Avoiding Local Optima in Single Particle Reconstruction

Marshall Bern[1], Jindong (JD) Chen[1], and Hao Chi Wong[1, 2]

[1] Palo Alto Research Center, 3333 Coyote Hill Rd., Palo Alto, CA 94304, USA
[2] Departamento de Ciência da Computação, UFMG, Belo Horizonte, MG, Brazil.
{bern,jchen,hcwong}@parc.com

Abstract. In single-particle reconstruction, a 3D structure is reconstructed from a large number of randomly oriented 2D projections, using techniques related to computed tomography. Unlike in computed tomography, however, the orientations of the projections must be estimated at the same time as the 3D structure, and hence the reconstruction process can be error-prone, converging to an incorrect local optimum rather than the true 3D structure. In this paper, we discuss and further develop a maximum-likelihood approach to reconstruction, and demonstrate that this approach can help avoid incorrect local optima for both 2D and 3D reconstructions.

1   Introduction

Cryo-Electron Microscopy (cryo-EM) uses a transmission electron microscope to acquire 2D projections of a specimen preserved in vitreous ice. A 3D electron density map can then be reconstructed from the 2D projections computationally. In "single-particle" cryo-EM, the specimen consists of many ostensibly identical copies of randomly oriented particles, and the reconstruction process must estimate the unknown orientations at the same time that it estimates the 3D structure. (Other types of cryo-EM specimens, such as 2D crystals or helical filaments, exhibit naturally ordered arrangements that greatly simplify the orientation problem.)

Cryo-EM is emerging as an important technique in structural biology, because it can determine 3D structures without the need for crystallization. It is especially well-suited to the study of large molecular complexes, which are vitally important but very difficult to study by other means [1]. Single particle cryo-EM [14] has been used quite extensively to study the ribosome, achieving about 10 Å resolution with 20,000–50,000 images. Cryo-EM can also be used in conjunction with x-ray crystallography, with x-ray data supplying atomic resolution (3 Å or better) for subunits whose overall arrangement in natural conditions is determined by cryo-EM. A recent demonstration [20] of the tremendous potential of such a combined approach is the atomic-resolution structure of an 11-subunit bacterial flagellar filament (a helical specimen).

The central issue in single particle reconstruction is the chicken-and-egg problem of determining image orientations without knowing the particle shape, and determining particle shape without knowing the orientations. The solution developed over the last 20 years [5, 18] is an iterative process that starts from an initial guess of the shape, perhaps as simple as an ellipsoid, and then iteratively aligns images to the current shape and reconstructs a new shape from the aligned images, as shown in Figure 1. In the
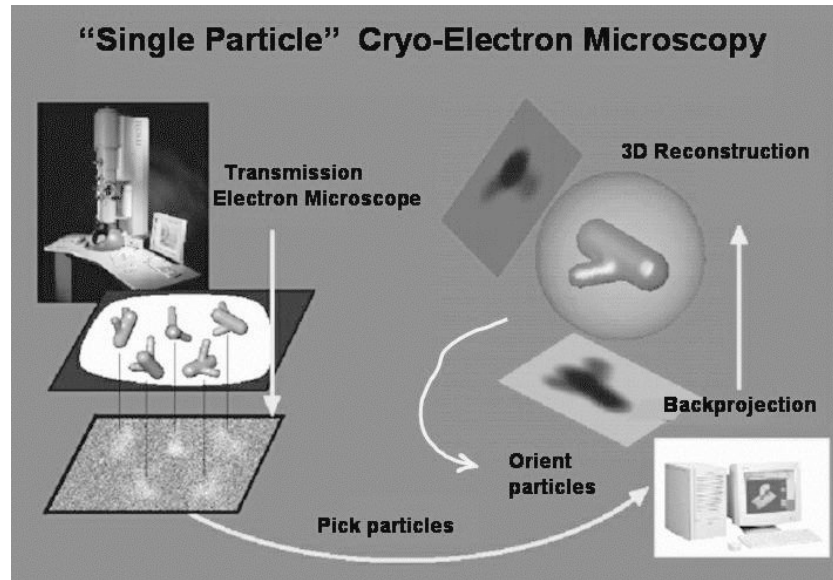
Fig. 1. The processing pipeline for single-particle cryo-EM includes particle picking, followed by iterations of orientation estimation and 3D reconstruction. (Figure adapted from [17])

standard approach [15], (sometimes called "angular reconstitution" [18]), each particle image contributes equally to the reconstruction, with a single orientation set by the image's best alignment, typically determined by maximum correlation. The standard approach—implemented in three major software packages, IMAGIC [18], SPIDER [6], and EMAN [10]—is remarkably successful, giving correct structures from images that to the human eye appear to be almost pure noise. Yet occasionally, depending upon the initial guess, the standard approach gives an incorrect structure, close to a fixed point for the iteration but nowhere near the true structure. See Figure 8(c).

