



**Repositorio Institucional de la Universidad Autónoma de Madrid**

<https://repositorio.uam.es>

Esta es la **versión de autor** de la comunicación de congreso publicada en:  
This is an **author produced version** of a paper published in:

Annals of Operations Research 148.1 (2006): 133-165

**DOI:** <http://dx.doi.org/10.1007/s10479-006-0078-8>

**Copyright:** © Springer-Verlag Berlin Heidelberg 2006

El acceso a la versión del editor puede requerir la suscripción del recurso  
Access to the published version may require subscription

# Optimization problems in electron microscopy of single particles

C. O. S. Sorzano · R. Marabini · A. Pascual-Montano ·  
S. H. W. Scheres · J. M. Carazo

© Springer Science + Business Media, LLC 2006

**Abstract** Electron Microscopy is a valuable tool for the elucidation of the three-dimensional structure of macromolecular complexes. Knowledge about the macromolecular structure provides important information about its function and how it is carried out. This work addresses the issue of three-dimensional reconstruction of biological macromolecules from electron microscopy images. In particular, it focuses on a methodology known as “single-particles” and makes a thorough review of all those steps that can be expressed as an optimization problem. In spite of important advances in recent years, there are still unresolved challenges in the field that offer an excellent testbed for new and more powerful optimization techniques.

**Keywords** Optimization · Electron microscopy · Single particles

**AMS Subject Classification** 92C55 · 90C90 · 44A12 · 78A15 · 65R32

## Introduction

The structure determination of large macromolecular complexes is having a tremendous impact in understanding the molecular machinery. This knowledge is crucial, for example, for

---

C. O. S. Sorzano (✉)

Department Ingeniería de Sistemas Electrónicos y de Telecomunicación, Univ. San Pablo-CEU, Campus Urb. Montepíncipe s/n, 28668 Boadilla del Monte, Madrid, Spain  
e-mail: coss.eps@ceu.es

R. Marabini

Escuela Politécnica Superior, Univ. Autónoma de Madrid, 28049 Cantoblanco, Madrid, Spain

A. Pascual-Montano

Department de Arquitectura de Computadores y Automática, Facultad de C.C. Físicas, Univ. Complutense, Ciudad Universitaria s/n, 28040, Madrid, Spain

S. H. W. Scheres · J. M. Carazo

Unidad de Biocomputación, Centro Nacional de Biotecnología (CSIC), Campus Univ. Autónoma s/n, 28049 Cantoblanco, Madrid, Spain

designing target-specific drugs and characterizing pathological vs. non-pathological situations (Sali et al., 2003; Thornton, Todd, and Milburn, 2000).

Three-dimensional electron microscopy (3DEM) allows imaging of large biological macromolecules nearly in their native state. 3DEM of single particles is able to produce 3D structures with resolutions between 6 and 30 Å ( $1 \text{ \AA} = 10^{-10} \text{ m}$ ). This range of resolution allows discerning biologically relevant information regarding molecular shape, domain architecture, and, on the high-resolution end, secondary structure (Frank, 2002). Substantial efforts have been invested in developing a technology that allows high resolution macromolecular imaging although it should be realized from the onset that micrographs are, in many cases, altered images of the native structure of the macromolecule, which are very noisy due to the random nature of the electron beam. Furthermore, the electron microscope introduces aberrations in the imaging process. And, finally, it must be taken into account that micrographs represent 2D projections of 3D dimensional objects, i.e., all depth information is lost during the image acquisition.

Different techniques have been developed within the last forty years to solve part of these problems. Not surprisingly, the success of EM particle analysis has been highly correlated with methodological advances and the continuous development of various EM image processing packages.

From an image processing point of view there are four main problems involved in 3D reconstruction using electron microscopy data:

- Biological material is very sensitive to radiation. So, in order to keep high resolution information, very low electron doses should be used to obtain the images. This results in extremely noisy images.
- The original 3-D structure can be reconstructed (up to a certain resolution), combining images of the molecule in various orientations. In order to combine these images we need to know their relative orientations. A task that is far from trivial.
- The aberrations introduced by the electron microscope are not negligible and need to be corrected so that the micrographs faithfully represent X-ray projections of the specimen under study.
- Due to the low quality of the images, it is very difficult to separate in a micrograph images coming from different kind of molecules. Even if the sample is biochemically pure (i.e., there is a single type of protein), biological molecules are flexible objects and the same protein may exhibit several conformations.

Although many methodologies have been proposed to achieve a complete 3D reconstruction, the problems involved in the process are very complex and, thus, only remain partially solved. In this review we aim at describing the 3D reconstruction workflow of single particles from A to Z making emphasis on those most critical steps in the workflow. We will concentrate on the formulation of the optimization problems involved. The spectrum of optimization techniques used in the field is very wide and ranges from discrete optimizers to continuous ones, from deterministic to stochastic, with or without constraints, with one or multiple objective functions, etc. Our intention is also to introduce the state-of-the-art of the optimization techniques employed in the field encouraging researchers to propose new methods that allow to improve the resolution of the 3D reconstructed particles. This will provide key structural information about the way in which macromolecular machines perform their functions in live cells.

The rest of the paper is organized as follows: Section 1 describes the electron microscope, the biological material that is imaged, and the image formation model. The image processing workflow commonly employed in the study of single particles is described at the end of this

section. Each of the steps of the workflow is described in detail in Section 2. For each step, we provide a general overview and formulate the optimization problem related to the topic. Then, we review the solutions currently devised. We conclude in Section 3 by indicating those problems in the field that in our opinion remain still open.

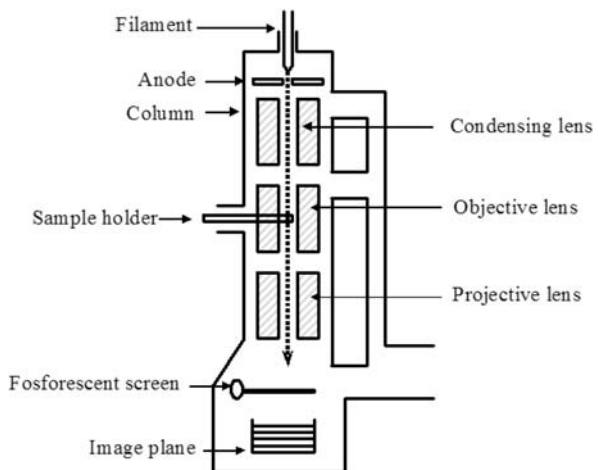
## 1. Electron microscopy principles

### 1.1. Electron microscope

The electron microscope is a device that uses highly accelerated electrons, focused with electromagnetic “lenses”, to obtain images of the specimen under study (Frank, 1996, 2006; Lenz, 1971). The disposition of the electron microscope is similar to that of the optical microscope (see Fig. 1), but upside down. The source of illumination is a filament (cathode) that emits the electrons. Since electrons are scattered by air molecules, the air must be removed by creating a high vacuum. The electrons are accelerated from the cathode to a nearby anode (electric potentials in the order of 200 kV or higher are typically used). Magnetic coils act as lenses and focus the electron beam crossing the specimen. The outgoing electron beam is recorded by a photographic plate or a CCD array. Most of the electrons never interact with the specimen and only contribute to form a background noise. A few electrons will interact elastically (without changing their energy) with the specimen and, finally, a negligible amount will interact strongly (inelastic scattering).

Under the weak-phase object approximation (i.e., inelastic scattering can be neglected compared to elastic scattering) the image acquired (projection image) can be modeled mathematically as the X-ray transform of the Coulomb potential of the specimen. The Coulomb potential is a three-dimensional function  $f(\mathbf{r})$  where  $\mathbf{r} \in \mathbb{R}^3$ , while its X-ray transform is a two-dimensional function. That is, the micrographs are projections obtained by projecting a rotated and translated version of the sample function  $f$  onto the XY plane. The translation is described by a 2D vector  $\mathbf{s}_0$ . The rotation is usually given by three Euler angles  $\phi, \theta, \psi$ . The projection at a given point  $\mathbf{s} \in \mathbb{R}^2$  is defined as the line integral of the rotated and translated

**Fig. 1** Schematic representation of an electron microscope (San Martín, 1996)



volume that passes through  $\mathbf{s}$ :

$$p_{\phi, \theta, \psi, \mathbf{s}_0}(\mathbf{s}) = \int_{\mathbb{R}} f(E_{\phi, \theta, \psi}^{-1}(H^t(\mathbf{s} - \mathbf{s}_0) + l\mathbf{e}_3)) dl, \quad (1)$$

where  $H$  is the projection onto the XY plane operator,  $H = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ ,  $E_{\phi, \theta, \psi} = R_z(\phi)R_y(\theta)R_z(\psi)$  is the Euler rotation matrix around Z, Y and new Z, and  $\mathbf{e}_3 = (0, 0, 1)$ .

As any other imaging device, the electron microscope introduces some distortion in the acquired images. This distortion is usually modeled in a first order approximation by the convolution with a Point Spread Function (PSF). Its dual function in Fourier space is called the Contrast Transfer Function (CTF) and, in the particular case of EM, it is real-valued. The CTF looks like a damped two-dimensional sine wave (Frank, 1996; Unwin, 1973). The effect of the CTF is twofold: it introduces zones of alternate contrast (some components are projected as white on a black background, while others are projected as black on a white background) and it introduces low pass and band pass filtration.

## 1.2. The sample biological material

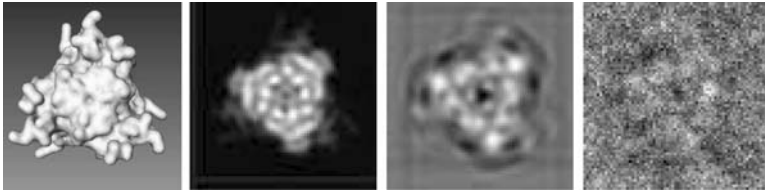
Before taking into account the reconstruction problem itself, we should discuss the kind of object to be reconstructed and its behavior during the recording process. Biological macromolecules are small. Their size ranges from 100 to 10,000 Å. This small size implies that a direct manipulation is extremely difficult, if at all possible, and can only be performed under rather restrained conditions, which represents an obstacle for their characterization.

The conditions inside the electron microscope, high vacuum and high electron radiation level, are very deleterious for the specimens, which should therefore be protected somehow (for example by embedding the sample in ice). This protection has as a side effect in that it decreases the signal-to-noise ratio (SNR, Jain (1989)). In addition, the problem of beam-induced damage is by no means negligible. Electron radiation induces intense ionization of the sample with the formation of free radicals and ions that produce important alterations of the structure. In order to minimize this damage, very low electron doses are used, which in turn produce images with extremely low SNR. Typically observed SNRs can be as low as 1/10.

The solution devised for improving the poor SNR in the micrographs has been to “average” over many (thousands) of identical copies of the specimen. This “averaging” process is mathematically known as the tomography problem (Herman, 1980; Kak and Slaney, 1987; Natterer and Wübbeling, 2001) in which 2D projections are combined in a 3D volume. This can be done directly in the case of 2D crystals, where particles are *a priori* ordered (a crystal is a structure made by an object that repeats itself following a regular pattern), or in the case of single particles (i.e., identical copies of a molecule that are recorded in random orientations inside the electron microscope) only after translational and rotational alignment.

## 1.3. Image formation model

As already introduced, the SNR in EM image processing is very low. Noise is generated by many sources. Among others the low, and possibly varying, electron dose, the random nature of the electron emission, the interaction of the electrons with the sample holder, the granular composition of the film where the image is recorded, the electronic noise of the scanner used to digitize the image, etc. (If a CCD camera is used instead of the film, then the electronic noise due to the CCD diodes must also be taken into account.) The resulting noise has



**Fig. 2** From left to right: top view of the isosurface of the bacteriorhodopsin, ideal projection (at  $1/7 \text{ \AA}^{-1}$  resolution) obtained from the top, projection obtained after applying the CTF, projection obtained after applying the CTF and adding noise

been shown to be additive and normally distributed (Frank and Al-Ali, 1975; Sorzano et al., 2004a). This helps simplifying the mathematical formulation of many of the optimization problems involved from the image acquisition step towards the 3D reconstruction of the macromolecule. For a complete description of the image formation model, the noise before and after the CTF needs to be considered ( $n_b(\mathbf{s})$  and  $n_a(\mathbf{s})$ , respectively) (Velázquez-Muriel et al., 2003).

Adding the noise dependent term to (1) and taking into account the effect of the PSF, the complete image formation model can be formulated as

$$p_{\phi, \theta, \psi, \mathbf{s}_0}(\mathbf{s}) = PSF(\mathbf{s}) \star \left( \int_{\mathbb{R}} f(E_{\phi, \theta, \psi}^{-1}(H^t(\mathbf{s} - \mathbf{s}_0) + l\mathbf{e}_3)) dl + n_b(\mathbf{s}) \right) + n_a(\mathbf{s}), \quad (2)$$

where  $\star$  represents the convolution operator. The effect of the noise and the CTF can be seen for the top view of the bacteriorhodopsin (a protein whose structure is known at atomic resolution) in Fig. 2. However, as will be seen in the following, many image processing algorithms assume a simpler image formation model where the CTF is not taken into account:

$$p_{\phi, \theta, \psi, \mathbf{s}_0}(\mathbf{s}) = \int_{\mathbb{R}} f(E_{\phi, \theta, \psi}^{-1}(H^t(\mathbf{s} - \mathbf{s}_0) + l\mathbf{e}_3)) dl + n(\mathbf{s}). \quad (3)$$

In this paper, the noisy projections will be referred to as  $y$ , the ideal projections as  $x$ , and this simple image formation model as additive-noise:  $y = x + n$ .

#### 1.4. 3D Reconstruction work-flow

Different approaches have been devised to reconstruct 3D structures from their EM projections. These approaches can be classified depending on the kind of data they work with, more specifically on the kind of symmetry that the imaged particle exhibits. In the case of helical filaments, a single view carries enough information to reconstruct the specimen up to certain resolution (De Rosier and Klug, 1968). Other types of symmetry that are typically encountered for biological macromolecules are: 2D-crystals (Ellis and Hebert, 2001) and icosahedral viruses Baker, Olson, and Fuller (1999). For the general case, however, we cannot count on symmetry. In the rest of this paper we will focus on the latter case, which is termed *single particle reconstruction*.

The process followed to obtain a 3D-reconstruction for single particles can be briefly described as follows (only those steps related to the digital image processing will be enumerated):

1. Images containing many identical copies of the specimen are recorded in the electron microscope and converted to digital form.
2. Micrographs may be preprocessed: (i) aberrations introduced by the microscope (CTF) are estimated and corrected, (ii) images are denoised.
3. Particle projections are identified and extracted from the micrographs.
4. Projections are normalized, aligned and classified (the particles are classified to distinguish possible structural variability, different projection directions or contaminating particles). This is an iterative process, the better the particles are aligned the better they may be classified, and vice-versa.
5. Finally, when a structurally homogeneous and aligned set of particles has been obtained, it can be combined to obtain a volume.

