Kam's Method for Single Particle Reconstruction

Amit Singer

Princeton University

Department of Mathematics and Program in Applied and Computational Mathematics

Third International Symposium on Cryo-Em 3D Image Analysis

March 22, 2018

National Institute of General Medical Sciences

Basic Discoveries for Better Health

GORDON AND BETTY MOORE FOUNDATION

SIMONS FOUNDATION

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

UNITED STATES AIR FORCE

3-D reconstruction is a statistical estimation problem

- The basic single particle reconstruction (SPR) problem is a statistical estimation problem:

  Estimate the 3-D structure from noisy, non-centered, CTF-effected 2-D tomographic projections taken at unknown viewing angles.

- The parameter of interest is the 3-D structure.

- All other parameters (viewing angles, shifts, etc.) are nuisance parameters.

Statistical Estimation Methods

1. The Method of Moments (Pearson, 1894)
   - Easy to compute
   - Consistent
   - Asymptotically normal ($1/\sqrt{n}$)
   - Not optimal (variance is not smallest among all possible estimators)
2. Maximum Likelihood (Fisher, 1912-1922)
   - Harder to compute
   - Consistent
   - Asymptotically normal ($1/\sqrt{n}$)
   - Optimal or efficient (has smallest variance for large sample size)
3. Bayesian inference
   - Incorporates a prior distribution

Statistical Estimation Methods in SPR

- Maximum likelihood was introduced to SPR by Sigworth in 1998, and together with Bayesian inference is the most popular and successful approach in the field (RELION, cryoSparc, and many more).
- A method of moments for SPR was proposed by Kam in 1980 but is not being used.
- Today:
    - Introduction to the Method of Moments
    - Kam's method
    - Limitations of Kam's original proposal
    - Potential advantages of the method of moments in SPR

The Method of Moments (univariate distributions)

- Suppose that the parameter $\theta = (\theta_1, \ldots, \theta_d)$ has d components.
- n data samples $x_1, \ldots, x_n$ are independently drawn from a probability distribution $F_\theta(x)$.
- The j'th moment of a distribution $F_\theta(x)$ is defined as

$$m_j(\theta) = \mathbb{E}_\theta[X^j] = \int x^j \, dF_\theta(x)$$

- The j'th sample moment is defined as

$$\hat{m}_j = \frac{1}{n} \sum_{i=1}^{n} x_i^j$$

- The Method of Moments Estimator $\hat{\theta}_n$ is the solution to the system of d equations in d unknowns:

$$m_1(\hat{\theta}_n) = \hat{m}_1, \; m_2(\hat{\theta}_n) = \hat{m}_2, \; \cdots, \; m_d(\hat{\theta}_n) = \hat{m}_d$$

The Method of Moments: Toy Example 1

- Estimate the rate parameter λ of a Poisson distribution (count data)

$$\Pr\{X = k\} = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

- The first moment is λ, so the Method of Moments Estimator (MME) is the sample mean

$$\hat{\lambda}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- The likelihood and log-likelihood functions are

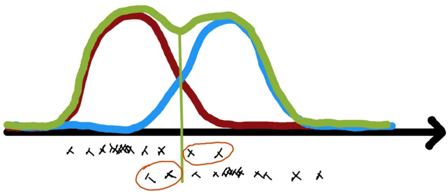
$$\mathcal{L}_n(\lambda) = \prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{x_i}}{x_i!}, \quad \ell_n(\lambda) = \sum_{i=1}^{n} -\lambda + x_i \log\lambda - \log(x_i!)$$

- First order condition for the Maximum Likelihood Estimator (MLE)

$$0 = \frac{d\ell_n}{d\lambda} = \sum_{i=1}^{n} -1 + \frac{x_i}{\lambda}$$

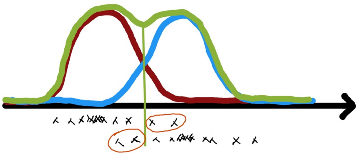
- MME and MLE coincide in this case.

The Method of Moments: Toy Example 2



- Estimate the parameters of a mixture of two Gaussians $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, p)$.
- The five moments m_1, m_2, m_3, m_4, m_5 are polynomials in the parameters. Pearson solved the polynomial system by hand, obtained finitely many solutions, and chose the solution that best fits the sixth moment.
- This was done in the 19th century.
- Just one pass over the data to compute sample moments, can be done on-the-fly, data need not be stored, then solve a small system independent of sample size.

