Class Averaging in Cryo-EM Single Particle Reconstruction

Zhizhen Zhao

Department of Physics,
Princeton University

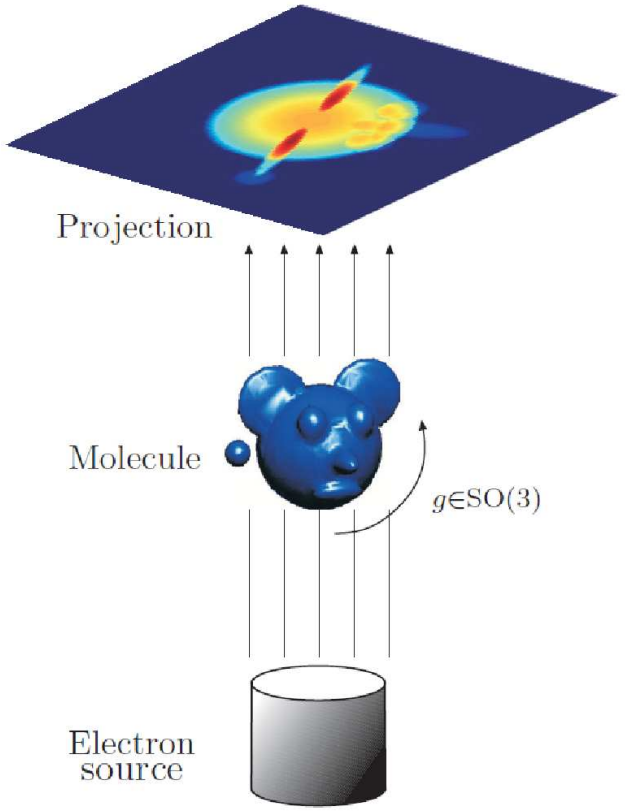
18 July 2013, Final Public Oral Examination

Outline

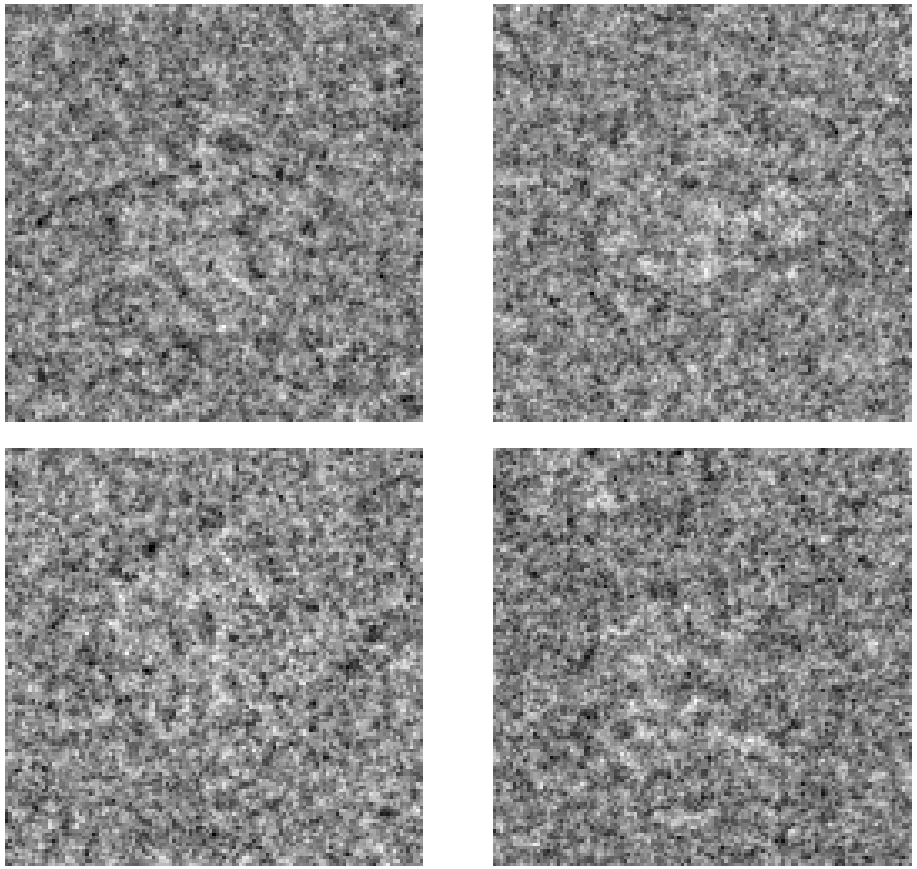
Cryo-EM

- Tens of thousands macromolecules are embedded in thin vitreous ice layer.
- Each particle resumes random unknown orientation.
- The projection images of macromolecules are of very low signal to noise ratio.

Goal: Find 3D electron density maps.



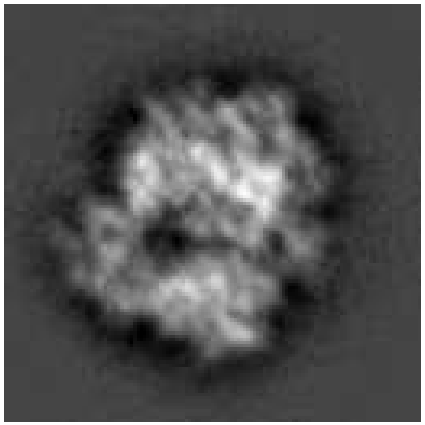
Cryo-EM Projection Images



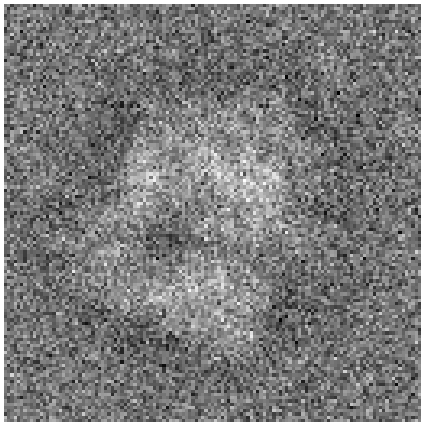
Algorithmic Procedure for Single Particle Reconstruction

- Particle selection
- Class averaging
- Orientation estimation
- 3D tomographic inversion
- Iterative refinement

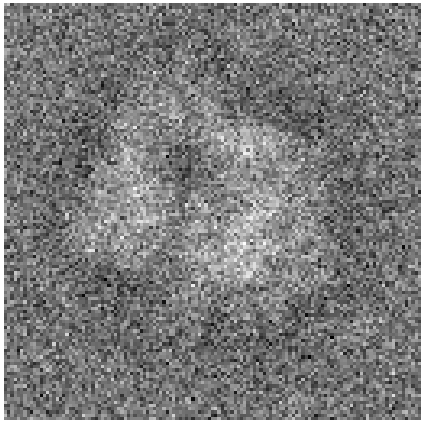
Class Averaging



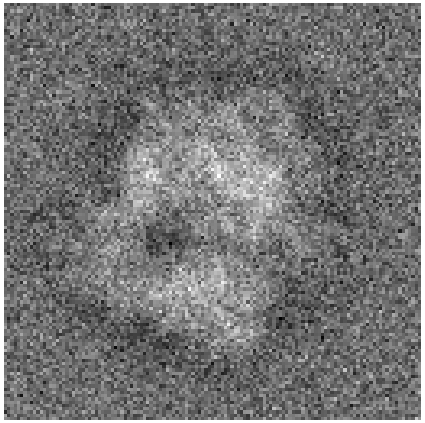
Clean



I_i



I_j



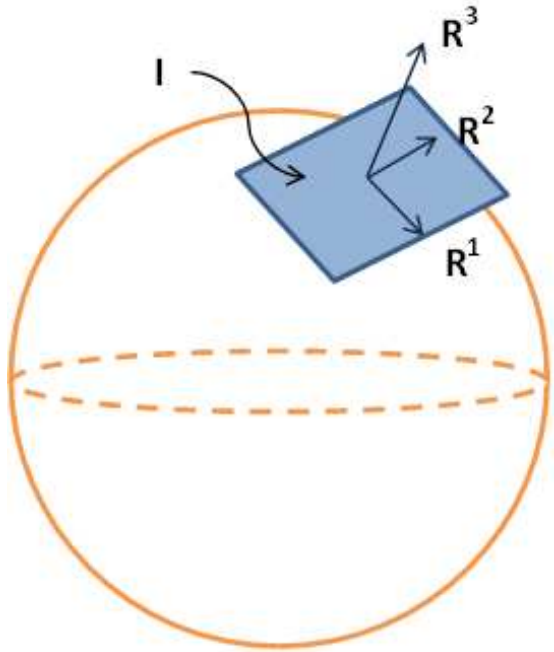
Average

- To each projection image I there corresponds a 3×3 unknown rotation matrix R describing its orientation.

$$R = \begin{pmatrix} | & | & | \\ R^1 & R^2 & R^3 \\ | & | & | \end{pmatrix}$$

satisfying
$RR^T = R^T R = \mathbb{I}, \det R = 1.$

- The projection image lies on a tangent plane to the two dimensional unit sphere S^2 at the viewing angle $v = v(R) = R^3$.



