PCA From Noisy Linearly Reduced Measurements

Joakim Andén
joint work with Amit Singer and Tejal Bhamre

The Program in Applied and Computational Mathematics
Princeton University

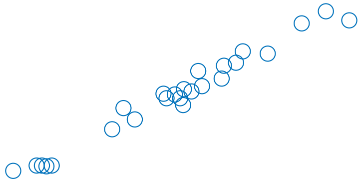
MMDS Workshop
June 24, 2016

PRINCETON UNIVERSITY

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 $\mathrm{Cov}[\mathbf{x}]$.

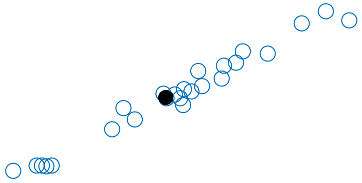
$$\boldsymbol{\mu}_n = \frac{1}{n}\sum_{k=1}^{n} \mathbf{x}_k \quad \boldsymbol{\Sigma}_n = \frac{1}{n}\sum_{k=1}^{n} (\mathbf{x}_k - \boldsymbol{\mu}_n)(\mathbf{x}_k - \boldsymbol{\mu}_n)^T$$



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 $\mathrm{Cov}[\mathbf{x}]$.

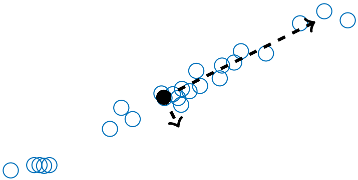
$$\boldsymbol{\mu}_n = \frac{1}{n}\sum_{k=1}^{n} \mathbf{x}_k \quad \boldsymbol{\Sigma}_n = \frac{1}{n}\sum_{k=1}^{n} (\mathbf{x}_k - \boldsymbol{\mu}_n)(\mathbf{x}_k - \boldsymbol{\mu}_n)^T$$



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 $\mathrm{Cov}[\mathbf{x}]$.

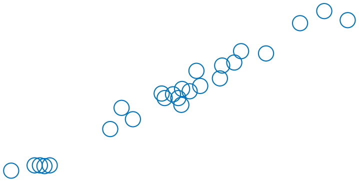
$$\boldsymbol{\mu}_n = \frac{1}{n}\sum_{k=1}^{n} \mathbf{x}_k \quad \boldsymbol{\Sigma}_n = \frac{1}{n}\sum_{k=1}^{n} (\mathbf{x}_k - \boldsymbol{\mu}_n)(\mathbf{x}_k - \boldsymbol{\mu}_n)^T$$



PCA from noisy linear measurements

- Measure with $A_k : \mathbb{R}^p \to \mathbb{R}^q$, noise variance $\mathbb{E}[|\mathbf{e}_k|^2] = \sigma^2$

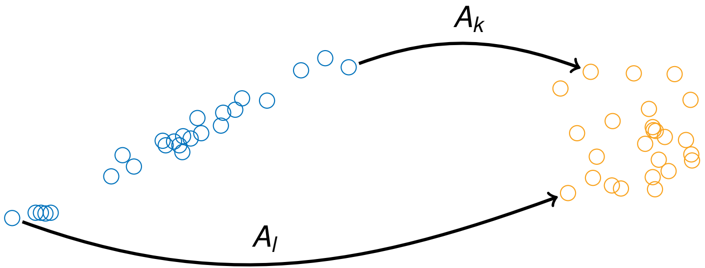
$$\mathbf{y}_k = A_k \mathbf{x}_k + \mathbf{e}_k.$$



PCA from noisy linear measurements

- Measure with $A_k : \mathbb{R}^p \to \mathbb{R}^q$, noise variance $\mathbb{E}[|\mathbf{e}_k|^2] = \sigma^2$

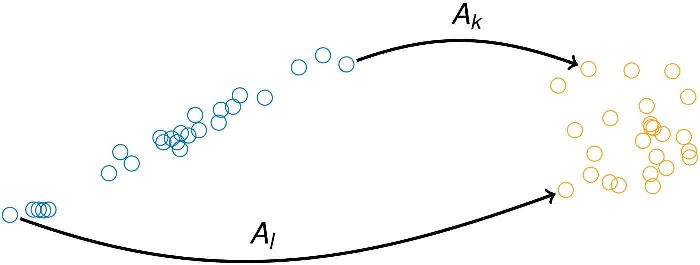
$$\mathbf{y}_k = A_k \mathbf{x}_k + \mathbf{e}_k.$$



