PCA from noisy linearly reduced measurements

Amit Singer

Princeton University
Department of Mathematics and Program in Applied and Computational Mathematics

September 27, 2016
Single particle reconstruction using cryo-EM

Schematic drawing of the imaging process:

The cryo-EM problem:
Precise knowledge of the structure of macromolecules in the cell is essential for understanding how they function. Structures of large macromolecules can now be obtained at near-atomic resolution by averaging thousands of electron microscope images recorded before radiation damage accumulates. This is what Amunts et al. have done in their research article on page 1485 of this issue (1), reporting the structure of the large subunit of the mitochondrial ribosome at 3.2 Å resolution by electron cryo-microscopy (cryo-EM). Together with other recent high-resolution cryo-EM structures (2–4) (see the figure), this achievement heralds the beginning of a new era in molecular biology, where structures at near-atomic resolution are no longer the prerogative of x-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy.

Ribosomes are ancient, massive protein-RNA complexes that translate the linear genetic code into three-dimensional proteins.
The 3.8 Å resolution cryo-EM structure of Zika virus

Devika Sirohi, Zhenguo Chen, Lei Sun, Thomas Klose, Theodore C. Pierson, Michael G. Rossmann, Richard J. Kuhn

1Markey Center for Structural Biology and Purdue Institute for Inflammation, Immunology and Infectious Disease, Purdue University, West Lafayette, IN 47907, USA. 2Viral Pathogenesis Section, Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA.
Single-particle cryo-electron microscopy (cryo-EM) is our choice for Method of the Year 2015 for its newfound ability to solve protein structures at near-atomic resolution. Featured is the 2.2-Å cryo-EM structure of β-galactosidase as recently reported by Bartesaghi et al. (Science 348, 1147–1151, 2015). Cover design by Erin Dewalt.
Big “movie” data, publicly available

http://www.ebi.ac.uk/pdbe/emdb/empiar/
Projection images \( y_k = T_kP_k\phi_k + \epsilon_k, \quad k = 1, 2, \ldots, n \).

\( \phi_k : \mathbb{R}^3 \to \mathbb{R} \) is the scattering potential of the \( k'th \) molecule.

\( P_k \) is 2D tomographic projection, after rotating \( \phi_k \) by \( R_k \).

\( T_k \) is a “filtering” operator due to microscope and detector.

Classical cryo-EM problem: Estimate \( \phi = \phi_1 = \phi_2 = \cdots = \phi_n \) (and \( R_1, \ldots, R_n \)) given \( y_1, \ldots, y_n \).

Heterogeneity problem: Estimate \( \phi_1, \phi_2, \ldots, \phi_n \) (and \( R_1, \ldots, R_n \)) given \( y_1, \ldots, y_n \).
Toy example
E. coli 50S ribosomal subunit

27,000 particle images provided by Dr. Fred Sigworth, Yale Medical School

3D reconstruction by S, Lanhui Wang, and Jane Zhao
Algorithmic and computational challenges

1. *Ab-initio* modelling and orientation assignment
2. Heterogeneity – resolving structural variability
3. Denoising and 2D class averaging
4. Particle picking
5. Symmetry detection and reconstruction
6. Motion correction
7. Estimation of contrast transfer function and noise statistics
8. Fast reconstruction and refinement
9. Validation and resolution determination
Principal components analysis (PCA)

- Given i.i.d. samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$ of some $\mathbf{x} \in \mathbb{R}^p$, estimate mean $\mathbb{E}[\mathbf{x}]$ and covariance $\text{Cov}[\mathbf{x}]$.

  $\mu_n = \frac{1}{n} \sum_{k=1}^{n} \mathbf{x}_k$
  $\Sigma_n = \frac{1}{n} \sum_{k=1}^{n} (\mathbf{x}_k - \mu_n)(\mathbf{x}_k - \mu_n)^T$

- Eigenvectors and eigenvalues of $\Sigma_n$ are used for dimensionality reduction, compression, and denoising.