The whole procedure is iterative, since a first rough reconstruction helps to better identify, classify and align the 2D projections. The newly aligned projections are then used to build a finer reconstruction which in turn is again used to align the 2D projections. This process is iterated until convergence (usually defined as no significant change of the projection alignment, or no significant improvement of the resolution achieved.)

## 2. Optimization problems

In this section we discuss each one of the image processing problems that are involved in the 3D reconstruction of single particles in EM from an optimization point of view. For each topic we formulate the optimization problem associated and review its current solutions. The section is divided in nine problems (image denoising, 2D alignment, image classification, angular assignment, 3D reconstruction, 3D reconstruction performance, CTF estimation, CTF amplitude correction, and the reconstruction process). These topics have been sorted by their appearance order in the image processing work-flow.

### Problem 1: Image denoising

Due to the extremely low SNR, one of the earliest steps that may be performed on the acquired images is image denoising. Denoised images serve various purposes, like image alignment, automatic particle picking, etc. At this early stage, a simple additive-noise image-formation model is considered:  $y = x + n$ , where  $n$  is assumed to be normally distributed random noise. In EM, this is a very reasonable approximation.

One of the most popular image denoising techniques employs the Fourier transform (Bracewell, 1986). Experimental images are usually filtered using either a low-pass or a band-pass filter since it is assumed that the power of noise is much higher than the power of signal at high frequencies (Jain, 1989). The solution adopted by Fourier filtering is the removal of all high frequency components. Although, in general, the overall SNR is increased, it can be easily seen that the high frequency components are always discarded disregarding whether they belong to the signal or not.

An alternative approach (Sorzano et al., 2006) that is formulated as an optimization problem takes advantage of the additive-noise image-formation model and the linearity of the Discrete Wavelet Transform (Mallat, 1999). Since the noise follows a Gaussian distribution

with zero mean and variance  $N$ ,  $p(n) = G(n, 0, N)$ , and the signal and noise are assumed to be independent, then  $p(y | x) = p(n) = G(y - x, 0, N)$ . The concept of scale in the wavelet transform is related to the concept of frequency in the Fourier transform. Finer scales are related to higher frequencies. For the kind of projection images EM deals with, white noise is a quite limiting assumption. On the contrary, it is reasonable to think that the noise power decreases as the scale becomes finer. In this case, it is better to use a different noise variance at each scale  $s$ ,  $N_s$ . A reasonable model for the wavelet coefficients at a given scale  $s$  states that their distribution is also Gaussian with zero mean and variance  $S_s$  (Bijaoui, 2002),  $p(x_s) = G(x_s, 0, S_s)$ . With these hypothesis, it can be easily shown (Bijaoui, 2002) that the *a posteriori* expected value of  $x$  is given by

$$E \{x_s | y_s\} = y_s \frac{S_s}{S_s + N_s}. \tag{4}$$

This expression weighs the wavelet components according to the probability of coming from the signal distribution or not. Then, the problem now is how to estimate the respective variances of signal and noise at each scale ( $S_s$  and  $N_s$ ).

If an orthogonal wavelet decomposition is used, then the signal and noise power are decomposed at each scale so that the total image power is the same as the power of its wavelet transform

$$\sum y^2 = \sum_s y_s^2 = \sum_s P_s, \tag{5}$$

where

$$P_s = n_s(S_s + N_s), \tag{6}$$

and  $n_s$  is the number of wavelet coefficients at scale  $s$ . This estimation problem must be restricted by some *a priori* knowledge about the problem at hand. First of all, the solutions for  $N_s$  and  $S_s$  must be positive since they represent the variance of some random variable. As has been already mentioned, the signal and noise usually power decrease with the scale  $N_s \leq N_{s+1}$ ,  $S_s \leq S_{s+1}$ . Furthermore, upper and lower bounds of the SNR can be provided

$$SNR_l \leq \frac{\sum_s n_s S_s}{\sum_s n_s N_s} \leq SNR_h. \tag{7}$$

It is also known that for the kind of signals present in EM, most of the energy is concentrated at low frequencies, therefore the SNR must increase as coarser scales are considered

$$\frac{\sum_{i=0}^s n_i S_i}{\sum_{i=0}^s n_i N_i} \leq \frac{\sum_{i=0}^{s+1} n_i S_i}{\sum_{i=0}^{s+1} n_i N_i}. \tag{8}$$

Due to the high number of constraints and the noisy nature of the measurements, it is very unlikely that the equation system in (6) can be solved exactly. Alternatively, a least-squares solution of the equation system is sought subject to a number of linear and nonlinear constraints. The cost function to minimize is  $\|C\mathbf{x} - \mathbf{P}\|^2$  subject to  $A\mathbf{x} \leq \mathbf{b}$ ,



$\sum_{i=0}^s \sum_{j=0}^{s+1} n_i n_j (S_i N_j - S_j N_i) \leq 0$ , and  $A_{eq} \mathbf{x} = \mathbf{b}_{eq}$  where

$$C = \begin{pmatrix} n_1 & 0 & 0 & 0 & \dots & n_1 & 0 & 0 & 0 & \dots \\ 0 & n_2 & 0 & 0 & \dots & 0 & n_2 & 0 & 0 & \dots \\ 0 & 0 & n_3 & 0 & \dots & 0 & 0 & n_3 & 0 & \dots \\ 0 & 0 & 0 & n_4 & \dots & 0 & 0 & 0 & n_4 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \\ \vdots \\ S_1 \\ S_2 \\ S_3 \\ S_4 \\ \vdots \end{pmatrix}, \quad \mathbf{P} = \begin{pmatrix} P_1 \\ P_2 \\ P_3 \\ P_4 \\ \vdots \end{pmatrix},$$

$$A = \begin{pmatrix} 1 & -1 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots \\ 0 & 1 & -1 & 0 & \dots & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 1 & -1 & \dots & 0 & 0 & 0 & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & \dots & 1 & -1 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 & -1 & 0 & \dots \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 & -1 & \dots \\ 0 & 0 & 0 & 0 & \dots & \dots & \dots & \dots & \dots & \dots \\ -1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots \\ 0 & -1 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & -1 & 0 & \dots & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & -1 & \dots & 0 & 0 & 0 & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & \dots & -1 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & \dots & 0 & -1 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & -1 & 0 & \dots \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & -1 & \dots \\ 0 & 0 & 0 & 0 & \dots & \dots & \dots & \dots & \dots & \dots \\ -SNR_h n_1 & -SNR_h n_2 & -SNR_h n_3 & -SNR_h n_4 & \dots & n_1 & n_2 & n_3 & n_4 & \dots \\ SNR_l n_1 & SNR_l n_2 & SNR_l n_3 & SNR_l n_4 & \dots & -n_1 & -n_2 & -n_3 & -n_4 & \dots \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix},$$

$A_{eq} = (n_1 \ n_2 \ \dots \ n_1 \ n_2 \ \dots)$ , and  $\mathbf{b}_{eq} = (\sum_s P_s)$ .

This problem is of the form

$$\begin{aligned} \mathbf{x}^* &= \arg \min_{\mathbf{x}} C(\mathbf{x}) = \arg \min_{\mathbf{x}} \|C\mathbf{x} - \mathbf{P}\|^2 \\ \text{s.t.} \quad & \mathbf{G}_{\text{eq}}(\mathbf{x}) = A_{\text{eq}}\mathbf{x} - \mathbf{b}_{\text{eq}} = \mathbf{0} \\ & \mathbf{G}_{\text{ineq}}(\mathbf{x}) = A\mathbf{x} - \mathbf{b} \leq \mathbf{0} \end{aligned} \tag{9}$$

where  $C(\mathbf{x})$  is a strictly convex function,  $\mathbf{G}_{\text{eq}}(\mathbf{x})$  is linear and  $\mathbf{G}_{\text{ineq}}(\mathbf{x})$  is a convex function if (8) is not considered. Under these circumstances, there exists a unique global solution to the problem. If (8) is taken into account, then the constraints are not convex any more, and local minima may exist.

A Successive Quadratic Programming (SQP) approach is followed to solve the estimation problem. This approach aims at solving local Quadratic Programming (QP) problems by finding solutions of the first order Karush-Kuhn-Tucker (KKT) conditions (Dennis and Schnabel, 1996). The gradient of each of the functions involved is approximated by a finite order approximation if (8) is considered and is analytically computed if it is not. The SQP method takes Newton-like steps and, therefore, it has a fast rate of convergence.

### Problem 2: 2D alignment

There are many complexes that show one or more preferential view directions (Boisset et al., 1998; San Martin et al., 1995; Valle et al., 2000). The analysis of the projections along these directions may reveal biologically relevant features. Image averaging of the individual experimental images is usually performed as a way to increase the SNR. If the image formation model is additive-noise,  $y = x + n$ , and the variance of the noise is  $\sigma^2$ , then averaging over  $N$  images reduces the power of the noise present in the average to  $\frac{\sigma^2}{N}$ . However, image averaging can only be performed if the corresponding experimental images have been previously aligned translationally and rotationally.

Given the ideal image  $x$  and the experimentally measured image  $y$ , the probability of measuring  $y$  after applying some shift  $\mathbf{s}_0 \in \mathbb{R}^2$  and some rotation  $\psi \in \mathbb{R}$  to the image  $x$  is given by the multivariate Gaussian distribution (Sigworth, 1998)

$$p(y | x, \mathbf{s}_0, \psi) = \frac{1}{(2\pi\sigma)^{\frac{N}{2}}} \exp\left(-\frac{1}{2\sigma^2} \|y - x(\mathbf{s}_0, \psi)\|^2\right) \tag{10}$$

It can be shown that the transformation parameters  $(\mathbf{s}_0, \psi)$  that maximize the likelihood of observing the experimental image are those that minimize  $\|y - x(\mathbf{s}_0, \psi)\|^2$ , which in turn are those maximizing the inner product  $\langle y, x(\mathbf{s}_0, \psi) \rangle$ . This inner product is computed as the sum of the product of the pixel values of both images

$$\langle y, x(\mathbf{s}_0, \psi) \rangle = \sum_i y_i x_i(\mathbf{s}_0, \psi), \tag{11}$$

where  $y_i$  denotes the  $i$ -th pixel of image  $y$ . This functional is usually referred to as the correlation between  $y$  and  $x(\mathbf{s}_0, \psi)$ . Thus, aligning two images can be formulated as the following optimization problem:

$$\begin{aligned} \mathbf{s}_0^*, \psi^* &= \arg \max_{\mathbf{s}_0, \psi} \langle y, x(\mathbf{s}_0, \psi) \rangle \\ \text{s.t.} \quad & (\mathbf{s}_0, \psi) \in \mathcal{S} \end{aligned} \tag{12}$$

where  $\mathcal{S}$  introduce our *a priori* knowledge about the limits of these parameters. Note that this criterion is sensitive to scaling differences in the image gray levels, which may be often the case. To avoid this dependency, the correlation index between the two images can be optimized instead. The correlation index normalizes the correlation of the two images by their respective variances ( $\frac{(x,y)}{\|\bar{x}\|\|\bar{y}\|}$ ).

The common approach in EM to optimize this functional is by an exhaustive search of all possible combinations of  $\mathbf{s}_0$  and  $\psi$  within a discrete grid (Frank, 1996). The search in the  $\mathbf{s}_0$  space can be greatly accelerated due to the shift property of the correlation function in Fourier space (Bracewell, 1986).

Maximizing the correlation not always produces the true transformation parameters since the noise creates local minima that would not be present if the measurements were noise-free. In addition, one way to diminish the search time is by proceeding from coarse to fine grids in the search space of  $\mathbf{s}_0$  and  $\psi$ . However, this may increase the chances of getting trapped in a local minimum.

So far, we assumed that the ideal model  $x$  is known. However, this is not true, since the whole point of the problem is to produce an average of the ensemble of experimental images ( $\bar{x}^*$ ) that allows better identification of the structural features visible from that point of view. It is common use to apply an Expectation-Maximization approach (Dempster, Laird, and Rubin, 1977) to solve this problem. In this approach, a current estimate of the ideal image  $\bar{x}^{(t)}$  is refined iteratively, until convergence to  $\bar{x}^*$ .

The Expectation-Maximization approach tries to find the estimate of  $\bar{x}$  that maximizes the joint probability  $p(\bar{x}, Y)$  where  $Y$  is the set of the  $N$  experimental images observed. Due to the presence of the hidden variables  $\mathbf{s}_0$  and  $\psi$ , this joint probability has to be maximized taking into account all the possibilities for these two variables

$$\bar{x}^* = \arg \max_{\bar{x}} \log p(\bar{x}, Y) = \arg \max_{\bar{x}} \sum_{\mathbf{s}_0, \psi} \log p(\bar{x}, Y, \mathbf{s}_0, \psi). \quad (13)$$

The Expectation-Maximization procedure builds a lower bound to  $p(\bar{x}, Y)$  based on the current estimate  $\bar{x}^{(t)}$  during the E-step. This bound will be referred to as  $B(\bar{x}; \bar{x}^{(t)})$ . Then, during the M-step the bound is optimized finding a new estimate  $\bar{x}^{(t+1)}$ . It can be shown that the best lower bound is given by

$$B(\bar{x}; \bar{x}^{(t)}) = \sum_{\mathbf{s}_0, \psi} p(\mathbf{s}_0, \psi | \bar{x}^{(t)}, Y) \log \frac{p(\bar{x}, Y, \mathbf{s}_0, \psi)}{p(\mathbf{s}_0, \psi | \bar{x}^{(t)}, Y)}, \quad (14)$$

and its optimizer is given by

$$\begin{aligned} \bar{x}^{(t+1)} &= \arg \max_{\bar{x}} B(\bar{x}; \bar{x}^{(t)}) = \arg \max_{\bar{x}} \log p(\bar{x}) \\ &\quad + \sum_{\mathbf{s}_0, \psi} p(\mathbf{s}_0, \psi | \bar{x}^{(t)}, Y) \log p(Y, \mathbf{s}_0, \psi | \bar{x}), \end{aligned} \quad (15)$$

It can be seen that in this method it is important to compute the probability distribution of the hidden variables in terms of the observed measurements and the current estimate of the ideal image. This probability can be computed as

$$p(\mathbf{s}_0, \psi | \bar{x}^{(t)}, Y) = \frac{p(Y | \bar{x}^{(t)}, \mathbf{s}_0, \psi) p(\mathbf{s}_0) p(\psi)}{p(Y)}. \quad (16)$$

Since no information is available about the *a priori* distribution of the transformation parameters, its distribution is taken as uniform over a range of plausible values. The same is true about  $p(Y)$ . In this way,  $p(\mathbf{s}_0, \psi | \bar{x}^{(t)}, Y)$  is proportional to  $p(Y | \bar{x}^{(t)}, \mathbf{s}_0, \psi)$  given in (10). Furthermore, the marginal distribution of  $\mathbf{s}_0, \psi$  given  $\bar{x}^{(t)}$  and  $Y$  is usually taken in EM as a delta function  $p(\mathbf{s}_0, \psi | \bar{x}^{(t)}, Y) = \delta(\mathbf{s}_0 - \mathbf{s}_0^*)\delta(\psi - \psi^*)$ , where  $\mathbf{s}_0^*, \psi^*$  are given in (12). Under this assumption and if no prior distribution is provided for  $x$ , the optimizer in (15) can be proved to be the average of the experimental images after application of the shift and rotations in (12).