PCA from noisy linear measurements

- Measure with $A_k : \mathbb{R}^p \to \mathbb{R}^q$, noise variance $\mathbb{E}[|\mathbf{e}_k|^2] = \sigma^2$

$$\mathbf{y}_k = A_k \mathbf{x}_k + \mathbf{e}_k.$$

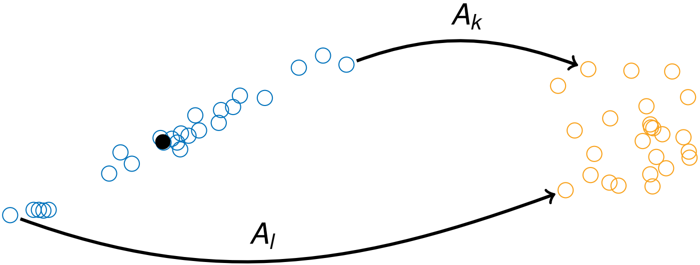


- **Goal:** Estimate $\mathbb{E}[\mathbf{x}]$ and $\mathrm{Cov}[\mathbf{x}]$ from $\mathbf{y}_1, \ldots, \mathbf{y}_n$.

PCA from noisy linear measurements

- Measure with $A_k : \mathbb{R}^p \to \mathbb{R}^q$, noise variance $\mathbb{E}[|\mathbf{e}_k|^2] = \sigma^2$

$$\mathbf{y}_k = A_k \mathbf{x}_k + \mathbf{e}_k.$$

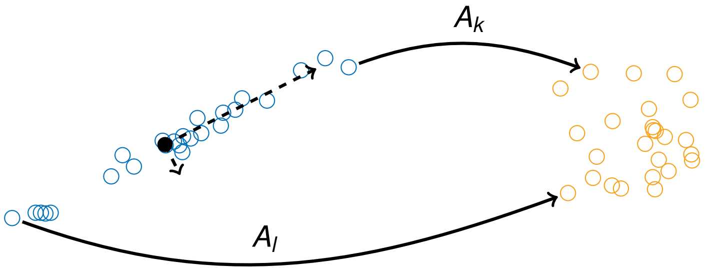


- **Goal:** Estimate $\mathbb{E}[\mathbf{x}]$ and $\mathrm{Cov}[\mathbf{x}]$ from $\mathbf{y}_1, \ldots, \mathbf{y}_n$.

PCA from noisy linear measurements

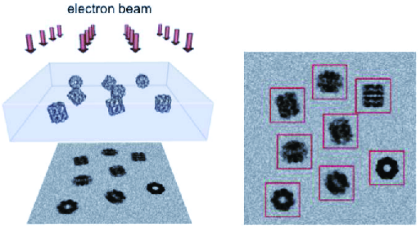
- Measure with $A_k : \mathbb{R}^p \to \mathbb{R}^q$, noise variance $\mathbb{E}[|\mathbf{e}_k|^2] = \sigma^2$

$$\mathbf{y}_k = A_k \mathbf{x}_k + \mathbf{e}_k.$$

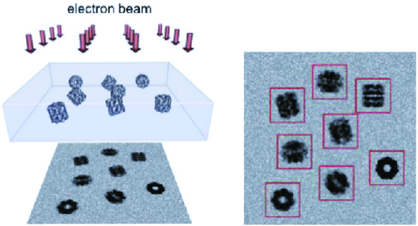


- **Goal:** Estimate $\mathbb{E}[\mathbf{x}]$ and $\mathrm{Cov}[\mathbf{x}]$ from $\mathbf{y}_1, \ldots, \mathbf{y}_n$.

Motivation: cryo-electron microscopy



Motivation: cryo-electron microscopy



Write image $\mathbf{y}_k$ as

$$\mathbf{y}_k = T_k P_k \phi_k + \mathbf{e}_k$$

where:

- ϕ_k is a three-dimensional molecular volume,
- P_k projects volume into an image and T_k filters it, and
- $\mathbf{e}_k$ is measurement noise.

