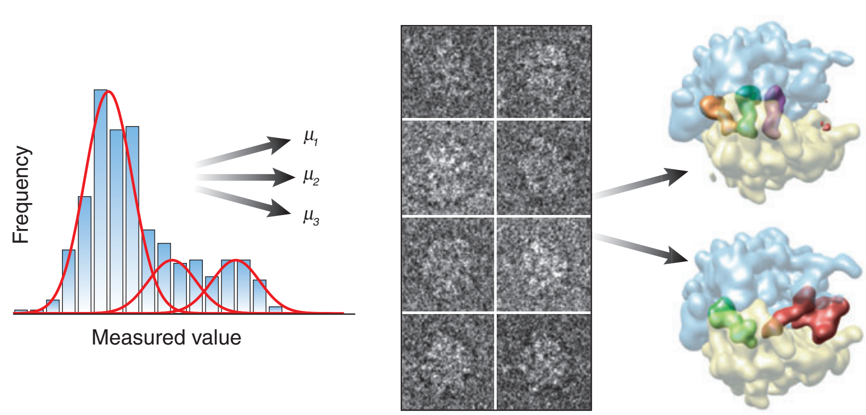
Maximum-likelihood methods

Fred Sigworth, Andrew Barthel, Hemant Tagare
Yale University

Maximum Likelihood solves all problems.



3D reconstruction as an optimization problem

Given a stack of images $\mathbf{X}_i = \{X_1, X_2, \ldots X_N\}$ find the best "model", that is the set of reconstructions and other parameters $\Theta = \{R_1, R_2, \ldots R_M, a_1, a_2 \ldots a_m, \sigma\}$.

What criterion should be used for the "best?"

How about maximizing the probability of the reconstructions given the data,

$$P(\Theta \mid \mathbf{X}).$$

$P(\Theta \mid \mathbf{X})$ is difficult to compute...or define...

However, we can compute $P(\mathbf{X} \mid \Theta)$.

define the likelihood as a function of Θ

$$\mathrm{Lik}(\Theta) = P(X \mid \Theta)$$

MLE and MAP Estimation

The probability of the model is related to the likelihood by Bayes' theorem,

$$p(\Theta \mid \mathbf{x}) = p(\mathbf{x} \mid \Theta) \frac{p(\Theta)}{p(\mathbf{x})}$$

The maximum-likelihood estimate (MLE) optimizes $p(\mathbf{x} \mid \Theta)$.

$$\text{Experiment} \longrightarrow \Theta$$

The maximum a posteriori estimate (MAP) optimizes $p(\mathbf{x} \mid \Theta) p(\Theta)$.

$$\underset{\textit{a priori}}{p(\Theta)} \longrightarrow \text{Experiment} \longrightarrow \underset{\textit{a posteriori}}{\Theta}$$

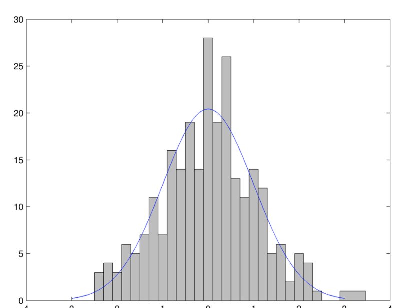
MLE Example 1: Gaussian random numbers

$$\text{Lik} = \varepsilon f(x_1) \cdot \varepsilon f(x_2) \cdot \ldots \cdot \varepsilon f(x_N)$$

where ε is the measurement resolution and the pdf is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2} \right).$$

x
0.285
0.826
-0.008
0.858
0.775
1.306
1.232
0.959
-1.655
-0.990



MLE Example 1: Gaussian random numbers

$$L = N\ln(\varepsilon) - \frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu)^2$$

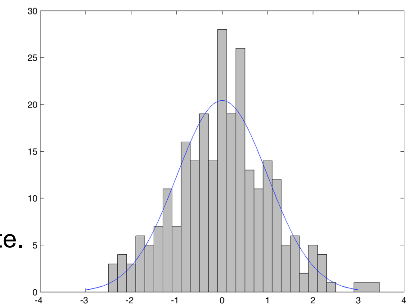
x
0.285
0.826
-0.008
0.858
0.775
1.306
1.232
0.959
-1.655
-0.990

To maximize *L* we set the derivatives to zero,

$$L = N\ln(\varepsilon) - \frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu)^2$$

$$\frac{\partial L}{\partial \mu} = 0 \quad \Rightarrow \quad \hat{\mu} = \frac{1}{\mathrm{N}}\sum_{i=1}^{N}x_i$$

$$\frac{\partial L}{\partial \sigma} = 0 \quad \Rightarrow \quad \sigma^2 = \frac{1}{\mathrm{N}}\sum_{i=1}^{N}(x_i-\mu)^2$$



In this case the MLE is equal to the least-squares estimate.