What can be done to prevent such a disaster? One possible solution is a better initial guess. If the particle has strongly preferred orientations, as many do, then it is possible with the "random conical tilt" approach [5] to obtain a set of projections with known relative orientations. Similarly, cryo-electron tomography [9] images the very same particle from a number of known orientations. In either case the result is a low-resolution (say 30–40 Å) 3D reconstruction that can be used as the initial guess for a subsequent single-particle reconstruction. (Why not use tomography for the entire problem? The answer is that accumulated electron dose destroys the particle, so the number of projections that can be gathered from a single particle is quite limited.) Obtaining tilted images, however, requires special equipment and set-up, and even with the better initial guess there is no guarantee that the orientation/reconstruction cycle will converge to the right answer. Hence we would also like an algorithmic solution to the problem of incorrect local optima. In this paper we investigate such a solution, based on an approach to single-particle reconstruction due to Sigworth [16].

Sigworth's approach assumes a probabilistic model of cryo-EM imaging and seeks the maximum-likelihood (ML) reconstruction. The well-known Expectation Maximization algorithm (EM applied to EM!) is used to maximize the likelihood [3]; this amounts to adding each image into the reconstruction for each possible orientation weighted according to the probability of that orientation (that is, "soft" rather than "hard" orientation assignments). Sigworth demonstrated his ML approach for an analogous 2D problem of independent interest: reconstructing an image from a number of noisy copies, randomly rotated and translated. The ML approach showed reduced sensitivity to the initial guess and recovered structures from images with lower signal-to-noise ratio. In this paper, we demonstrate that the ML approach can also avoid incorrect local optima, first with synthetic 2D examples (Section 3), and then with real 2D and 3D data (Sections 4 and 5).

2   Reconstruction Algorithms

Following Sigworth [16], we describe reconstruction algorithms for an analogous 2D problem, rather than for the full 3D problem. The analogy between the two problems is not perfect, because the 2D problem does not involve projections (integrals of the density along parallel viewing lines) of the original structure; nevertheless both include the central chicken-and-egg problem. The 2D problem arises as a subproblem in 3D reconstruction in each of the major software packages [5, 6, 10].

Let the particle images—the data—be denoted $D_1, D_2, \ldots, D_n$. Each image D_i is assumed to be a noisy copy of the same original image Φ, rotated and translated by some unknown amounts. We are assuming that the D_i's have been located by a particle picking process that approximately locates the center of each particle, but leaves the rotation and the precise translation unknown [16]. Let $\theta_i = (\alpha_i, x_i, y_i)$ denote the *alignment parameters* for image D_i, with α_i the rotation angle and (x_i, y_i) the translation vector. That is,

$$D_i = z \cdot (-\theta_i(\Phi)) + (1 - z) \cdot G_i, \tag{1}$$

where G_i is the noise, z is a constant with $0 \leq z \leq 1$, and the notation $(-\theta_i(\Phi))$ means Φ transformed by the rotation and translation $-\theta_i()$ (so that the alignment parameters $\theta_i()$ return D_i to its original orientation). Assuming Φ and G_i have the same variance, $z^2/(1-z)^2$ gives the signal-to-noise ratio (SNR), defined as the ratio of signal variance to noise variance [5]. When we refer to images as 95% noise we mean that $1 - z = 0.95$.

2.1   The Standard Method

The standard method [12, 15] attempts to directly determine the alignment parameters. Given the reconstruction $\Phi^{(k)}$ after k iterations, this method computes a number of *reference images*, denoted $R_1^{(k)}, R_2^{(k)}, \ldots, R_m^{(k)}$, all possible rotations and translations of $\Phi^{(k)}$ quantized to some pre-determined level of precision.[3] To determine the alignment

---

[3] In practice, references are produced only for different rotations, and translations are determined by convolution using the Fast Fourier Transform.

parameters for image D_i, the standard method finds the j that maximizes the cross-correlation $D_i \cdot R_j^{(k)}$ and sets $\theta_i^{(k)}$ to be equal to the rotation and translation values of the maximizing $R_j^{(k)}$. As usual the cross-correlation is the inner product of the images, treated as vectors.

Once the alignment parameters are determined, then the next reconstruction is simply the aligned average of the data images,

$$\Phi^{(k+1)} \;=\; \frac{1}{n} \sum_{i=1}^{n} \theta_i^{(k)}(D_i). \tag{2}$$

For continuous values of the alignment parameters, this iteration converges to a local maximum of the Euclidean norm of the reconstruction $\|\Phi\|_2$, in other words the strongest-contrast reconstruction [12, 16]. With quantized alignment parameters, however, the iteration does not necessarily converge but rather reaches a small limit cycle. Limit cycles are not usually reached in real biological projects, which use 1000's of data images, 1000's of reference images, and fewer than 50 iterations, but they are indeed observable in our small 2D experiments (Figure 4).

2.2 Maximum-Likelihood Method

For low-SNR images, random correlation peaks cause mistakes in the alignment parameters, and the standard method can give poor results. Sigworth's method [16] never actually determines the alignment parameters, but rather treats them as hidden random variables with probability distributions over possible values, and then attempts to maximize the likelihood of the data images D_i as a function of the reconstruction and the parameters of the probability distributions. Sigworth derives his algorithm, a version of the Baum-Welch or EM algorithm [3], assuming that the data arises from a specific generative model: additive Gaussian pixel noise (independent and identically distributed), normally distributed image translations, and uniformly distributed image rotations. In this section, we rederive the ML approach in a more realistic setting. In Section 2.3 we discuss the key step of the algorithm in more detail. Then in Section 2.4 we develop three variations on the ML approach that address some of the practical issues.