Application: deconvolution in noise

- **Goal:** Estimate unfiltered projection $\mathbf{x}_k = P_k \phi_k$ from

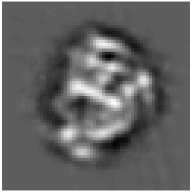
$$\mathbf{y}_k = T_k \mathbf{x}_k + \mathbf{e}_k.$$

Application: deconvolution in noise

- **Goal:** Estimate unfiltered projection $\mathbf{x}_k = P_k \phi_k$ from

$$\mathbf{y}_k = T_k \mathbf{x}_k + \mathbf{e}_k.$$

$\mathbf{x}_k$

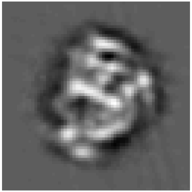


Application: deconvolution in noise

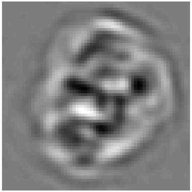
- **Goal:** Estimate unfiltered projection $\mathbf{x}_k = P_k \phi_k$ from

$$\mathbf{y}_k = T_k \mathbf{x}_k + \mathbf{e}_k.$$

$\mathbf{x}_k$



$T_k \mathbf{x}_k$

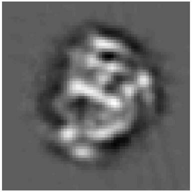


Application: deconvolution in noise

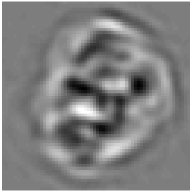
- **Goal:** Estimate unfiltered projection $\mathbf{x}_k = P_k \phi_k$ from

$$\mathbf{y}_k = T_k \mathbf{x}_k + \mathbf{e}_k.$$

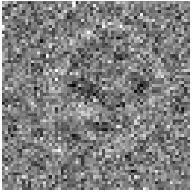
$\mathbf{x}_k$



$T_k \mathbf{x}_k$



$T_k \mathbf{x}_k + \mathbf{e}_k$

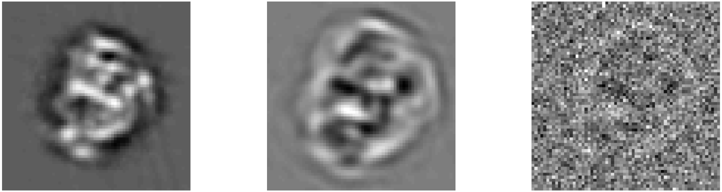


Application: deconvolution in noise

- **Goal:** Estimate unfiltered projection $\mathbf{x}_k = P_k \phi_k$ from

$$\mathbf{y}_k = T_k \mathbf{x}_k + \mathbf{e}_k.$$

$\mathbf{x}_k$ $\qquad$ $T_k \mathbf{x}_k$ $\qquad$ $T_k \mathbf{x}_k + \mathbf{e}_k$

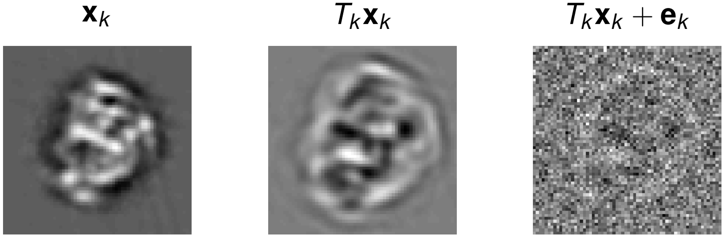


- Filtering $T_k : \mathbb{R}^p \to \mathbb{R}^p$ is singular or ill-conditioned.

Application: deconvolution in noise

- **Goal:** Estimate unfiltered projection $\mathbf{x}_k = P_k \phi_k$ from

$$\mathbf{y}_k = T_k \mathbf{x}_k + \mathbf{e}_k.$$

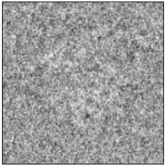


- Filtering $T_k : \mathbb{R}^p \to \mathbb{R}^p$ is singular or ill-conditioned.
- Using low-rank $\mathrm{Cov}[\mathbf{x}]$, we can invert T_k and reduce effect of noise $\mathbf{e}_k$.

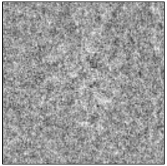
Application: heterogeneity

How to classify images from different molecular structures?

Class A



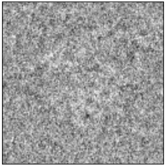
Class B[1]

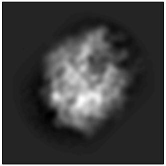


[1] Images courtesy Joachim Frank (Columbia)

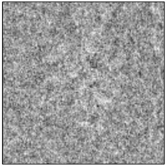
How to classify images from different molecular structures?

