom.

of $F(c^{(k)})$ and $\eta^{(k)}$ for $k = 1, 2, \ldots$. The corresponding values for the initial guess are $F(c^{(0)}) = -1.067 \times 10^{+11}$ and $\eta^{(0)} = 0.8089$. This example of a comparison between Chahine's method and ART in the unrealistic noiseless and aberrationless case is representative of a large set of such comparisons that we have performed.

Next, in figure 5, we present reconstructions from noisy data in the presence of CTFs using three different methods: ART with no CTF correction, ART with CTF phase flipping and Chahine's method. The synthetic data were generated from projections from a random set of directions that was subjected to the CTFs described in figure 1 and further corrupted by noise. The SNR is 1/3 on average. The relaxation parameter used for ART is 0.06. ART results are reported after two cycles through all the projections, while the results of Chahine's method are shown after the sixth iteration.

The evolution of the objective function and of the squared relative residual error in Chahine's method is presented in figure 6. The corresponding values for the initial 1-valued estimate $c^{(0)}$ are $F(c^{(0)}) = -4.448 \times 10^{+11}$ and $\eta^{(0)} = 0.4229$, which means that most of the improvement was achieved in the first iteration.

In order to assess objectively the advantage of using Chahine's method over the other two algorithms, we present in figure 7 a plot of the Fourier shell correlation (FSC) between the phantom and the reconstructions. The FSC between two volumes is defined as follows. Given the Fourier transform of the phantom $\mathcal{F}[p](\boldsymbol{R})$ and of the reconstruction $\mathcal{F}[r](\boldsymbol{R})$, the FSC at a given frequency $\boldsymbol{R}$ with a shell width $\Delta\boldsymbol{R}$ is

$$\mathrm{FSC}(\boldsymbol{R}, \Delta\boldsymbol{R}) = \frac{\sum_{R \leqslant R' \leqslant R+\Delta R} \mathcal{F}[p](\boldsymbol{R}')\mathcal{F}^*[r](\boldsymbol{R}')}{(\sum_{R \leqslant R' \leqslant R+\Delta R} |\mathcal{F}[p](\boldsymbol{R}')|^2)^{1/2} (\sum_{R \leqslant R' \leqslant R+\Delta R} |\mathcal{F}[r](\boldsymbol{R}')|^2)^{1/2}}.$$

See [31] for more information about the FSC. As can be seen in figure 7, the correlation between the Chahine reconstruction and the phantom is better than the correlation between either of the

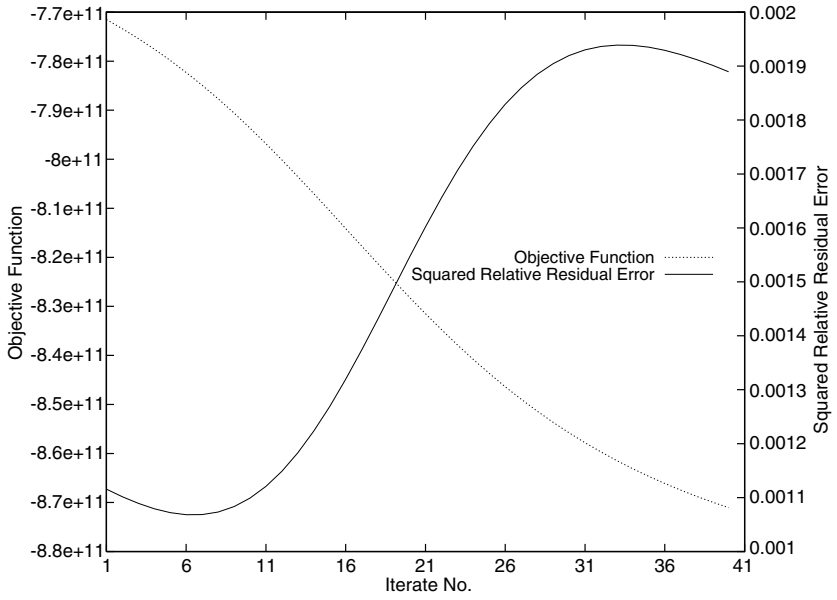

Figure 6. Evolution of the objective function $F(c^{(k)})$ and of the squared relative residual error $\eta^{(k)}$, with $k = 1, 2, \ldots$ for the noisy ϕ-29 example by Chahine's method.