The likelihood is the probability that the observed data set $D_1, D_2, \ldots, D_n$ arises from a given model. Our model assumes an original image Φ and alignment parameters $\theta_1, \theta_2, \ldots, \theta_n$ for the data images. Assume that the alignment parameters, rather than being continuous random variables as they are in real life, are discrete variables that take exactly the same values as the rotation angles and translations of the reference images $R_1, R_2, \ldots, R_m$. Assume that the data images $D_1, D_2, \ldots, D_n$ are produced by a random process that first selects a reference R_j $(= R_j(\Phi))$ according to any probability distribution and then sets D_i to be a copy of R_j with added noise. This *mixture model* is more realistic than Sigworth's classical parametric model, because cryo-EM target particles usually have preferred orientations and are not uniformly distributed over 3D rotations. Let π_j denote the probability of selecting reference R_j, with $\sum_j \pi_j = 1$. For the model just described, the likelihood is a function of Φ (through the reference copies $\{R_j\}$) and $\Pi = (\pi_1, \pi_2, \ldots, \pi_m)$. Assuming independence of the random variables associated with the individual data images, we can express the logarithm of the likelihood

---

1. Compute references $R_1^{(k)}, R_2^{(k)}, \ldots, R_n^{(k)}$ from the current reconstruction $\Phi^{(k)}$
2. **for** each data image D_i, $i = 1, 2, \ldots, n$ **do**
   2.1. **for** $j = 1, 2, \ldots, m$, compute $\gamma_{ij}^{(k)}$, an estimate of probability $P[D_i \mid R_j^{(k)}]$
   2.2. Normalize $\gamma_{ij}^{(k)}$ values so they sum to 1, that is, $\gamma_{ij}^{(k)} \leftarrow \gamma_{ij}^{(k)} / \left(\sum_j \gamma_{ij}^{(k)} \right)$
   2.3. **for** $j = 1, 2, \ldots, m$, add D_i to the next reconstruction $\Phi^{(k+1)}$
        with orientation j and weight $\gamma_{ij}^{(k)}$
3. $k \leftarrow k + 1$ and go to 1.

Fig. 2. The EM algorithm applied to single particle reconstruction.

(log likelihood) $\mathcal{L}(\Phi, \Pi)$ as a sum of the log likelihoods of the individual images.

$$\mathcal{L}(\Phi, \Pi) \;=\; \sum_{i=1}^{n} \log \left(\sum_{j=1}^{m} \pi_j \, P[D_i \mid R_j] \right). \tag{3}$$

The expression in equation (3) is difficult to maximize directly, due to the sum of terms inside the logarithm. So we introduce unobserved "latent" variables δ_{ij}, with $\delta_{ij} = 1$ if D_i came from R_j and $\delta_{ij} = 0$ otherwise. (See [7], pp. 236–241, for a similar development of the EM algorithm.) This trick allows the sum to be pulled outside the logarithm, and the log likelihood becomes

$$\mathcal{L}(\Phi, \Delta) \;=\; \sum_{i=1}^{n} \sum_{j=1}^{m} \delta_{ij} \, \log P[D_i \mid R_j]. \tag{4}$$

The EM algorithm given in Figure 2 is the standard way to maximize a likelihood expression of this form [7]. In its "E-step" (Steps 2.1 and 2.2), this algorithm makes a soft assignment to δ_{ij}; that is, rather than using a value in $\{0, 1\}$ for δ_{ij}, the algorithm uses a fractional value $\gamma_{ij}^{(k)}$, the expectation of δ_{ij} given the current reconstruction $\Phi^{(k)}$ and references $R_j^{(k)}$. Rather remarkably, this algorithm for maximizing $\mathcal{L}(\Phi, \Delta)$ also monotonically increases the original log likelihood $\mathcal{L}(\Phi, \Pi)$ until it reaches a local maximum [7]; indeed the two log likelihoods agree at a joint maximum.

2.3 Estimating the Image Probabilities

The algorithm as given in Figure 2 is not fully specified. We must give a way to compute $\gamma_{ij}^{(k)}$, which is an estimate of the probability that image D_i is a noisy copy of reference $R_j^{(k)}$. One possibility is simply to set $\gamma_{ij}^{(k)} = 1$ if $R_j^{(k)}$ maximizes $D_i \cdot R_j^{(k)}$ and $\gamma_{ij}^{(k)} = 0$ otherwise. With these crude estimates of $\gamma_{ij}^{(k)}$, the EM algorithm reduces to the standard method of Section 2.1. A more principled estimate of $\gamma_{ij}^{(k)}$ assumes a probabilistic model of image formation. Sigworth's model [16] assumes independent

