Estimating low-rank covariance matrices from tomographic projections for Cryo-EM classification

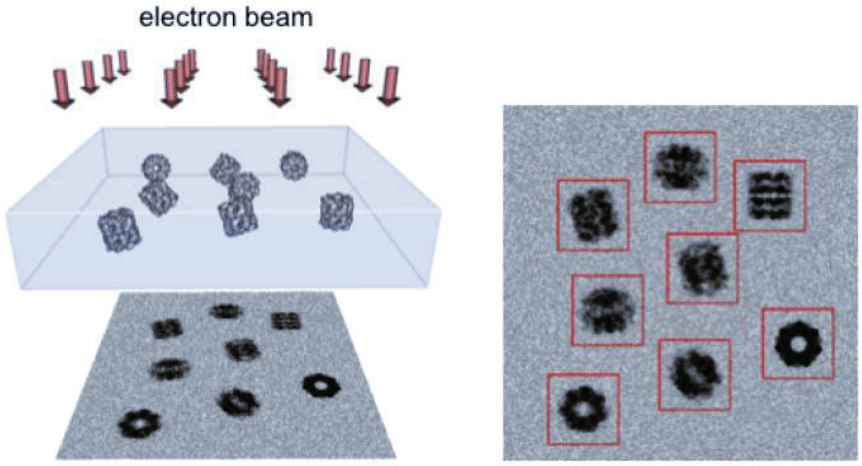
Joakim Andén, Eugene Katsevich, Amit Singer