MSA/MRA Classification

(Dube et al. 1993)

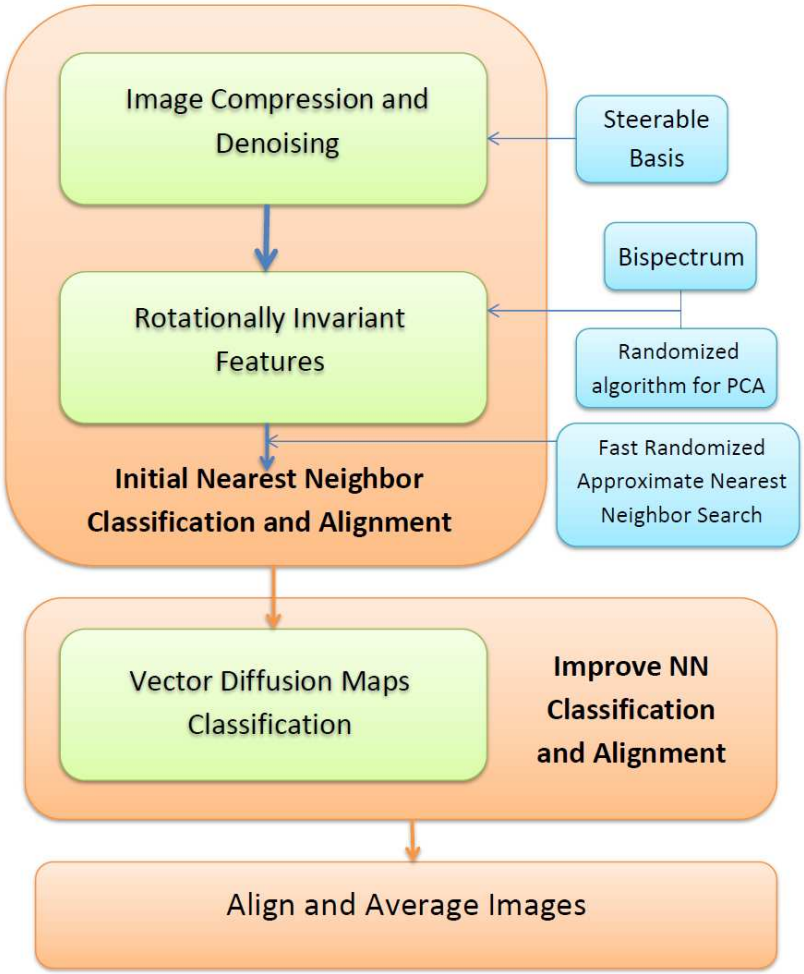
- Use PCA to compress and denoise images (termed MSA in cryo-EM).
- Use hierachical ascending classification to classify the data into N groups and generate N references.
- Use multi-reference alignment (MRA) to align and classify the data into the different references.
- Iterate MSA compression, HAC and MRA until convergence.

Iterative Reference-free Alignment and K-means Clustering

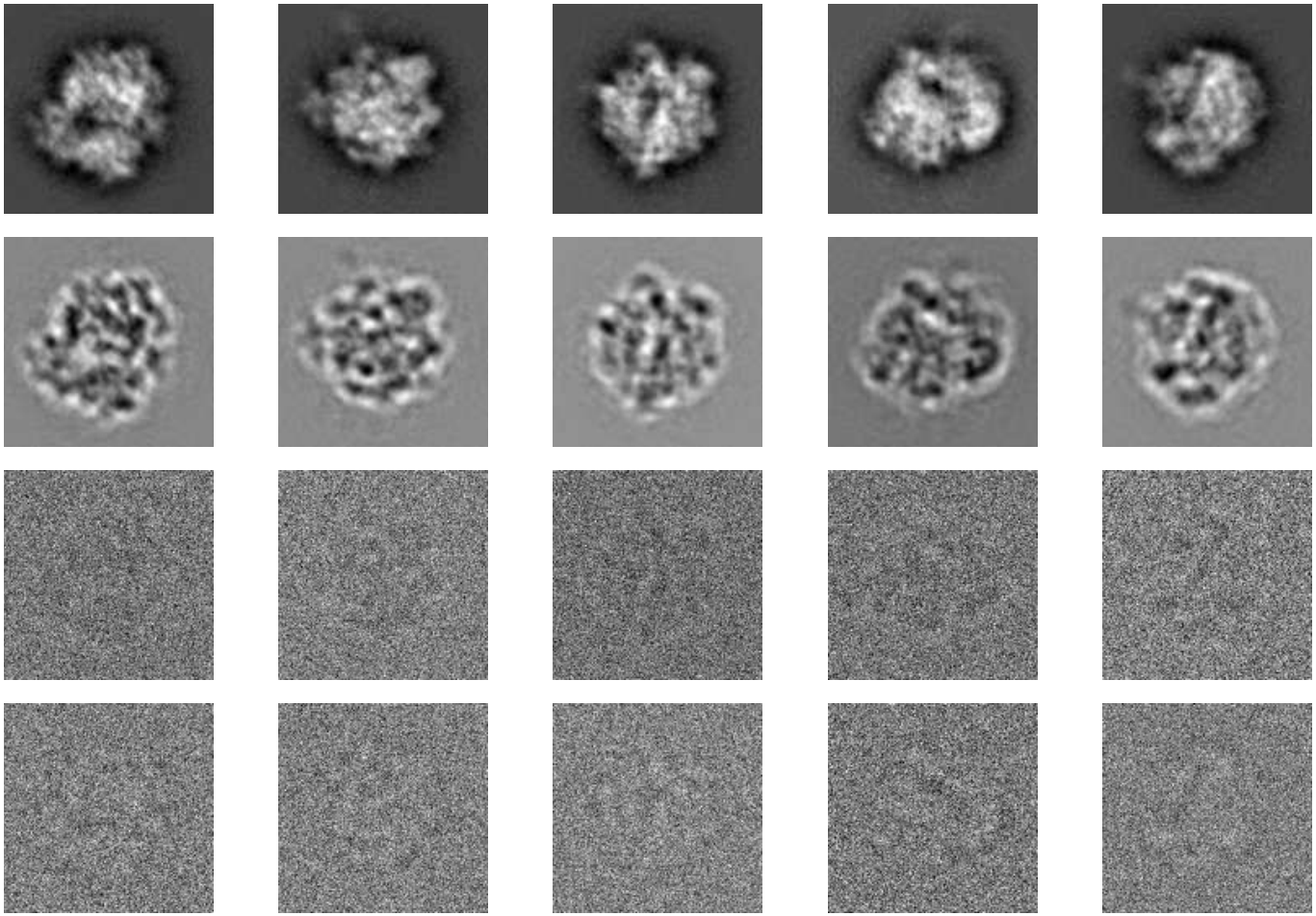
(Penczek et al. 1992, 1996)

- Global alignment is achieved by sequentially aligning images to a changing global average.
- Apply K-means clustering on the aligned images to classify the data set into K groups. The class averages are the means of each group.

Our Class Averaging Procedure

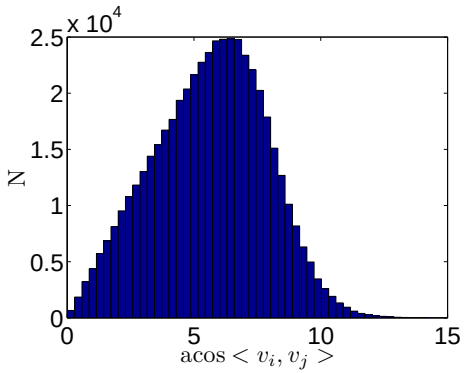


Simulated Projection Images



Comparison with Existing Methods I

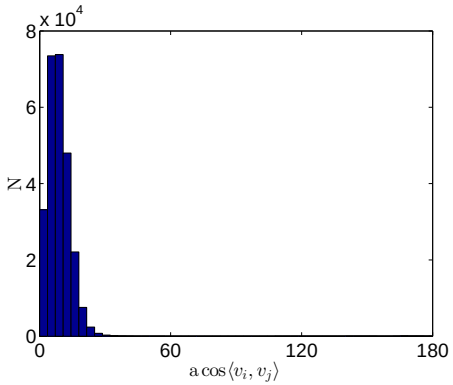
10^4 clean centered simulated projection images of 70S ribosome. $k = 50$.