Gaussian noise with variance σ^2 in the image pixels, and hence

$$P[D_i \mid R_j^{(k)}] \; \propto \; \exp\left(-\sum_p \frac{(D_i(p) - R_j^{(k)}(p))^2}{\sigma^2}\right), \tag{5}$$

where $D_i(p)$ is the p-th pixel of D_i, and $R_j^{(k)}(p)$ is the p-th pixel of $R_j^{(k)}$. Since multiplying all the $\gamma_{ij}^{(k)}$'s by the same constant does not change the reconstruction, we can simply set $\gamma_{ij}^{(k)}$ in Step 2.1 to the expression on the right in equation (5). To do so, we must also have an estimate of the noise variance σ^2. In our synthetic-data experiments, we simply plugged in the true value for σ^2. In attempting to add an EM option to EMAN, however, Ludtke (personal communication) found that setting a value for σ^2 that works for real data images was the major stumbling block. The exponential in equation (5) makes $\gamma_{ij}^{(k)}$ quite sensitive to incorrect σ^2 and to deviations of the real data from the imaging model. The problem is not estimating the noise variance, which is relatively easy, but rather the assumption of independent pixels. Another issue is numerical underflow, as $\gamma_{ij}^{(k)}$ values are small, something like e^{-2400} for our synthetic images, so calculations must be done with logarithms. Assuming these difficulties can be overcome, a well-known shortcut to computing the expression in equation (5) is to use the cross-correlation $D_i \cdot R_j^{(k)}$ and the relation

$$-\sum_p (D_i(p) - R_j^{(k)}(p))^2 \; = \; 2D_i \cdot R_j^{(k)} - \sum_p (D_i(p))^2 - \sum_p (R_j^{(k)}(p))^2. \tag{6}$$

Step 2.2 also has an impact on the image probabilities. This step sets $\sum_j \gamma_{ij}^{(k)}$ equal to one, which has the consequence that each D_i contributes equally to the reconstruction $\Phi^{(k+1)}$. Sigworth's version of the EM algorithm omits this step, because it is derived from a continuous probabilistic model, rather than from a discrete mixture model that assumes that one of the reference orientations is indeed correct. Step 2.2 turns out to be essential. Due to the huge range of $\gamma_{ij}^{(k)}$ values, if Step 2.2 is omitted a small subset of data images—those that happen to be closest to reference image angles—will dominate the reconstruction.

We can imagine interesting ways to modify Step 2.2; for example, if we had an independent measurement of the probability $p_i^{(k)}$ that D_i is indeed a good data image, we could normalize the $\gamma_{ij}^{(k)}$ values to sum to $p_i^{(k)}$ rather than to one. Such a variable normalization could be useful, as cryo-EM particle images vary quite substantially in quality (see Section 4), and the quality of an image cannot be easily assessed *a priori*.

2.4   Three Variations

In this section, we describe three variations of the EM algorithm. The first variation, which we call EM-BASIC, is essentially the same as Sigworth's original algorithm using Equations (5) and (6). The original algorithm, however, is very slow (over an hour for a reconstruction from 200 images), so EM-BASIC incorporates some practical tricks that speed up the processing by a factor of 20, yet give reconstructions almost visually

indistinguishable from those of the original algorithm. An obvious speed-up is to round the smallest $\gamma_{ij}^{(k)}$ values to zero, and thus skip Step 2.3 for most i, j pairs. We rounded $\gamma_{ij}^{(k)} < 0.001$ to zero. A further speed-up, found empirically, is to set all except the single largest of the $\gamma_{ij}^{(k)}$ values corresponding to the same rotation angle to zero (before the normalization in Step 2.2), and then to low-pass filter D_i (Gaussian filter with standard deviation $\sigma_F = 1.0$ pixels) before adding it into the reconstruction with orientation $\gamma_{ij}^{(k)}$ to approximate the effect of using multiple translations. This speed-up works because $\gamma_{ij}^{(k)}$ tends to vary in a smooth and predictable way over small translations, but in a much less predictable way over rotations.

Running time is a significant issue in cryo-EM reconstruction, as large-scale reconstructions (say 20,000 images) can take more than a day on a compute cluster. Indeed, large-scale reconstruction almost always includes speed-ups to the standard method, for example, only testing a data image D_i against reference images $R_j^{(k)}$ for orientations that D_i had matched fairly well in previous iterations.

The second variation is called EM-MULTI for multiresolution. EM-MULTI is the same as EM-BASIC, with the addition that it low-pass filters data images heavily in early iterations, and gradually reduces this filtering in later iterations. The standard deviation of the filter is $\sigma_F = \max\{1.0, 12/(k+1)\}$ pixels, where k is the iteration number, running from 1 to 20 in our experiments. Multiresolution optimization is common in pose estimation problems in computer vision [2]. Low-pass filtering can change the SNR of the images, because the signal (the particle) has more low-spatial-frequency content than the noise, so for fair comparison, all our algorithms (including our implementation of the standard method) low-pass filtered images with $\sigma_F = 1.0$, the final value used by EM-MULTI.

Our third variation, EM-ROBUST, addresses the problem reported by Ludtke, that is, the difficulty of robustly estimating $P[D_i \mid R_j^{(k)}]$ for real data images. EM-ROBUST is the same as EM-MULTI, but for each i, EM-ROBUST blindly sets $\gamma_{ij}^{(k)} = \frac{1}{2}$ for the j maximizing $D_i \cdot R_j^{(k)}$, and $\gamma_{ij}^{(k)} = \frac{1}{4}$ for the second best j, and $\gamma_{ij}^{(k)} = \frac{1}{8}$ for the third best, and so forth. Thus EM-ROBUST does not even attempt to map image correlations to probabilities. EM-ROBUST could be used in the case that the matching scores for data and reference images cannot be interpreted as probabilities. For example, EMAN uses a two-stage approach that would be difficult to treat as probabilities: it first matches data images to projections $R_j^{(k)}$ and then reclusters data images, as in the k-means algorithm, by matching them to the average of the data images that matched each $R_j^{(k)}$. In Section 6, we further discuss the problem of setting the $\gamma_{ij}^{(k)}$'s. We are not advocating EM-ROBUST as the solution, but rather we show that even a "deliberately dumb" algorithm such as EM-ROBUST can give reasonable results.