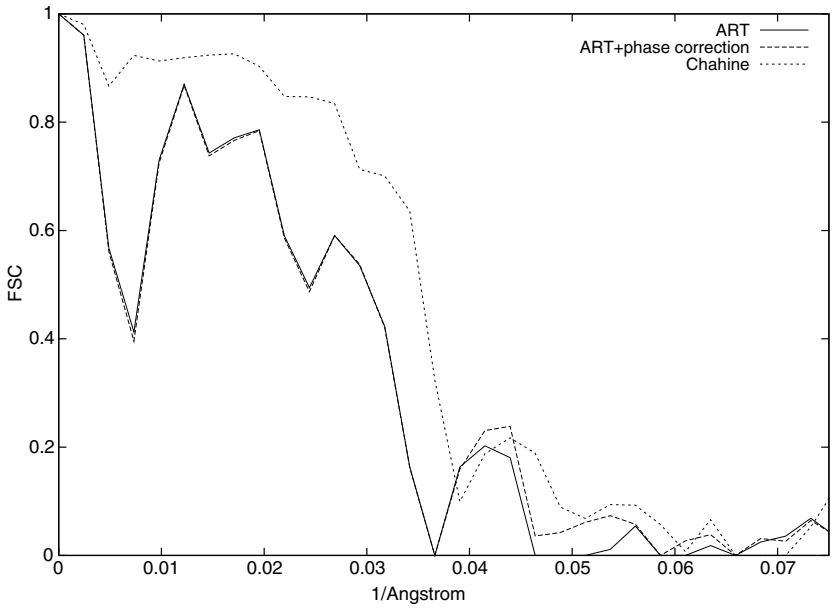

Figure 7. Correlation between the Fourier transform of the phantom and of the reconstruction as a function of frequency shell radius.

ART-based reconstructions and the phantom for all the frequencies at which the phantom and the reconstructions are consistent. (The term 'consistent' does not have a standard definition in the field of electron microscopy, but usually a reconstruction is considered consistent with a phantom at those frequencies at which the FSC is greater than 0.5.)

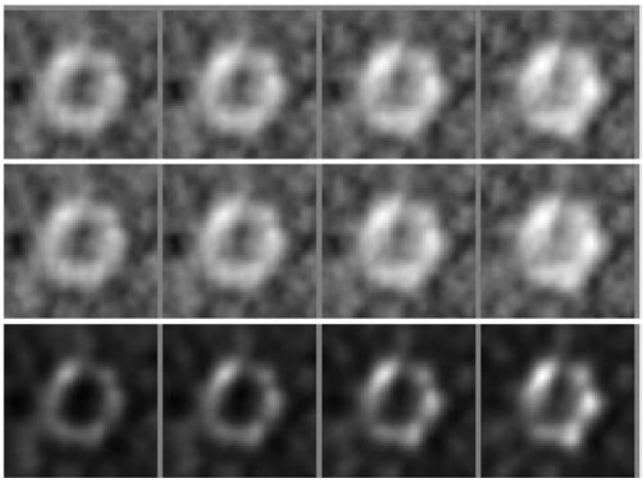

Figure 8. Central slices of reconstructions from experimental data. Top: ART reconstruction from uncorrected measurements. Middle: ART with CTF phase flipping. Bottom: Chahine's method after 11 iterations.