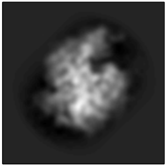
Class A





Class B[1]





[1] Images courtesy Joachim Frank (Columbia)

Application: heterogeneity (cont.)

- **Goal:** Classify images

$$\mathbf{y}_k = T_k P_k \phi_k + \mathbf{e}_k,$$

according to molecular state ϕ_k.

[2] Penczek et al. (2009)

Application: heterogeneity (cont.)

- **Goal:** Classify images

$$\mathbf{y}_k = T_k P_k \phi_k + \mathbf{e}_k,$$

according to molecular state ϕ_k.

- Cannot invert since $T_k P_k : \mathbb{R}^p \to \mathbb{R}^q$ and $q < p$.

[2] Penczek et al. (2009)

Application: heterogeneity (cont.)

- **Goal:** Classify images

$$\mathbf{y}_k = T_k P_k \phi_k + \mathbf{e}_k,$$

according to molecular state ϕ_k.

- Cannot invert since $T_k P_k : \mathbb{R}^p \to \mathbb{R}^q$ and $q < p$.
- Volumes ϕ_k live in low-dimensional space – can be estimated from top eigenvectors of $\mathrm{Cov}[\phi_k]$.

[2] Penczek et al. (2009)

Application: heterogeneity (cont.)

- **Goal:** Classify images

$$\mathbf{y}_k = T_k P_k \phi_k + \mathbf{e}_k,$$

according to molecular state ϕ_k.

- Cannot invert since $T_k P_k : \mathbb{R}^p \to \mathbb{R}^q$ and $q < p$.
- Volumes ϕ_k live in low-dimensional space – can be estimated from top eigenvectors of $\mathrm{Cov}[\phi_k]$.
- Restrict $T_k P_k$ to this space, invert, cluster and reconstruct[2].

[2] Penczek et al. (2009)

Application: heterogeneity (cont.)

- **Goal:** Classify images

$$\mathbf{y}_k = T_k P_k \phi_k + \mathbf{e}_k,$$

according to molecular state ϕ_k.

- Cannot invert since $T_k P_k : \mathbb{R}^p \to \mathbb{R}^q$ and $q < p$.
- Volumes ϕ_k live in low-dimensional space – can be estimated from top eigenvectors of $\mathrm{Cov}[\phi_k]$.
- 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)

Least-squares estimator (mean)

- From $\mathbf{y}_k = A_k \mathbf{x}_k + \mathbf{e}_k$, we have

$$\mathbb{E}[\mathbf{y}_k] = A_k \mathbb{E}[\mathbf{x}].$$

Least-squares estimator (mean)

- From $\mathbf{y}_k = A_k \mathbf{x}_k + \mathbf{e}_k$, we have

$$\mathbb{E}[\mathbf{y}_k] = A_k \mathbb{E}[\mathbf{x}].$$

- On the other hand

$$\mathbb{E}[\mathbf{y}_k] \approx \mathbf{y}_k,$$

so

$$\mathbf{y}_k \approx A_k \mathbb{E}[\mathbf{x}].$$

Least-squares estimator (mean)

- From $\mathbf{y}_k = A_k \mathbf{x}_k + \mathbf{e}_k$, we have

$$\mathbb{E}[\mathbf{y}_k] = A_k \mathbb{E}[\mathbf{x}].$$

- On the other hand

$$\mathbb{E}[\mathbf{y}_k] \approx \mathbf{y}_k,$$

so

$$\mathbf{y}_k \approx A_k \mathbb{E}[\mathbf{x}].$$

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

$$\boldsymbol{\mu}_n = \arg\min_{\mu} \frac{1}{n} \sum_{k=1}^{n} \|\mathbf{y}_k - A_k \mu\|^2.$$

Least-squares estimator (covariance)

- Again $\mathbf{y}_k = A_k \mathbf{x}_k + \mathbf{e}_k$ gives

$$\mathrm{Cov}[\mathbf{y}_k] = A_k \mathrm{Cov}[\mathbf{x}] A_k^T + \sigma^2 \mathrm{I}_q.$$

Least-squares estimator (covariance)