This work

$p_{20^\circ} = 1$

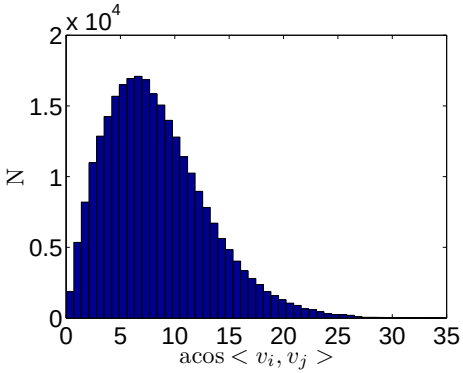
run time: 10min



MSA/MRA

$p_{20^\circ} = 0.975$

1 hr



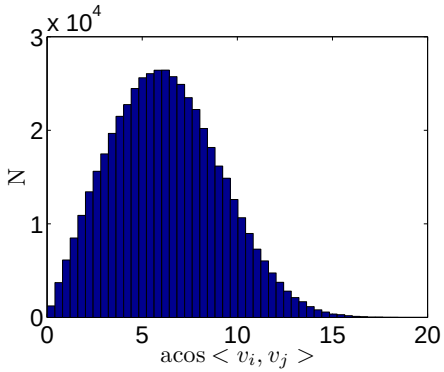
RFA

$p_{20^\circ} = 0.998$

1 hr

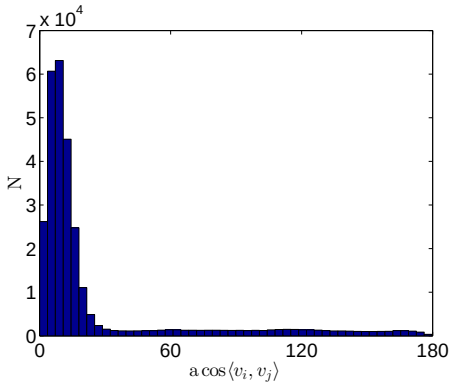
Comparison with Existing Methods II

10^4 noisy (SNR $= 1/50$) centered simulated projection images of 70S ribosome. $k = 50$.



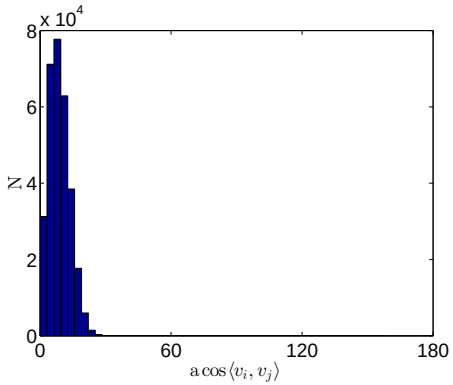
This work

$p_{20^\circ} = 1$



MSA/MRA

$p_{20^\circ} = 0.78$

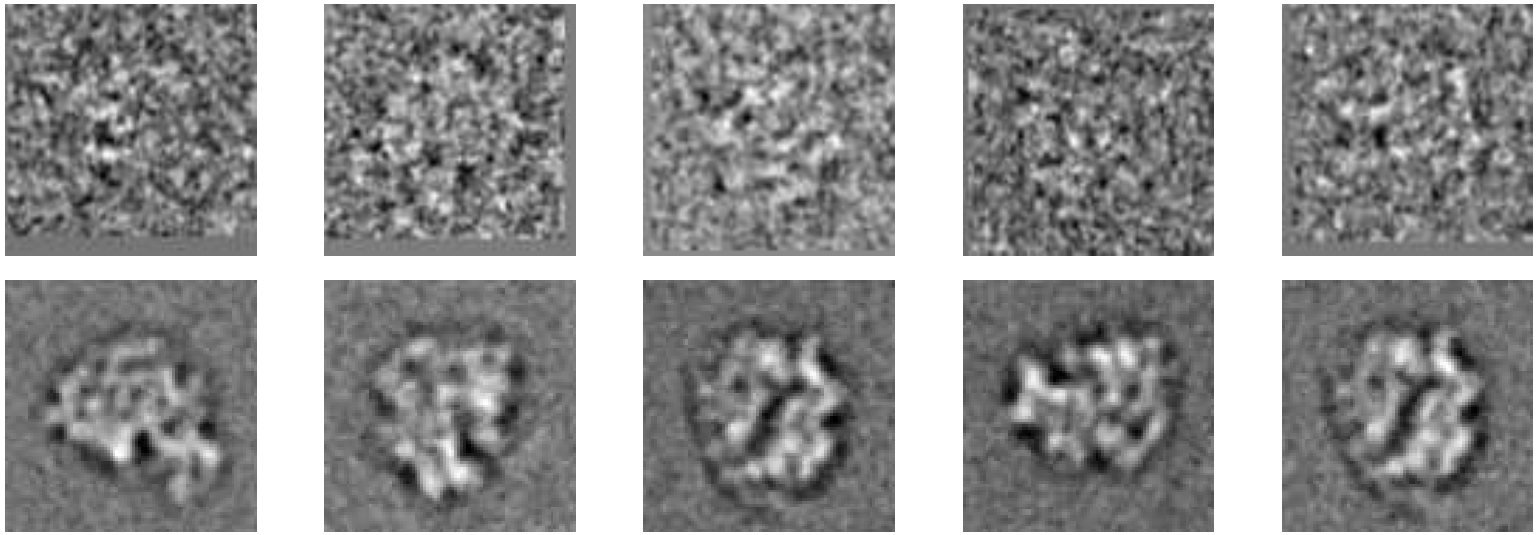


RFA

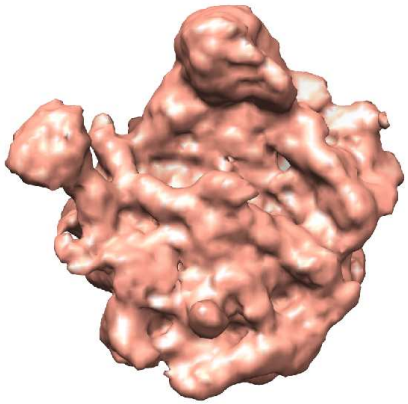
$p_{20^\circ} = 0.997$

Ab initio Reconstruction of Experimental Data Sets I

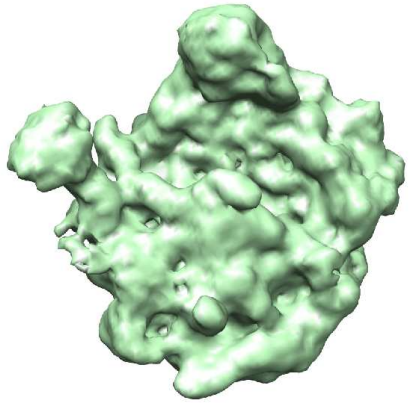
$27,121$ images of 50S ribosomal subunit (Courtesy of Dr Fred Sigworth)
Image size: 90×90 pixels. Pixel size: 3.36 Å.



The data was split into two groups to estimate class averages (k=50), and used common-lines based algorithm (Wang, et al. 2013) for orientation estimation.