We stress once more that in the above (and in the following subsection) the experiments were carried out not for the purpose of testing the effectiveness of algorithms (ART or Chahine's) to solve a system of equations, but rather the effectiveness of different systems of equations (due to the choices of $\mathcal{D}$ and $\mathcal{E}$ in section 1), each combined with an appropriate algorithm for solving the equations. Thus the data used in the comparisons are the same for the three methods that are compared, but the methods differ both in the systems of equations to be solved and in the algorithms used for their solution. However, the methods are designed to solve the same physical problem: the estimation of c from the measured data y.

5.2. Real data reconstructions

We now present central slices of reconstructions using experimental data from cryomicroscopic projections of the Large T-antigen of the virus SV40. The angle information was obtained by *model based orientation search* [26] and the CTF information was obtained using the technique in [32]. We compare the reconstruction obtained by Chahine's method with the results obtained by ART with no CTF correction and with CTF phase flipping. The results are reported in figures 8 and 9.

6. Conclusions and suggestions for further research

We have shown that it is possible to replace the normal equations

$$\mathcal{A}c = \mathcal{B}\mathcal{H}^{\mathrm{T}}y \qquad \text{with} \qquad \mathcal{A} = \mathcal{B}\mathcal{H}^{\mathrm{T}}\mathcal{H}\mathcal{P}$$

by processing the data y in such a way that the resulting system

$$\tilde{\mathcal{A}}c = \tilde{d}$$

satisfies the hypotheses of theorem 1, thus ensuring the convergence of the Chahine's algorithm iterates. The resulting method was implemented and was successfully tested with simulated as well as with experimental data. The reconstructed images that we have presented here are typical of a much larger set of experiments we performed. Based on the experiments and on

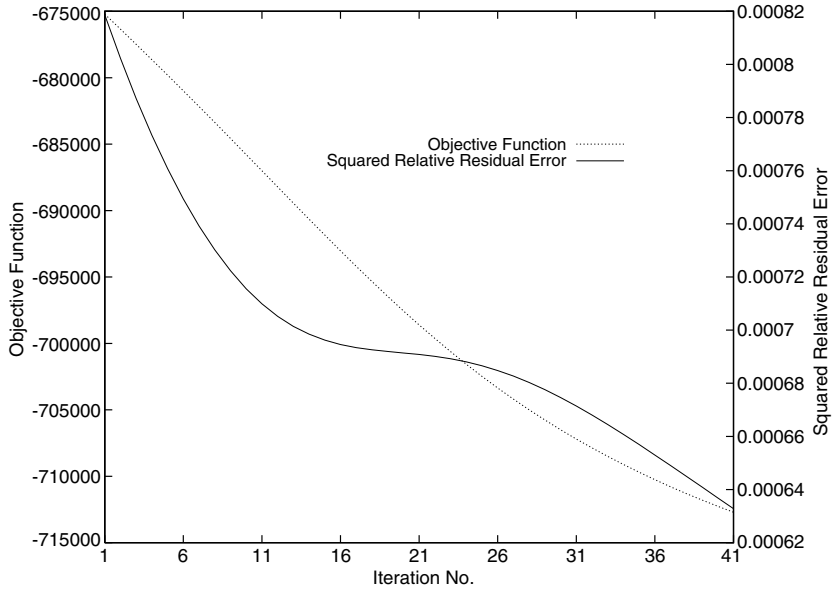

Figure 9. Evolution of the objective function $F(c^{(k)})$ and of the squared relative residual error $\eta^{(k)}$, with $k = 1, 2, \ldots, 41$ for the real data reconstruction using Chahine's method. The corresponding values for the initial unit-valued estimate were $F(c^{(0)}) = -242\,114$ and $\eta^{(0)} = 0.641\,133$.

the FSC curves we have a strong indication that Chahine's algorithm is a promising technique for three-dimensional image reconstruction in electron microscopy when implemented with the data processing described in sections 3 and 4. Furthermore, the results in figure 7 indicate that the correlation between the Chahine reconstruction and the phantom is better than the correlation between any of the ART based reconstructions for a substantial frequency range, including those for which the phantom and the reconstructions are consistent.