This result justifies the common practice in EM where an initial guess of the ideal image is chosen arbitrary by the user. Then, one finds the best translation and rotation parameters according to this initial guess. These transformation parameters are applied to the experimental images, and the average of the resulting images is taken as the next estimate of the ideal image. This procedure is iterated until convergence. As was already pointed out, the computations in (12) may be prone to errors due to the elevated levels of noise in the measurements. If the SNR of the experimental images is very low, then this process may not converge to the right solution in practical cases.

Alternatively, the maximum of (13) can be achieved by the following iterative step (Sigworth, 1998)

$$\bar{x}^{(t+1)} = \frac{1}{N} \sum_{i=1}^N \frac{\sum_{\mathbf{s}_0, \psi} y_i(\mathbf{s}_0, \psi) p(y_i | \bar{x}^{(t)}, \mathbf{s}_0, \psi) p(\mathbf{s}_0) p(\psi)}{\sum_{\mathbf{s}_0, \psi} p(y_i | \bar{x}^{(t)}, \mathbf{s}_0, \psi) p(\mathbf{s}_0) p(\psi)} \quad (17)$$

where now  $y_i$  represents the  $i$ -th image in the ensemble. This process can be shown to converge to a local minimum of the likelihood of observing the set of experimental images given the model. In the work of Sigworth (1998) the distribution of  $\mathbf{s}_0$  is assumed to be a multivariate (2D) Gaussian whose parameters are also estimated during the optimization process, as well as the power of the noise in the experimental images  $\sigma^2$ . Simulated as well as experimental data shows that this procedure is capable of yielding better averages than the pseudo Expectation-Maximization approach traditionally followed in single-particles (Sigworth, 1998; Scheres et al., 2005). This alignment procedure can be further extended to the multireference case (Scheres et al., 2005) in which it is assumed that there is not a single average but several ones corresponding to different proteins or protein conformations.

### Problem 3: Image classification

3D Reconstruction algorithms usually assume that the input data is homogenous, that is, a set of projections of the very same object. Unfortunately, this is not true since, even for biochemically pure samples, many proteins present a certain degree of flexibility. Therefore, before applying the reconstruction algorithm, the input data need to be classified into homogenous subsets.

The general approach to image classification first extracts a number of features from each image. These features are arranged in a vector and, then, the classification is performed on this set of vectors. Pixel gray values are usually used as features, but other values might be as well like spectral features (color or tone, gradient, spectral parameter, etc.), geometric features (edge, shape, size, etc.), or textural features (pattern, spatial frequency, homogeneity, etc.). The main particularity of EM data is that images are extremely noisy, and the feature vectors have a high variability even if they belong to the same class. In this context, classification methods are pushed to their limits and very robust algorithms are required.

Different methods for single particle image classification have been proposed. The most popular approach used in this field are based on a combination of Multivariate Statistical Analysis (MSA) (Frank and van Heel, 1982; van Heel and Frank, 1981) and Hierarchical ascendant classification (HAC) (van Heel, 1984). MSA is used here in order to reduce the number of variables characterizing an image while HAC is then performed for unsupervised image classification (clustering) in this low-dimensional space. Hybrid (C-means and HAC) classification approach (Frank, Chiu, and Degn, 1988b), Fuzzy C-means (Carazo et al., 1990), Self-Organizing Maps (Marabini and Carazo, 1994) and some of its variants (Pascual-Marqui et al., 2001; Pascual-Montano et al., 2001) are also good techniques widely used in the EM field. In this review we will focus our attention on the most popular methods in the field.

### *Hierarchical ascendant classification*

One of the most spread classification algorithms in 3DEM is HAC (van Heel, 1984; Frank et al., 1988a). This classification algorithm partitions the data in a set of nested spaces represented in a binary tree (also called dendogram). Sectioning the tree at a given level, splits the data in  $g$  disjoint groups. The sectioning level (or threshold) represents the maximum within-group distance that is allowed. In other words, two groups are different at a given threshold if their distance is greater than the threshold.

There are two ways of computing the dendogram: divisive or agglomerative. In the first approach, the whole dataset is successively split in two halves until there are as many groups as images (each group formed by a single image). The algorithms available for this approach are quite inefficient. The second approach starts with as many groups as images. The two closest groups are joined forming a single group, thus, reducing the number of groups. This process is iterated until only one group is left. This procedure is said to be ascendant and is the common choice in 3DEM.

The key to build the hierarchical tree is how to measure the distance between two groups. This is done in 3DEM by adopting a merging criterion due to Ward (1963) that minimizes the inner dissimilarity of each cluster.

Given a set of feature vectors  $\mathcal{Y} = \{y_i \in \mathbb{R}^v\}$  ( $v$  is the number of features available,  $i = 1, \dots, N$  being  $N$  the number of images), the problem is to find the partition  $\Pi^*$  of  $\mathcal{Y}$  minimizing

$$\Pi^* = \arg \min_{\Pi} \sum_{C \in \Pi} \sum_{y_i \in C} \|y_i - \bar{C}\|^2 \quad (18)$$

where  $C$  is each of the elements (classes) of the partition  $\Pi$ , and  $\bar{C}$  is the centroid (arithmetic mean) of the  $y_i$  vectors assigned to the class  $C$ .

In Electron Microscopy, images are usually projected onto a small number of Principal Components (Frank and van Heel, 1982; van Heel and Frank, 1981; van Heel, 1984). In this way, each image is represented by a feature vector formed by its projection onto the selected number of eigenvectors (between 5 and 15 eigenvectors are used). Euclidean distance is employed in Electron Microscopy to measure the intra-cluster dissimilarity. However, other measures can be used giving raise to the generalized Ward clustering problem (Batagelj, 1988).

In general, there are no efficient and exact algorithms to solve the clustering problem (Batagelj, 1988). Instead, suboptimal algorithms are employed. The algorithm in Electron Microscopy corresponds to a greedy heuristic that leads to an almost optimal partition of

the data space. The key steps of the agglomerative algorithm is the update of the matrix of distances between groups after the joining two clusters and the computation of the joint centroid. Joining group  $i$  with group  $j$  results in a new group  $i + j$  whose centroid is given by Frank et al. (1988a)

$$\bar{C}_{i+j} = \frac{n_i \bar{C}_i + n_j \bar{C}_j}{n_i + n_j}, \quad (19)$$

where  $n_i$  is the number of elements in the class  $i$  and  $\bar{C}_i$  is their centroid. The new distance matrix can be efficiently updated by Webb (2002, Chap. 10).

$$d_{i+j,k} = \frac{(n_k + n_i)d_{i,k} + (n_k + n_j)d_{j,k} - n_k d_{i,j}}{n_k + n_i + n_j}, \quad (20)$$

where  $d_{i+j,k}$  is the distance between the new group  $i + j$  and any other group  $k$ . Actually, this update formula is a particular case of the Lance-Williams-Jambu formula for updating dissimilarities between clusters (Batagelj, 1988).

The agglomerative procedure can be represented in a hierarchical tree (also called dendrogram) whose root is the single cluster with all data. The tree is carefully drawn so that the point at which two branches are joined represents the distance between the two joining groups. This dendrogram is used to divide the data into classes by setting a threshold on the maximum distance allowed between groups. This threshold implies a cut in the dendrogram that gives  $C$  different classes.

The user is often asked to cut the dendrogram by selecting the appropriate threshold based on her expertise and previous knowledge. However, there are a number of heuristics helping her decision. In Electron Microscopy the so-called "aggregation index" is used (Frank et al., 1988a). The aggregation index is defined as

$$\Delta I_{i,j} = \frac{n_i n_j}{n_i + n_j} \|\bar{C}_i - \bar{C}_j\|^2. \quad (21)$$

Nodes are ranked by aggregation index as a measure of the significance of the partitioning at that level. The cutting of the tree is done on the lowest level associated with a significant aggregation index (Zupan, 1982).

### *Self-organizing maps*

We start our discussion with the well-known C-means algorithm (Bishop, 1995). Given a set of feature vectors  $\mathcal{Y} = \{y_i \in \mathbb{R}^v\}$  ( $v$  is the number of features available,  $i = 1, \dots, N$  being  $N$  the number of images), we try to find some vectors  $\mathcal{V} = \{v_j \in \mathbb{R}^v\}$  ( $j = 1, \dots, C$ ) that better discriminate between data classes. Implicitly, we have assumed that there exist exactly  $C$  classes in the data. The  $v_j$  are referred to as class representatives, class centroids, or code vectors. The input data is assigned to the class whose code vector minimizes the Euclidean distance to the given input vector. Thus, finding the code vectors is reduced to the following minimization problem

$$\mathcal{V}^* = \arg \min_{\mathcal{V}} \sum_{i=1}^N \sum_{j=1}^C u_{ji} \|y_i - v_j\|^2, \quad (22)$$

where  $u_{ji} = 1$  if the  $j$ -th code vector is the closest code vector to the  $i$ -th data vector, or zero otherwise. In fact, this problem can be seen as a discrete optimization problem where the class assignments ( $u_{ji}$ ) have to be found and the centroids of each class are calculated *a posteriori*. The most common implementation of the algorithm converges to a “local” minimum (indeed it cannot be truly called local due to the discrete nature of the problem). The algorithm consists of a simple re-estimation procedure as follows. First, the data points are assigned at random to the  $C$  sets (classes). Then, the centroid is computed for each set, and finally the data points are reassigned to each class according to the minimum distance to its centroid. These latest two steps are alternated until a stopping criterion is met, for instance, when there is no further change in the assignment of the data points. From a statistical viewpoint, the clusters obtained by  $C$ -means can be interpreted as the Maximum Likelihood Estimates for the cluster means if we assume that each cluster comes from a set of spherical Gaussian distributions with different means but identical covariance matrices.

The previous problem was extended to consider fuzzy memberships (Dunn, 1973; Bezdek, 1981). The idea is to express our uncertainty that a data vector belongs to a specific cluster. Alternatively, the “probability” of belonging to all existing clusters is taken into account by allowing the  $u_{ji}$  coefficients to range between 0 and 1. In this way, the optimization problem becomes

$$\begin{aligned} \mathcal{V}^*, U^* = \arg \min_{\mathcal{V}, U} & \sum_{i=1}^N \sum_{j=1}^C u_{ji}^m \|y_i - v_j\|^2 \\ \text{s.t.} & 0 \leq u_{ji} \leq 1 \\ & \forall i : \sum_{j=1}^C u_{ji} = 1, \end{aligned} \quad (23)$$

where  $m$  is a parameter that controls the fuzziness of the assignment (if  $m$  is close enough to 1, then the algorithm tends to a crisp  $C$ -means), and  $U$  is a matrix whose  $i, j$ -th element is equal to  $u_{ij}$ . The last set of constraints expresses our certainty that each input vector must belong to a specific cluster. The following iterative algorithm can be shown to converge to a local minimum of the objective function:

1. Initialize  $\mathcal{V}$  and  $U$  randomly.
2. Update the code vectors as follows:

$$v_j^{(k+1)} = \frac{\sum_{i=1}^N (u_{ji}^{(k)})^m y_i}{\sum_{i=1}^N (u_{ji}^{(k)})^m} \quad (24)$$

3. Update the fuzzy membership values as follows:

$$u_{ji}^{(k+1)} = \frac{1}{\sum_{l=1}^C \left( \frac{\|y_i - v_j^{(k)}\|}{\|y_i - v_l^{(k)}\|} \right)^{\frac{2}{m-1}}}. \quad (25)$$

4. Repeat 2 and 3 until convergence.

This algorithm is derived by optimizing two surrogate problems: one in which the functional is optimized for  $U$  leaving  $\mathcal{V}$  as constant; and another in which the functional is optimized for  $\mathcal{V}$  given the previously calculated  $U$ .

Pascual-Marqui et al. (2001) proposed a modification of the functional shown in (23) by imposing two important constraints: the code vectors are distributed on a regular low-dimensional grid and a penalization term is added in order to impose a smooth distribution for the values of the code vectors on the grid. This new modification to the classical Fuzzy C-means functional described previously is intended to create a new variant of the classical Self-Organizing Maps (Kohonen, 1982) by keeping two of its most important properties: faithfulness to the data and smooth distribution of the code vectors on the grid, allowing to orderly map high-dimensional input data onto a low-dimensional map while conserving quite consistently the original topological and metric relationships. The new algorithm is called Smoothly distributed Fuzzy C-means, or FuzzySOM for short. The functional to minimize is

$$\begin{aligned}
 V^*, U^* &= \arg \min_{V,U} \sum_{i=1}^N \sum_{j=1}^C u_{ji}^m \|y_i - v_j\|^2 - \kappa \text{tr} V S V^t \\
 \text{s.t.} \quad & 0 \leq u_{ji} \leq 1 \\
 & \forall i : \sum_{j=1}^C u_{ji} = 1,
 \end{aligned}
 \tag{26}$$

where  $V$  is a matrix with all the code vectors  $v$  as columns, and  $S$  is a symmetric, positive definite matrix introducing the smoothness constraints and  $\kappa$  is a multiplier that controls the importance of the smoothness term.  $S$  is usually derived as  $S = G^t G$  where  $G$  is some discrete gradient-like operator (therefore,  $S$  is in that case a discrete Laplacian-like operator). Optimizing (26) in a similar way to (23) provides the same solution for the fuzzy membership values,  $u_{ji}^{(k+1)}$ , as in (25), while the update step for the code vectors is given by

$$v_j^{(k+1)} = \frac{\sum_{i=1}^N (u_{ji}^{(k)})^m y_i - \kappa \sum_{l=1, l \neq j}^C S_{jl} v_l^{(k)}}{\sum_{i=1}^N (u_{ji}^{(k)})^m + \kappa S_{jj}}.
 \tag{27}$$

It can be seen that the fuzzy SOM problem as well as its solution reduces to fuzzy C-means if  $\kappa = 0$ .