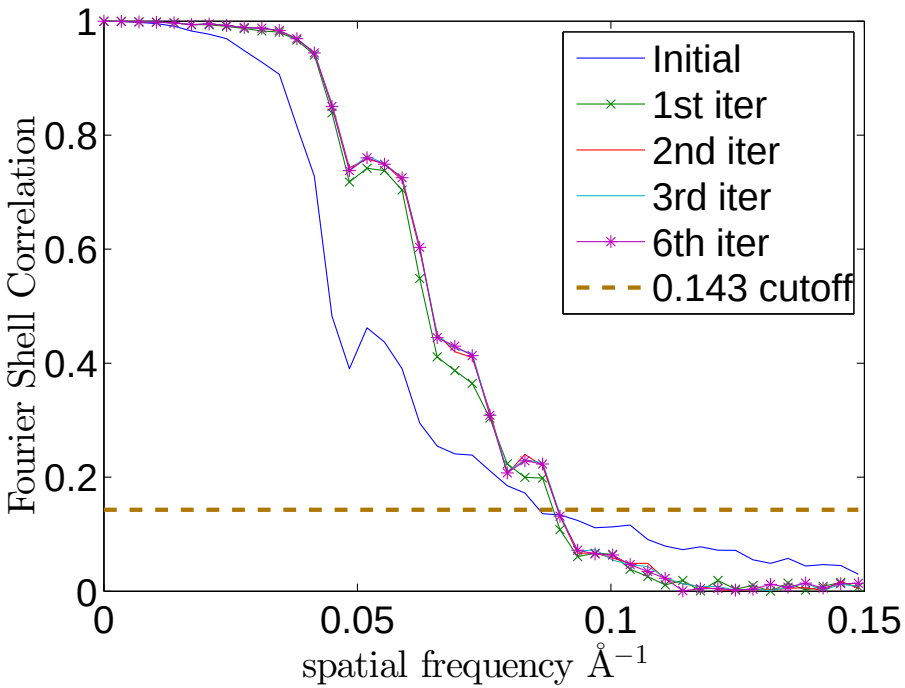


a) 50S Recon. 1



b) 50S Recon. 2

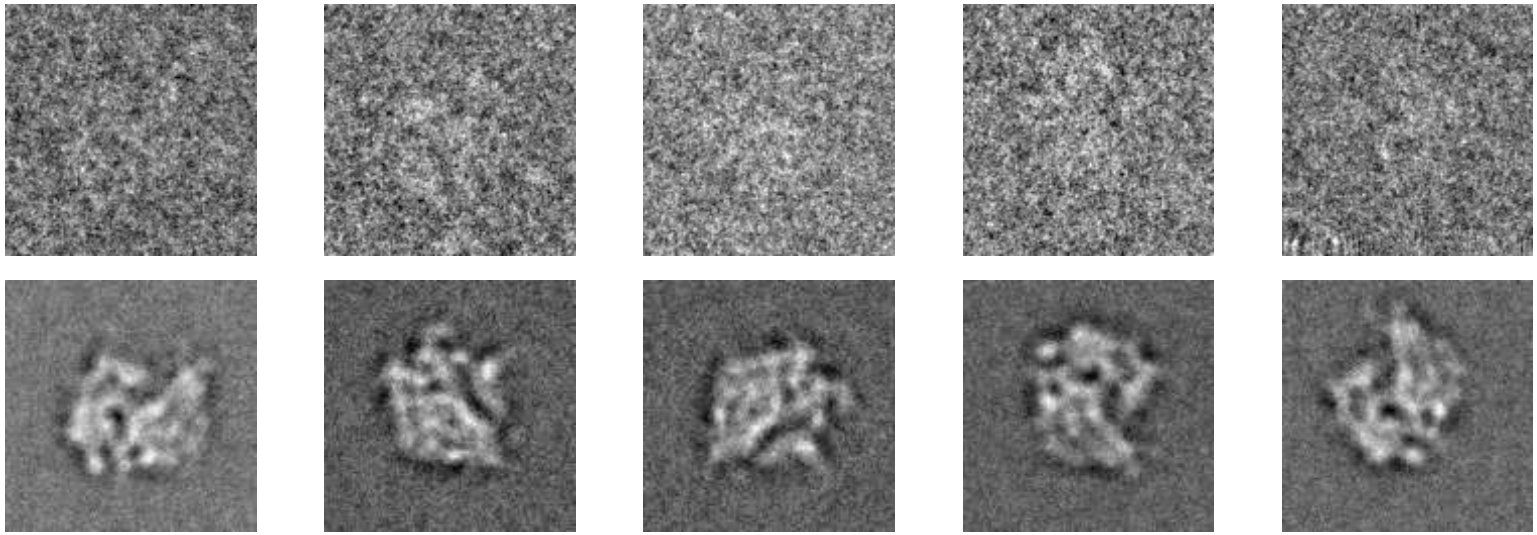
Movie of 50S ribosomal subunit



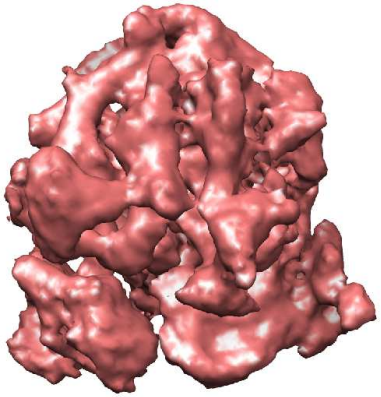
Generating class averages takes 3.5 hours. 200 class averages were selected for reconstructions. Ab initio reconstruction from the class averages takes about 20 minutes. Each refinement iteration takes 4 hours.

Ab initio Reconstruction of Experimental Data Sets II

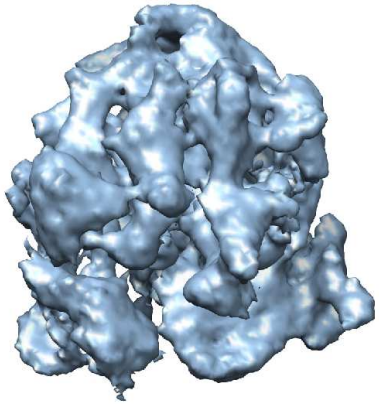
40,778 images of 70S ribosome (Courtesy of Dr Joachim Frank)
Image size: 125×125 pixels. Pixel size: 3 Å



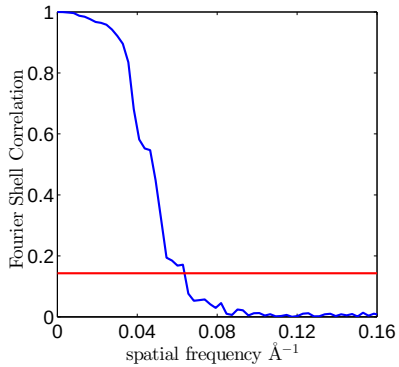
The data was split into two groups to estimate class averages (k=20). Used common-lines based algorithm for orientation estimation.



a) 70S Recon. 1



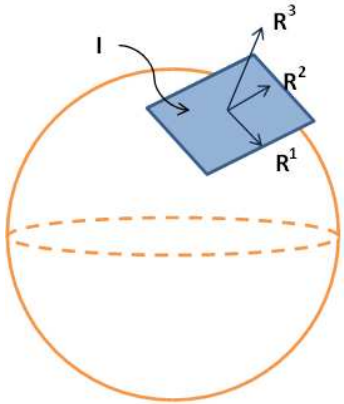
b) 70S Recon. 2

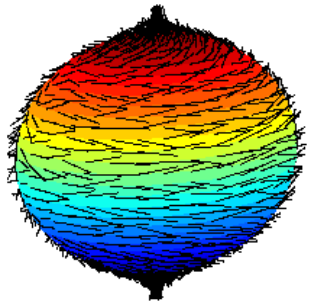


c) FSC

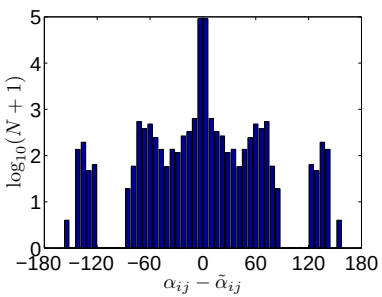
Movie of 70S ribosome

Hairy Ball Theorem and Global Alignment

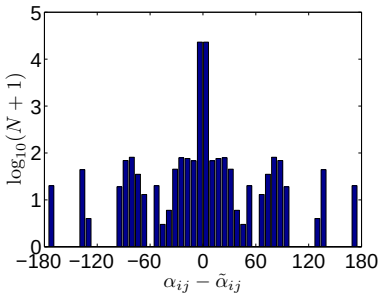




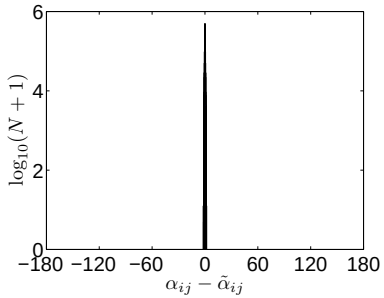
Errors in alignment for similar images after reference-free alignment:



whole sphere



60 °



20 °

Outline

Steerable PCA

(Zhao and Singer 2013)

- PCA is a classical method for dimensionality reduction, compression and denoising.
- Inclusion of the rotated and reflected images for PCA in some cases is advantageous.
- When all rotated images are included, the eigenimages have a special separation of variables form in polar coordinates in terms of radial functions and angular Fourier modes (Perona 1995, Hilai and Rubinstein 1994, Uenohara 1998).
- Efficient algorithms: Jogan et al. 2003, Ponce and Singer 2011.

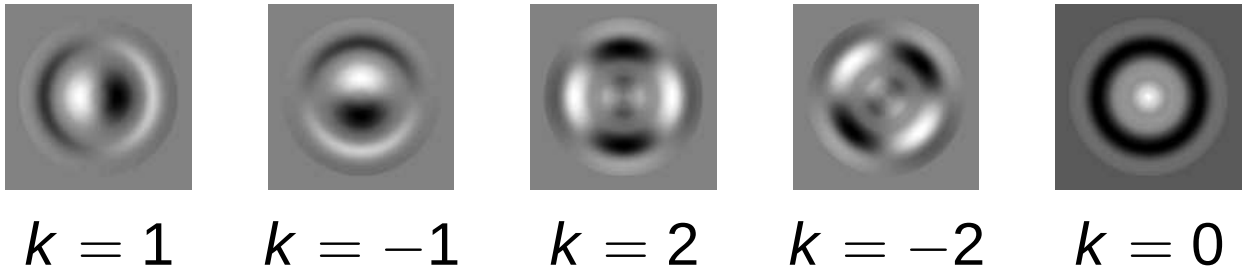
Challenges

- Challenge I: Mitigate the increased computational cost associated with replicating each image multiple times.
- Challenge II: Non-unitary transformation from Cartesian to polar grid changes the noise statistics.
- Challenge III: If we build the sample covariance matrix by simply enlarging our dataset by the number of rotations and reflections, the realizations of noise are no longer independent.