MLE for Mixture of Gaussians



Two versions of maximum likelihood:

- Estimate the mixture parameters and the class labels.
  - A constant positive fraction of the sampled are mislabeled.
  - The MLE is guaranteed to be consistent only when the number of parameters does not grow indefinitely with the sample size (e.g., the Neyman-Scott "paradox", 1948).
- Estimate only the mixture parameters by marginalizing over class labels.
  - The MLE is consistent.
  - The MLE is more accurate than the MME.
  - How to compute the MLE? Expectation-maximization (EM), stochastic gradient descent (SGD).
  - MLE requires multiple passes over the data until convergence (?), slower than MME and must store the entire data.

Lessons for SPR

- The number of nuisance parameters (viewing angles, shifts, etc.) grow indefinitely with the number of images.
- Estimating nuisance parameters is not recommended:
  - MLE (also of the parameter of interest, i.e., 3-D structure) becomes inconsistent.
  - The nuisance parameters are going to be badly estimated.
  - Should avoid validation using estimated nuisance parameters.
- MME can be computed much faster than MLE, in a streaming fashion.
- MME can be used to initialize MLE refinement.

Method of Moments for SPR

- Suppose $I_1, \dots, I_n$ are noisy 2-D projection images of a 3-D structure.
- Naïve sample estimators for the moments are of the form

$$\hat{m}_1(x, y) = \frac{1}{n} \sum_{i=1}^{n} I_i(x, y)$$

$$\hat{m}_2(x_1, y_1; x_2, y_2) = \frac{1}{n} \sum_{i=1}^{n} I_i(x_1, y_1) I_i(x_2, y_2)$$

$$\hat{m}_3(x_1, y_1; x_2, y_2; x_3, y_3) = \frac{1}{n} \sum_{i=1}^{n} I_i(x_1, y_1) I_i(x_2, y_2) I_i(x_3, y_3)$$

- Can we estimate the 3-D structure from the moments?
- How many moments are required?
- How to estimate the 3-D structure from the moments?
- How to deal with the increased dimensionality of the moments? m_1 seems to be 2-D, m_2 is 4-D, m_3 is 6-D.
- How to compute the moments?

Method of Moments for SPR

- The first order moment is a radially symmetric image: either there is no preferred in-plane rotation, or we can enforce uniform distribution of in-plane rotations by angular averaging

$$\hat{m}_1(r) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{2\pi}\int_0^{2\pi} I_i(r,\alpha)\, d\alpha$$

  (CTF can be easily incorporated, not included here for simplicity; also non-perfect centering is ignored)
- The first order moment is just 1-D.
- Similarly, the second order moment is just 3-D (rather than 4-D):

$$\hat{m}_2(r_1, r_2, \psi) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{2\pi}\int_0^{2\pi} I_i(r_1, \alpha_1+\alpha) I_i(r_2, \alpha_2+\alpha)\, d\alpha, \quad \psi = \alpha_2 - \alpha_1.$$

- Comparing the number of parameters to be estimated (3-D structure) and number of moment equations (also 3-D), we might be lucky.
- Computation of the second order moment is essentially PCA of the 2-D images.
- Is there enough information in the mean image and eigenimages and eigenvalues to determine the 3-D structure?

J. theor. Biol. (1980) **82,** 15–39

The Reconstruction of Structure from Electron Micrographs of Randomly Oriented Particles

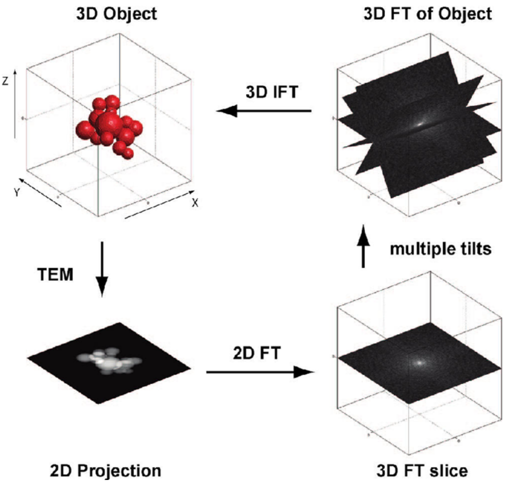
ZVI KAM

Polymer Department The Weizmann Institute of Science Rehovot, Israel

(*Received* 19 *August* 1977, *and in revised form* 20 *April* 1979)

A new method for enhancing and reconstructing the three dimensional structure of randomly oriented particles from their electron micrographs is developed. The method requires as an input many pictures of randomly oriented identical particles. The analysis is based on the calculation and accumulation of the spatial correlation of the densities on the electron micrographs, from which the spherical harmonic coefficients of the structure can be found. The process of enhancement of the spatial correlation and the averaging out of background noise enables reconstructions by use of pictures with low signal-to-noise ratio. The theory is presented and implemented in a computer program package. Simulated electron micrographs of ellipses, rods and a model of hexameric glutamate dehydrogenase are analyzed to demonstrate reconstructions using the computer programs.