The previous SOM can be further generalized Pascual-Montano et al. (2001) by introducing the concept of the probability density function of the input data  $p(y)$  and its approximation by a kernel model

$$p(y) \approx \frac{1}{c} \sum_{j=1}^C K_\alpha(y - v_j),$$

where  $K_\alpha$  can be any kernel function used to estimate the probability density function and  $\alpha$  is a parameter defining the shape of the kernel. The general underlying idea is to find the set of  $C$  surrogate data (code vectors) whose probability density function resembles as best as possible the density of the input data.

Maximizing the log-likelihood of the observed data and regularizing the problem by imposing the code vectors to lay on a low-dimensional, smooth map, we arrive to the following



unconstrained optimization problem

$$\mathcal{V}^*, \alpha^* = \arg \max_{\mathcal{V}, \alpha} \sum_{i=1}^N \log \left( \frac{1}{C} \sum_{j=1}^C K_{\alpha}(\mathbf{y}_i - \mathbf{v}_j) \right) - \frac{\kappa}{2\alpha} \text{tr}(V S V^t). \quad (28)$$

This method is usually referred to as KerDenSOM (standing for Kernel Probability Density Self-Organizing Map). The solution of this problem depends on the approximating kernel used. If the Gaussian kernel ( $K_{\alpha}(\mathbf{y}) = \frac{1}{(2\pi\alpha)^{D/2}} \exp(-\frac{\|\mathbf{y}\|^2}{2\alpha})$ ) is used, then taking partial derivatives of (28) with respect to  $\mathcal{V}$  and  $\alpha$  and setting them to zero gives

$$\alpha^{(k+1)} = \frac{1}{Nf} \sum_{i=1}^N \sum_{j=1}^C u_{ji}^{(k)} \| \mathbf{y}_i - \mathbf{v}_j^{(k)} \|^2 + \frac{\kappa}{Nf} V S V^t, \quad (29)$$

where

$$u_{ji}^{(k)} = \frac{K_{\alpha^{(k)}}(\mathbf{y}_i - \mathbf{v}_j^{(k)})}{\sum_{l=1}^C K_{\alpha^{(k)}}(\mathbf{y}_i - \mathbf{v}_l^{(k)})}, \quad (30)$$

and the update equation for the code vector  $\mathbf{v}_j^{(k+1)}$  is provided by (27) with  $m = 1$ .

Unlike other SOM algorithms, KerDenSOM belongs can be very sensitive to the initial conditions. To achieve a better maximum of (28) a deterministic annealing approach is followed. In this approach the KerDenSOM algorithm is run until convergence for a set of  $\kappa$  values (remind that  $\kappa$  controls the smoothness of the output map) between  $\kappa_0$  (very smooth) and  $\kappa_F$  (very sharp). The output of one  $\kappa$ -run is used as initialization for the next.  $\kappa$  is varied as  $\kappa^{(t)} = \exp(\log \kappa_F - \frac{\log \kappa_F - \log \kappa_0}{T} t)$ , where  $t$  is the index of the current  $\kappa$ -run and  $T$  is the total number of  $\kappa$ -runs. The value of  $\kappa$  is varied from  $\kappa_0$  down to  $\kappa_F$  and the best value is chosen by calculating the randomized generalized cross-validation criteria of Wahba et al. (1994).

#### Problem 4: Angular assignment

Before getting the experimental images into a reconstruction algorithm, their relative orientation and shift must be determined, i.e., the corresponding Euler angles and shifts (see (1)) must be estimated so that the different 2D images can be combined as projections of the same 3D volume. This problem is actually an extension of the 2D alignment problem from three to five parameters. The mathematics developed for the 2D case is still valid and most of the algorithms are based on the maximization of the correlation between two images. Many of the different approaches differ on the space (real space, Radon space, wavelet space, or Fourier space) in which they perform the correlation maximization. Another difference of some existing algorithms is the explicit use of the Central Section Theorem (CST) (Kak and Slaney, 1987; Natterer and Wübbeling, 2001) that states that the 2D Fourier transform of a projection image is a central slice of the 3D Fourier transform of the volume from which the projection was taken.

Extending the 2D alignment problem to 3D is straightforward and leads to the following optimization problem:

$$\begin{aligned} \mathbf{s}_0^*, \phi^*, \theta^*, \psi^* &= \arg \max_{\mathbf{s}_0, \phi, \theta, \psi} \langle y, x(\mathbf{s}_0, \phi, \theta, \psi) \rangle \\ \text{s.t.} \quad & \mathbf{s}_0 \in \mathcal{S} \\ & \phi_0 \leq \phi \leq \phi_F \\ & \theta_0 \leq \theta \leq \theta_F \\ & \psi_0 \leq \psi \leq \psi_F \end{aligned} \quad (31)$$

Now  $x(\mathbf{s}_0, \phi, \theta, \psi)$  involves the projection along a direction determined by  $\phi$  and  $\theta$  of a reference volume (supposed to be ideal at the moment) and the translation and rotation (according to  $\mathbf{s}_0$  and  $\psi$ ) of the projected image.

Penczek, Grasucci, and Frank (1994) proposed an algorithm that performs an exhaustive search in the parameter space to identify the parameters that maximize the likelihood of observing each of the experimental images. The exhaustive search is performed on a regular grid in the parameter space. For the sake of implementation, the reference volume is projected onto a set of fixed directions obtaining, thus, a library of reference projections. Each experimental image is 2D aligned to each of the reference projections identifying the best 2D transformation parameters if that experimental image were coming from the reference projection at hand. The reference projection and the 2D transformation parameters maximizing the correlation between the two images define the optimal 3D transformation parameters. The 2D alignment is performed taking advantage of the Fourier transform correlation property that greatly accelerates the search.

Radermacher (1994) proposed an algorithm that performs the exhaustive search in the Radon space. The 3D Radon transform of a volume  $f(\mathbf{r})$  is defined as

$$\hat{f}(p, \mathbf{d}) = \int_{\mathbb{R}^3} f(\mathbf{r}) \delta(p - \langle \mathbf{d}, \mathbf{r} \rangle) d\mathbf{r}, \quad (32)$$

i.e., the input volume is integrated over the plane perpendicular to  $\mathbf{d}$  and defined by  $p = \langle \mathbf{d}, \mathbf{r} \rangle$ . In this way, the 3D Radon transform maps a volume into a set of 1D projections indexed by the vector  $\mathbf{d}$ . Similarly, the 2D Radon transform (or sinogram) of an image maps that image into a set of 1D projections indexed by a vector  $\mathbf{d}$  simply by integrating over lines. The computation of the correlation is not so intuitive as in the real-space case (Penczek, Grasucci, and Frank, 1994). However, it can be computed as the sum of the correlations of lines within the 3D Radon transform of the reference volume and the lines of the 2D Radon transform of the experimental images.

More recently, a new approach has been proposed in which the correlation between the reference library images defined by Penczek, Grasucci, and Frank (1994) and the experimental images are computed in the wavelet space Sorzano et al. (2004b). This space offers the advantage of being able to compute the correlation in a coarse-to-fine fashion making use of a decomposition of the correlation formula similar to that employed in (5). In this way, images are first compared (correlated) at a coarse resolution. If they match sufficiently, they progress to the comparison in the next stage in which the resolution is increased including finer details. This process continues until all the image details have been considered. It has been shown that this procedure is more robust to the presence of local minima in the objective function.

Library based algorithms are by far the most commonly used in EM. They all perform an exhaustive search of the 3D transformation parameters by optimizing the correlation between the experimental image at hand and a set of reference images in some space. However, there are other algorithms that exploit the CST. A compact formulation of this theorem is given by

$$X(\boldsymbol{\omega}; \phi, \theta, \psi, \mathbf{s}_0) = F \left( E_{\phi, \theta, \psi}^t \boldsymbol{\omega} \right) e^{-i(\boldsymbol{\omega}, \mathbf{s}_0)}, \quad (33)$$

where  $X(\boldsymbol{\omega}; \phi, \theta, \psi, \mathbf{s}_0)$  is the 2D Fourier transform of the projection taken with 3D transformation parameters  $\phi, \theta, \psi, \mathbf{s}_0$ .  $\boldsymbol{\omega} \in \mathbb{R}^2$  is the 2D frequency vector.  $H$  is defined as in (1), and  $F$  is the 3D Fourier transform of  $f$ . This formula indicates how to extract the 2D Fourier transform of a projection from a volume  $f$  given a set of 3D transformation parameters. Note that Fourier transforms are usually complex functions, and therefore their real and imaginary part can be treated independently. Because of the CST, finding for the optimal 3D transformation parameters can be done by minimizing the objective function (Jonic et al., 2005)

$$E(\phi, \theta, \psi, \mathbf{s}_0) = \sum_{\boldsymbol{\omega} \in \mathbb{Z}^2 - \{(0,0)\}} w(\boldsymbol{\omega}) \left[ \left( \frac{\operatorname{Re} \{X(\boldsymbol{\omega}; \phi, \theta, \psi, \mathbf{s}_0)\}}{\sigma_x} - \frac{\operatorname{Re} \{Y(\boldsymbol{\omega})\}}{\sigma_y} \right)^2 + \left( \frac{\operatorname{Im} \{X(\boldsymbol{\omega}; \phi, \theta, \psi, \mathbf{s}_0)\}}{\sigma_x} - \frac{\operatorname{Im} \{Y(\boldsymbol{\omega})\}}{\sigma_y} \right)^2 \right], \quad (34)$$

where, Re and Im stand for the real and imaginary parts of the Fourier transform respectively,  $w(\boldsymbol{\omega})$  is a weighting function in Fourier space, and  $\sigma_x$  and  $\sigma_y$  accounts for possible different scaling factors,  $\sigma_y^2 = \sum_{\boldsymbol{\omega} \in \mathbb{Z}^2 - \{(0,0)\}} \|Y(\boldsymbol{\omega})\|^2$ , and  $\sigma_x^2 = \sum_{\boldsymbol{\omega} \in \mathbb{Z}^2 - \{(0,0)\}} \|X(\boldsymbol{\omega}; \phi, \theta, \psi, \mathbf{s}_0)\|^2$ .

The computation of the Fourier transform of the theoretical projection  $X(\boldsymbol{\omega}; \phi, \theta, \psi, \mathbf{s}_0)$  implies that the Fourier transform of the volume must be interpolated. This can be solved thanks to a continuous approximation of the Fourier transform of the volume using B-splines (Unser, Aldroubi, and Eden, 1991; Unser, 1999). This fact is key since from now on, the 3D transformation parameters can take any value and not only those in a predefined grid. At the same time, we can resort to continuous optimization techniques much more powerful than exhaustive search.

An optimization algorithm that is inspired by the Levenberg-Marquardt nonlinear regression (Thévenaz, Ruttiman, and Unser, 1998) is used for this minimization. This algorithm achieves a gradual transition between quasi-Newton and gradient-descent steps. The specific tradeoff depends on the goodness of a local model of the function being minimized. Briefly, this method updates the transformation coefficients  $\mathbf{c} = (\phi, \theta, \psi, \mathbf{s}_0)$  in an iterative fashion  $\mathbf{c}^{(k+1)} = \mathbf{c}^{(k)} + \Delta \mathbf{c}^{(k)}$ , where  $\Delta \mathbf{c}$  is the solution of the equation system  $\tilde{\mathbf{H}} \Delta \mathbf{c}^{(k)} = \nabla E(\mathbf{c}^{(k)})$ . There,  $\nabla E(\mathbf{c}^{(k)})$  is the gradient of the objective function with respect to the transformation coefficients evaluated at  $\mathbf{c}^{(k)}$ , and  $\tilde{\mathbf{H}}$  is a modified version of the Hessian  $\mathbf{H}$  such that the component  $[\tilde{\mathbf{H}}]_{ij} = (1 + \lambda \delta_{ij}) [\mathbf{H}]_{ij}$ , where  $\delta_{ij} = 1 - |\operatorname{sign}(i - j)|$  is Kronecker's delta. When  $\lambda$  is large the step is more steepest-descent-like while, for small values of  $\lambda$ , it is more Newton-like. The value of  $\lambda$  is adaptively modified according to successes or failures of  $\mathbf{c}^{(k)}$  to minimize the given objective.

The authors of Jonic et al. (2005) used a diagonal approximation to the true Hessian based only on first derivatives of the objective function. This is a fair approximation close to the minimum. However, the procedure can benefit from a BFGS estimation of the Hessian (Press et al., 1992). An advantage of the BFGS approximation to the Hessian is that it is always

positive semi-definite, which is a mandatory requirement for the stability of a quasi-Newton optimization algorithm. BFGS results in the following iterative approximation to the Hessian:

$$\Delta \mathbf{c}^{(k)} = \mathbf{c}^{(k+1)} - \mathbf{c}^{(k)} \tag{35}$$

$$\Delta \mathbf{f}^{(k)} = \nabla E(\mathbf{c}^{(k+1)}) - \nabla E(\mathbf{c}^{(k)}) \tag{36}$$

$$\Delta \mathbf{H} \mathbf{c}^{(k)} = \mathbf{H}^{(k)} \Delta \mathbf{c}^{(k)} \tag{37}$$

$$\mathbf{H}^{(k+1)} = \mathbf{H}^{(k)} + \frac{\Delta \mathbf{f}^{(k)} \otimes \Delta \mathbf{f}^{(k)}}{\langle \Delta \mathbf{f}^{(k)}, \Delta \mathbf{c}^{(k)} \rangle} - \frac{\Delta \mathbf{H} \mathbf{c}^{(k)} \otimes \Delta \mathbf{H} \mathbf{c}^{(k)}}{(\Delta \mathbf{c}^{(k)})^\top \mathbf{H}^{(k)} \Delta \mathbf{c}^{(k)}}, \tag{38}$$

where  $\mathbf{x} \otimes \mathbf{y} = \mathbf{x} \mathbf{y}^t$ . The Hessian estimate is initialized by the diagonal approximation of the Hessian found in Thévenaz, Ruttiman, and Unser (1998) and is updated only on successful estimates  $\mathbf{c}^{(i)}$ ,  $k \neq i$ . For keeping the semi-positive-definite quality of the estimate, this update can only be performed if  $\langle \Delta \mathbf{f}^{(k)}, \Delta \mathbf{c}^{(k)} \rangle > 0$  (Dennis and Schnabel, 1996, Chapter 9). The combination of this Levenberg-Marquardt inspired algorithm with the BFGS approximation of the Hessian was successfully tested by Sorzano et al. (2005b).