Outline of Solutions

How to compute steerable PCA efficiently and accurately?

- Fourier-Bessel transform.
- Sampling criterion.
- Block diagonal structure of the sample covariance matrix.


$k = 1$
$k = -1$
$k = 2$
$k = -2$
$k = 0$
$k = 5$
$k = -5$
$k = 3$
$k = -3$
$k = 1$
$k = -1$
$k = 3$
$k = -3$
$k = 2$
$k = -2$

Fourier-Bessel Basis

- We assume that the set of images correspond to spatially limited objects. By appropriate scaling of the pixel size, we can assume that the images vanish outside a disk of radius 1.
- The Fourier-Bessel functions are given by

$$\psi^{kq}(r,\theta) = \begin{cases} N_{kq} J_k(R_{kq} r) e^{\imath k\theta}, & r \leq 1 \\ 0, & r > 1. \end{cases}$$

- J_k is the Bessel function of integer order k and R_{kq} is the q^{th} root of the equation $J_k(R_{kq}) = 0$.

Sampling Criterion

- The 2D Fourier transform of the Fourier-Bessel functions, denoted $\mathcal{F}(\psi^{kq})$, is given in polar coordinates as

$$\mathcal{F}(\psi^{kq})(k_0, \phi_0) = 2\sqrt{\pi}(-1)^q(-\imath)^k R_{kq} \frac{J_k(2\pi k_0)}{(2\pi k_0)^2 - R_{kq}^2} e^{\imath k \phi_0}.$$

- The Fourier transform $\mathcal{F}(\psi^{kq})(k_0, \phi_0)$ vanishes on concentric rings of radii $k_0 = \frac{R_{kq'}}{2\pi}$ with $q' \neq q$. The maximum of $|\mathcal{F}(\psi^{kq})(k_0, \phi_0)|$ is obtained near the ring $k_0 = \frac{R_{kq}}{2\pi}$.

Sampling Criterion

- Due to the Nyquist criterion, the Fourier-Bessel expansion requires components for which

$$\frac{R_{kq}}{2\pi} \leq \frac{L}{2},$$

because other components represent features beyond the resolution and their inclusion would result in aliasing.

- With the asymptotic expansion of the Bessel function, $R_{kq} \sim \frac{\pi}{2}(k + 2q - \frac{1}{2})$ asymptotically, and we get

$$k + 2q \leq 2L + \frac{1}{2}.$$

Sampling Criterion

The cut-off criterion $k + 2q \leq 2L + \frac{1}{2}$ is similar to the criterion derived by Klug and Crowther (1972, Nature), $k + 2q \leq \pi L + 2$, but is not identical to it. Specifically, their criterion has a different cut-off value for $k + 2q$. The difference stems from the fact that Klug and Crowther studied a different problem, namely, the 2D reconstruction problem of an image from its 1D line projections.

Expansion

- We denote by $\tilde{I}_i$ the continuous approximation of the i'th image in terms of a truncated Fourier-Bessel expansion including only components satisfying the sampling criterion.

$$\tilde{I}_i(r,\theta) = \sum_{k=-2L}^{2L} \sum_{q=1}^{p_k} a^i_{k,q} \psi^{kq}(r,\theta),$$

- The coefficient vector a^i of the i'th image is the solution of $\min_{a^i} \|\Psi a^i - I_i\|^2$, given by $a^i = (\Psi^\dagger \Psi)^{-1} \Psi^\dagger I_i$.

Properties of Expansion Coefficients

- Under in-plane rotation,

$$\tilde{I}_i^{\alpha}(r,\theta) = \tilde{I}_i(r,\theta-\alpha) = \sum_{k,q} a^i_{k,q} e^{-\imath k\alpha} \psi^{kq}(r,\theta).$$

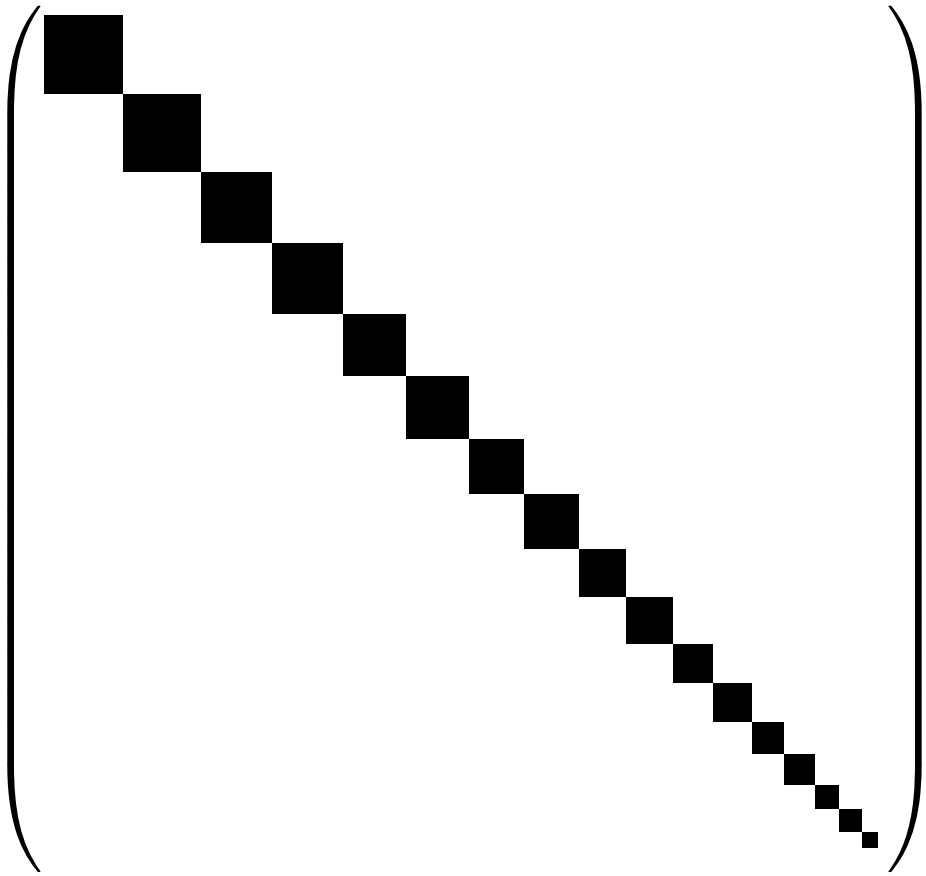
- Under reflection $a^i_{k,q}$ changes to $a^i_{-k,q}$, and the reflected image, denoted $\tilde{I}_i^r$, is given by

$$\tilde{I}_i^r(r,\theta) = \tilde{I}_i(r,\pi-\theta) = \sum_{k,q} a^i_{-k,q} \psi^{kq}(r,\theta).$$

Covariance Matrix

In the Fourier-Bessel basis the covariance matrix C of the data with all rotations and reflection is given by

$$\begin{aligned} & C_{(k,q),(k',q')} \\ = & \frac{1}{4\pi n} \sum_{i=1}^{n} \int_0^{2\pi} \left[a^i_{k,q} \overline{a^i_{k',q'}} + a^i_{-k,q} \overline{a^i_{-k',q'}} \right] e^{-\imath(k-k')\alpha} d\alpha \\ = & \delta_{k,k'} \frac{1}{n} \sum_{i=1}^{n} \Re\{ a^i_{k,q} \overline{a^i_{k,q'}} \}. \end{aligned}$$



Each small block $C^{(k)}_{q,q'} = \frac{1}{n}\sum_{i=1}^{n} \Re\{a^i_{k,q}\overline{a^i_{k,q'}}\}$ is by itself a sample covariance matrix of size $p_k \times p_k$. The block size p_k decreases as the angular frequency k increases.

FBsPCA Basis

- For each block C_k, the associated eigenvalues and eigenvectors are, in descending order, $\lambda_k^1 \geq \lambda_k^2 \cdots \geq \lambda_k^{p_k}$ and the associated eigenvectors $h_k^1, h_k^2, \ldots, h_k^{p_k}$.

- The eigenimages V^{kl} ($0 \leq k \leq 2L$, $1 \leq l \leq p_k$) are linear combinations of the Fourier-Bessel functions, i.e., $V^{kl} = \Psi^{(k)} h_k^l$, where $\Psi^{(k)} = [\psi^{k1}, \ldots, \psi^{kp_k}]$.