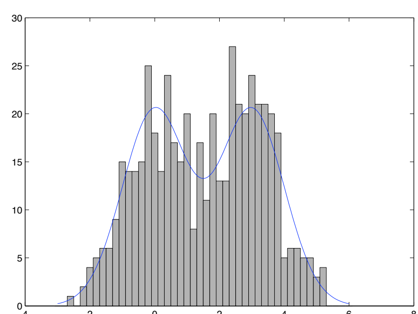
Example 2: Mixture of Gaussians

x
2.9699
3.0242
-0.7529
3.0639
0.4348
1.3791
0.6150
3.1327
2.9938
1.1833

$$f(x) = a_1 f_1(x) + a_2 f_2(x), \qquad f_j(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-(x-\mu_1)^2}{2\sigma^2}\right)$$

$$L = \frac{-N}{2}\ln(2\pi\sigma^2) - \sum_{i=1}^{N}\ln\left[a_1\exp\left(\frac{-(x-\mu_1)^2}{2\sigma^2}\right) + a_2\exp\left(\frac{-(x-\mu_2)^2}{2\sigma^2}\right)\right]$$

Taking the derivatives of *L* is not going to be easy. How to maximize it?



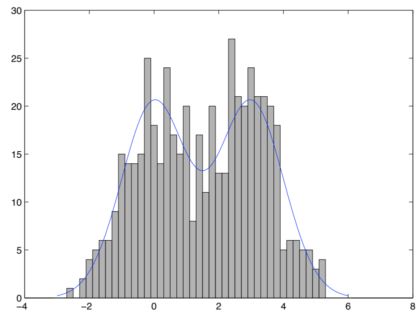
Example 2: Mixture of Gaussians

x	z
0.9863	0
2.9980	1
1.8384	0
0.2488	0
1.6752	0
2.6736	1
2.0572	1
2.6596	1
3.6584	1
2.4223	1

Suppose we had extra information in the form of "switch variables" z_i. Then estimating the two mean values and the weights would be really easy:

$$\hat{\mu}_1 = \frac{\sum_i z_i x_i}{\sum_i z_i} \qquad \hat{\mu}_2 = \frac{\sum_i (1 - z_i) x_i}{\sum_i (1 - z_i)}$$

$$a_1 = \frac{1}{N} \sum_i z_i \qquad a_2 = \frac{1}{N} \sum_i (1 - z_i)$$



Example 2: Mixture of Gaussians

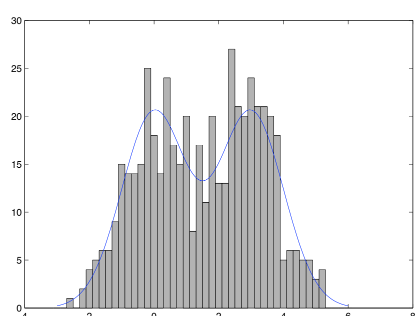
x	z	z'
0.9863	0	0.8236
2.9980	1	0.0111
1.8384	0	0.2660
0.2488	0	0.9771
1.6752	0	0.3716
2.6736	1	0.0287
2.0572	1	0.1582
2.6596	1	0.0299
3.6584	1	0.0015
2.4223	1	0.0591

The EM algorithm

Given estimates of the unknown model variables μ_1, μ_2, a_1 and a_2, compute expectation values of the z's:

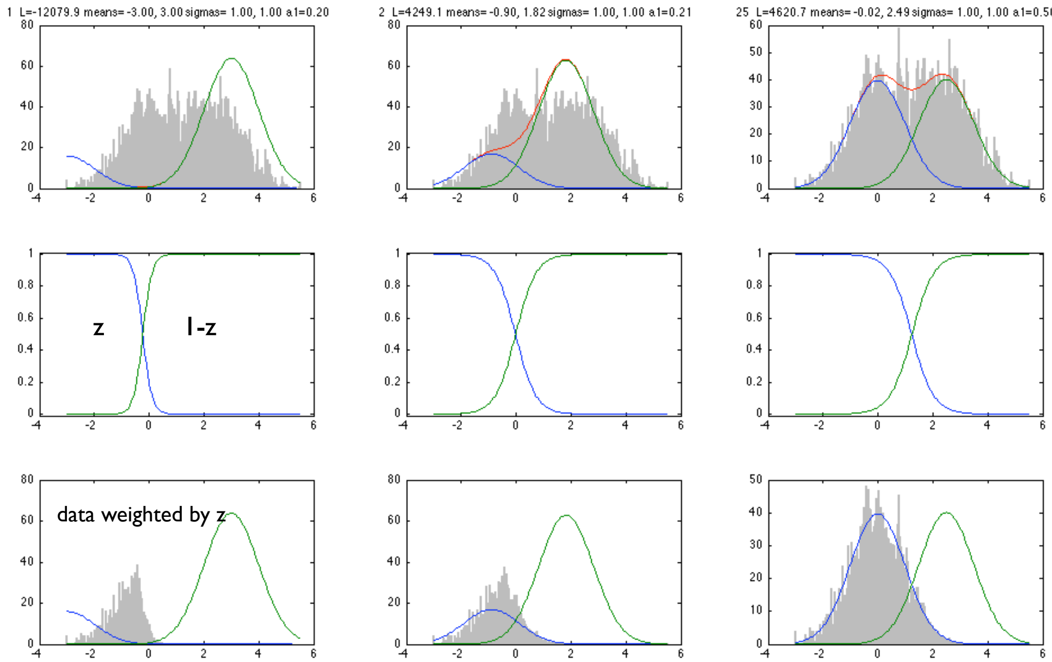
$$\hat{z}_i = \frac{a_1 f_1(x_i)}{a_1 f_1(x_i) + a_2 f_2(x_i)} \qquad \hat{z}_2 = \frac{a_2 f_2(x_i)}{a_1 f_1(x_i) + a_2 f_2(x_i)}$$

Use these instead of the true z's!



Example 2: Mixture of Gaussians

1 iteration | 2 iterations | 25 iterations



Likelihood with hidden variables

The general problem is formulated this way.

$\mathbf{x} = \{x_1, x_2 ... x_N\}$ is the set of <u>observed</u> data;
$\mathbf{z} = \{z_1, z_2, ... z_M\}$ is a set of <u>hidden</u> variables.

Suppose it's easy to compute the probability $p(\mathbf{x},\mathbf{z} \mid \Theta)$ of the complete data set $\{x_1, x_2 ... x_N, z_1, z_2 ... z_M\}$.

By doing an integral over all values of the *z*'s we can get the log likelihood:

$$L = \ln \int p(\mathbf{x},\mathbf{z} \mid \Theta)\, p(\mathbf{z} \mid \mathbf{x},\Theta)\, d\mathbf{z}$$

Our Problem

$\mathbf{x} = \{x_1, x_2 ... x_N\}$ is the set of images;
$\mathbf{z} = \{z_1, z_2, ... z_M\}$ is the set of alignment parameters (Euler angles and translations) for each image.

It is easy to compute the probability $p(\mathbf{x}, \mathbf{z} \mid \Theta)$ of the images, with the alignment parameters known.

In the end we don't care about the alignment parameters, and just integrate them out:

$$L = \ln \int p(\mathbf{x}, \mathbf{z} \mid \Theta)\, p(\mathbf{z} \mid \mathbf{x}, \Theta)\, d\mathbf{z}$$

The EM Algorithm

One way to increase the likelihood is to use the Expectation Maximization algorithm, which has two conceptual steps.

1. Given a previous estimate of the model parameters $\Theta^{(\mathrm{old})}$, compute the expectation

$$Q(\Theta) = \int \ln p(\mathbf{x}, \mathbf{z} \mid \Theta) p(\mathbf{z} \mid \mathbf{x}, \Theta^{(\mathrm{old})}) d\mathbf{z}$$

2. Maximize each parameter of the model by solving

$$\frac{\partial Q(\Theta^{(\mathrm{new})})}{\partial L} = 0$$

$p(\mathbf{x},\mathbf{z} \mid \Theta)$ is easy to compute??