In future work, we plan to address in detail the issue of regularization. We point out that Chahine's method is naturally amenable to regularization with a minimal additional cost. Indeed, following the standard procedure of regularizing the problem $\mathcal{A}c = b$ by means of finding $c^{(\epsilon)}$ such that $(\mathcal{A} + \epsilon I)c^{(\epsilon)} = b$, a moment's thought indicates that the iteration $c_j^{n+1} = b_j/(\epsilon + ((\mathcal{A}c^n)_j)/c_j^n))$ performs such regularization with only one additional operation per coefficient.

Concerning the need to find the stopping point of the iterative procedure, we did not make this automatic in the code. We instead followed the plot of $F(c^{(n)})$. This is the function that is being minimized by Chahine's method. As it turns out a few iterations sufficed as shown in figures 4 and 6 as well as in a large number of other experiments we performed.

The technique described in section 3 can also be applied to other reconstruction problems of similar structure; for example, if the data y were treated by Wiener filtering. In this case, the PSF h_i would be substituted by

$$\mathcal{F}^{-1}\left(\frac{H_i}{\sum_j H_j^2 + \epsilon}\right),$$

where ϵ is the signal to noise ratio. This approach is a natural one for following up the present work.

Acknowledgments

This work was supported by CNPq (Brazil) through grant 30.1003, by the Comisión Interministerial de Ciencia y Tecnología (Spain) through grants BIO 98-0761 and BIO 2001-1237, and by the National Institute of Health (USA) through grants HL67465 and HL740472. The implementation in the present work made extensive use of the X-Windows-based Microscopy Image Processing Package (XMIPP) which is available from http://www.cnb.uam.es/~bioinfo/. The authors thank Professor Alvaro De Pierro for providing a number of useful references. The experimental data used in the reconstructions were kindly provided by Dr M Gómez-Lorenzo at the Biocomputing Unit, Centro Nacional de Biotecnología, CSIC, Spain.

References

[1] Barcena M, Ruiz T, Donate L E, Brown S, Dixon N E, Radermacher M and Carazo J 2001 The DnaB.DnaC complex: a structure based on dimers assembled around an occluded channel *EMBO J.* **20** 1462–8
[2] Barcilon V 1975 On Chahine's relaxation method for the radiative transfer equation *J. Atmos. Sci.* **32** 1626–30
[3] Chahine M 1975 Inverse problems in radiative transfer *J. Atmos. Sci.* **27** 960–7
[4] Daube-Witherspoon M and Muehllehner G 1987 An iterative image space reconstruction algorithm for volume ECT *IEEE Trans. Med. Imaging* **5** 61–6
[5] De Pierro A R 1987 On the convergence of the iterative image space reconstruction algorithm for volume ECT *IEEE Trans. Med. Imaging* **6** 174–5
[6] De Pierro A R 1990 Nonlinear relaxation methods for solving symmetric linear complementarity problems *J. Opt. Theor. Appl.* **64** 87–99
[7] Eggermont P P B 1990 Multiplicative iterative algorithms for convex programming (Linear algebra in image reconstruction from projections) *Linear Algebra Appl.* **130** 25–42
[8] Eggermont P P B 1991 On Volterra–Lotka differential equations and multiplicative algorithms for monotone complementarity problems *Mathematical Methods in Tomography (Oberwolfach, 1990) (Springer Lecture Notes in Mathematics vol 1497)* (Berlin: Springer) pp 141–52
[9] Frank J 1996 *Three-Dimensional Electron Microscopy of Macromolecular Assemblies* (New York: Academic)
[10] Frank J 2002 Single-particle imaging of macromolecules by cryo-electron microscopy *Ann. Rev. Biophys. Biomol. Struct.* 303–19
[11] Frank J and Penczek P 1995 On the correction of the contrast function in biological electron microscopy *Optik* **98** 125–9
[12] Gold R and Scofield N 1960 Iterative solution for matrix representation of detection systems *Bull. Am. Phys. Soc.* **2** 276
[13] Grigorieff N 1998 Three-dimensional structure of bovine NADH: ubiquinone oxidoreductase (complex I) at 22 A in ice *J. Mol. Biol.* **277** 1033–46
[14] Herman G T 1980 *Image Reconstruction from Projections: the Fundamentals of Computerized Tomography* (New York: Academic)
[15] Herman G T, Levkowitz H, McCormick S and Tuy H K 1984 Multilevel image reconstruction *Multiresolution Image Processing and Analysis* ed A Rosenfeld (New York: Springer)
[16] Lewitt R M 1990 Multidimensional image representations using generalized Kaiser–Bessel window functions *J. Opt. Soc. Am.* A **7** 1834–46
[17] Lewitt R M 1992 Alternatives to voxels for image representation in iterative reconstruction algorithms *Phys. Med. Biol.* **37** 705–16
[18] Ludtke S, Baldwin P R and Chiu W 1999 EMAN: semiautomated software for high-resolution single-particle reconstructions *J. Struct. Biol.* **128** 82–97
[19] Ludtke S, Jakana J, Song J, Chuang D and Chiu W 2001 A 11.5 A single-particle reconstruction of GroEL using EMAN *J. Mol. Biol.* **314** 82–97
[20] Marabini R, Herman G T and Carazo J M 1998 3D reconstruction in electron microscopy using ART with smooth spherically symmetric volume elements (blobs) *Ultramicroscopy* **72** 53–65
[21] Marabini R, Rietzel E, Schröder R, Herman G T and Carazo J M 1997 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 *J. Struct. Biol.* **120** 363–71