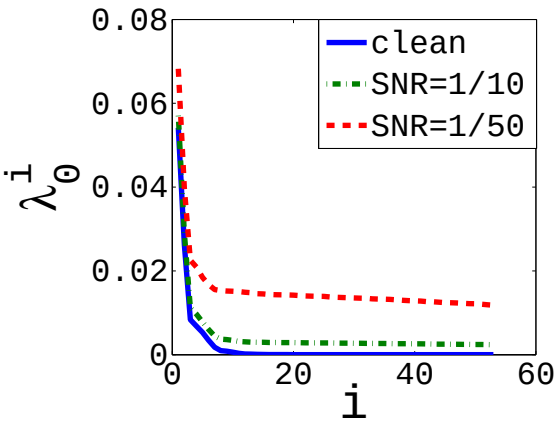
- Computational complexity: FBsPCA $O(nr_{max}^4 + r_{max}^5)$, PCA $O(nr_{max}^4 + r_{max}^6)$.

Selecting Meaningful Components for Noisy Images

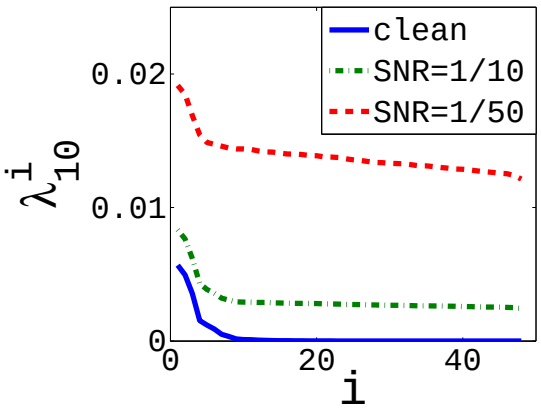
Given a good estimation of the noise variance $\hat{\sigma}^2$, we select the number of components by the following criteria: for the sample covariance matrix eigenvalues λ_k^i,

$$\lambda_k^i > \hat{\sigma}^2(1+\sqrt{\gamma_k})^2$$

where γ_k is determined by the number of images n and the size of the small covariance matrix p_k.

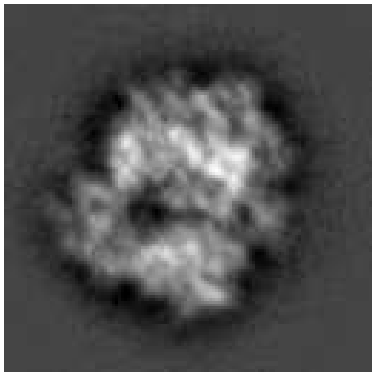


$k = 0$

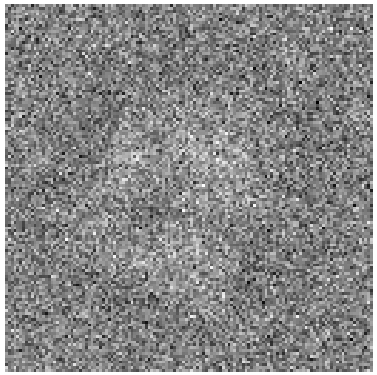


$k = 10$

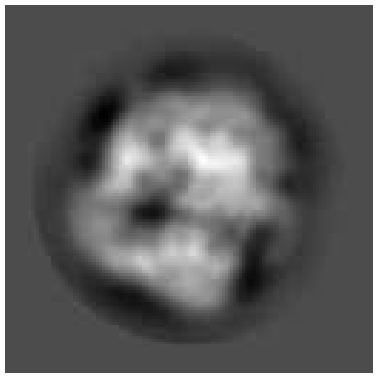
Experiments



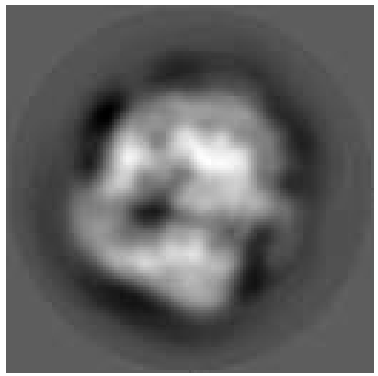
clean



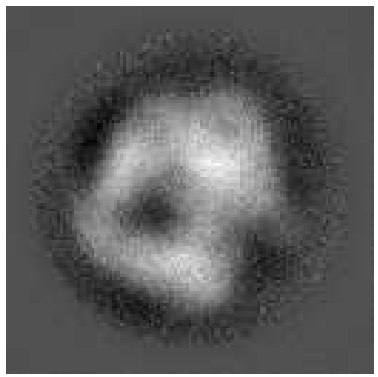
SNR$= 1/20$



FBsPCA



PTsPCA



PCA

	MSE (10^{-5})	PSNR(dB)	1-SSIM (10^{-6})
FBsPCA	6.6	20.2	4.4
PTsPCA	7.7	19.5	5.9
PCA	10.1	18.3	5.9

Denoising effect: FBsPCA, PTsPCA and PCA

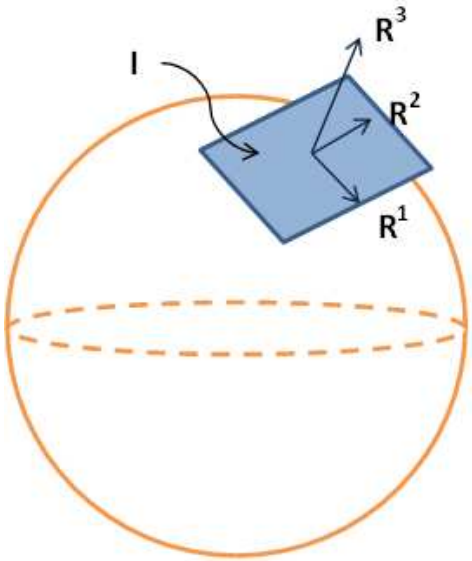
Summary

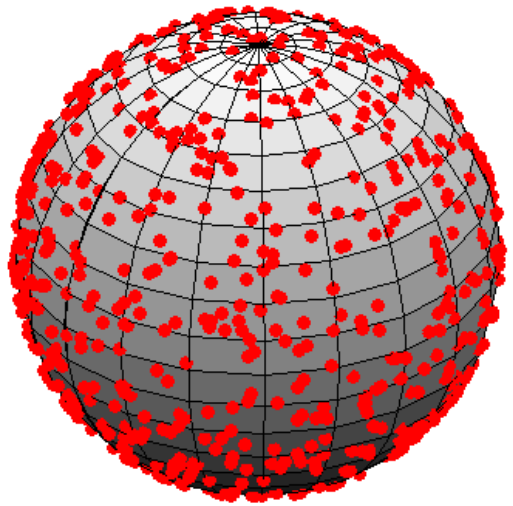
- FBsPCA obtains the maximum information without the introduction of spurious detail and amplification the noise by effects which are beyond resolution.
- The eigen-images are invariant to reflection and rotation of the input dataset.
- Computational complexity is smaller than the traditional PCA without duplicated copies, $O(nr_{max}^4 + r_{max}^5)$.

Outline

Group Invariant Features

Why do we need group invariant features?





Group Invariant Features

How do we get in-plane rotationally invariant features?

- Suppose we have a 1D periodic discrete signal $f(x)$. The discrete Fourier transform of f is defined as $\hat{f}(k) = \sum_{x=1}^{n} f(x) e^{-i\frac{2\pi}{n}kx}$.
- Bispectrum is defined as $b(k_1, k_2) = \hat{f}(k_1)\hat{f}(k_2)\overline{\hat{f}(k_1 + k_2)}$.
- Bispectrum is unbiased, complete and translationally invariant.
- Extend the definition of bispectrum over 1D periodic signal to the above expansion coefficients, we get

$$b^i_{k_1,k_2,q_1,q_2,q_3} = a^i_{k_1,q_1} a^i_{k_2,,q_2} \overline{a^i_{k_1+k_2,q_3}}.$$

k-NN Search and Alginment

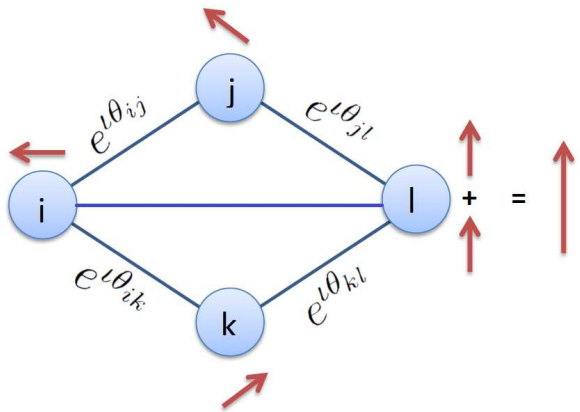
- The dimensionality of the feature vectors **b** is reduced by a randomized algorithm for low rank matrix approximation.
- The normalized cross-correlation between the reduced feature vectors define the in-plane rotationally invariant affinity between two images.
- k nearest neighbors are chosen and aligned.
- For very large data set, a randomized nearest neighbor search algorithm is used to estimate nearest neighbors in almost linear time with n.