The Program in Applied and Computational Mathematics
Princeton University

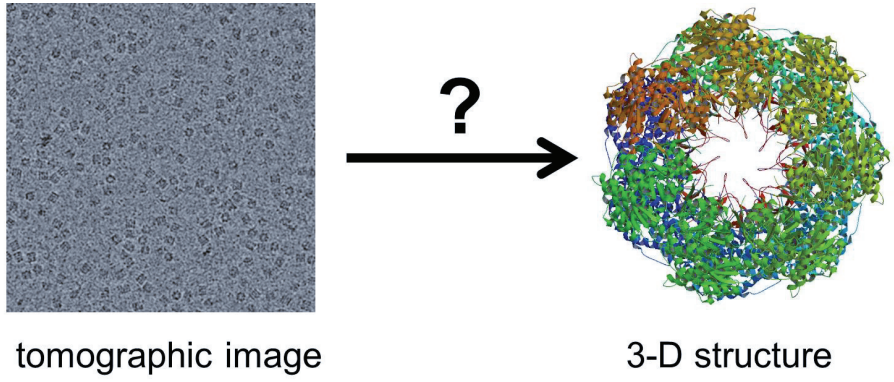
SPARS, July 9, 2015

Single particle reconstruction in Cryo-EM

Imaging process in cryo-electron microscopy

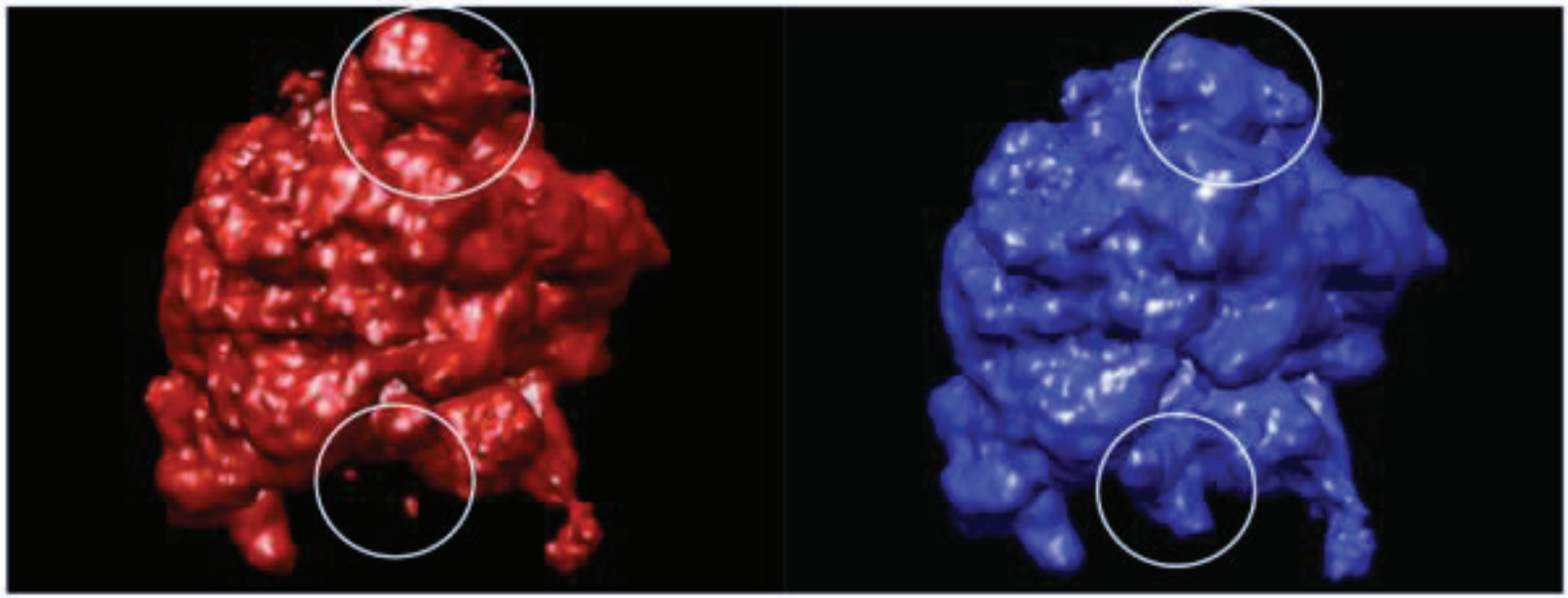


The cryo-EM problem



The heterogeneity problem

What if the molecule has more than one possible structure?

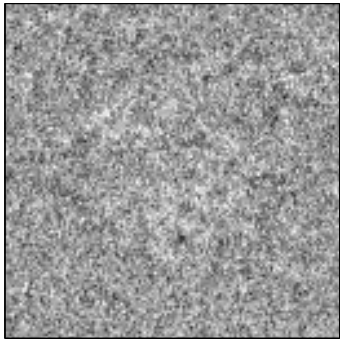


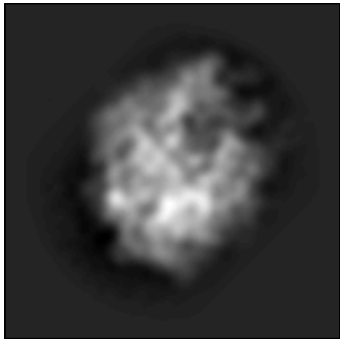
(Image source: H. Liao and J. Frank, Classification by bootstrapping in single particle methods, *Proceedings of the 2010 IEEE international conference on biomedical imaging*, 2010.)

Experimental data: 70S Ribosome

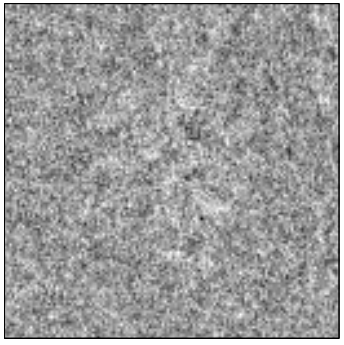
10000 image dataset (130-by-130), courtesy Joachim Frank (Columbia University)

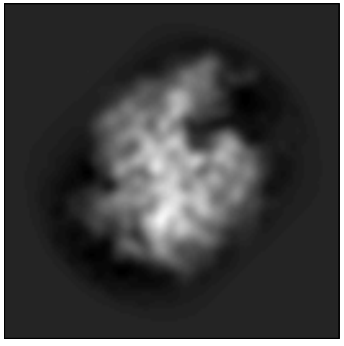
Class 1





Class 2





Morphing video

Problem formulation

- Let $\mathbf{X}_1, \ldots, \mathbf{X}_n \in \mathbb{R}^p$ be a set of i.i.d. discrete random variables representing volumes in a Cryo-EM dataset. The number of possible states is $C \ll p$.
- Let $M_s : \mathbb{R}^p \to \mathbb{R}^q$ project a volume along a viewing direction $R_s \in SO(3)$ and filter an image by the contrast transfer function h_s.
- Set of projection images given by