3 Experiments with Synthetic 2D Data

For our 2D experiments we generated synthetic data, grayscale images with values 0–255 and size 73×73 pixels. We used uniformly distributed rotation angles α_i, and normally distributed x- and y-translations x_i and y_i with zero mean and $\sigma = 2.0$ pixels;
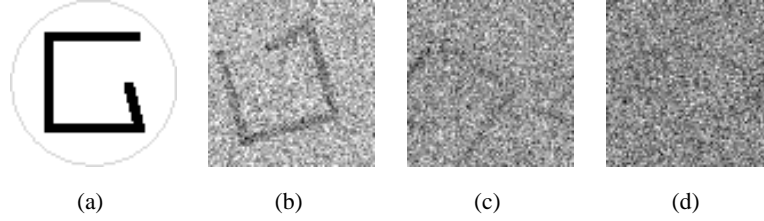
(a) (b) (c) (d)

Fig. 3. Synthetic images: (a) the original particle, (b) with 80% noise, (c) with 90% noise, and (d) with 95% noise, the value used in most of our experiments. The particle shape was chosen to have a likely incorrect local optimum: a square. We used periodic boundary conditions (wrap-around) as seen in (c), to avoid edge effects.

we used bilinear interpolation for subpixel accuracy. The pixel noise, denoted G_i in equation (1), was independent Gaussian noise with zero mean and $\sigma = 36$ gray levels. We imposed periodic boundary conditions (wrap-around), so that a a pixel moved off the right boundary reappeared on the left, and similarly for top and bottom; wrap-around neatly handles the issue of filling in missing pixels [16]. We did not quantize the rotation angle; hence our data images were not exact matches to any of the reference images, giving a more realistic test of the EM algorithm than was done previously [16].

In all experiments, we used references with rotations of $0°, 5°, 10°, \ldots, 355°$. Each reference image is the average image over three rotations in order to represent "patches" of orientations. For example, the reference image for $10°$ is the average of the rotations at $5°$, $15°$, and two copies of $10°$. Rotations every $5°$ may seem rather coarse, but it would take over 1000 reference images to cover the sphere of views of a 3D particle this densely (not counting planar rotations as separate orientations), and hence this level of quantization is realistic for most biological studies.

We used one additional data image, with α_i, x_i and y_i all equal to zero, as the initial guess $\Phi^{(0)}$. After each iteration, we linearly transformed the grayscale of the reconstruction $\Phi^{(k)}$ so that it had mean 128 and standard deviation 36. This normalization helps with some numerical issues and also enhances the contrast in the reconstructions, making them easier to compare by eye.

Figure 3 shows some examples of synthetic data. Our synthetic particles were deliberately chosen to pose difficult reconstruction problems. For example, the particle shown in Figure 3 has no rotational symmetry, yet is close to a particle—a square—with $90°$ rotational symmetry. Incorrect orientations, approximately $90°$ or $180°$ away from the correct rotation, can easily lead the reconstruction algorithm to converge to a square.

Results are given in Figures 4–6. STANDARD is the standard algorithm of Section 2.1, which uses maximum correlation to give a single hard orientation assignment to each data image. We found that all the algorithms including STANDARD perform well with high-SNR images, images that are at most 92% noise.[4] As the noise level

---

[4] For black-and-white particles, 92% noise corresponds to SNR of $\sigma_S^2/\sigma_N^2 = (0.08 \times 128)^2/(0.92 \times 36)^2 \approx 0.095$, where σ_S^2 is signal variance and σ_N^2 is noise variance. Sim-
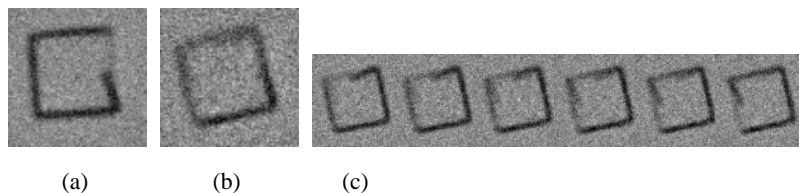
(a)   (b)   (c)

Fig. 4. The success of the STANDARD algorithm depends upon the signal-to-noise ratio. (a) With 93% noise (SNR = 0.072) the algorithm gives a correct reconstruction. (b) With 95% noise (SNR = 0.035) the algorithm falls into an incorrect local optimum. (c) With 94% noise (SNR = 0.051) the algorithm finds a limit cycle with a rotating particle that alternately resembles reconstructions (a) and (b). All reconstructions were made using 200 data images and 20 iterations, but reconstructions changed little after the first 10–12 iterations, except in the case of the limit cycle.
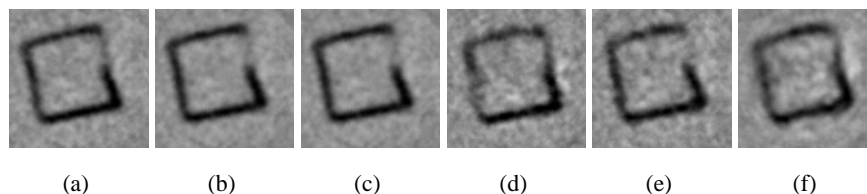