- Again $\mathbf{y}_k = A_k \mathbf{x}_k + \mathbf{e}_k$ gives

$$\mathrm{Cov}[\mathbf{y}_k] = A_k \mathrm{Cov}[\mathbf{x}] A_k^T + \sigma^2 \mathrm{I}_q.$$

- On the other hand

$$\mathrm{Cov}[\mathbf{y}_k] \approx (\mathbf{y}_k - A_k \boldsymbol{\mu}_n)(\mathbf{y}_k - A_k \boldsymbol{\mu}_n)^T,$$

so

$$(\mathbf{y}_k - A_k \boldsymbol{\mu}_n)(\mathbf{y}_k - A_k \boldsymbol{\mu}_n)^T \approx A_k \mathrm{Cov}[\mathbf{x}] A_k^T + \sigma^2 \mathrm{I}_q.$$

Least-squares estimator (covariance)

- Again $\mathbf{y}_k = A_k \mathbf{x}_k + \mathbf{e}_k$ gives

$$\mathrm{Cov}[\mathbf{y}_k] = A_k \mathrm{Cov}[\mathbf{x}] A_k^T + \sigma^2 \mathrm{I}_q.$$

- On the other hand

$$\mathrm{Cov}[\mathbf{y}_k] \approx (\mathbf{y}_k - A_k \boldsymbol{\mu}_n)(\mathbf{y}_k - A_k \boldsymbol{\mu}_n)^T,$$

so

$$(\mathbf{y}_k - A_k \boldsymbol{\mu}_n)(\mathbf{y}_k - A_k \boldsymbol{\mu}_n)^T \approx A_k \mathrm{Cov}[\mathbf{x}] A_k^T + \sigma^2 \mathrm{I}_q.$$

- Form least-squares estimator for $\mathrm{Cov}[\mathbf{x}]$

$$\boldsymbol{\Sigma}_n = \arg\min_{\Sigma} \frac{1}{n} \sum_{k=1}^{n} \left\| (\mathbf{y}_k - A_k \boldsymbol{\mu}_n)(\mathbf{y}_k - A_k \boldsymbol{\mu}_n)^T - (A_k \Sigma A_k^H + \sigma^2 \mathrm{I}_q) \right\|_F^2.$$

Normal equations

- Mean estimator $\boldsymbol{\mu}_n$ satisfies

$$\left(\frac{1}{n}\sum_{k=1}^{n} A_k^T A_k\right)\boldsymbol{\mu}_n = \frac{1}{n}\sum_{k=1}^{n} A_k^T \mathbf{y}_k.$$

Normal equations

- Mean estimator $\boldsymbol{\mu}_n$ satisfies

$$\left(\frac{1}{n}\sum_{k=1}^{n} A_k^T A_k\right)\boldsymbol{\mu}_n = \frac{1}{n}\sum_{k=1}^{n} A_k^T \mathbf{y}_k.$$

- For covariance, $L_n(\boldsymbol{\Sigma}_n) = B_n$, with $L_n : \mathbb{R}^{p\times p} \to \mathbb{R}^{p\times p}$ given by

$$L_n(\Sigma) = \frac{1}{n}\sum_{k=1}^{n} A_k^T A_k \Sigma A_k^T A_k$$

and

$$B_n = \frac{1}{n}\sum_{k=1}^{n} A_k^T \left(\mathbf{y}_k - A_k\boldsymbol{\mu}_n\right)\left(\mathbf{y}_k - A_k\boldsymbol{\mu}_n\right)^T A_k - \sigma^2 A_k^T A_k.$$

Normal equations

- Mean estimator $\boldsymbol{\mu}_n$ satisfies

$$\left(\frac{1}{n}\sum_{k=1}^{n} A_k^T A_k\right)\boldsymbol{\mu}_n = \frac{1}{n}\sum_{k=1}^{n} A_k^T \mathbf{y}_k.$$

- For covariance, $L_n(\boldsymbol{\Sigma}_n) = B_n$, with $L_n : \mathbb{R}^{p\times p} \to \mathbb{R}^{p\times p}$ given by

$$L_n(\Sigma) = \frac{1}{n}\sum_{k=1}^{n} A_k^T A_k \Sigma A_k^T A_k$$

and

$$B_n = \frac{1}{n}\sum_{k=1}^{n} A_k^T \left(\mathbf{y}_k - A_k\boldsymbol{\mu}_n\right)\left(\mathbf{y}_k - A_k\boldsymbol{\mu}_n\right)^T A_k - \sigma^2 A_k^T A_k.$$

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

Estimator properties

- Taking $A_k = \mathrm{I}_p$ and $\sigma^2 = 0$, recover sample mean and covariance.