Another angular assignment approach that makes use of the CST is the so-called common-lines search (Frank, 1996). This approach does not use any reference volume. On the contrary, all the angular assignment is performed on the information of the sole set of experimental images. Let us consider two different projections of the same volume. Since, due to the CST, both are central slices of the 3D Fourier transform of the volume, then both slices must share a line in Fourier space. Two different algorithms have been proposed for the detection of the common lines (van Heel, 1987; Penczek, Grasucci, and Frank, 1994). Both approaches are based on the maximization of the correlation index between any two lines of the 2D Radon transform of the two projections. The main drawback of this stage of detecting common lines between experimental images is that, due to the low SNR, it is not easy to reach global maxima, and therefore, many false matches are returned. However, for high-symmetry particles (like, icosahedral viruses), this is the standard approach.

For three projections, there are three pairs of common lines and their relative position is uniquely determined. van Heel (1987) solved the angular assignment problem for three projections using the directions of the common lines. However, it was not until (Penczek, Zhu, and Frank, 1996) that the problem was practically solved for more than three projections (assuming that the shift parameters were already solved). If there are  $N$  projections, there are  $3(N - 1)$  free Euler angles to determine and  $\frac{N(N-1)}{2}$  pairs of common lines. Penczek, Zhu, and Frank (1996) maximizes the following functional

$$E(\phi_1, \theta_1, \psi_1, \dots, \phi_N, \theta_N, \psi_N) = \int_{\mathbb{R}^3} (f_i(E_{\phi_i, \theta_i, \psi_i} \mathbf{r}))^2 d\mathbf{r} \tag{39}$$

where  $f_i(\mathbf{r})$  is the volume whose 3D Fourier transform is formed by a single slice (on the XY plane) whose content is the  $i$ -th projection, and  $\phi_1 = \theta_1 = \psi_1 = 0$ , the actual objective function implemented is in Fourier space expressed in polar coordinates, however it is equivalent to the one presented which is much more comprehensive. An exhaustive search is performed on a regular grid for each of the Euler angles. The optimization procedure proceeds as follows: the three Euler angles corresponding to a given projection are optimized assuming that the rest of projections are fixed. Then, the next projection is selected and optimized. When all projections have been visited, it is said that a cycle has been completed. This procedure

is iterated until none of the projections changed its orientation during a full cycle. Lindhal (2001) extended this procedure to include the shift parameters during the optimization.

Finally, there is another approach that performs the angular assignment (without shifts) by comparing the reference image 2D Radon transforms, and those of the experimental images. However, before comparing the corresponding sinograms, each line of the sinogram is projected onto a lower-dimensional space using Correspondence Analysis (Bellon, Cantele, and Lanzavecchia, 2001; Bellon et al., 2002). The projection of all the lines of a sinogram defines a closed trajectory in the projected space  $\tilde{y}_{ij}$ ,  $i$  spans the number of lines in the sinogram while  $j$  the number of dimensions of the low dimension space. These trajectories are periodic in the sense  $\tilde{y}_{ij} = \tilde{y}_{i+kN,j}$  for any integer  $k$  and being  $N$  is the number of lines in the sinogram. The trajectory of the experimental image is denoted as  $\tilde{y}_{ij}$  and the trajectory of the reference projection at given  $\phi, \theta$  angles as  $\tilde{x}_{ij}(\phi, \theta)$ . The projection direction is assigned by comparing the trajectory of the experimental image at hand with those obtained in the reference library. This comparison is performed by minimizing

$$\phi^*, \theta^*, m^* = \arg \min_{\phi, \theta, m} \sum_i \left( \sum_j (\dot{y}_{ij} - \dot{x}_{i+m,j}(\phi, \theta))^2 \right)^{\frac{1}{2}}, \quad (40)$$

where  $\dot{y}_{ij} = \frac{\tilde{y}_{ij} - \tilde{y}_{i-1,j}}{\|\tilde{y}_i - \tilde{y}_{i-1}\|}$ , correspondingly for  $\dot{x}$ .  $m$  is a parameter that takes advantage of the periodic nature of the closed trajectories to determine the  $\psi$  angle. The previous objective function is optimized by exhaustive search.

### Problem 5: 3D reconstruction

The problem of 3D reconstruction consists of producing a volume that is compatible with the experimental projections acquired by the microscope. For this, it is absolutely mandatory to have already assigned the point of view of each projection (angular assignment) and have aligned the center of the projections. The most widely used reconstruction algorithm employed in the field is called Weighted Back Projection (WBP) (Radermacher, 1992). This algorithm sums the set of volumes obtained by backprojecting each of the experimental images. It will not be further described in this review since it is not based on the optimization of any functional.

There is a whole family of reconstruction algorithms called ‘‘Series Expansion Methods’’, i.e., it is assumed that a volume  $f$  can be approximated by a linear combination of a finite set of known basis functions  $b_j$ , each one of which is just the same function  $b$  shifted to one of  $J$  grid points (denoted by  $\mathbf{g}_j$ ), as in

$$f(\mathbf{r}) \approx \sum_{j=1}^J c_j \cdot b_j(\mathbf{r}) = \sum_{j=1}^J c_j \cdot b(\mathbf{r} - \mathbf{g}_j). \quad (41)$$

The task of the algorithm is to estimate the unknown coefficients  $c_j$ . A consequence of the volume series expansion is an image formation model of the form:

$$y_i \approx \sum_{j=1}^J l_{i,j} c_j$$

where  $y_i$  is the  $i$ th measurement of the volume to be reconstructed (that is, a pixel value in the experimental data) and  $l_{i,j}$  is the corresponding line integral of the basis function  $b_j$ . The values  $y_i$  and  $c_j$  form a  $MN$ -dimensional vector and a  $J$ -dimensional vector respectively (which we will denote by  $\mathbf{y}$  and  $\mathbf{c}$ ), where  $N$  is the number of projections and  $M$  is the number of pixels per projection. The  $y_i$  elements are arranged in such a way that all pixels belonging to a projection are consecutive. Therefore, reconstructing the volume  $f$  is equivalent to solving for  $\mathbf{c}$  in the linear equation system

$$\mathbf{y} = L\mathbf{c}. \quad (42)$$

The  $i$ -th row of this equation system defines an hyperplane in the  $J$ -dimensional space given by  $y_i = \mathbf{l}_i \mathbf{c}$ . Solving for  $\mathbf{c}$  in the previous equation system is the same as finding a point  $\mathbf{c}$  belonging to all the hyperplanes specified by the equation system. For this reason, the reconstruction algorithms presented below are said to solve the ‘‘feasibility problem’’.

One possible way of solving this equation system is by an Algebraic Reconstruction Technique (ART, Herman (1980, 1998)). This is an iterative algorithm that updates a current guess of the solution after correcting for the mismatching between the projection of the current guess onto a given pixel ( $(\mathbf{l}_{i(k)}, \mathbf{c}^{(k)})$ ) and the actual measurement obtained at that pixel ( $y_{i(k)}$ )

$$\mathbf{c}^{(k+1)} = \mathbf{c}^{(k)} + \lambda^{(k)} \frac{y_{i(k)} - (\mathbf{l}_{i(k)}, \mathbf{c}^{(k)})}{\|\mathbf{l}_{i(k)}\|^2} \mathbf{l}_{i(k)}. \quad (43)$$

$i(k)$  is a sequence of indexes providing the order in which the measured pixels are accessed, and  $\lambda^{(k)}$  is a sequence of positive numbers called ‘‘relaxation factors’’.

Let  $\mathcal{S}$  be the set of points obtained by linear combination of the rows of the matrix  $L$ ,  $\mathcal{S} = \{\mathbf{x} \mid \mathbf{x} = \sum_{i=1}^{MN} \beta_i \mathbf{l}_i\}$ . Let  $\mathcal{L}$  be the set of feasible points, i.e., the set of points satisfying the linear equation system. If  $\mathbf{c}^{(0)} \in \mathcal{S}$ , then the sequence  $\{\mathbf{c}^{(k)}\}$  converges to  $\mathbf{c}^* = \arg \min_{\mathbf{c} \in \mathcal{L}} \|\mathbf{c}\|$ . In other words, among all the possible solutions of the equation system, the one with smallest norm is chosen. This is interesting from the point of view of signal since the reconstruction produced is the one with smallest variance, and therefore, the one with fewest features that is compatible with the projection set. It is common practice to start the algorithm with  $\mathbf{c}^{(0)} = \mathbf{0}$  since  $\mathbf{0}$  belongs to  $\mathcal{S}$ . Because of the noise present in the measurements and if the number of measurements is larger than the number of unknowns in  $\mathbf{c}$ , it is likely that  $\mathcal{L}$  is empty, that is, there is no volume that is at the same time compatible with all the projections. In this case, the ART sequence converges to a cycle of points  $\{\mathbf{c}^{*(k)} : k = 1, 2, \dots, MN\}$  with no special property from the point of view of optimization (Herman, 1998).

Notice that (43) updates the current guess of the solution after considering each of the pixels individually. This procedure has very interesting properties from the point of view of rate of convergence (ART is, in general, one of the fastest algorithms for solving linear equation systems). An alternative approach would update the current solution only after considering all the updates of all the pixels at the same time

$$\mathbf{c}^{(k+1)} = \mathbf{c}^{(k)} + \lambda^{(k)} \sum_{i=1}^{MN} (y_i - (\mathbf{l}_i, \mathbf{c}^{(k)})) \mathbf{l}_i = \mathbf{c}^{(k)} + \lambda^{(k)} L^t (\mathbf{y} - L\mathbf{c}^{(k)}). \quad (44)$$

Notice that the normalizing factor  $\|\mathbf{l}_i\|^2$  has been removed. This algorithm is known as ‘‘Simultaneous Iterative Reconstruction Technique’’ (SIRT) and it belongs to a class of optimizers

very useful in 3D reconstruction as will be shown below. The general class of optimizers is defined as

$$\mathbf{c}^* = \arg \min_{\mathbf{c} \in \mathcal{K}} \|D^{-1}\mathbf{c}\|, \quad (45)$$

where

$$\mathcal{K} = \left\{ \mathbf{c} \mid \mathbf{c} = \arg \min_{\mathbf{c} \in \mathbb{R}^J} a(\mathbf{y} - L\mathbf{c})^T A(\mathbf{y} - L\mathbf{c}) + (\mathbf{c} - \mathbf{c}_0)^T (bB + cC^{-1})(\mathbf{c} - \mathbf{c}_0) \right\}, \quad (46)$$

$A$  is a  $MN \times MN$  matrix,  $B$  and  $C$  are  $J \times J$  matrices, and  $a, b, c$  are nonnegative scalars.

If  $bB + cC^{-1}$  is a positive definite matrix, then the following sequence can be shown to converge to the unique minimum of (45)

$$\mathbf{c}^{(k+1)} = \mathbf{c}^{(k)} + \lambda^{(k)} (aCL^T A(\mathbf{y} - L\mathbf{c}^{(k)}) + (bcB + cI)(\mathbf{c}_0 - \mathbf{c}^{(k)})), \quad (47)$$

where  $I$  is the identity  $J \times J$  matrix. Otherwise, if  $bB + cC^{-1}$  is a matrix full of zeros, then the following sequence can be shown to converge to the minimum norm minimum of (45)

$$\mathbf{c}^{(k+1)} = \mathbf{c}^{(k)} + \lambda^{(k)} D^2 L^T A(\mathbf{y} - L\mathbf{c}^{(k)}). \quad (48)$$

It can be easily seen that the SIRT-type algorithm presented in (44) is a particular case of (48) where  $D$  and  $A$  are identity matrices. Iterations converge to  $\mathbf{c}^* = \arg \min_{\mathbf{c} \in \mathcal{K}} \|\mathbf{c}\|$  with  $\mathcal{K} = \{\mathbf{c} \mid \mathbf{c} = \arg \min_{\mathbf{c} \in \mathbb{R}^J} \|\mathbf{y} - L\mathbf{c}\|^2\}$ . In fact, if the image formation model is assumed to be additive-noise, then this corresponds to the maximum likelihood solution if the noise values are independent, identically and normally distributed and no prior information is available about the distribution of  $\mathbf{c}$ . Many other assumptions or constraints in 3D reconstruction like imposing the solution to be smooth or the knowledge that it is normally distributed around some known volume can be expressed in terms of an optimization problem like (45) (Herman, 1980). The most used implementation of this algorithm is the one proposed by Penczek, Radermacher, and Frank (1992).

Skoglund et al. (1996) give a completely different view of the reconstruction problem formulating it from a constrained optimization perspective. Their goal is to produce the minimally informative volume that is compatible with the projections. The idea is to avoid overfitting the projections since they are known to be noisy. This is accomplished by maximizing the entropy of the reconstructed volume subject to a normalization constraint and another constraint that avoids the overfitting

$$\begin{aligned} f^*(\mathbf{x}) &= \arg \max_{f(\mathbf{x})} - \int f(\mathbf{x}) \log \left( \frac{f(\mathbf{x})}{f_0(\mathbf{x})} \right) d^3 \mathbf{x} \\ \text{s.t.} \quad & \int f(\mathbf{x}) d^3 \mathbf{x} = 1 \\ & C_r^{(i)} = 1 \quad \forall i \in \{1, \dots, N\} \\ & f(\mathbf{x}) \geq 0 \end{aligned} \quad (49)$$

where  $f_0(\mathbf{x})$  is a volume known *a priori* to which the reconstruction resembles,  $C_r^{(i)} = \frac{1}{M} \sum_{j=1}^M \frac{1}{\sigma_j^2} (y_j^{(i)} - (P^{(i)} f)_j)$  is called the reduced  $\chi^2$  statistic,  $y_j^{(i)}$  is the

$j$ -th pixel of the  $i$ -th experimental image,  $(P^{(i)}f)_j$  is the corresponding pixel of the theoretical projection in the same direction as  $y_i$ , and  $\sigma_j^2$  is the local variance of the experimental image at pixel  $j$ . The projection operator  $P^{(i)}$  includes the CTF and some scaling factor to adapt to the experimental projection values. The optimization technique employed in this problem is iterative and approximates the objective function as well as the constraints at the current estimate of the reconstructed volume by quadratic models, and then these models are optimized yielding a new estimate of the reconstructed volume. This process is iterated until convergence.

### Problem 6: 3D Reconstruction performance

Most 3D reconstruction algorithms have free parameters to optimize like the relaxation factor  $\lambda$  in the ART or SIRT-type algorithms. The problem is how to choose these parameters so that the performance of the algorithm is optimal. In fact, a wrong choice for this kind of parameters can lead to wrong conclusions. For instance, Boisset et al. (1998) reported an elongation along a direction of overabundant projections when reconstructing with WBP or SIRT. This elongation was indeed associated to the fact of the oversampling in the projection space although it can be overcome by the appropriate selection of the free parameters of the reconstruction algorithm (Sorzano et al., 2001).