Assuming independent Gaussian noise in each of *P* pixels in an image, the probability for one image is

$$p(x_i,\phi_i,\kappa_i \mid \Theta) = \left(\frac{\varepsilon}{\sqrt{2\pi}\sigma}\right)^P \exp\left(-\frac{\left\|x_i - R_{\varphi_i}(V_\kappa)\right\|^2}{2\sigma^2}\right)$$

where

x_i is the image
φ_i,κ_i are the corresponding hidden parameters (alignment, conformation)
R_φ is the projection operator
V_κ is the k^{th} reconstruction volume (an element of Θ)

The other quantity we need for the EM algorithm is the probability of the hidden variables

$$p(\phi,\kappa \mid x_i,\Theta) = \frac{p(x_i,\phi,\kappa \mid \Theta)}{\sum_\kappa \int p(x_i,\phi,\kappa \mid \Theta)\,d\phi}$$

ML 3D Reconstruction

ML 3D reconstruction reduces to this problem:

Maximize the quantity (I've left out some constants)

$$Q = \sum_{i=1}^{N} \frac{\sum_\kappa \int \left\|x_i - R_{\varphi_i}(V_\kappa)\right\|^2 p(x_i,\phi,\kappa \mid \Theta^{(old)})\,d\phi}{\sum_\kappa \int p(x_i,\phi,\kappa \mid \Theta^{(old)})\,d\phi}$$

Q is maximized with respect to each voxel of each reconstructed volume, plus a few other parameters (e.g. σ). The maximization was done using an algebraic reconstruction technique by Scheres, Carazo et al.

Notice that both the numerator and denominator involve an integral over all five alignment parameters and a sum over the conformations.

Example 3: 1D alignment

Suppose we have many instances x_i of a 1D signal buried in noise, and its relative time of arrival is z_i. Can we reconstruct it?