The Fourier Projection-Slice Theorem



Kam's theory: uniform viewing angle distribution

- The second order moment of the Fourier transformed images is the autocorrelation function of the Fourier transformed 3-D structure $\hat{\phi}$ with itself over the rotation group SO(3).
- A function cannot be uniquely determined from its autocorrelation function.
- The autocorrelation function determines the magnitudes of the Fourier transform but not its phases.
- The Fourier transform over the rotation group SO(3) calls for expansion of $\hat{\phi}$ in spherical harmonics:

$$\hat{\phi}(k,\theta,\varphi) = \sum_{l=0}^{L} \sum_{m=-l}^{l} A_{lm}(k) Y_l^m(\theta,\varphi)$$

- The second moment determines for each l the matrix C_l given by

$$C_l(k_1,k_2) = \sum_{m=-l}^{l} A_{lm}(k_1) A_{lm}^*(k_2), \quad \text{or} \quad C_l = A_l A_l^*$$

- From C_l we can get the A_{lm}'s only up to an orthogonal matrix of size $(2l+1) \times (2l+1)$.
- The missing orthogonal matrix is the analog of the missing phase.

Estimating the second order moment

- Considering the important role of the second order moment we developed an accurate and efficient procedure for its estimation in SPR:

  Zhao, Shkolnisky, Singer, IEEE Trans. Computational Imaging, 2016.

  Bhamre, Zhang, Singer, Journal of Structural Biology, 2016.

- Improvements over standard Principal Component Analysis (PCA) or Multivariate Statistical Analysis (MSA):

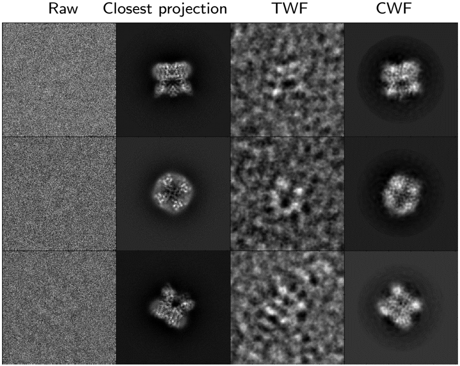
- Accuracy:

  - Eigenvalue shrinkage turned out to be extremely important (using high dimensional statistics and random matrix theory)
  - Full CTF correction, not just phase flipping.

- Computational complexity: For images of size $L \times L$, standard PCA takes $O(nL^4 + L^6)$ and $O(L^4)$ storage, whereas our steerable PCA procedure takes $O(nL^3 + L^4)$ and $O(L^3)$ storage.

Improved Image Restoration without Averaging

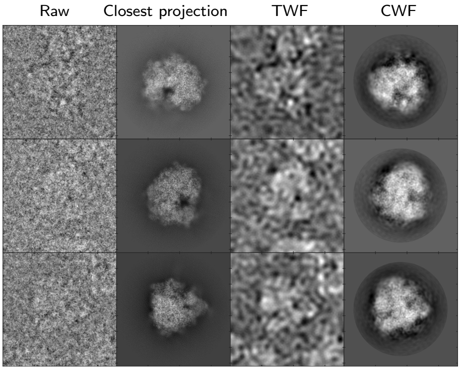
Bhamre, Zhang, Singer (*Journal Structural Biology*, 2016)



- CWF = Covariance Wiener Filter, TWF = Traditional Wiener Filter
- TRPV1, K2 direct electron detector
- 35645 motion corrected, picked particle images of 256×256 pixels belonging to 935 defocus groups (Liao et al., Nature 2013)

Improved Image Restoration without Averaging

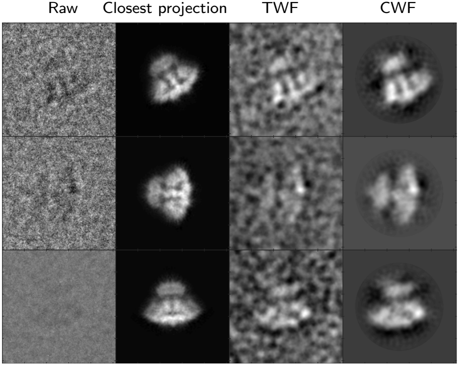
Bhamre, Zhang, Singer (*Journal Structural Biology*, 2016)



- 80S Ribosome, FALCON II 4k$\times$4k direct electron detector
- 105247 motion corrected, picked particle images of 360$\times$360 pixels, 290 defocus groups (Wong et al., eLife 2014)

Improved Image Restoration without Averaging

Bhamre, Zhang, S (*Journal Structural Biology*, 2016)



- IP3R1, Gatan 4k×4k CCD
- 37382 picked particle images of 256×256 pixels, 851 defocus groups (Ludtke et al., Structure 2011)