The approach usually followed to optimize this performance can be briefly described as follows (Furuie et al., 1994). Several realizations from a statistically defined set of phantoms (artificial volumes) are created. These phantoms must resemble, in some way, the real objects of interest (macromolecules). After that, the algorithm under study is run several times varying randomly the set of noise variables (i.e., those variables that cannot be controlled in a real-life experiment like the angular distribution or the noise realization) and using each time a different free parameter. The experimental 3D reconstruction process is simulated by projecting the phantoms in a realistic manner and making a 3D reconstruction from the simulated data. Finally, the degree of accomplishment of the defined tasks is measured using numerical observers called Figures of Merit (FOM). These FOMs compare the input of the simulation process (the phantoms) with its outputs (the reconstructed volumes). Several FOMs can be defined to measure the usefulness of the reconstruction for various tasks.

Some approaches to optimize the performance of the algorithm optimize a single FOM (called training FOM) with respect to the free parameters (Furuie et al., 1994; Matej, Furuie, and Herman, 1996; Matej et al., 1994; Sorzano et al., 2001). Because of the noisy nature of the FOM measurements (images are noisy and the FOM measurement is subject to the particular noise realization), the identification of the FOM optimum is handled statistically. The free parameter with the largest average of the training FOM is selected. Then, an interval is chosen such that the average training FOM of all points within the interval is statistically indistinguishable (usually measured as a t-Student test) from that with maximum training FOM except in the interval extremes which must show an statistically-significant inferior performance.

This approach provides useful results. However, it considers the information provided by a single FOM although many of them can be defined. Other FOMs may capture some information that cannot be observed with a single training FOM. If several FOMs are considered, the problem becomes a multiobjective optimization problem with noisy measurements. A possible solution is to pick a value from the set of those  $\lambda$  that are not inferior to any other choice of the parameter (a point  $\lambda_1$  is inferior to  $\lambda_2$  if and only if, for  $1 \leq i \leq n$ ,  $FOM_i(\lambda_1) \leq FOM_i(\lambda_2)$  and, for some  $i$ ,  $FOM_i(\lambda_1) < FOM_i(\lambda_2)$ .) This set is termed the



Pareto-optimal set. In other words, if a point is a Pareto optimum, one cannot increase any of the FOMs selecting a surrounding point without decreasing some other FOM. However, it can be shown (Sorzano et al., 2005a) that for some algorithms, for instance block ART, all free parameters are Pareto optima. An algorithm is needed to select an interval within the Pareto-optimal set. Again, a statistical approach was followed (Sorzano et al., 2005a) although this time multivariate. Briefly, the optimization procedure can be summarized as follows:

1. FOM normalization: Normalize the FOMs so that they have comparable values.
2. Removal of irrelevant FOMs: Remove all those FOMs that cannot detect differences among the various values of the parameters.
3. FOM clustering: Cluster all those FOMs showing a similar dependency with the parameters. This was done by hierarchical clustering and validated by Principal Component Analysis.
4. Cluster dimensionality reduction: Reduce the dimensionality of the clusters obtaining a single representative of each cluster. This was accomplished by a weighted sum of the PCA decomposition of each cluster.
5. Interval selection: Select an optimal region for the parameters using the information contained in the cluster representatives. This is performed by constructing an interval for each cluster representative as was already done for the training FOM, and then intersecting all the intervals.

#### Problem 7: CTF estimation

As has been shown, most of the algorithms used in EM do not make use of the complete image formation model given in Section 1. In common practice, the phase of CTF is corrected at the beginning of the image processing, while its amplitude is corrected at the end. In order to correct for the CTF, the CTF must be firstly estimated from the micrographs. This is done in two steps. First, the Power Spectral Density (PSD) is estimated from the image. The result of this stage is a real image in Fourier space  $PSD(\boldsymbol{\omega})$ . Second, this PSD is decomposed into two theoretical models: one coming from the signal that is passing through the CTF, and another one coming from the two noise sources (the one before CTF and the one after CTF, see the image formation model in Section 1). The two theoretical models are parametrized with a set of parameters  $\Phi$  that includes things like the microscope voltage, the defocus, different kind of aberrations, etc. For a complete description of the CTF model, the interested reader is referred to Velázquez-Muriel et al. (2003), Zhu et al. (1997), and Ludtke, Baldwin, and Chiu (1999). Thus, for a given set of parameters  $\Phi$  we can express the power spectrum of a projection view  $p_{\phi, \theta, \psi, s_0}(\mathbf{s})$  as

$$PSD_{\phi, \theta, \psi, s_0}(\boldsymbol{\omega}) = |CTF_{\Phi}(\boldsymbol{\omega})|^2 F_{\phi, \theta, \psi, s_0}(\boldsymbol{\omega}) + |CTF_{\Phi}(\boldsymbol{\omega})|^2 PSD_{N_b, \Phi}(\boldsymbol{\omega}) + PSD_{N_a, \Phi}(\boldsymbol{\omega}),$$

where  $\boldsymbol{\omega}$  is the two dimensional spatial frequency,  $PSD_{\phi, \theta, \psi, s_0}(\boldsymbol{\omega})$  is the Power Spectrum Density of the experimental image at the frequency  $\boldsymbol{\omega}$ ,  $|CTF_{\Phi}(\boldsymbol{\omega})|^2$  is the Contrast Transfer Function whose parameters are defined by  $\Phi$ ,  $F_{\phi, \theta, \psi, s_0}(\boldsymbol{\omega})$  is the power spectrum density of the ideal projection (also called structure factor),  $PSD_{N_b, \Phi}(\boldsymbol{\omega})$  is the power spectrum density of the noise before CTF and whose actual parameters are also defined by the parameter set  $\Phi$ , and finally  $PSD_{N_a, \Phi}(\boldsymbol{\omega})$  is the power spectrum density of the noise after

CTF with parameters given by  $\Phi$ . The two noise terms can be combined into a single noise term  $PSD_{N,\Phi}(\omega)$ . The problem of the CTF estimation consists of estimating the vector  $\Phi$  that better fits the observed PSD (which has to be estimated from the experimental image itself) given that the actual structure factors are unknown. It is difficult to have good estimates of the PSD from a single projection image, and therefore the PSD is obtained as the average of several individual PSDs as long as the noise characteristics, the CTF and the structure factors can be assumed to be approximately the same. Under these conditions we can drop the subindices depending on a specific orientation and shift of the projection to yield

$$PSD(\omega) = |CTF_{\Phi}(\omega)|^2 F(\omega) + PSD_{N,\Phi}(\omega), \tag{50}$$

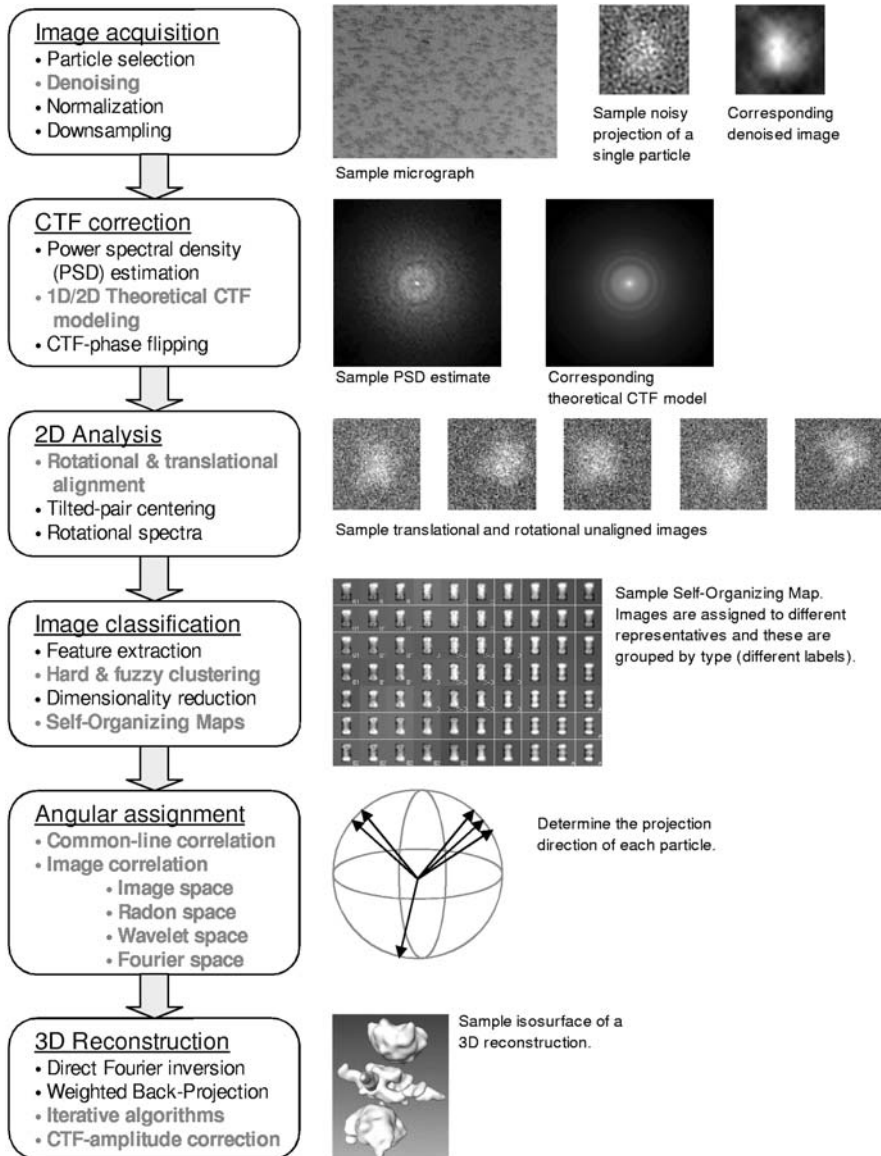
In this review we will discuss two different approaches that are relevant from an optimization point of view: that of Velázquez-Muriel et al. (2003) (who assume that  $F(\omega) = 0$ ) and that of Sinkevich et al. (2000) (who assume that  $PSD_{N,\Phi}(\omega) = 0$ ). None of these assumptions is practically fulfilled in real-life experiments:  $F(\omega)$  would mean that there is no particle in the micrograph field (which is not so untrue if the protein concentration is low), while  $PSD_{N,\Phi}(\omega) = 0$  would assume that there is no noise affected by the CTF (what is clearly not true since most of the image is contaminated by the projection of the embedding ice or the supporting carbon film). However, these two simplifications allow deriving practical algorithms whose results are still of use.

Velázquez-Muriel et al. (2003) solves the estimation problem by minimizing the distance between the observed PSD and the PSD that would be observed with parameters  $\Phi$  from the pure noise from a variety of point of views:

$$\begin{aligned} \Phi^* = \arg \min_{\Phi} & \quad k_1 \sum_{\omega \in \Omega(\Phi)} W(|\omega|) \frac{|PSD(\omega) - PSD_{N,\Phi}(\omega)|}{|PSD(\omega)|} \\ & + k_2 \sum_{\omega \in \Omega(\Phi)} |\widehat{PSD}(|\omega|) - \widehat{PSD}_{N,\Phi}(|\omega|)| \\ & + k_3 \sum_{\omega \in \Omega(\Phi)} |\widehat{PSD}'(|\omega|) - \widehat{PSD}'_{N,\Phi}(|\omega|)| \\ & + k_4 \sum_{s \in \mathcal{R}} \frac{|\log_{10} |psd(s)| - \log_{10} |psd_{N,\Phi}(s)||}{|\log_{10} |psd(s)||} \\ & - k_5 \frac{\sum_{\omega \in \Omega(\Phi)} W(|\omega|)(PSD(\omega) - \overline{PSD})(PSD_{N,\Phi}(\omega) - \overline{PSD}_{N,\Phi})}{\sqrt{\sum_{\omega \in \Omega(\Phi)} W(|\omega|)(PSD(\omega) - \overline{PSD})^2} \sqrt{\sum_{\omega \in \Omega(\Phi)} W(|\omega|)(PSD_{N,\Phi}(\omega) - \overline{PSD}_{N,\Phi})^2}} \end{aligned} \tag{51}$$

s.t.  $\Phi \in \Phi_{\text{feasible}}$ .

The first term in the objective measures the percentage distance between the two estimates in Fourier space. This distance is weighted by a window function  $W(|\omega|)$  that gives more importance to low frequencies (small  $|\omega|$ ). This distance is measured over a region  $\Omega(\Phi)$  that encompasses the region between the estimated first and fourth zeros of the CTF (see the rings in the CTF image in Fig. 3). The second term measures the distance between the corresponding radial averages ( $\widehat{PSD}$  and  $\widehat{PSD}_{N,\Phi}$ ). The third term measures the distance between the derivatives of the radial averages. The derivative is computed using a discrete approximation with five points due to Richardson. The fourth term measures the percentage distance between both functions in real space within a fixed region  $\mathcal{R}$ . The last term measures the weighted correlation between both functions in Fourier space ( $\widehat{PSD}$  and  $\widehat{PSD}_{N,\Phi}$  are the



**Fig. 3** Schematic work-flow of EM image processing analysis. Those areas with relevant optimization problems are highlighted in red and bold font

average values of the observed PSD and the PSD of the noise in Fourier space). Finally, the set of parameters  $\Phi$  (up to 23 parameters) must have a physical meaning. For instance, it makes no sense to have a negative acceleration voltage within the microscope or an acceleration voltage that is not in the range of currently built microscopes (between 80 kVolts and 1 M Volts). This need for physical meaning restricts the set of feasible solutions.

The main features of this optimization problem is that it is medium-size, with a highly nonlinear objective function, with boundary conditions ( $\Phi_{\text{feasible}}$ ), and with highly nonlinear

constraints ( $\Omega(\Phi)$ ) depending on the current estimate of the solution. None of the derivatives of the functions involved are known. This problem is currently solved in Velázquez-Muriel et al. (2003) using Powell's direction set method (Press et al., 1992, 10.5). This method iteratively minimizes the objective function by minimizing it in as many successive directions as the dimension of the problem (in this case 23). In each direction a line optimization is performed. The set of directions is chosen so that each new direction added to the set is conjugate of the set of already existing directions in the set. An interesting property of conjugate directions is that minimizing along one of them does not spoil the minimization achieved so far by minimizing in previous directions. It can be proved that this method converges quadratically to the minimum of the function.

Other optimization techniques have been tried on this problem like trust-region methods (Conn, Gould, and Toint, 2000). The main assumption of these methods is that the evaluation of the objective function is very expensive (computation time, resources, etc.) For this reason, the main optimization work is performed on a surrogate problem instead of the original problem itself. A model of the objective function is built based on a number of measurements of it. This initial measurements define the original trust-region (the region within which the model is trusted). Then, the model is minimized within the trust-region. The evaluation of the model is assumed to be much cheaper than the evaluation of the original objective function. Once a minimum of the model is achieved, the original objective function is evaluated at the minimum of the model. The new measurement is added to the model (one of the previous ones must be discarded), the trust-region is increased or decreased depending whether the minimum of the model corresponds to a decrease or increase of the objective function, and the new model is again optimized. This process is iterated until convergence.