Outline

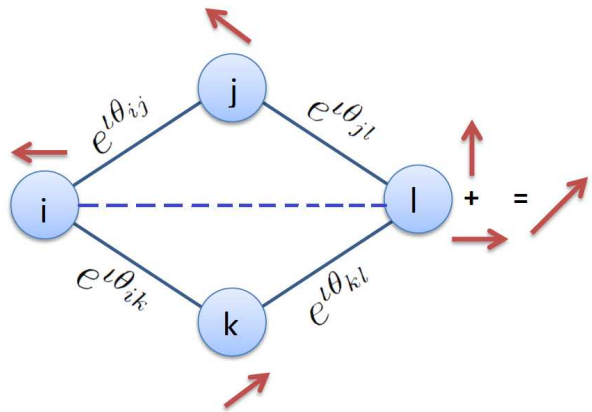
Vector Diffusion Maps

(Singer et al. 2011; Singer and Wu 2012)

Vector Diffusion Maps take into account the consistency of in-plane rotational transformations. The affinity between two nodes in a graph is defined as the consistency of the transformation summed over different connected paths between the nodes.



a) Higher affinity



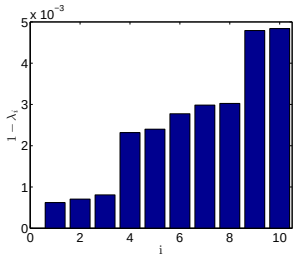
b) lower affinity

VDM

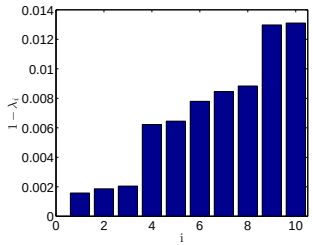
- From the initial nearest neighbor search and alignment, we build a sparse $n \times n$ Hermitian matrix H,

$$H_{ij} = \begin{cases} w_{ij} e^{\iota \alpha_{ij}} & \{i,j\} \in E, \\ 0. & \{i,j\} \notin E, \end{cases}$$

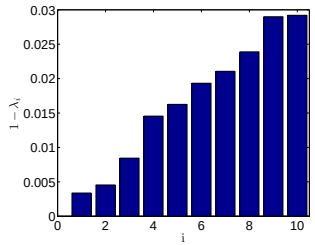
- $S = D^{-1/2} H D^{-1/2}$ is a Hermitian matrix similar to H normalized by the degree of each node.
- If the viewing angles are uniform over the sphere. The eigenvalues of the matrix has multiplicities $3, 5, 7, \ldots$.



SNR$= 1/2$



SNR $= 1/16$



SNR$= 1/32$

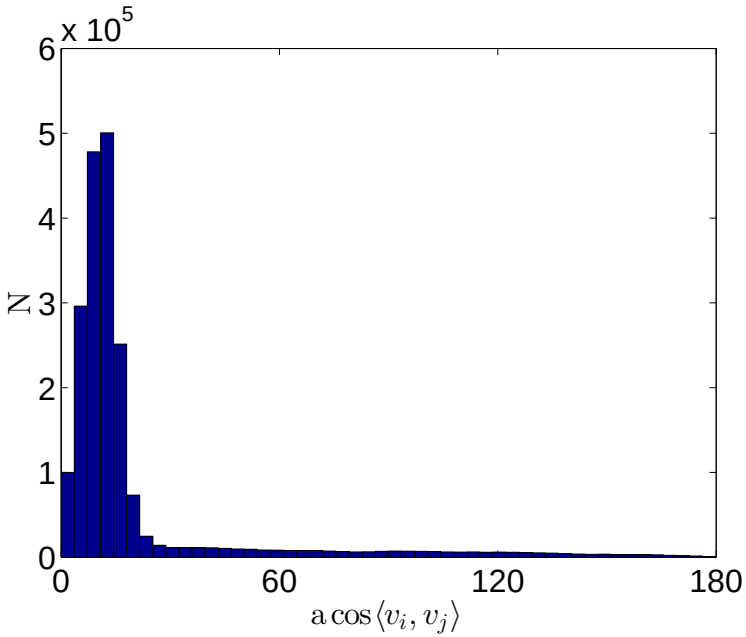
- The normalized affinity between i and j is defined as

$$\frac{|S^{2t}(i,j)|^2}{\sqrt{|S^{2t}(i,i)|^2|S^{2t}(j,j)|^2}}.$$

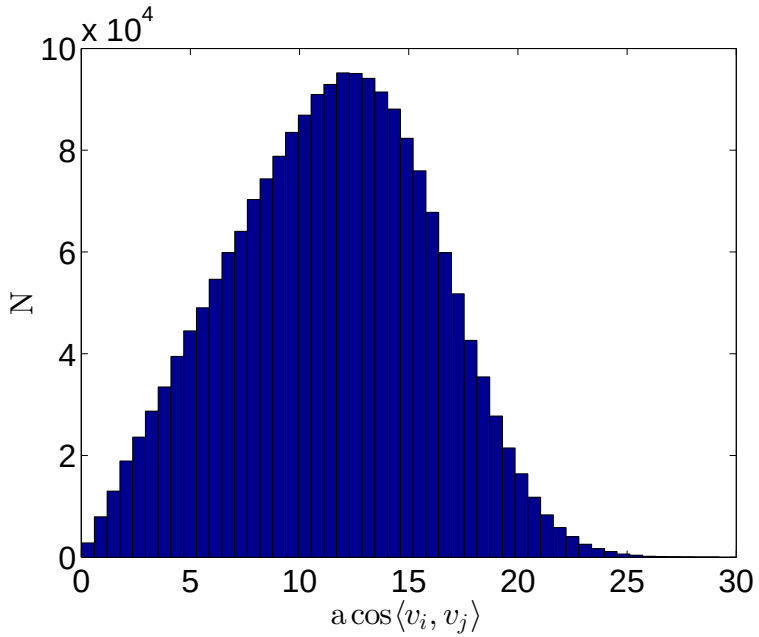
- The algorithm is efficient, because we use the eigen-decomposition of the sparse matrix S.

Initial vs VDM classification

Noisy (SNR $= 1/50$) centered 10^4 simulated projection images of 70S ribosome. $k = 200$



Initial Classification



VDM

Outline

Summary

- Computational framework for class averaging in cryo-EM SPR.
- More efficient and accurate for dealing with large cryo-EM data set.
- Introduced Fourier-Bessel steerable PCA basis for data denoising and compression.
- Constructed in-plane rotationally invariant features for fast nearest neighbor search.
- Use the knowledge of graph structure and linear transformations to get better estimation of nearest neighbors and alignment parameters.
- Work not covered in the talk: background noise estimation, translational alginment for non-centered images.

Cryo-EM and Other Applications

- Estimate the covariance matrix that does not have CTF modulation.
- Define metrics over images with different CTF.
- Construct invariants that are invariant to rotation and small shifts.
- Methods described here can be applied to X-ray free electron Laser (XFEL) single particle reconstruction.
- Steerable PCA and rotational invariant features can also be applied in studying fMRI volume data.

Acknowledgement

Amit Singer (Princeton)
Michael Aizenman (Princeton)
Yoel Shkolnisky (Tel-Aviv)
Ronny Hadani (U. Texas)
Shamgar Gurevich (U. Wisconsin)
Fred Sigworth (Yale)
Peter Jones (Yale)
Andrei Osipov (Yale)
Joachim Frank (Columbia)
Hstau Y. Liao (Columbia)

Lanhui Wang
Xiuyuan Cheng
Hau-Tieng Wu
Mihai Cucuringu
Kunal Chaudhury
Teng Zhang

Questions?

Thank You!

Low Rank Matrix Approximation

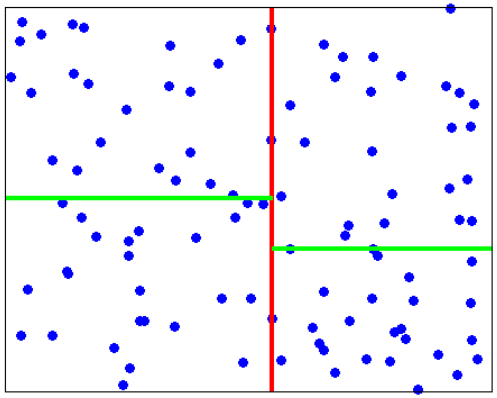
- We use a randomized algorithm for low rank matrix approximation (Rokhlin et. al. 2008) to reduce the dimensionality of the bispectrum.
- For a data matrix A of size $m \times n$ we seek a rank k approximation matrix B.
- The algorithm first forms a real $l \times m$ matrix G whose entries are i.i.d. Gaussian random variables of zero mean and unit variance, and then forms $R = GA$.
- Use a pivoted QR-decomposition algorithm on R^T to form Q whose columns are orthonormal.
- Form an SVD of an $m \times l$ matrix $T = AQ$, $T = U\Sigma W^T$. Then the low rank approximation of A is $B = U_k \Sigma_k V_k^T$, where $V = QW$.