- Useful when most variability is captured by a few leading principal components: low rank approximation.
Measure with $A_k : \mathbb{R}^p \rightarrow \mathbb{R}^q$, noise variance $\text{Cov}[e_k] = \sigma^2 I_q$

$$y_k = A_k x_k + e_k.$$ 

**Goal:** Estimate $E[x]$ and $\text{Cov}[x]$ from $y_1, \ldots, y_n$. 
Write image $y_k$ as

$$y_k = T_k P_k \phi_k + e_k, \quad k = 1, \ldots, n$$

where:

- $\phi_k$ is a three-dimensional molecular volume,
- $P_k$ projects volume into an image and $T_k$ filters it, and
- $e_k$ is measurement noise.
Goal: Estimate unfiltered projection $x_k = P_k \phi_k$ from

$$y_k = T_k x_k + e_k.$$
Goal: Estimate unfiltered projection $x_k = P_k \phi_k$ from

$$y_k = T_k x_k + e_k.$$
**Goal:** Estimate unfiltered projection $x_k = P_k \phi_k$ from

$$y_k = T_k x_k + e_k.$$
**Goal:** Estimate unfiltered projection $x_k = P_k \phi_k$ from

$$y_k = T_k x_k + e_k.$$
Goal: Estimate unfiltered projection $x_k = P_k \phi_k$ from

$$y_k = T_k x_k + e_k.$$ 

Filtering $T_k : \mathbb{R}^p \rightarrow \mathbb{R}^p$ is singular or ill-conditioned.
**Goal:** Estimate unfiltered projection $x_k = P_k \phi_k$ from

$$y_k = T_k x_k + e_k.$$ 

**Filtering** $T_k : \mathbb{R}^p \to \mathbb{R}^p$ is singular or ill-conditioned.

**Using low-rank $\text{Cov}[x]$, we can invert $T_k$ and reduce effect of noise $e_k$.**
Contrast transfer function: $T_k$

CTF suppresses information and inverts contrast
CTF cannot be trivially inverted (zero crossings)
Additional envelope Gaussian decay of high frequencies
Information lost from one defocus group could be recovered from another group that has different zero crossings.

Application II: heterogeneity

How to classify images from different molecular structures?¹

Class A

Class B

¹Images courtesy Joachim Frank (Columbia)
Goal: Classify images

\[ y_k = T_k P_k \phi_k + e_k, \]

according to molecular state \( \phi_k \).

Cannot invert since \( T_k P_k : \mathbb{R}^p \rightarrow \mathbb{R}^q \) and \( q < p \).

Volumes \( \phi_k \) live in low-dimensional space – can be estimated from top eigenvectors of \( \text{Cov}[\phi_k] \).

Restrict \( T_k P_k \) to this space, invert, cluster and reconstruct\(^2\).

Aggregating large-scale datasets give accurate estimates – despite high noise.

---

\(^2\)Penczek et al. (2009)
From $y_k = A_kx_k + e_k$, we have

$$\mathbb{E}[y_k] = A_k \mathbb{E}[x].$$

On the other hand

$$\mathbb{E}[y_k] \approx y_k,$$

so

$$y_k \approx A_k \mathbb{E}[x].$$

Form least-squares estimator for $\mathbb{E}[x]$

$$\mu_n = \arg \min_{\mu} \frac{1}{n} \sum_{k=1}^{n} \| y_k - A_k \mu \|^2.$$
Least-squares estimator (covariance)

- Again \( y_k = A_k x_k + e_k \) gives
  \[
  \text{Cov}[y_k] = A_k \text{Cov}[x] A_k^T + \sigma^2 I_q.
  \]

- On the other hand
  \[
  \text{Cov}[y_k] \approx (y_k - A_k \mu_n) (y_k - A_k \mu_n)^T,
  \]
  so
  \[
  (y_k - A_k \mu_n) (y_k - A_k \mu_n)^T \approx A_k \text{Cov}[x] A_k^T + \sigma^2 I_q.
  \]