[3] Katsevich et al (2015)

Estimator properties

- Taking $A_k = \mathrm{I}_p$ and $\sigma^2 = 0$, recover sample mean and covariance.
- Does not rely on a particular distribution of $\mathbf{x}$.

[3] Katsevich et al (2015)

Estimator properties

- Taking $A_k = \mathrm{I}_p$ and $\sigma^2 = 0$, recover sample mean and covariance.
- Does not rely on a particular distribution of $\mathbf{x}$.
- With prior information on $\mathbf{x}$, can regularize by adding terms to least-squares objective.

[3] Katsevich et al (2015)

Estimator properties

- Taking $A_k = \mathrm{I}_p$ and $\sigma^2 = 0$, recover sample mean and covariance.
- Does not rely on a particular distribution of $\mathbf{x}$.
- With prior information on $\mathbf{x}$, can regularize by adding terms to least-squares objective.
- Both $\boldsymbol{\mu}_n$ and Σ_n are consistent estimators[3]. For $n \to \infty$

$$\boldsymbol{\mu}_n \xrightarrow{a.s.} \mathbb{E}[\mathbf{x}]$$
$$\Sigma_n \xrightarrow{a.s.} \mathrm{Cov}[\mathbf{x}]$$

[3] Katsevich et al (2015)

High-dimensional PCA: Marčenko-Pastur

For pure Gaussian white noise of unit variance

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

we have sample covariance $\frac{1}{n}\sum_{k=1}^{n} \mathbf{y}_k \mathbf{y}_k^T$.

High-dimensional PCA: Marčenko-Pastur

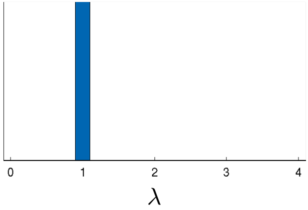
For pure Gaussian white noise of unit variance

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

we have sample covariance $\frac{1}{n}\sum_{k=1}^{n} \mathbf{y}_k \mathbf{y}_k^T$.

$n \gg p$

Central limit theorem



High-dimensional PCA: Marčenko-Pastur

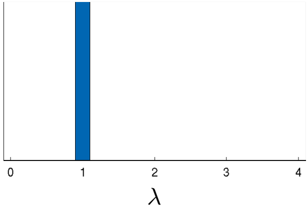
For pure Gaussian white noise of unit variance

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

we have sample covariance $\frac{1}{n}\sum_{k=1}^{n} \mathbf{y}_k \mathbf{y}_k^T$.

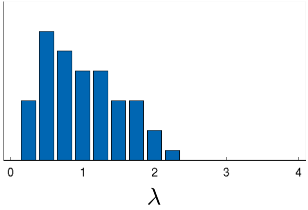
$n \gg p$

Central limit theorem



$n \asymp p$

Marčenko-Pastur law



High-dimensional PCA: spiked covariance model

Both signal and noise

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

[4] Johnstone (2001)

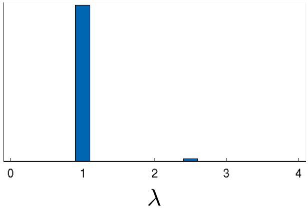
High-dimensional PCA: spiked covariance model

Both signal and noise

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

$n \gg p$

Central limit theorem



[4] Johnstone (2001)

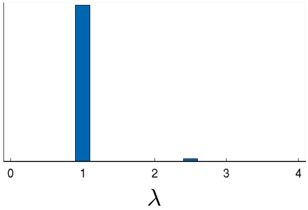
High-dimensional PCA: spiked covariance model

Both signal and noise

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

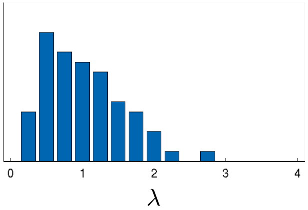
$n \gg p$

Central limit theorem



$n \asymp p$

Spiked covariance model[4]



[4] Johnstone (2001)

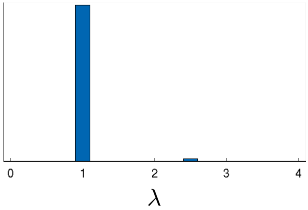
High-dimensional PCA: spiked covariance model