$$\mathbf{I}_s = M_s \mathbf{X}_s + \mathbf{E}_s, \quad s = 1, \ldots, n,$$

  where $\mathbf{E}_1, \ldots, \mathbf{E}_n \in \mathbb{R}^q$ is a set of i.i.d. Gaussian white noise variables of variance σ^2.
- Given realizations $I_1, \ldots, I_n$ of $\mathbf{I}_1, \ldots, \mathbf{I}_n$, goal is to cluster according to corresponding state of $X_1, \ldots, X_n$.

Subspace approach

- If $\mathbf{X}_s$ has C possible values, volumes reside in a $C-1$-dimensional affine space $\mathbf{U}$ centered in $\mathbb{E}[\mathbf{X}_s]$ and spanned by $\mathrm{Var}[\mathbf{X}_s]$[1].
- Basic algorithm:
  1. Estimate viewing directions R_s for all images assuming single molecule.
  2. Find $\mathbb{E}[\mathbf{X}_s]$ and $\mathrm{Var}[\mathbf{X}_s]$; construct $\mathbf{U}$.
  3. Project I_s onto $M_s\mathbf{U}$ and use coordinates α_s to cluster.
  4. Reconstruct a volume for each cluster using standard techniques.
  5. Re-estimate viewing directions and repeat (optional).
- Given realizations $I_1, \ldots, I_n$, estimate $\mathbb{E}[\mathbf{X}_s]$ and $\mathrm{Var}[\mathbf{X}_s]$.

Principal component analysis from noisy projections

[1] Penczek et al. (2009), Liao & Frank (2010)

Least-squares estimators

- From $\mathbf{I}_s = M_s \mathbf{X}_s + \mathbf{E}_s$, we have

$$\mathbb{E}[\mathbf{I}_s] = M_s \mathbb{E}[\mathbf{X}_s], \quad \mathrm{Var}[\mathbf{I}_s] = M_s \mathrm{Var}[\mathbf{X}_s] M_s^H + \sigma^2 \mathrm{I}_q.$$

- Given a set of realizations $I_1, \ldots, I_n$, define

$$\mu_n = \arg\min_{\mu} \frac{1}{n} \sum_{s=1}^{n} \|I_s - M_s \mu\|^2 + \lambda \|\mu\|^2,$$

and

$$\Sigma_n = \arg\min_{\Sigma} \frac{1}{n} \sum_{s=1}^{n} \|(I_s - M_s \mu_n)(I_s - M_s \mu_n)^H - (M_s \Sigma M_s^H + \sigma^2 \mathrm{I}_q)\|_F^2 + \lambda \|\Sigma\|^2.$$

- For uniform distribution of viewing directions, $h_s(\omega) = 1$, and $\lambda = 0$, $\Sigma_n \xrightarrow{a.s.} \mathrm{Var}[\mathbf{X}_s]$ as $n \to \infty$.

Normal equations

Mean estimator satisfies $A_n\mu_n = b_n$, where

$$A_n = \frac{1}{n}\sum_{s=1}^{n} M_s^H M_s + \lambda \mathrm{I}_q, \text{ and } b_n = \frac{1}{n}\sum_{s=1}^{n} M_s^H I_s.$$

For covariance, $L_n(\Sigma_n) = B_n$, where $L_n : \mathbb{R}^{p\times p} \to \mathbb{R}^{p\times p}$ is defined by

$$L_n(\Sigma) = \frac{1}{n}\sum_{s=1}^{n} M_s^H M_s \Sigma M_s^H M_s + \lambda\Sigma$$