The experiments performed on this problem with a trust-region method showed a much faster convergence to a minimum (smaller than the one found by Powell's direction set method) in terms of number of iterations and function evaluations (results unpublished). However, despite of the complexity of the cost function in (51), the hypothesis that evaluating the model was cheaper than evaluating the cost function was violated in this case, and in practical terms the trust-region method is not applied for detecting the CTF in the EM field.

Sinkevich et al. (2000) solve the CTF estimation problem assuming there is no noise before CTF, and therefore, the PSD of the observed micrograph is given by

$$PSD(\omega) = |CTF_{\Phi}(\omega)|^2 F(\omega) + PSD_{N_s, \Phi}(\omega).$$

Furthermore, they work only on radial averages boiling down to the model

$$\widehat{PSD}(\omega) = |\widehat{CTF}_{\Phi}(\omega)|^2 \widehat{F}(\omega) + \widehat{PSD}_{N_s, \Phi}(\omega),$$

where  $\omega = |\omega|$ .

This simplification automatically precludes the work with astigmatic images. Since  $\widehat{F}(\omega)$  is an unknown function that is common to all micrographs, these authors propose the estimation of the CTF of  $M$  micrographs simultaneously and, as a side-product, to estimate the structure factor  $\widehat{F}(\omega)$ . That is, to solve the following optimization problem:

$$\begin{aligned} \Phi^*, \widehat{F}^*(\omega) = \arg \min_{\Phi, \widehat{F}(\omega)} & \sum_{m=1}^M \sum_{\omega \in \Omega} |PSD^{(m)}(\omega) - |\widehat{CTF}_{\Phi}^{(m)}(\omega)|^2 \widehat{F}(\omega) - \widehat{PSD}_{N_s, \Phi}^{(m)}(\omega)|^2 \\ \text{s.t.} & \Phi \in \Phi_{\text{feasible}} \\ & \widehat{F}(\omega) \geq 0 \end{aligned} \quad (52)$$

where the superscript  $(m)$  denotes the evaluation of the corresponding quantity on the micrograph  $m$ . Note that in this case  $\Omega$  is a fixed set of frequencies independent from the parameters being estimated. Formulated like this, the CTF estimation becomes a constrained nonlinear least squares problem. To solve it, Sinkevich et al. (2000) use a primal-dual interior-point method (Wright, 1997); particularly, a Newton interior-point method (El-Bakry et al., 1996) due to the ability of these methods to treat inequality constraints since in their formulation the set  $\Phi_{\text{feasible}}$  is defined as the intersection of a number of inequalities.

An interesting result from Sinkevich et al. (2000) is that the solution for  $\widehat{F}(\omega)$  for a given  $\Phi$  is

$$\widehat{F}_{\Phi}^*(\omega) = \frac{\sum_{m=1}^M |\widehat{CTF}_{\Phi}^{(m)}(\omega)|^2 (PSD^{(m)}(\omega) - \widehat{PSD}_{N_d, \Phi}^{(m)}(\omega))}{\sum_{m=1}^M |\widehat{CTF}_{\Phi}^{(m)}(\omega)|^4} \quad (53)$$

Further information on this latter method can be found at Sinkevich (2000).

#### Problem 8: CTF amplitude correction

Once the CTF is identified, the next step is to correct for it. There exist several approaches to do this in EM. Among them, only two are expressed as the solution of a minimization problem.

Zhu et al. (1997) proposed to include the CTF in the projection operator  $L$  of the SIRT-type reconstruction method described in (44). This is done by matrix multiplication with a matrix  $C$  accounting for the convolution in real space with the inverse Fourier transform of the 3D CTF. They assumed that the 3D CTF is well defined for a volume by radially symmetrizing the 1D CTF. For doing this, it is necessary that the CTF is the same for all the projections used in the reconstruction. In this version of the algorithm the solution is regularized by imposing smoothness on the Laplacian of the volume. The Laplacian of the volume is computed by a discrete approximation expressed by a matrix  $B$ . Thereby, the reconstruction algorithm becomes

$$\mathbf{c}^{(k+1)} = \mathbf{c}^{(k)} + \lambda^{(k)}((1 - \kappa)C^t L^t (\mathbf{y} - L\mathbf{c}^{(k)}) - \kappa B^t B\mathbf{c}^{(k)}), \quad (54)$$

where  $\kappa$  is a factor between 0 and 1 that balances the importance of the regularization term with respect to the data term. This iterative method converges to the solution of  $\min_{\mathbf{c} \in \mathcal{K}} \|\mathbf{c}\|$  where  $\mathcal{K} = \{\mathbf{c} \mid \mathbf{c} = \arg \min_{\mathbf{c} \in \mathbb{R}^J} (1 - \kappa) \|\mathbf{y} - L\mathbf{c}\|^2 + \kappa \|B\mathbf{c}\|^2\}$ . This algorithm provides a regularized least-squares solution of the equation system  $L\mathbf{c} = \mathbf{y}$ . Zhu et al. (1997) shows how to consider the case in which the projections have different CTFs. Basically, it consists on grouping the set of projections by similar CTFs, and then applying a step like the one in (54) for each group.

If the 2D CTF of each image is used instead of the 3D CTF, then the equation system to solve is  $C\mathbf{L}\mathbf{c} = \mathbf{y}$  where now  $C$  is a matrix accounting for the spatial convolution of each image with its own CTF. Instead of solving this equation system, Zubelli et al. (2003) proposed to solve the normal equation

$$L^t C^t C L \mathbf{c} = L^t C^t \mathbf{y} \quad (55)$$

The problem in EM is that because of the decay of the CTF in high-frequency, the matrix  $L^t C^t C L$  is singular or nearly singular, and therefore the equation system is ill-posed. As a solution, Zubelli et al. (2003) proposed to modify the data as well as the CTF operator so

that the new equation system

$$L^t \tilde{C}^t \tilde{C} L \mathbf{c} = L^t \tilde{C}^t \tilde{\mathbf{y}} \quad (56)$$

is not ill-conditioned. The solutions of this equation system are shown to be close to the ones of (55). The following iterative algorithm

$$\mathbf{c}_j^{(t+1)} = \mathbf{c}_j^{(t)} \frac{(L^t \tilde{C}^t \tilde{\mathbf{y}})_j}{(L^t \tilde{C}^t \tilde{C} L \mathbf{c}^{(t)})_j} \quad (57)$$

can be shown to converge to  $\mathbf{c}^* = \arg \min_{\mathbf{c} \in \mathbb{R}^J} \|L^t C^t C L \mathbf{c} - L^t C^t \mathbf{y}\|^2$ . This algorithm is due to Chahine (1970) and one of its strengths is its simplicity since  $\tilde{C}$  is a symmetric matrix, and therefore,  $\tilde{C}^t = \tilde{C}$ . Thus, all that needs to be done is to apply the modified CTF (this can be very efficiently in Fourier Space) twice while projecting the volume and update the current solution in a multiplicative manner.

**Problem 9:** The reconstruction process

In the 2D alignment problem, the common EM procedure to iteratively refine the transformation parameters was related to a pseudo Expectation-Maximization approach. The idea was to start with a 2D model, to refine the projection translation and rotation parameters with respect to that model, and to average the aligned projections. The new average served as a model for the next iteration.

The 3D reconstruction process is carried out in a very much alike fashion. A first 3D model of the macromolecule under study is proposed. Then, the 3D transformation parameters are found using any of the techniques available for angular assignment. The aligned projections are “averaged” using a 3D reconstruction algorithm. And the reconstructed volume serves as a model for the next iteration of this process. Sometimes, the correction for the CTF amplitude effects is included within the loop. This iterative process is a pseudo Expectation-Maximization approach of the same kind as the one found in 2D alignment.

### 3. Conclusions

In this paper we have reviewed a wide range of optimization problems found in currently used techniques in image processing of electron microscopy single-particles. As can be seen, the set of optimization techniques needed to solve them is quite rich, picking methods of nearly all kinds. However, there are still many open problems facing the EM community. We list here a short list indicating some of them:

- As was already pointed out in the “Classification problem” (see Problem 3), particles are flexible objects. Hence, images from two biochemically and functionally identical macromolecules from the same projection direction may not be exactly the same (except for the noise) since the two molecules may be subjected to different forces within the sample holder. For this reason, the two particles may have deformed slightly differently. This fact severely affects the maximum achievable resolution since averaging flexible particles results in a high-quality volume up to the minimum size of the flexible parts. Models of the particle flexibility appeared very recently (Chacón, Tama, and Wriggers, 2003; Ming et al., 2002; Tama, Wriggers, and Brooks, 2002) and have not yet been incorporated in the image processing steps.

- The structure of several thousands of macromolecules is already known and stored in public databases (Berman et al., 2000; Boutselakis et al., 2003; Golovin et al., 2004). Unfortunately, no EM reconstruction algorithm incorporates this information, that is, produces a 3D reconstruction compatible with the micrographs that maximizes the probability of being a particle (or what is the same: it shares the characteristics of the already reconstructed molecules that perform a similar function).
- Due to the presence of a high level of noise, usually reconstructed volumes are contaminated by undesired artifacts that are still compatible with the acquired data. However, they are due to the presence of the noise, CTF, lack of projection directions, etc. Reconstruction algorithms should be regularized to minimize the presence of this kind of artifacts. Finding the minimum norm solution (as shown in the Problem 5: 3D reconstruction) is an attempt in this line. However, more powerful regularization terms are needed to further improve the quality of the reconstructed volumes.
- Despite of the fact that computers are becoming more and more powerful, processing time is becoming a bottle-neck in EM (Fernández et al., 2002; 2005) The amount of data is increasing exponentially and, nowadays, it is normal to work with several dozens of thousands images to achieve high resolution. Convergence rates of the different algorithms and execution times are, therefore, currently even more important if the structure of a macromolecule is to be solved in a reasonable time.

We hope the problematic described in this review encourage optimization researchers and mathematicians to study and propose new methods that outperform the existing ones in a unified attempt to increase the resolution of the 3D reconstruction of biological macromolecules.

**Acknowledgments** We acknowledge partial support from the “Comunidad Autónoma de Madrid” through grants CAM-07B-0032-2002, GR/SAL/0653/2004 and GR/SAL/0342/2004, the “Comisión Interministerial de Ciencia y Tecnología” of Spain through grants BIO2001-1237, BIO2001-4253-E, BIO2001-4339-E, BIO2002-10855-E, BFU2004-00217/BMC, the Spanish FIS grant (G03/185), the European Union through grants QLK2-2000-00634, QLRI-2000-31237, QLRT-2000-0136, QLRI-2001-00015, FP6-502828 and the NIH through grant HL70472. Alberto Pascual and Roberto Marabini acknowledge support by the Spanish Ramon y Cajal Program.

## References

- Baker, T., N. Olson, and S. Fuller. (1999). “Adding the Third Dimension to Virus Life Cycles: Three-Dimensional Reconstruction of Icosahedral Viruses from Cryo-Electron Micrographs.” *Microbiol. Mol. Biol.*, 63, 862–922.
- Batagelj, V. (1988). “Generalized Ward and Related Clustering Problems.” In H.H. Bock (Ed.), *Classification and Related Methods of Data Analysis*, pp. 67–74. North-Holland.
- Bellon, P., F. Cantele, and S. Lanzavecchia. (2001). “Correspondence Analysis of Sinogram lines. Sinogram trajectories in Factor Space Replace raw Images in the Orientation of Projections of Macromolecular Assemblies.” *Ultramicroscopy*, 87, 187–197.
- Bellon, P.L., F. Cantele, S. De Carlo, and S. Lanzavecchia. (2002). “A Trajectory-Based Algorithm to Determine and Refine Euler Angles of Projections in Three-Dimensional Microscopy. Improvements and Tests.” *Ultramicroscopy*, 93, 111–121.
- Berman, H., J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, H. Weissig, I. Shindyalov, and P. Bourne. (2000). “The Protein Data Bank.” *Nucleic Acids Research*, 28, 235–242.
- Bezdek, J.C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum.
- Bijaoui, A. (2002). “Wavelets, Gaussian Mixtures and Wiener Filtering.” *Signal Processing*, 82, 709–712.
- Bishop, C.M. (1995). *Neural Networks For Pattern Recognition*. Oxford University Press.
- Boisset, N., P. Penczek, J.C. Taveau, V. You, F. de Haas, and J. Lamy. (1998). “Overabundant Single-Particle Electron Microscope Views Induce a Three-Dimensional Reconstruction Artifact.” *Ultramicroscopy*, 74, 201–207.