[22] Matej S and Lewitt R M 1995 Efficient 3D grids for image reconstruction using spherically-symmetric volume elements *IEEE Trans. Nucl. Sci.* **42** 1361–70
[23] Matej S and Lewitt R M 1996 Practical considerations for 3-D image reconstruction using spherically symmetric volume elements *IEEE Trans. Med. Imaging* **15** 68–78
[24] Natterer F 2001 *The Mathematics of Computerized Tomography* (Philadelphia, PA: SIAM) (reprint of the 1986 original)
[25] Natterer F and Wübbeling F 2001 *Mathematical Methods in Image Reconstruction* (Philadelphia, PA: SIAM)
[26] Radermacher M 1994 3-dimensional reconstruction from random projections—orientational alignment via Radon transforms *Ultramicroscopy* **53** 121–36
[27] Skoglund U, Ofverstedt L, Burnett R and Bricogne G 1996 Maximum-entropy three-dimensional reconstruction with deconvolution of the contrast transfer function: a test application with adenovirus *J. Struct. Biol.* **117** 173–88
[28] Stark H, Orlova E, Rinke-Appel J, Junke N, Mueller F, Rodnina M, Wintermeyer W, Brimacomber R and van Heel M 1997 Arrangement of tRNAs in pre- and posttranslocational ribosomes revealed by electron cryomicroscopy *Cell* **88** 19–29
[29] Thon F 1966 Zur defokussierungsabhngigkeit des phasenkontrastes bei der elektronenmikroskopischen abbildung *Z. Naturf.* a **21** 476–8
[30] Unwin P N T 1973 Phase contrast electron microscopy of biological materials *J. Microsc.* **98** 299–312
[31] van Heel M 1987 Similarity measures between images *Ultramicroscopy* **21** 95–100
[32] Velázquez-Muriel J A, Sorzano C O S, Fernández J J and Carazo J M 2002 A method for estimating the CTF in electron microscopy based on ARMA models and parameter adjusting *Ultramicroscopy* at press
[33] Wade R 1996 A brief look at imaging and contrast transfer *Ultramicroscopy* **46** 145–56
[34] Zhu J, Penczek P, Schröder R and Frank J 1997 Three-dimensional reconstruction with contrast transfer function correction from energy-filtered cryoelectron micrographs: procedure and application to the 70s *Escherichia coli* ribosome *J. Struct. Biol.* **118** 197–219