and

$$B_n = \frac{1}{n}\sum_{s=1}^{n} M_s^H (I_s - M_s\mu_n)(I_s - M_s\mu_n)^H M_s$$

$$- \sigma^2 \frac{1}{n}\sum_{s=1}^{n} M_s^H M_s \ .$$

Inversion of L_n

- Direct inversion of matrix intractable; costs $O(p^6) = O(N^{18})$, since $p = O(N^3)$ for resolution N.
- Katsevich et al (2015) designed a basis in which L_n is approximated by a block-diagonal operator. This requires uniform distribution of viewing directions and no CTF.
- Another approach is to use iterative methods to invert L_n, such as the algebraic reconstruction technique (ART), also known as the Kaczmarz method[2].
- We use conjugate gradient (CG) method to calculate $L_n^{-1}(B_n)$ through repeated application of L_n.
- Non-uniform fast Fourier transforms allow us to calculate L_n in $O(nN^6 \log N)$, so T iterations cost $O(nTN^6 \log N)$.

[2] Liao et al. (2014)

Convolution operator

- The operator $M_s^H M_s$ projects, then backprojects, so it can be written as a convolution. As a result

$$\Sigma \mapsto M_s^H M_s \Sigma M_s^H M_s,$$

is a convolution along the rows and columns of Σ.

- Plugging this into the expression for L_n, we have

$$\begin{aligned} L_n(\Sigma) &= \frac{1}{n} \sum_{s=1}^{n} M_s^H M_s \Sigma M_s^H M_s + \lambda \Sigma \\ &= \Sigma * \mathrm{Ker} + \lambda \Sigma \end{aligned}$$

where Ker is a convolution kernel in six dimensions.

Convolution operator (cont.)

- Precalculating Ker in one pass through the dataset takes $O(nN^6 \log N)$, but once calculated, applying L_n amounts to a convolution costing only $O(N^6 \log N)$.

- Solving $L_n(\Sigma_n) = B_n$ then has computational complexity $O((n+T)N^6 \log N)$.

- We can also approximate convolution by circular convolution. In the Fourier domain, we then solve

$$\widehat{\mathrm{Ker}} \circ \widehat{\Sigma}_n = \widehat{B}_n,$$

where $\circ$ is the elementwise matrix product. Complexity for this method is just $O(nN^6 \log N)$.

Low-rank approximation

- However, these methods all have complexity of at least $O(N^6 \log N)$, severely limiting resolution.
- Katsevich et al (2015) showed that

$$\lim_{n\to\infty} \widehat{\mathrm{Ker}}(\vec{\xi_1}, \vec{\xi_2}) = \frac{2}{\|\vec{\xi_1} \times \vec{\xi_2}\|},$$

  for $h_s(\omega) = 1$ and uniformly distributed R_s.
- The eigenvalues decay as $1/k$, so $\widehat{\mathrm{Ker}}$ is approximately low-rank.
- For a small C, $\mathrm{Var}[\mathbf{X}_s]$ has low rank $C - 1$. Therefore B_n is also of low rank, since $\widehat{B}_n$ is the elementwise product of two low-rank matrices $\widehat{\mathrm{Var}[\mathbf{X}_s]}$ and $\widehat{\mathrm{Ker}}$.

Low-rank approximation (cont.)

- Being the sum of rank-one matrices, B_n can be applied to vectors fast.
- Probablistic matrix decompositions techniques thus allow for efficient low-rank approximation[3].
- Asymptotically, $1/\widehat{\mathrm{Ker}}(\vec{\xi_1}, \vec{\xi_2}) \approx \frac{1}{2}\|\vec{\xi_1} \times \vec{\xi_2}\|$, so we expect

$$\widehat{\mathrm{iKer}}(\vec{\xi_1}, \vec{\xi_2}) = \frac{1}{\widehat{\mathrm{Ker}}(\vec{\xi_1}, \vec{\xi_2}) + \lambda}$$

  to be smooth and therefore low-rank. The Nyström method provides a low-rank approximation of $\widehat{\mathrm{iKer}}$ by calculating a subset of its columns and interpolating[4].

---

[4] Halko et al (2011)

[4] Williams & Seeger (2001)

Low-rank approximation (cont.)

- We can now calculate

$$\widehat{\Sigma}_n = \widehat{\mathrm{iKer}} \circ \widehat{B}_n,$$

  where $\widehat{\Sigma}_n$ is a sum of rank-one matrices, since the elementwise product of two rank-one matrices is also rank-one.