(a)   (b)   (c)   (d)   (e)   (f)

Fig. 5. The max-likelihood approach is more resistant to local optima. (a) EM-BASIC is just starting to fail at 95% noise (SNR = 0.035), but (b) EM-MULTI and (c) EM-ROBUST give correct reconstructions. All these results (from 200 images and 20 iterations) are significantly better than Figure 4(b). (d) EM-BASIC fails at 96% noise (SNR = 0.023), but (e) EM-MULTI is still correct and (f) EM-ROBUST is borderline.

grows larger, STANDARD is always the first algorithm to fail, with the various versions of the EM algorithm giving reasonable results with somewhat higher noise percentages. Typically STANDARD fails with noise percentage at about 94% (SNR = 0.052) and the EM algorithms fail with noise percentage at about 96% (SNR = 0.023). We found that results were surprisingly insensitive to initial guess: reconstructions that failed with a data image as the initial guess always failed with the original black-and-white particle as the initial guess as well. Because of its hard orientation assignments, STANDARD also fails in a qualitatively different way from the EM algorithms: it often falls into a noticeable limit cycle with an unstable reconstruction as shown in Figure 4. The EM algorithms reach more stable reconstructions with only slight variations from iteration to iteration.

The noise percentage failure thresholds for the various EM algorithms are typically close, differing by less than 1%. Generally EM-MULTI is the best algorithm. For example, Figure 5(e) not only looks better than Figure 5(d), but also achieves 1.1% larger value on an objective measure of reconstruction quality, the sum of maximum image

ilarly 95% noise corresponds to SNR = 0.035, and 96% corresponds to SNR = 0.023. Realistic SNR values for cryo-EM are generally in the range from 0.01 to 0.5, depending upon the particle and the defocus.
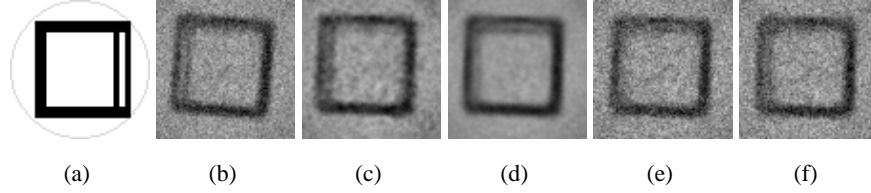
Fig. 6. (a) This particle is an attempt to fool EM-MULTI and EM-ROBUST, which will blur away the doubled edge until late iterations. Yet (b) EM-BASIC, (c) EM-MULTI, and (d) EM-ROBUST achieve fairly similar results at 94% noise. They all start to fail at 95% noise; shown are (e) EM-BASIC and (f) EM-MULTI.
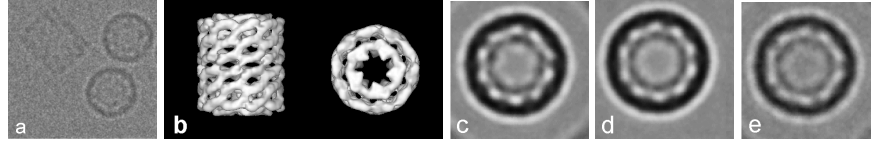

Fig. 7. (a) A piece of a micrograph showing images of hemocyanin particles (at about $1~\mu$m defocus). (b) An isosurface (side and top views) of a 3D reconstruction of hemocyanin made from about 1100 particle images. (c) A 2D reconstruction of the top view made by STANDARD from 250 images. (d) A 2D reconstruction made by EM-MULTI from the same 250 images. (e) A 2D reconstruction made by EM-MULTI from the 25 "best" images.

correlation scores: $S^{(k)} = \sum_i \max_j D_i \cdot R_j^{(k)}$ for $k = 20$. EM-ROBUST falls between the other two EM algorithms, as its $S^{(20)}$ score is 0.6% better than that of EM-BASIC.

4   Real 2D Data

We also applied the algorithms described above to a real 2D data set. The particle is keyhole limpet hemocyanin [11], a didecamer cylindrical "cage" about 400 Å long. This particle shows strongly preferred orientations; for our 2D reconstruction experiment we used 250 automatically selected, 100×100 pixel, "top" views [19] with the same nominal defocus, ostensibly identical up to rotation and translation. The question we asked was whether the 2D reconstruction would find the 5-fold rotational symmetry of the particle, or whether it would fall into a circularly symmetric local optimum. As above, we used reference images every $5°$. For σ^2 in Equation (5) we used the image variance, in effect treating the image as independent pixels of pure noise.