In this case Θ=y is the reconstructed signal and the Q function is

$$Q(y) = \sum_i \sum_z \left\| x_i - T_z y \right\|^2 p(z \mid x_i, y^{(old)})$$

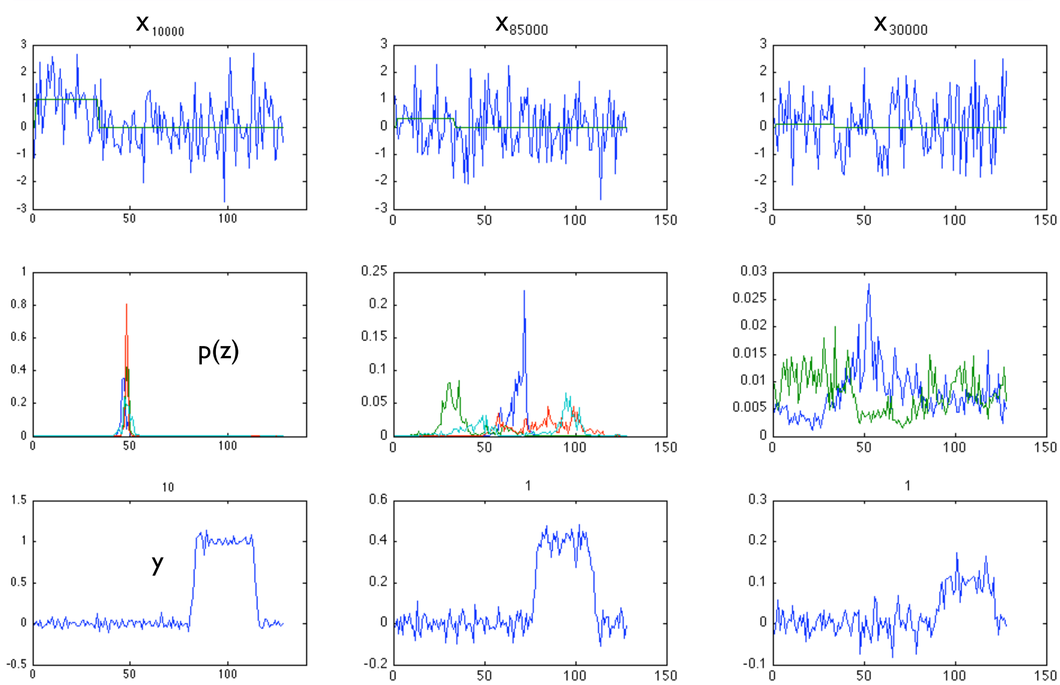
and the EM iteration is

$$y^{(new)} = \sum_i \sum_z T_z x_i p(z \mid x_i, y^{(old)})$$

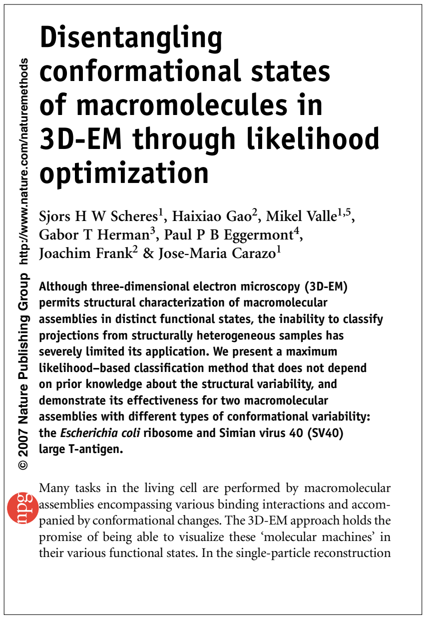
with the “switch variable” probability

$$p(z \mid x, y) = \frac{\exp\left(\dfrac{\left\| T_z x - y \right\|}{2\sigma^2} \right)}{\sum_z \exp\left(\dfrac{\left\| T_z x - y \right\|}{2\sigma^2} \right)}$$

Example 3: 1D alignment



ML 3D Reconstruction

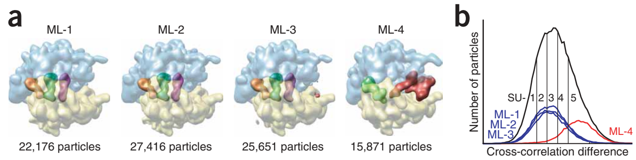


Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization

Sjors H W Scheres[1], Haixiao Gao[2], Mikel Valle[1,5], Gabor T Herman[3], Paul P B Eggermont[4], Joachim Frank[2] & Jose-Maria Carazo[1]

Although three-dimensional electron microscopy (3D-EM) permits structural characterization of macromolecular assemblies in distinct functional states, the inability to classify projections from structurally heterogeneous samples has severely limited its application. We present a maximum likelihood–based classification method that does not depend on prior knowledge about the structural variability, and demonstrate its effectiveness for two macromolecular assemblies with different types of conformational variability: the *Escherichia coli* ribosome and Simian virus 40 (SV40) large T-antigen.

Many tasks in the living cell are performed by macromolecular assemblies encompassing various binding interactions and accompanied by conformational changes. The 3D-EM approach holds the promise of being able to visualize these 'molecular machines' in their various functional states. In the single-particle reconstruction



Because it involves integration over all five degrees of freedom, ML reconstruction is extremely computation-intensive. Involved in the ribosome reconstruction of Scheres et. al. (2007):

90,000 64 x 64 pixel images

10 degree steps

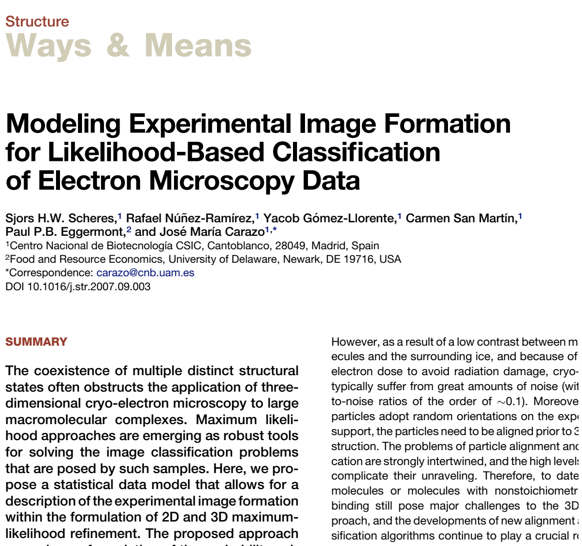
No CTF correction

4 volumes reconstructed

18 CPU-days per iteration (full integration)

6 CPU-months total

ML 3D Reconstruction



Structure

Ways & Means

Modeling Experimental Image Formation for Likelihood-Based Classification of Electron Microscopy Data