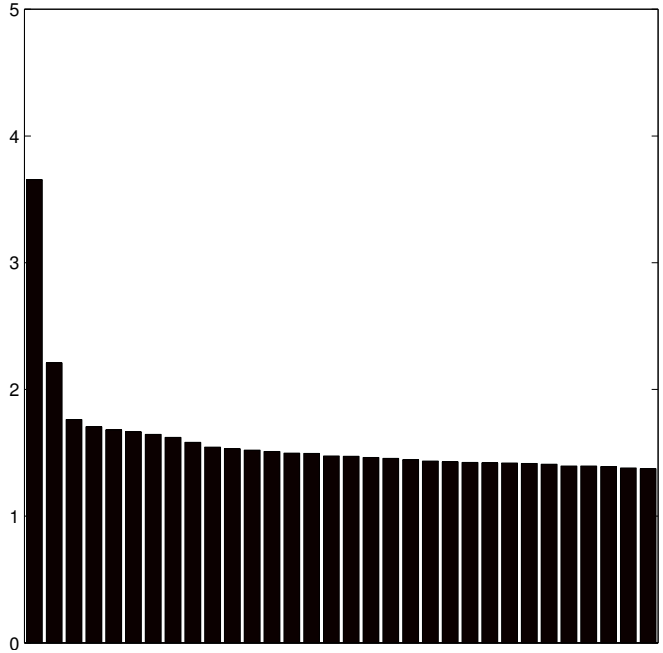
- Using this representation, we can apply standard eigendecomposition algorithms to $\widehat{\Sigma}_n$ to capture its principal eigenvectors.

- Letting r be the rank of B_n and iKer, complexity is now $O(rnN^3 \log N)$. This is a significant improvement over previous methods.

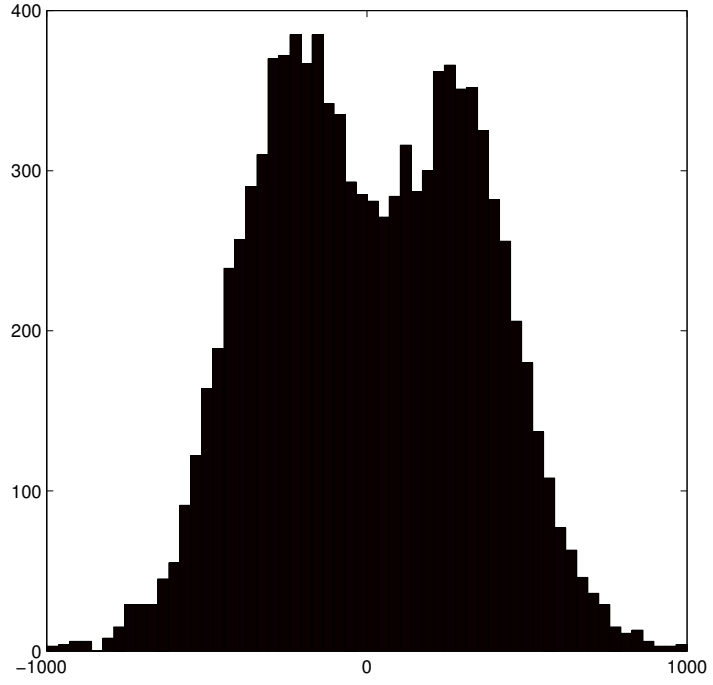
Conjugate gradient deconvolution: 70S ribosome

Dataset of 10000 images (130-by-130), 2 classes, courtesy Joachim Frank (Columbia University). Downsampled to $N = 17$.