As it turned out, none of the algorithms fell into a circularly symmetric local optimum, even when we used as few as 10 data images. There was a slight difference, however, between STANDARD and the EM algorithms. As seen in Figure 7(c) and (d), STANDARD gives a less symmetric solution, with some sides of the knobby, white, inner ring more clearly dumbbell-shaped than others. (The strongest dumbbell side, bottom left in Figure 7(c), also "rotates" as in Figure 4.) We believe that many of the top views
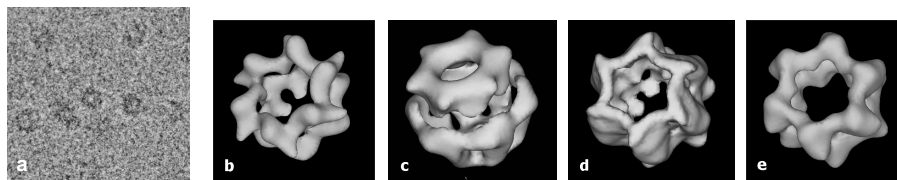
Fig. 8. (a) A piece of a micrograph showing images of p97 AAA ATPase particles (the dark blobs). (b) An isosurface of a correct reconstruction of p97 made from about 4000 particle images. (c) An incorrect reconstruction made from 617 images, using EMAN's default initial guess, an extruded shape. What went wrong is that some of the "top" and "bottom" views (hexagonal rings) of the particle were mistaken for "side" views. (d) A better reconstruction using EMAN with the same initial guess, along with multiresolution refinement and "annealing" of image orientations. (e) A "best-possible" reconstruction from the 617 images using the reconstruction from (b) as the initial guess.

are slightly tipped, and STANDARD, with its single orientation per image, tends to align the images by tip, whereas EM-MULTI spreads the asymmetry around.

A major difference between synthetic and real data is variation in the quality of the particle images. Synthetic images are essentially all of the same quality, differing only in pseudorandom noise and in how well the random rotation fits the references. Real images, on the other hand, vary due to particle conformations, defocus values, ice thickness, sample charging, and electron-beam-induced motion, to name only some of the known and suspected phenomena. We used a delete-half jackknife [4] to estimate the variance of reconstruction quality, as measured by the sum of maximum correlations, $S^{(10)} = \sum_i \max_j D_i \cdot R_j^{(10)}$, over all the images. (For speed we ran only 10 iterations.) We computed 20 reconstructions, each time leaving out a randomly chosen half of the images. For 200 synthetic images (those used in Figure 5), $S^{(10)}$ for the delete-half subsamples ranged from 97.2% to 99.1% of the $S^{(10)}$ achieved by the entire data set. For 250 real images, $S^{(10)}$ for the delete-half subsamples ranged from 95.4% to 101.1%.

We also ran an experiment to see if we could sense the high-quality images. The algorithm used to make Figure 7(e) considers all the images, but at each iteration chooses only the best 10% (by maximum correlations) to go into the reconstruction. The best 10% of the real images give a reconstruction with $S^{(10)}$ that is 97.1% of the $S^{(10)}$ given by all 250 images; whereas, the same procedure with the synthetic images gave $S^{(10)}$ that is 95.5% of the $S^{(10)}$ given by all the images. Interestingly, the "best" 50% of the real images gave $S^{(10)}$ with 99.4% of the $S^{(10)}$ given by all the images, which seems quite good until compared with the 101.1% achieved by the best delete-half subsample. This experiment suggests that maximum correlation gives only a weak assessment of image quality.

5   Real 3D Data

Finally we applied our ideas to a 3D reconstruction problem from a real biological study, shown in Figure 8. The particle is p97 AAA ATPase [13], a 6-fold symmetric

particle about 150 Å across. A large p97 data set converges to the correct optimum (Figure 8(b)), whereas a small data set converges to an incorrect local optimum (Figure 8(c)). The incorrect optimum seems to occur more often with EMAN than with SPIDER, perhaps due to EMAN's reclassification of image orientations, which tends to accelerate convergence.

We modified EMAN in an attempt to correct the problem. Actually implementing some version of the ML algorithm would have required extensive modifications to EMAN, so we tried an approximation instead: we added a random component to EMAN's matching score for each image-orientation pair. Thus suboptimal asssignments occasionally win by chance, as in a Monte Carlo sampling algorithm. If each particle orientation is represented by a number of images in the data set, these randomized assignments approximate the soft assignments of the ML algorithm, but with rather arbitrary "probabilities" as in EM-ROBUST. Specifically, at each iteration k, for $k = 1, 2, \ldots, 20$, for each image i, we added a random component, uniform in the range $[0, (1/4^k) \cdot MaxCorr_i(k)]$ to each correlation score $D_i \cdot R_j^{(k)}$, where $MaxCorr_i(k) = \max_j\{D_i \cdot R_j^{(k)}\}$ is the maximum correlation score for image i at iteration k.

The size of the random component is reduced in later iterations, so that the best assignment becomes relatively better as the reconstruction grows more reliable. This reduction step, akin to simulated annealing [8], enables the sampling algorithm to reach a stable reconstruction. Reduction is unnecessary for stability in the 2D versions of the EM algorithm, because they use continuous weights for the soft assignments.

Our modified version of EMAN also includes the multiresolution idea of EM-MULTI. We blur images very heavily at first and gradually reduce the amount of blurring in later iterations. Both Monte Carlo sampling with annealing and multiresolution seem to be necessary to avoid the local optimum for the 617-image p97 data set, as either one alone gives a reconstruction more closely resembling Figure 8(c) than (d).