Sjors H.W. Scheres,[1] Rafael Núñez-Ramírez,[1] Yacob Gómez-Llorente,[1] Carmen San Martín,[1] Paul P.B. Eggermont,[2] and José María Carazo[1,*]
[1]Centro Nacional de Biotecnología CSIC, Cantoblanco, 28049, Madrid, Spain
[2]Food and Resource Economics, University of Delaware, Newark, DE 19716, USA
*Correspondence: carazo@cnb.uam.es
DOI 10.1016/j.str.2007.09.003

SUMMARY

The coexistence of multiple distinct structural states often obstructs the application of three-dimensional cryo-electron microscopy to large macromolecular complexes. Maximum likelihood approaches are emerging as robust tools for solving the image classification problems that are posed by such samples. Here, we propose a statistical data model that allows for a description of the experimental image formation within the formulation of 2D and 3D maximum-likelihood refinement. The proposed approach

However, as a result of a low contrast between m ecules and the surrounding ice, and because of electron dose to avoid radiation damage, cryo- typically suffer from great amounts of noise (wit to-noise ratios of the order of ~0.1). Moreove particles adopt random orientations on the exp support, the particles need to be aligned prior to 3 struction. The problems of particle alignment an cation are strongly intertwined, and the high level complicate their unraveling. Therefore, to date molecules or molecules with nonstoichiometr binding still pose major challenges to the 3D proach, and the developments of new alignment sification algorithms continue to play a crucial r

A new paper (Oct. 2007) describes ML reconstruction with CTF correction and a model including nonwhite noise. This algorithm runs somewhat more slowly, but performs better.

Why do we expect so much from MLE?

This is easiest to explain in the 2D case. The iteration to improve a 2D reconstruction (like a class average) is computed as

$$A = \sum_{i=1}^{N} \frac{\int T_{-\phi}(x_i)\, p(\phi \mid x_i, \Theta^{(old)})\, d\phi}{\int p(\phi \mid x_i, \Theta^{(old)})\, d\phi}$$

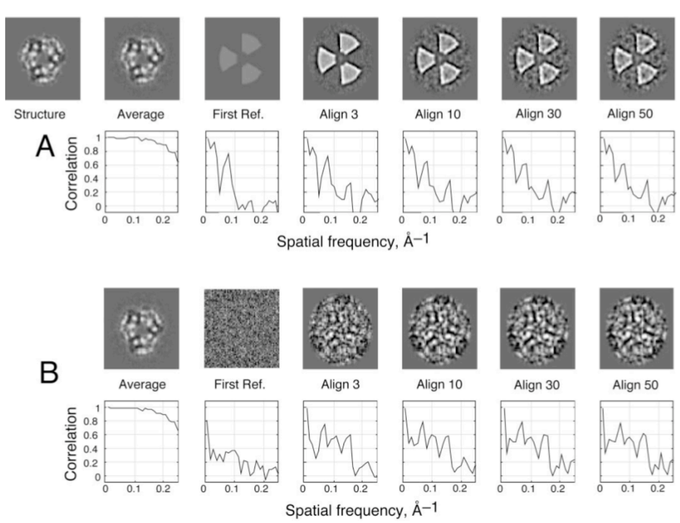
where

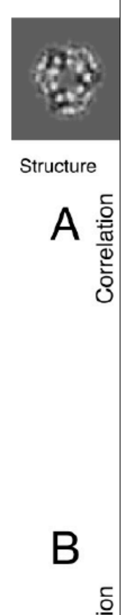
$$\begin{aligned} p(\phi \mid x_i, \Theta) &= \left(\frac{\varepsilon}{\sqrt{2\pi}\sigma}\right)^{P} \exp\left(-\frac{\left\| x_i - T_\phi(A)\right\|^2}{2\sigma^2}\right) \\ &= \text{const } \times \exp\left(\frac{x_i \bullet T_\phi(A)}{\sigma^2}\right) \end{aligned}$$

Thus the average image A is obtained as a weighted average of transformed data images, with the weights being an exponential function of the cross-correlation.

Why do we expect so much from MLE?