Kam's theory: uniform viewing angle distribution

- 2nd order moment is insufficient to determine the structure.
- Kam proposed using a slice of higher order moments and empirically observed uniqueness.
- 3rd order moment guarantees uniqueness (or more precisely, a finite number of possible 3-D structures):

  Bandeira, Blum-Smith, Perry, Weed, Wein, arXiv preprint, 2017

- 3rd order moment is more difficult to estimate than 2nd order moment:

  – Requires more images (because noise variance is amplified cubically rather than quadratic)

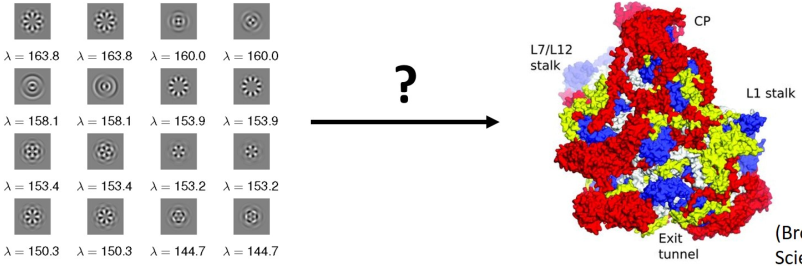
  – Has higher dimensionality (5-D instead of 3-D)

  – No existing eigenvalue shrinkage procedures for tensors

- Can we avoid 3rd order moments?

Can we avoid 3rd order moment? Non-uniform viewing angles

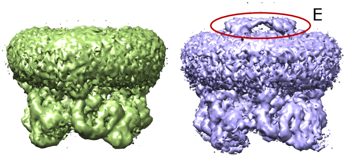
Conjecture: 2nd order moment is sufficient to guarantee uniqueness (up to finitely many possibilities) for non-uniform distribution of viewing angles (work in progress).



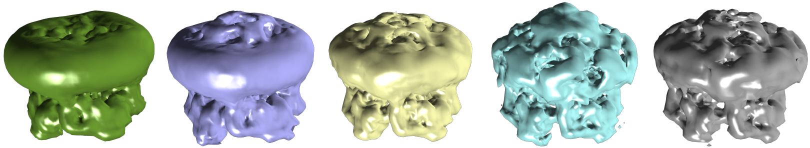
(Brown et al., Science 2014)

Can we avoid 3rd order moment? Homology modeling

Estimate orthogonal matrices from another similar previously solved structure
Bhamre, Zhang, Singer, IEEE Symp. Biomedical Imaging 2015
Bhamre, Zhang, Singer, arXiv, 2017



EMDB 8118 (a) EMDB 8117 (b)



Homologous structure (a) Ground Truth (b) Least Squares (c) Twicing (d) Anisotropic Twicing (e)

Synthetic Dataset: TRPV1 with DxTx and RTX, SNR$= 1/40$, 26000 images, 10 defocus groups.

Can we avoid 3rd order moment? Add a couple of "clean" projections

- 2nd order moment + 2 "clean" projections determine the 3-D structure uniquely

  Levin, Bendory, Boumal, Kileel, Singer, IEEE Symp. Biomedical Imaging 2018

- High quality projections could be obtained if there is a preferred orientation under some sample preparation conditions, or from dominant class averages.

Why was Kam's method mostly ignored?

- Idea that was ahead of its time: There was not enough data to accurately calculate second and third order statistics.

- Requires uniform distribution of viewing directions.

- Maximum likelihood framework prevailed.

Summary

- The Method of Moments cannot compete with Maximum Likelihood in terms of resolution, but can be much faster, and can be used to initialize refinement procedures.
- Extending Kam's method to non-uniform distributions of viewing directions may yield an ultra-fast *ab-initio* modeling technique.
- Just one pass over the data for PCA type computation, completely sidesteps rotation estimation, no need to worry about initial model and convergence.
- From the theoretical standpoint: SPR is possible at any SNR, as long as sufficiently many particles can be picked.
- At low SNR, rotations (nuisance parameters) cannot be estimated accurately. Rotation-based validation methods are not very informative at low SNR.

Princeton–Nature Conference:

Frontiers in Electron Microscopy for the Physical and Life sciences

Princeton, NJ, July 11-13, 2018

nature research

https://www.nature.com/natureconferences/fempl2018/index.html

The ASPIRE team at Princeton University

ASPIRE (Algorithms for Single Particle Reconstruction):

- Tamir Bendory
- Nicolas Boumal
- Ayelet Heimowitz
- Joe Kileel
- Yuehaw Khoo (Stanford)
- Roy Lederman
- Will Leeb
- Eitan Levin
- Nir Sharon

http://spr.math.princeton.edu/