- Boutselakis, H., D. Dimitropoulos, J. Fillon, A. Golovin, K. Henrick, A. Hussain, J. Ionides, M. John, P.A. Keller, E. Krissinel, P. McNeil, A. Naim, R. Newman, T. Oldfield, J. Pineda, A. Rachedi, J. Copeland, A. Sitnov, S. Sobhany, A. Suarez-Uruena, J. Swaminathan, M. Tagari, J. Tate, S. Tromm, S. Velankar, and W. Vranken. (2003). "E-MSD: The European Bioinformatics Institute Macromolecular Structure Database." *Nucleic Acids Research*, 31, 458–462.
- Bracewell, R. (1986). *The Fourier Transform and Its Applications*, Edn. 2. Electrical and Electronic Engineering. McGrawHill International.
- Carazo, J.M., F.F. Rivera, E.L. Zapata, M. Radermacher, and J. Frank. (1990). "Fuzzy Sets-Based Classification of Electron Microscopy Images of Biological Macromolecules with an Application to Ribosomal Particles." *J. Microscopy*, 157, 187–203.
- Chacón, P., F. Tama, and W. Wriggers. (2003). "Mega-Dalton Biomolecular Motion Captured from Electron Microscopy Reconstructions." *J. Molecular Biology*, 326, 485–492.
- Chahine, M.T. (1970). "Inverse Problems in Radiative Transfer: Determination of Atmospheric Parameters." *J. Atmospheric Science*, 27, 960–967.
- Conn, A.R., N.I.M. Gould, and P.L. Toint. (2000). *Trust-Region Methods*. SIAM.
- De Rosier, D.J. and A. Klug. (1968). "Reconstruction of Three Dimensional Structures from Electron Micrographs." *Nature*, 217, 130–134.
- Dempster, A., N. Laird, and D. Rubin. (1977). "Maximum Likelihood from Incomplete Data Via the EM Algorithm." *J. of the Royal Statistical Society, Series B* 39, 1–38.
- Dennis, J.E. and R.B. Schnabel. (1996). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Philadelphia: SIAM.
- Dunn, J.C. (1973). "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters." *J. of Cybernetics*, 3, 32–57.
- El-Bakry, A., R.A. Tapia, T. Tsuchiya, and Y. Zhang. (1996). "On the Formulation and Theory of the Newton Interior Point Method for Nonlinear Programming." *J. Optimization Theory and Applications*, 89, 507–541.
- Ellis, M. and H. Hebert. (2001). "Structure Analysis of Soluble Proteins using Electron Crystallography." *Micron*, 32, 541–550.
- Fernández, J., A. Lawrence, J. Roca, I. García, M. Ellisman, and J. Carazo. (2002). "High Performance Electron Tomography of Complex Biological Specimens." *J. Structural Biology*, 138, 6–20.
- Fernández, J., J. Bilbao-Castro, R. Marabini, J. Carazo, and I. Garcia. (2005). "Biological Structure Determination by EM is Well Suited to Grid Computing." *New Generat. Comput. (Special Issue on Grid Systems for Life Science)*, 23, 101–112.
- Frank, J. (1996). *Three Dimensional Electron Microscopy of Macromolecular Assemblies*. San Diego, CA: Academic Press.
- Frank, J. (2002). "Single-Particle Imaging of Macromolecules by Cryo-Electron Microscopy." *Annual Review of Biophysics & Biomolecular Structure*, 31, 303–319.
- Frank, J. (2006). *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State*. USA: Oxford Univ. Press.
- Frank, J. and L. Al-Ali. (1975). "Signal-to-Noise Ratio of Electron Micrographs Obtained by Cross Correlation." *Nature*, 256, 376–379.
- Frank, J., J.P. Breaudiere, J.M. Carazo, A. Verschoor, and T. Wagenknecht. (1988a). "Classification of Images of Biomolecular Assemblies: A Study of Ribosomes and Ribosomal Subunits of Escherichia Coli." *J. of Microscopy*, 150, 99–115.
- Frank, J., W. Chiu, and L. Degn. (1988b). "The Characterization of Structural Variations within a Crystal Field." *Ultramicroscopy*, 26, 345–360.
- Frank, J. and M. van Heel. (1982). "Correspondence Analysis of Aligned Images of Biological Particles." *J. Molecular Biology*, 161, 134–137.
- Furuie, S.S., G.T. Herman, T.K. Narayan, P.E. Kinahan, J.S. Karp, R.M. Lewitt, and S. Matej. (1994). "A Methodology for Testing for Statistically Significant Differences Between Fully 3D PET Reconstruction Algorithms." *Physics in Medicine & Biology*, 39, 341–354.
- Golovin, A., T. Oldfield, J.G. Tate, S. Velankar, G.J. Barton, H. Boutselakis, D. Dimitropoulos, J. Fillon, A. Hussain, J.M. Ionides, M. John, P.A. Keller, E. Krissinel, P. McNeil, A. Naim, R. Newman, A. Pajon, J. Pineda, A. Rachedi, J. Copeland, A. Sitnov, S. Sobhany, A. Suarez-Uruena, G.J. Swaminathan, M. Tagari, S. Tromm, W. Vranken, and K. Henrick. (2004). "E-MSD: An Integrated Data Resource for Bioinformatics." *Nucleic Acids Research*, 32, D211–D216.
- Herman, G.T. (1980). *Image Reconstruction From Projections: The Fundamentals of Computerized Tomography*. New York: Academic Press.



- Herman, G.T. (1998). "Algebraic Reconstruction Techniques in Medical Imaging." In C.T. Leondes (Ed.), *Medical Imaging, Systems Techniques and Applications*, vol. 6: Computational Techniques, PP. 1–42. Amsterdam: Gordon and Breach Science Publishers.
- Jain, A.K. (1989). *Fundamentals of Digital Image Processing*. Prentice-Hall.
- Jonic, S., C.O.S. Sorzano, P. Thévenaz, C. El-Bez, S. De Carlo, and M. Unser. (2005). "Spline-Based Image-to-Volume Registration for Three-Dimensional Electron Microscopy." *Ultramicroscopy*, 103/104, 303–317.
- Kak, A. and M. Slaney. (1987). *Principles of Computerized Tomographic Imaging*. IEEE Press.
- Kohonen, T. (1982). "Self-Organized Formation of Topologically Correct Feature Maps." *Biol. Cybernet.* 43, 59–69.
- Lenz, F.A. (1971). "Transfer of Image Information in the Electron Microscope." In U. Valdrè (Ed.), *Electron Microscopy in Material Sciences*, pp. 540–569. Academic Press.
- Lindhal, M. (2001). "Strul—A Method for 3D Alignment of Single-Particle Projections Based on Common Line Correlation in Fourier Space." *Ultramicroscopy*, 87, 165–175.
- Ludtke, S.J., P.R. Baldwin, and W. Chiu. (1999). "EMAN: Semiautomated Software for High-Resolution Single-Particle Reconstructions." *J. Structural Biology*, 128, 82–97.
- Mallat, S. (1999). *A Wavelet Tour of Signal Processing*. Academic Press.
- Marabini, R. and J.M. Carazo. (1994). "Pattern Recognition and Classification of Images of Biological Macromolecules Using Artificial Neural Networks." *Biophysical Journal*, 66, 1804–1814.
- Matej, S., S.S. Furuie, and G.T. Herman. (1996). "Relevance of Statistically Significant Differences Between Reconstruction Algorithms." *IEEE Trans. Image Processing*, 5, 554–556.
- Matej, S., G.T. Herman, T.K. Narayan, S.S. Furuie, R.M. Lewitt, and P.E. Kinahan. (1994). "Evaluation of Task-Oriented Performance of Several Fully 3D PET Reconstruction Algorithms." *Physics in Medicine & Biology*, 39, 355–367.
- Ming, D., Y. Kong, S. Wakil, J. Brink, and J.Ma. (2002). "Domain Movements in Human Fatty Acid Synthase by Quantized Elastic Deformational Model." In *Proc. Natl. Acad. Sci. USA* 99, 7895–7899.
- Natterer, F. and F. Wübbeling. (2001). *Mathematical Methods in Image Reconstruction*. Philadelphia: SIAM.
- Pascual-Marqui, R.D., A. Pascual-Montano, K. Kochi, and J.M. Carazo. (2001). "Smoothly Distributed Fuzzy c-Means: A New Self-Organizing Map." *Pattern Recognition*, 34, 2395–2402.
- Pascual-Montano, A., L.E. Donate, M. Valle, M. Bárcena, R.D. Pascual-Marqui, and J.M. Carazo. (2001). "A Novel Neural Network Technique for Analysis and Classification of EM Single-Particle Images." *J. Structural Biology*, 133, 233–245.
- Penczek, P., M. Radermacher, and J. Frank. (1992). "Three-Dimensional Reconstruction of Single Particles Embedded in Ice." *Ultramicroscopy*, 40, 33–53.
- Penczek, P.A., R.A. Grassucci, and J. Frank. (1994). "The Ribosome at Improved Resolution: New Techniques for Merging and Orientation Refinement in 3D Cryo-Electron Microscopy of Biological Particles." *Ultramicroscopy*, 53, 251–270.
- Penczek, P.A., J. Zhu, and J. Frank. (1996). "A Common-Lines Based Method for Determining Orientations for N>3 Particle Projections Simultaneously." *Ultramicroscopy*, 63, 205–218.
- Press, W., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. (1992). *Numerical Recipes in C*, 2 edition. Cambridge University Press.
- Radermacher, M. (1992). "Weighted Back-Projection Methods." In J. Frank (ed.), *Electron Tomography*, pp. 91–115. Plenum.
- Radermacher, M. (1994). "Three-Dimensional Reconstruction from Random Projections—Orientational Alignment via Radon Transforms." *Ultramicroscopy*, 53, 121–136.
- Sali, A., R. Glaeser, T. Earnest, and W. Baumeister. (2003). "From Words to Literature in Structural Proteomics." *Nature*, 422, 216–225.
- San Martín, C. (1996). *Procesamiento de Imágenes de Microscopía Electrónica Aplicado Al Estudio de la Estructura de dos ADN-Helicadas Hexaméricas Representativas*. Ph.D. thesis, Univ. de Santiago de Compostela.
- San Martín, M.C., N.P.J. Stamford, N. Dammerova, N.E. Dixon, and J.M. Carazo. (1995). "A Structural Model for the *Escherichia coli* DnaB helicase based on electron-microscopy data." *J. Structural Biology*, 114, 167–176.
- Scheres, S.H.W., M. Valle, R. Núñez, C.O.S. Sorzano, R. Marabini, G.T. Herman, and J.M. Carazo. (2005). "Maximum-Likelihood Multi-Reference Refinement for Electron Microscopy Images." *J. Molecular Biology*, 348, 139–149.
- Sigworth, F.J. (1998). "A Maximum-Likelihood Approach to Single-Particle Image Refinement." *J. Structural Biology*, 122, 328–339.
- Sinkevich, O. (2000). *Optimization for Parameter Estimation with Application to Transmission Electron Microscopy*. Ph.D. thesis, Rice University.

- Sinkevich, O., R.A. Tapia, Y. Zhang, and S.J. Ludtke. (2000). "Simultaneous Structure Factor and Contrast Transfer Function Parameter Determination in Transmission Electron Microscopy." Technical Report, Dept. of Computational and Applied Mathematics, Rice Univ. TR00-36.
- Skoglund, U., L.G. Ofverstedt, R. Burnett, and G. Bricogne. (1996). "Maximum-Entropy Three-Dimensional Reconstruction with Deconvolution of the Contrast Transfer Function: A Test Application with Adenovirus." *J. Structural Biology*, 117, 173–188.
- Sorzano, C.O.S., L.G. de la Fraga, R. Clackdoyle, and J.M. Carazo. (2004a). "Normalizing Projection Images: A Study of Image Normalizing Procedures for Single Particle Three-Dimensional Electron Microscopy." *Ultramicroscopy*, 101, 129–138.
- Sorzano, C.O.S., S. Jonic, C. El-Bez, J.M. Carazo, S. De Carlo, P. Thévenaz, and M. Unser. (2004b). "A Multiresolution Approach to Pose Assignment in 3-D Electron Microscopy of Single Particles." *J. Structural Biology*, 146, 381–392.
- Sorzano, C.O.S., R. Marabini, N. Boisset, E. Rietzel, R. Schröder, G.T. Herman, and J.M. Carazo. (2001). "The Effect of Overabundant Projection Directions on 3D Reconstruction Algorithms." *J. Structural Biology*, 133, 108–118.
- Sorzano, C.O.S., R. Marabini, G.T. Herman, and J.M. Carazo. (2005a). "Multiobjective Algorithm Parameter Optimization Using Multivariate Statistics." *Pattern Recognition*, 38, 2587–2601.
- Sorzano, C.O.S., E. Ortiz, M. López, and J. Rodrigo. (2006). "Improved Bayesian Image Denoising Based on Wavelets with Applications to Electron Microscopy." *Pattern Recognition* (in press).
- Sorzano, C.O.S., P. Thévenaz, and M. Unser. (2005b). "Elastic Registration of Biological Images Using Physically Constrained Regularization Term." *IEEE Trans. on Biomedical Eng.*, 52, 652–663.
- Tama, F., W. Wriggers, and C. Brooks. (2002). "Exploring Global Distortions of Biological Macromolecules and Assemblies from Low-Resolution Structural Information and Elastic Network Theory." *J. Molecular Biology*, 321, 297–305.
- Thornton, J.M., A.E. Todd, and D. Milburn. (2000). "From Structure to Function: Approaches and Limitations." *Nature Structural Biology*, 7 (Suppl.), 991–994.
- Thévenaz, P., U.E. Ruttiman, and M. Unser. (1998). "A Pyramid Approach to Subpixel Registration Based on Intensity." *IEEE Trans. Image Processing*, 7, 27–41.
- Unser, M. (1999). "Splines: A Perfect Fit for Signal and Image Processing." *IEEE Signal Processing Magazine*, 22–38.
- Unser, M., A. Aldroubi, and M. Eden. (1991). "Fast B-spline Transforms for Continuous Image Representation and Interpolation." *IEEE Trans. Pattern Analysis & Machine Intelligence*, 13, 277–285.
- Unwin, P. (1973). "Phase Contrast Electron Microscopy of Biological Materials." *J. Microscopy*, 98, 299–312.
- Valle, M., C. Gruss, L. Halmer, J.M. Carazo, and L.E. Donate. (2000). "Large T-Antigen Double Hexamers Imaged at the Simian Virus 40 Origin of Replication." *Molecular and Cellular Biology*, 20, 34–41.
- van Heel, M. (1984). "Multivariate Statistical Classification of Noisy Images (Randomly Oriented Biological Macromolecules)." *Ultramicroscopy*, 13, 165–183.
- van Heel, M. (1987). "Angular Reconstitution: A *Posteriori* Assignment of Projection Directions for 3D Reconstruction." *Ultramicroscopy*, 21, 111–124.
- van Heel, M. and J. Frank. (1981). "Use of Multivariate Statistical Statistics in Analysing the Images of Biological Macromolecules." *Ultramicroscopy*, 6, 187–194.
- Velázquez-Muriel, J.A., C.O.S. Sorzano, J.J. Fernández, and J.M. Carazo. (2003). "A Method for Estimating the CTF in Electron Microscopy Based on ARMA Models and Parameter Adjusting." *Ultramicroscopy*, 96, 17–35.
- Wahba, G., D.R. Johnson, F. Gao, and J. Gong. (1994). "Adaptive Tuning of Numerical Weather Prediction Models. I Randomized GCV and Related Methods in Three and Four Dimensional Data Assimilation." Technical Report TR-920, Dept. Statistics (Univ. of Wisconsin-Madison).
- Ward, J.H. (1963). "Hierarchical Grouping to Optimize an Objective Function." *Am. Statist. Assoc. J*, 58, 236–244.
- Webb, A.R. (2002). *Statistical Pattern Recognition*, 2 edition. New York, U.S.A.: John Wiley and Sons.
- Wright, S. (1997). *Primal-Dual Interior Point Methods*. Philadelphia, PA, USA: SIAM.
- Zhu, J., P.A. Penczek, R. Schröder, and J. Frank. (1997). "Three-Dimensional Reconstruction with Contrast Transfer Function Correction from Energy-Filtered Cryoelectron Micrographs: Procedure and Application to the 70S *Escherichia coli* Ribosome." *J. Structural Biology*, 118, 197–219.
- Zubelli, J.P., R. Marabini, C.O.S. Sorzano, and G.T. Herman. (2003). "Three-Dimensional Reconstruction by Chahine's Method from Electron Microscopic Projections Corrupted by Instrumental Aberrations." *Inverse Problems*, 19, 933–949.
- Zupan, J. (1982). *Clustering of Large Datasets*. Letchworth, U.K.: Research Studies Press.