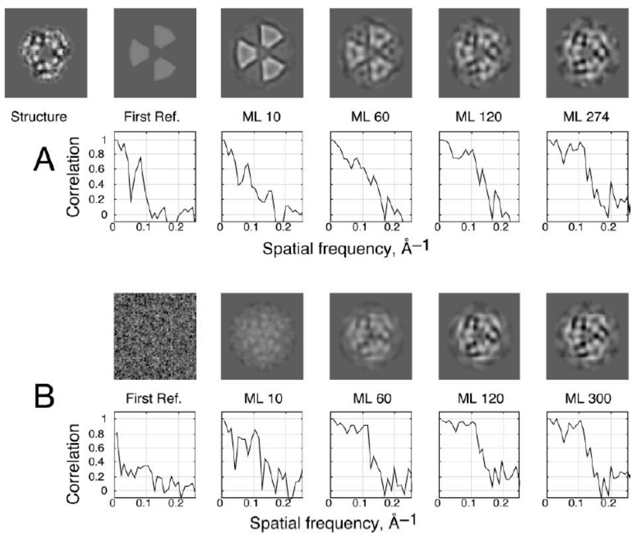
This means that in situations where, at low SNR, conventional alignment shows reference bias...

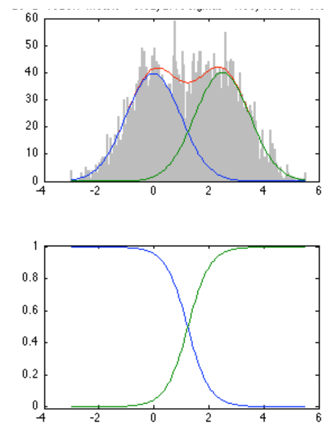




Why do we expect so much from MLE?

...ML estimation shows much less reference bias. This comes from the fact that hard assignments of alignment parameters are not made.





Speeding up the 2D EM algorithm

$$\underline{\text{Finite Mixture}}: \;\; p(I \mid \mu, \alpha, \sigma) = \sum_{j=1}^{M} \alpha_j \int p_g(T_{-t}(I) \mid \mu_j, \sigma) p(\tau) d\tau,$$

where,
I is the image,
M is the number of components,
p_g is a normal distribution with mean u_j, std. σ, $\mu = \{\mu_j\}$.
α_j is the component weight, $\alpha = \{\alpha_j\}$
τ are transformation parameters (1 rot.+2 trans.) distributed as $p(\tau)$.

$$\underline{\text{Max. Likelihood}}: \;\; \mu, \alpha, \sigma = \arg\max \sum_{i=1}^{N} \log p(I_i \mid \mu, \alpha, \sigma).$$

H. Tagare

Observations

<u>EM Loop</u>:

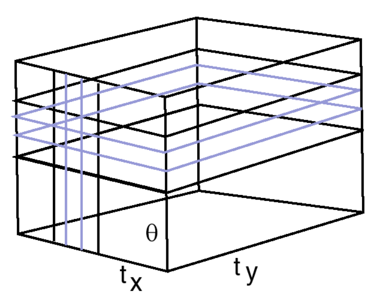
$$p(y_i, \tau \mid I_i, \{\mu, \alpha, \sigma\}^{[n]}) = \frac{\alpha_{y_i}^{[n]} p_g(T_{-\tau}(I_i \mid \mu_{y_i}, \sigma))}{\sum_{j=1}^{M} \alpha_j \int p_g(T_{-\tau}(I_i) \mid \mu_j, \sigma) d\tau},$$

$$\mu_j^{[n+1]} = \frac{\sum_{i=1}^{N} \int T_{-\tau_i}(I_i) p(y_i = j, \tau_i \mid I_i, \{\mu, \alpha\}^{[n]}, \sigma) d\tau_i}{\sum_{i=1}^{N} \int p(y_i = j, \tau_i \mid I_i, \{\mu, \alpha\}^{[n]}, \sigma) d\tau_i}.$$

Note:

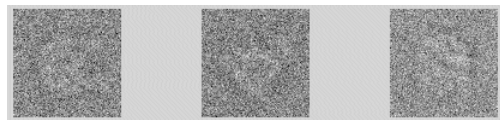
1. All integrals on 3-d space (1 rot. + 2 trans.)
2. Integral in first equation can be evaluated on a coarse grid. Only a few grid points contribute.
3. Integral in the second equation needs a fine grid.