Both signal and noise

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

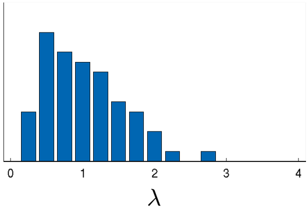
$n \gg p$

Central limit theorem



$n \asymp p$

Spiked covariance model[4]



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

[4] Johnstone (2001)

High-dimensional PCA: shrinkage

To estimate $\mathrm{Cov}[\mathbf{x}]$ 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 et al. (2013)

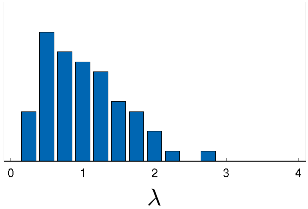
High-dimensional PCA: shrinkage

To estimate $\mathrm{Cov}[\mathbf{x}]$ from

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

when $n \asymp p$, we apply shrinkage to eigenvalues of sample covariance[5]

Original



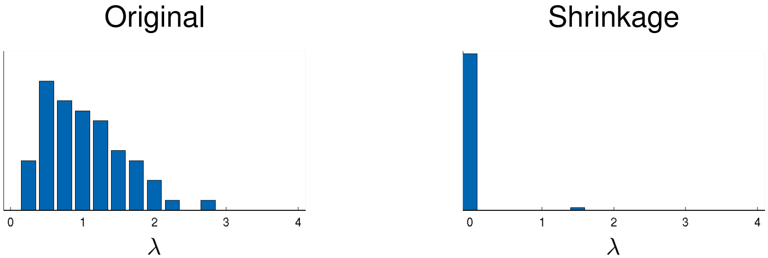
[5] Donoho et al. (2013)

High-dimensional PCA: shrinkage

To estimate $\mathrm{Cov}[\mathbf{x}]$ 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 et al. (2013)

Shrinkage of B_n

- In clean case, $\mathbf{y}_k = A_k \mathbf{x}_k$ and

$$B_n = \frac{1}{n} \sum_{k=1}^{n} A_k^T A_k \left(\mathbf{x}_k - \boldsymbol{\mu}_n\right) \left(\mathbf{x}_k - \boldsymbol{\mu}_n\right)^T A_k^T A_k,$$

Shrinkage of B_n

- In clean case, $\mathbf{y}_k = A_k \mathbf{x}_k$ and

$$B_n = \frac{1}{n} \sum_{k=1}^{n} A_k^T A_k \left(\mathbf{x}_k - \boldsymbol{\mu}_n\right) \left(\mathbf{x}_k - \boldsymbol{\mu}_n\right)^T A_k^T A_k,$$

so

$$\mathbb{E}[B_n] = \frac{1}{n} \sum_{k=1}^{n} \mathrm{Cov}[A_k^T A_k \mathbf{x}_k].$$

Shrinkage of B_n

- In clean case, $\mathbf{y}_k = A_k \mathbf{x}_k$ and

$$B_n = \frac{1}{n} \sum_{k=1}^{n} A_k^T A_k \left(\mathbf{x}_k - \boldsymbol{\mu}_n\right) \left(\mathbf{x}_k - \boldsymbol{\mu}_n\right)^T A_k^T A_k,$$

so

$$\mathbb{E}[B_n] = \mathrm{Cov}[\mathbf{A}_k^T \mathbf{A}_k \mathbf{x}_k]$$

if the A_k matrices are realizations of random matrices $\mathbf{A}_k$.

Shrinkage of B_n

- In clean case, $\mathbf{y}_k = A_k \mathbf{x}_k$ and

$$B_n = \frac{1}{n} \sum_{k=1}^{n} A_k^T A_k \left(\mathbf{x}_k - \boldsymbol{\mu}_n\right) \left(\mathbf{x}_k - \boldsymbol{\mu}_n\right)^T A_k^T A_k,$$

so

$$\mathbb{E}[B_n] = \mathrm{Cov}[\mathbf{A}_k^T \mathbf{A}_k \mathbf{x}_k]$$

if the A_k matrices are realizations of random matrices $\mathbf{A}_k$.