Largest 32 eigenvalues



Coordinate histogram



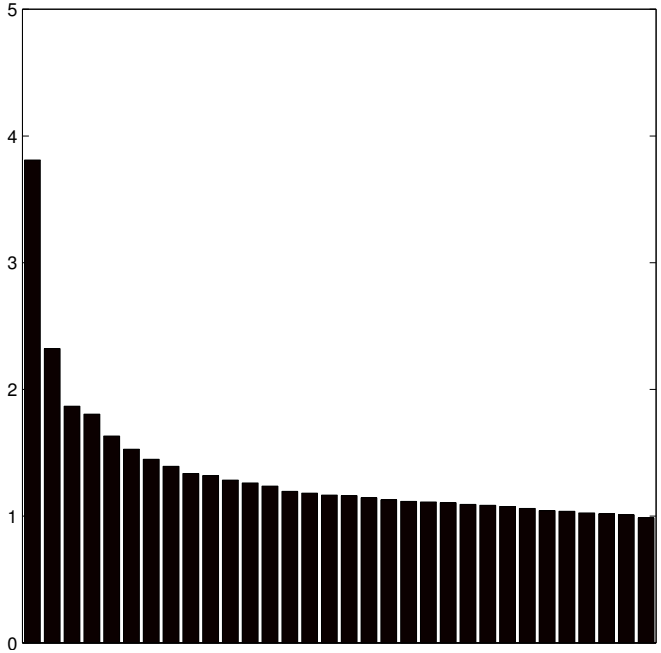
Took 18 min on a 2.9 GHz, 16-core CPU with 96 GB of memory.

Clustering accuracy 88% with respect to dataset labeling.

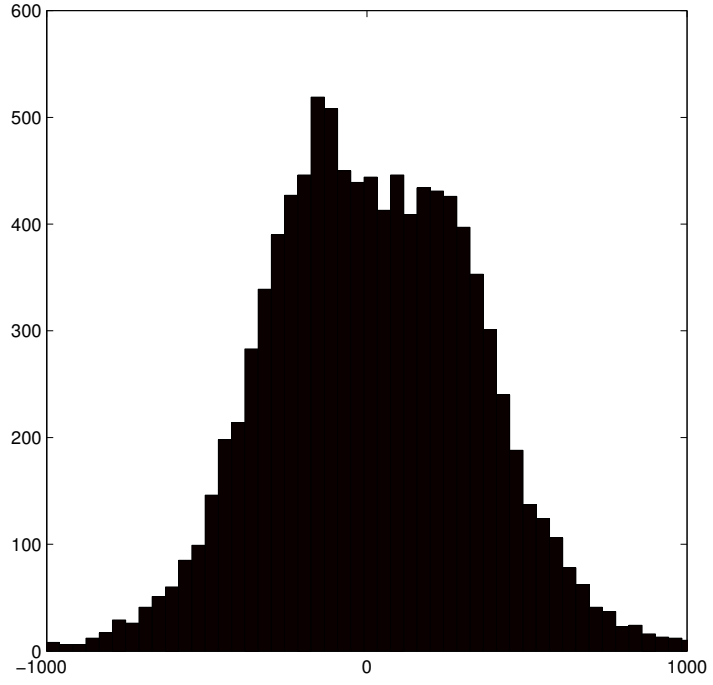
Low-rank deconvolution: 70S ribosome

Dataset of 10000 images (130-by-130), 2 classes, courtesy Joachim Frank (Columbia University). Downsampled to $N = 17$.

Largest 32 eigenvalues



Coordinate histogram



Took 5 min on a 2.9 GHz, 16-core CPU with 96 GB of memory.

Clustering accuracy 84% with respect to dataset labeling.

Conclusions

- Principal component analysis of volumes from noisy projections using a least-squares estimator provides a powerful measure of variability in Cryo-EM volumes.
- Precalculating convolutional kernel coupled with conjugate gradient allows for efficient calculating of least-squares covariance estimator.
- Leveraging the low rank of the covariance matrix and convolution kernel significantly reduces complexity in time and memory.

Bibliography

J. Andén, E. Katsevich, and A. Singer, ”Estimating low-rank covariance matrices from tomographic projections for Cryo-EM classification,” *Signal Processing with Adaptive Sparse Structured Representations workshop*, 2015.

J. Andén, E. Katsevich, and A. Singer, ”Covariance estimation using conjugate gradient for 3D classification in Cryo-EM,” *12th IEEE International Symposium on Biomedical Imaging*, 2015.

E. Katsevich, A. Katsevich, and A. Singer, “Covariance matrix estimation for the Cryo-EM heterogeneity problem,” *SIAM Journal on Imaging Sciences*, vol. 8, no. 1, pp. 126–185, 2015.