- Form least-squares estimator for \( \text{Cov}[x] \)
  \[
  \Sigma_n = \arg \min_{\Sigma} \frac{1}{n} \sum_{k=1}^{n} \| (y_k - A_k \mu_n) (y_k - A_k \mu_n)^T - (A_k \Sigma A_k^H + \sigma^2 I_q) \|_F^2.
  \]
Normal equations

- Mean estimator $\mu_n$ satisfies
  \[
  \left( \frac{1}{n} \sum_{k=1}^{n} A_k^T A_k \right) \mu_n = \frac{1}{n} \sum_{k=1}^{n} A_k^T y_k.
  \]

- For covariance, $L_n(\Sigma_n) = B_n$, with $L_n : \mathbb{R}^{p \times p} \rightarrow \mathbb{R}^{p \times p}$ given by
  \[
  L_n(\Sigma) = \frac{1}{n} \sum_{k=1}^{n} A_k^T A_k \Sigma A_k^T A_k
  \]

  and

  \[
  B_n = \frac{1}{n} \sum_{k=1}^{n} A_k^T (y_k - A_k \mu_n) (y_k - A_k \mu_n)^T A_k - \sigma^2 A_k^T A_k.
  \]

- Resulting linear system in $p^2$ variables can be solved efficiently using the conjugate gradient method.
Estimator properties

- Taking $A_k = I_P$ and $\sigma^2 = 0$, recover sample mean and covariance.

- Does not rely on a particular distribution of $x$.

- With prior information on $x$, can regularize by adding terms to least-squares objective.

- Both $\mu_n$ and $\Sigma_n$ are consistent estimators$^3$. For $n \to \infty$

\[
\begin{align*}
\mu_n & \xrightarrow{a.s.} \mathbb{E}[x] \\
\Sigma_n & \xrightarrow{a.s.} \text{Cov}[x]
\end{align*}
\]

---

$^3$Katsevich, Katsevich, S (2015)
High-dimensional PCA: Marčenko-Pastur

For pure Gaussian white noise of unit variance

\[ y_k = e_k, \]

we have sample covariance

\[ \frac{1}{n} \sum_{k=1}^{n} y_k y_k^T. \]

\[ n \gg p \]

Central limit theorem

\[ n \approx p \]

Marčenko-Pastur law
Both signal and noise

\[ y_k = x_k + e_k. \]

### Central limit theorem

\[ n \gg p \]

### Spiked covariance model

\[ n \approx p \]

Signal eigenvalues pop out of bulk when SNR > \( \sqrt{p/n} \).

---

\(^4\)Johnstone (2001)
To estimate $\text{Cov}[\mathbf{x}]$ from

$$\mathbf{y}_k = \mathbf{x}_k + \mathbf{e}_k,$$

when $n \asymp p$, we apply shrinkage to eigenvalues of sample covariance\(^5\)

\(^5\)Donoho, Gavish, Johnstone (2013)
In clean case, $y_k = A_k x_k$ and

$$B_n = \frac{1}{n} \sum_{k=1}^{n} A_k^T A_k (x_k - \mu_n) (x_k - \mu_n)^T A_k^T A_k,$$
In clean case, $y_k = A_k x_k$ and

$$B_n = \frac{1}{n} \sum_{k=1}^{n} A_k^T A_k (x_k - \mu_n) (x_k - \mu_n)^T A_k^T A_k,$$

so

$$\mathbb{E}[B_n] = \frac{1}{n} \sum_{k=1}^{n} \text{Cov}[A_k^T A_k x_k].$$
In clean case, $y_k = A_k x_k$ and

$$B_n = \frac{1}{n} \sum_{k=1}^{n} A_k^T A_k (x_k - \mu_n) (x_k - \mu_n)^T A_k^T A_k,$$

so

$$\mathbb{E}[B_n] = \text{Cov}[A_k^T A_k x_k]$$

if the $A_k$ matrices are realizations of random matrices $A_k$. 
Shrinkage of $B_n$

- In clean case, $y_k = A_k x_k$ and

$$B_n = \frac{1}{n} \sum_{k=1}^{n} A_k^T A_k (x_k - \mu_n) (x_k - \mu_n)^T A_k^T A_k,$$

so

$$\mathbb{E}[B_n] = \text{Cov}[A_k^T A_k x_k]$$

if the $A_k$ matrices are realizations of random matrices $A_k$.