RANN

- We use a randomized algorithm for approximate nearest neighbor search (RANN) (Jones and Osipov 2011) for fast nearest neighbor search.

- Given n points $\{x_i\}$ in $\mathbb{R}^d$ the brute force algorithm is $O(n^2 d)$. In approximate nearest neighbor search, we would like to find, for each x_i, k points whose distances are statistically similar to the true k NN's.

- RANN is an iterative algorithm. In each iteration the data is randomly rotated. Based on the first coordinate (after rotation), the points are divided by their median into two bins. The splitting process is repeated for each bin based on the other coordinates until k points remain in each bin.



- The time complexity of the algorithm is $Tn(d\log d + k(d + \log(k))\log(n)) + nk^2(d + \log k)$, where T is the number of iterations. The red term is rotation/hashing and the blue term is graph search.

VDM

Since S is Hermitian, it has a complete set of eigenvectors $v_1, v_2, \ldots, v_n$ and real eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$. We order the eigenvalues in decreasing order of magnitude. The spectral decomposition of S and S^{2t} are given by

$$S(i,j) = \sum_{l=1}^{n} \lambda_l v_l(i)\overline{v_l(j)}, \quad \text{and} \quad S^{2t}(i,j) = \sum_{l=1}^{n} \lambda_l^{2t} v_l(i)\overline{v_l(j)}.$$

VDM

$|S^{2t}(i,j)|^2$ is an inner product for the finite dimensional Hilbert space $\mathbb{C}^{n^2}$ via the mapping V_t :

$$V_t : i \mapsto ((\lambda_l \lambda_r)^t \langle v_l(i), v_r(i) \rangle)_{l,r=1}^{n}.$$

That is,

$$|S^{2t}(i,j)|^2 = \langle V_t(i), V_t(j) \rangle.$$

The normalized affinity can be expressed using the mapping V_t as

$$\frac{|S^{2t}(i,j)|^2}{\sqrt{|S^{2t}(i,i)|^2|S^{2t}(j,j)|^2}} = \left\langle \frac{V_t(i)}{|V_t(i)|}, \frac{V_t(j)}{|V_t(j)|} \right\rangle .$$

VDM

we can approximate the normalized affinity by truncating the mapping V_t to its leading m^2 coordinates (instead of n^2) as

$$V_t^m : i \mapsto ((\lambda_l \lambda_r)^t \langle v_l(i), v_r(i) \rangle)_{l,r=1}^m.$$

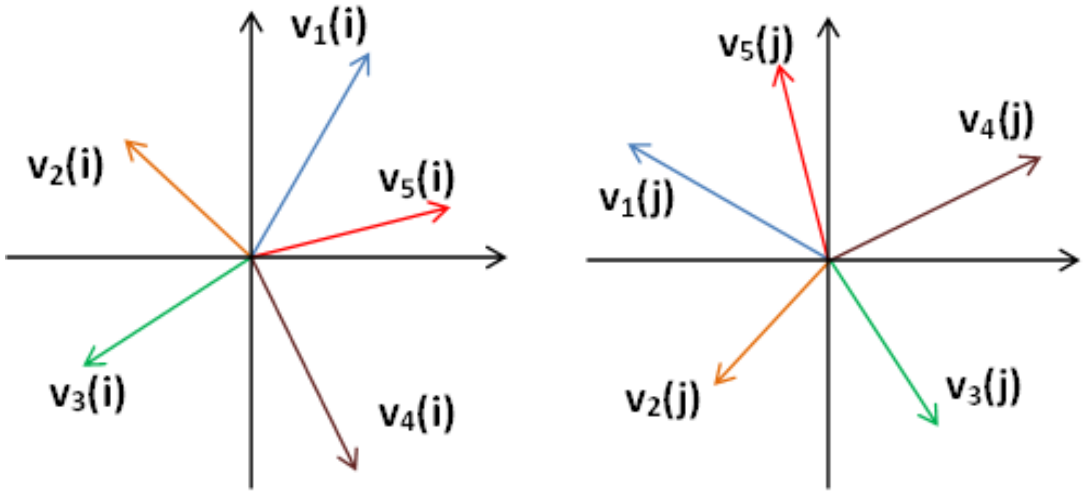
where m is the largest integer satisfying $\lambda_m^{2t} > \delta$ for some δ much smaller than 1. The approximate normalized affinity becomes

$$\left\langle \frac{V_t^m(i)}{|V_t^m(i)|}, \frac{V_t^m(j)}{|V_t^m(j)|} \right\rangle.$$

Alignment using eigenvectors

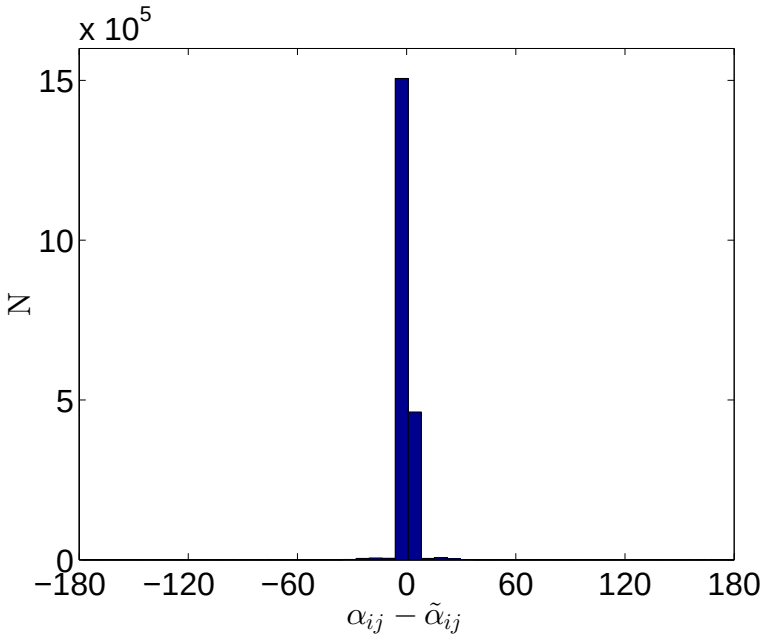
If i and j are of the same viewing angle, the eigenvectors of S satisfy the relation, $v_l(i) = e^{\iota\alpha_{ij}} v_l(j), \quad \forall l = 1, \ldots, n$. We can use the eigenvectors to estimate the rotational alignment,

$$v_l(i) = e^{\iota\alpha_{ij}} v_l(j), \quad \forall l = 1, ..., n.$$

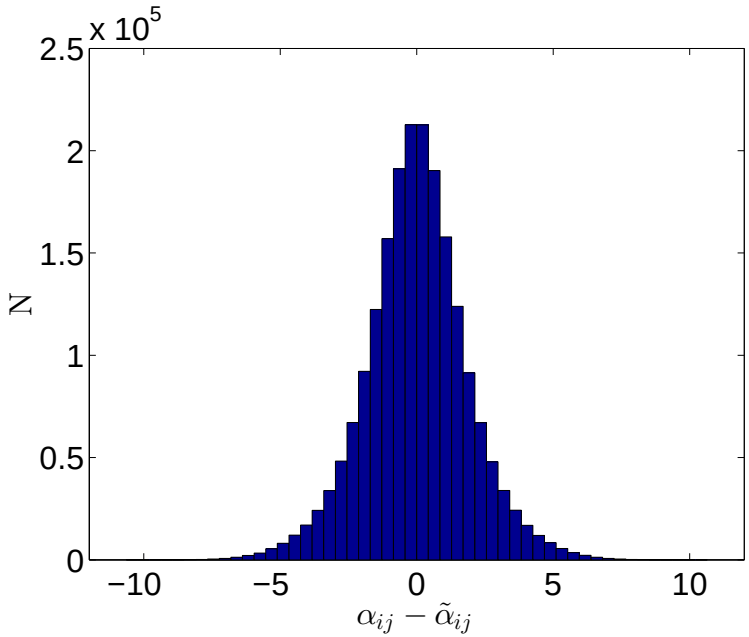


The alignment angles are estimated by

$$e^{\iota\alpha_{ij}} = \frac{\sum_{l=1}^{m} \lambda_l^{2t} v_l(i)\overline{v_l(j)}}{|\sum_{l=1}^{m} \lambda_l^{2t} v_l(i)\overline{v_l(j)}|}.$$



Initial Classification



VDM