We specified C_6 symmetry in EMAN, so the symmetry of the reconstructions in Figure 8 is imposed rather than discovered as in Figure 7. Following EMAN's guidelines, we used 49 reference orientations—not counting planar rotations as separate orientations—for all the reconstructions from 617 images. The 617 images are ones picked automatically by a model-based particle picker we developed [19]. Figure 8(d) shows the results from our modified EMAN.

6 Discussion

In this paper we have confronted the most vexing problem of single-particle reconstruction: convergence to incorrect local optima. In experiments with synthetic 2D data, we found that variations of Sigworth's ML approach can indeed avoid incorrect local optima at lower signal-to-noise ratios than the standard approach. We have left open the question of how to estimate the probabilities $\gamma_{ij}^{(k)}$ in a robust and effective way. Accurate estimates may require good models of both the signal (projections of the target particle) and the noise, including the dependencies between neighboring pixels, which are largely due to the point spread, or contrast transfer function (CTF), of the imaging system. Such detailed models could be obtained, at least in principle; for example, the CTF is often estimated using the power spectrum of an entire micrograph. It would be

more convenient, however, if a very simple method, such as that used in EM-ROBUST, could give reliable results. Here we are imagining an empirically determined schedule of probabilities, rather than the exponentially decreasing sequence used by EM-ROBUST. Such a method would have to be trained on a variety of real data sets, so it is beyond the scope of the present work.

Although the 2D experimentation described in this paper involved very simple signal and noise models, it apparently served as a reasonable guide for algorithm development, as the multiresolution idea, first implemented in EM-MULTI, proved effective on a real 3D data set when incorporated into EMAN. We now believe that the key to solving the problem of incorrect local optima is to slow convergence and allow the initial iterations to explore the space of reconstructions. Besides multiresolution, other techniques from areas such as machine learning and computer vision, could find useful application in cryo-EM.

Acknowledgements

References

1. A. Abbott. The society of proteins. *Nature* 417, 27 June 2002, 894–896.
2. L.G. Brown. A survey of image registration techniques. *ACM Computing Surveys*, 24 (1992), 325–376.
3. A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Royal Statistical Soc. B* 39 (1977), 1–38.
4. B. Efron. *The Jackknife, the Bootstrap and Other Resampling Plans*. CBMS-NSF Regional Conference Series in Applied Mathematics #38, SIAM, 1982.
5. J. Frank. *Three-dimensional electron microscopy of macromolecular assemblies*. Academic Press, 1996.
6. J. Frank, M. Radermacher, P. Penczek, J. Zhu, Y. Li, M. Ladjadj, and A. Leith. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* 116 (1996), 190–199. http://www.wadsworth.org/spider_doc/spider/docs/
7. T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2001.
8. S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi. Optimization by simulated annealing. *Science* 220 (1983), 671–680.
9. A.J. Koster, R. Grimm, D. Typke, R. Hegerl, A. Stoschek, J. Walz, and W. Baumeister. Perspectives of molecular and cellular electron tomography. *J. Struct. Biology* 120 (1997), 276–308.
10. S.J. Ludtke, P.R. Baldwin, and W. Chiu. EMAN: Semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* 128 (1999), 82–97. http://ncmi.bcm.tmc.edu/homes/stevel/EMAN/doc/home.html

11. E.V. Orlova, P. Dube, J.R. Harris, E. Beckman, F. Zemlin, and J. Markl. Structure of keyhole limpet hemocyanin type 1 (KLH1) at 15 Å resolution by electron cryomicroscopy and angular reconstitution. *J. Molecular Biology* 271 (1997) 417–437.

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

13. I. Rouiller, B. DeLaBarre, A.P. May, W.I. Weis, A.T. Brunger, R. Milligan, E.M. Wilson-Kubalek. Conformational changes of the multifunction p97 AAA ATPase during its ATPase cycle. *Nature Struct. Biol.* 9 (2002), 950–957.

14. J. Ruprecht and J. Nield. Determining the structure of biological macromolecules by transmission electron microscopy, single particle analysis and 3D reconstruction. *Progress in Biophysics and Molecular Biology*, 75 (2001), 121–164.

15. W.O. Saxton and J. Frank. Motif detection in quantum noise-limited electron micrographs by cross-correlation. *Ultramicroscopy* 2 (1977), 219–227.

16. F.J. Sigworth. A maximum-likelihood approach to single-particle image refinement. *J. Struct. Biology* 122 (1998), 328–339.

17. H. Stark. 3D Electron Cryomicroscopy. http://www.mpibpc.gwdg.de/abteilungen/103/index.html

18. M. van Heel, *et al.* Single-particle electron microscopy: towards atomic resolution. *Quarterly Reviews of Biophysics* 33 (2000), 307–369.

19. H.C. Wong, J. Chen, F. Mouche, I. Rouiller, and M. Bern. Model-based particle picking for cryo-electron microscopy. *J. Struct. Biology* 145 (2004), 157–167.

20. K. Yonekura, S. Maki-Yonekura, and K. Namba. Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. *Nature* 424 (2003), 643–650.