- For large $n$, we then have

$$B_n \approx \text{Cov}[A_k^T A_k x_k].$$
Shrinkage of $B_n$

- In clean case, $y_k = A_k x_k$ and

$$B_n = \frac{1}{n} \sum_{k=1}^{n} A_k^T A_k (x_k - \mu_n) (x_k - \mu_n)^T A_k^T A_k,$$

so

$$\mathbb{E}[B_n] = \text{Cov}[A_k^T A_k x_k]$$

if the $A_k$ matrices are realizations of random matrices $A_k$.

- For large $n$, we then have

$$B_n \approx \text{Cov}[A_k^T A_k x_k].$$

- How do we estimate $\text{Cov}[A_k^T A_k x_k]$ given noisy $y_k$?
Colored noise

\[
A_k^T y_k = A_k^T A_k x_k + A_k^T e_k.
\]
Colored noise

\[ A_k^T y_k = A_k^T A_k x_k + A_k^T e_k. \]

Whiten by calculating

\[ S = \frac{1}{n} \sum_{k=1}^{n} A_k^T A_k \approx \mathbb{E}[A_k^T A_k] \]

and setting

\[ z_k = S^{-1/2} A_k^T (y_k - A_k \mu_n). \]
Colored noise

\[ \text{covariance } \mathbb{E}[\mathbf{A}_k^T \mathbf{A}_k] \]

\[ \mathbf{A}_k^T y_k = \mathbf{A}_k^T \mathbf{A}_k x_k + \mathbf{A}_k^T \mathbf{e}_k. \]

Whiten by calculating

\[ \mathbf{S} = \frac{1}{n} \sum_{k=1}^{n} \mathbf{A}_k^T \mathbf{A}_k \approx \mathbb{E}[\mathbf{A}_k^T \mathbf{A}_k] \]

and setting

\[ \mathbf{z}_k = \mathbf{S}^{-1/2} \mathbf{A}_k^T (y_k - \mathbf{A}_k \mu_n). \]

Calculate sample covariance
Colored noise

\[
\text{covariance } \mathbb{E}[\mathbf{A}_k^T \mathbf{A}_k]
\]

\[
\mathbf{A}_k^T \mathbf{y}_k = \mathbf{A}_k^T \mathbf{A}_k \mathbf{x}_k + \mathbf{A}_k^T \mathbf{e}_k.
\]

Whiten by calculating

\[
\mathbf{S} = \frac{1}{n} \sum_{k=1}^{n} \mathbf{A}_k^T \mathbf{A}_k \approx \mathbb{E}[\mathbf{A}_k^T \mathbf{A}_k]
\]

and setting

\[
\mathbf{z}_k = \mathbf{S}^{-1/2} \mathbf{A}_k^T (\mathbf{y}_k - \mathbf{A}_k \mathbf{\mu}_n).
\]

Calculate sample covariance

\[
\sum_{k=1}^{n} \mathbf{z}_k \mathbf{z}_k^T
\]
Colored noise

\[
A_k^T y_k = A_k^T A_k x_k + A_k^T e_k.
\]

Whiten by calculating

\[
S = \frac{1}{n} \sum_{k=1}^{n} A_k^T A_k \approx \mathbb{E}[A_k^T A_k]
\]

and setting

\[
z_k = S^{-1/2} A_k^T (y_k - A_k \mu_n).
\]

Calculate sample covariance, shrink

\[
\rho \left( \sum_{k=1}^{n} z_k z_k^T \right)
\]
Colored noise

\[ \text{covariance } \mathbb{E}[A_k^T A_k] \]

\[ A_k^T y_k = A_k^T A_k x_k + A_k^T e_k. \]