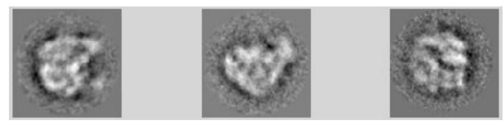
Adaptive Integration



1. Do the first integral over a coarse grid (20 deg., 2pix., 2pix.)
2. Interpolate $\log p$ (the correlation function).
3. Identify the largest grid elements contributing to 99.99% of the integral.
4. Do the last integral on a fine grid over these elements.

First results of adaptive integration





SNR= -17db

Projections from random directions
in 5 deg. cones around x,y,z axis

24 images/mean

10 EM iterations

Speed gain: X 20~30

Fourier shell energy indistinguishable from exact EM.

Extensions are limited only by the imagination

...and computation time!

The likelihood framework allows much flexibility in defining models.

For example, a hidden variable could be the angle of a rigid-body motion. The model would contain parameters say for the mean angle and its S.D. of random variation. The EM algorithm can handle this, and in principle can provide a "sharp" structure, along with a description of the motion.

$\mathbf{x} = \{x_1, x_2 \ldots x_N\}$ is the set of images;
$\mathbf{z} = \{z_1, z_2, \ldots z_M\}$ is the set of alignment parameters (Euler angles and translations) for each image.

Concluding remarks

1. Everything gets harder at low SNR or high ambiguity: EM convergence is slower, and the quality of the noise model becomes more important.

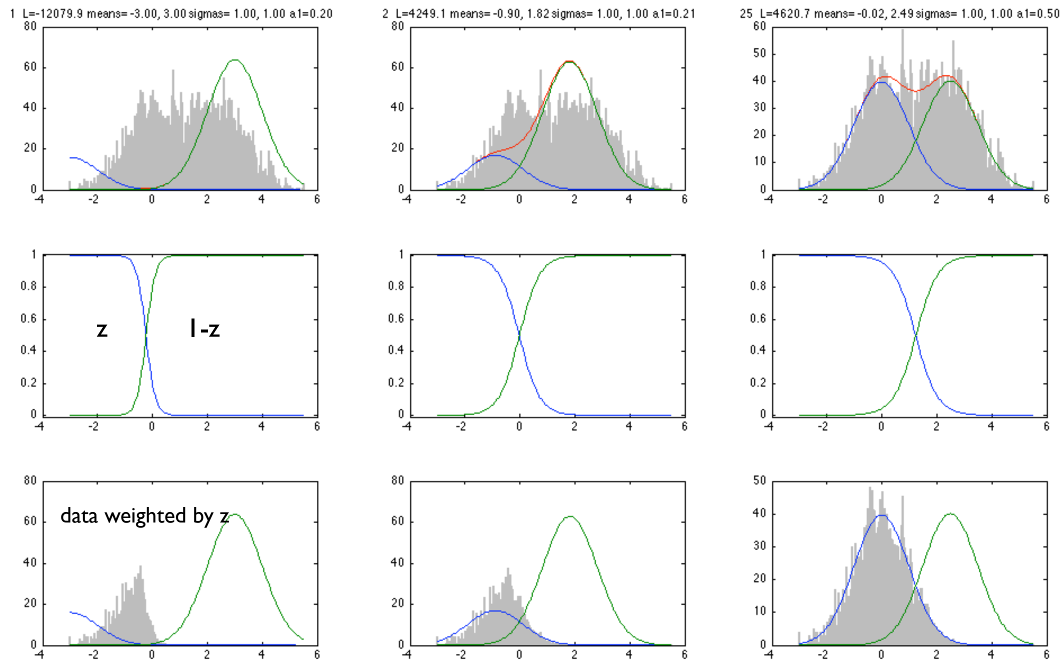
2. The EM algorithm is guaranteed only to find a local maximum.

3. MLE is asymptotically optimal: in the limit of infinite data it should provide correct estimates even with low-quality data. However, the required amount of data grows very rapidly as SNR decreases.

4. MLE does not solve all problems.

Thanks to

Andrew Barthel (Biomedical Engineering)
Hemant Tagare (EE and Diagnostic Radiology)
Yale University



1 L=-12079.9 means= -3.00, 3.00 sigmas= 1.00, 1.00 a1=0.20
2 L=4249.1 means= -0.90, 1.82 sigmas= 1.00, 1.00 a1=0.21
25 L=4620.7 means= -0.02, 2.49 sigmas= 1.00, 1.00 a1=0.50
z
1-z
data weighted by z