Whiten by calculating

\[ S = \frac{1}{n} \sum_{k=1}^{n} A_k^T A_k \approx \mathbb{E}[A_k^T A_k] \]

and setting

\[ z_k = S^{-1/2} A_k^T (y_k - A_k \mu_n). \]

Calculate sample covariance, shrink, and unwhiten

\[ B_n = S^{1/2} \rho \left( \frac{1}{n} \sum_{k=1}^{n} z_k z_k^T \right) S^{1/2}. \]
Recall noisy, filtered, image

\[ y_k = T_k x_k + e_k \]

given clean projection \( x_k \).

Using covariance estimation method, estimate \( \text{Cov}[x_k] \).

Construct classical Wiener filter and invert \( T_k \) – estimate \( x_k \) from \( y_k \):

\[
\Sigma_n T_k^T (T_k \Sigma_n T_k^T + \sigma^2 I_q)^{-1} (y_k - T_k \mu_n) + \mu_n.
\]
Applying shrinkage to right-hand side $B_n$ in $L_n(\Sigma_n) = B_n$ results in significant increase in accuracy for the covariance estimation.
Experimental results: TRPV1 (EMPIAR-10005)


Input
Experimental results: TRPV1 (EMPIAR-10005)

Experimental results: TRPV1 (EMPIAR-10005)


Input

Filtered

Closest projection
K2 direct electron detector

35645 motion corrected, picked particle images of $256 \times 256$ pixels
### Experimental data - 80S ribosome

- **FALCON II** $4k \times 4k$ direct electron detector
- 105247 motion corrected, picked particle images of $360 \times 360$ pixels
Gatan 4k×4k CCD
37382 picked particle images of 256×256 pixels
Steerable PCA: covariance matrix commutes with in-plane rotation, block diagonalized in Fourier-Bessel basis, computational cost $O(nN^3)$ (Zhao, Shkolnisky, S, IEEE TCI 2016)

Wiener filtering: taking into account that estimated covariance differs from population covariance (S and Wu, SIAM Imaging 2013)

Visualization of underlying particles without class averaging

Outlier detection of picked particles

Improved class averages
Given images

\[ y_k = \underbrace{A_k}_{T_k P_k} x_k + e_k \]

classify \( y_k \) according to molecular state \( x_k \).

We use proposed method to estimate covariance \( \text{Cov}[x_k] \).

Reduce computational and memory costs by writing \( L_n \) as a convolution and circulant preconditioning.

Dominant eigenvectors of \( \Sigma_n \) define low-dimensional subspace where volumes \( x_k \) live. For \( C \) classes, \( \text{rank Cov}[x_k] = C - 1 \).

We can invert \( T_k P_k \) to find coordinates in subspace and cluster.
Experimental results: 70S ribosome

Dataset of 10000 images (130-by-130), 2 classes, courtesy Joachim Frank (Columbia University). Downsampling to $N = 16$.

Largest 32 eigenvalues

Coordinate histogram

Clustering accuracy 89% with respect to dataset labeling.

Took 26 min. on a 2.7 GHz, 12-core CPU.
Conclusions

- Least-squares approach to covariance estimation from noisy partial measurements is a simple but powerful method to characterize the distribution of the underlying data.

- In the high-dimensional regime, eigenvalue shrinkage can significantly improve the performance of the estimator.

- Covariance information can be leveraged to great effect in denoising and classification tasks such as those encountered in cryo-EM.
Current and future work

- Incorporating constraints on covariance $\Sigma_n$, such as sparsity.

- Replacing the least squares cost with a robust cost function.

- Theoretical understanding of bulk noise distribution in $\Sigma_n$, not just in $B_n$.

- Application to other tasks, such as embryo development, magnetic resonance imaging, and matrix completion.
Open source toolbox, publicly available:
http://spr.math.princeton.edu/
References