- For large n, we then have

$$B_n \approx \mathrm{Cov}[\mathbf{A}_k^T \mathbf{A}_k \mathbf{x}_k].$$

Shrinkage of B_n

- In clean case, $\mathbf{y}_k = A_k \mathbf{x}_k$ and

$$B_n = \frac{1}{n} \sum_{k=1}^{n} A_k^T A_k \left(\mathbf{x}_k - \boldsymbol{\mu}_n \right) \left(\mathbf{x}_k - \boldsymbol{\mu}_n \right)^T A_k^T A_k,$$

  so

$$\mathbb{E}[B_n] = \mathrm{Cov}[\mathbf{A}_k^T \mathbf{A}_k \mathbf{x}_k]$$

  if the A_k matrices are realizations of random matrices $\mathbf{A}_k$.

- For large n, we then have

$$B_n \approx \mathrm{Cov}[\mathbf{A}_k^T \mathbf{A}_k \mathbf{x}_k].$$

- How do we estimate $\mathrm{Cov}[\mathbf{A}_k^T \mathbf{A}_k \mathbf{x}_k]$ given noisy $\mathbf{y}_k$?

Shrinkage of B_n (cont.)

- Colored noise

$$\mathbf{A}_k^T \mathbf{y}_k = \mathbf{A}_k^T \mathbf{A}_k \mathbf{x}_k + \overbrace{\mathbf{A}_k^T \mathbf{e}_k}^{\text{covariance } \mathbb{E}[\mathbf{A}_k^T \mathbf{A}_k]}.$$

Shrinkage of B_n (cont.)

- Colored noise

$$\mathbf{A}_k^T \mathbf{y}_k = \mathbf{A}_k^T \mathbf{A}_k \mathbf{x}_k + \overbrace{\mathbf{A}_k^T \mathbf{e}_k}^{\text{covariance } \mathbb{E}[\mathbf{A}_k^T \mathbf{A}_k]} .$$

- 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

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

Shrinkage of B_n (cont.)

- Colored noise

$$\mathbf{A}_k^T \mathbf{y}_k = \mathbf{A}_k^T \mathbf{A}_k \mathbf{x}_k + \overbrace{\mathbf{A}_k^T \mathbf{e}_k}^{\text{covariance } \mathbb{E}[\mathbf{A}_k^T \mathbf{A}_k]} .$$

- 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

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

- Calculate sample covariance

Shrinkage of B_n (cont.)

- Colored noise

$$\mathbf{A}_k^T \mathbf{y}_k = \mathbf{A}_k^T \mathbf{A}_k \mathbf{x}_k + \overbrace{\mathbf{A}_k^T \mathbf{e}_k}^{\text{covariance } \mathbb{E}[\mathbf{A}_k^T \mathbf{A}_k]} .$$

- 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

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

- Calculate sample covariance

$$\sum_{k=1}^{n} \mathbf{z}_k \mathbf{z}_k^T$$

Shrinkage of B_n (cont.)

- Colored noise

$$\mathbf{A}_k^T \mathbf{y}_k = \mathbf{A}_k^T \mathbf{A}_k \mathbf{x}_k + \overbrace{\mathbf{A}_k^T \mathbf{e}_k}^{\text{covariance } \mathbb{E}[\mathbf{A}_k^T \mathbf{A}_k]}.$$

- 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

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

- Calculate sample covariance, shrink

$$\rho \left(\sum_{k=1}^{n} \mathbf{z}_k \mathbf{z}_k^T \right)$$

Shrinkage of B_n (cont.)

- Colored noise

$$\mathbf{A}_k^T \mathbf{y}_k = \mathbf{A}_k^T \mathbf{A}_k \mathbf{x}_k + \overbrace{\mathbf{A}_k^T \mathbf{e}_k}^{\text{covariance } \mathbb{E}[\mathbf{A}_k^T \mathbf{A}_k]}.$$

- 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

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

- Calculate sample covariance, shrink, and unwhiten

$$B_n = \mathbf{S}^{1/2} \rho \left(\frac{1}{n} \sum_{k=1}^{n} \mathbf{z}_k \mathbf{z}_k^T \right) \mathbf{S}^{1/2}.$$

Deconvolution in noise

- Recall noisy, filtered, image

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

given clean projection $\mathbf{x}_k$.

Deconvolution in noise

- Recall noisy, filtered, image

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

given clean projection $\mathbf{x}_k$.

- Using covariance estimation method, estimate $\mathrm{Cov}[\mathbf{x}_k]$.

Deconvolution in noise

- Recall noisy, filtered, image

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

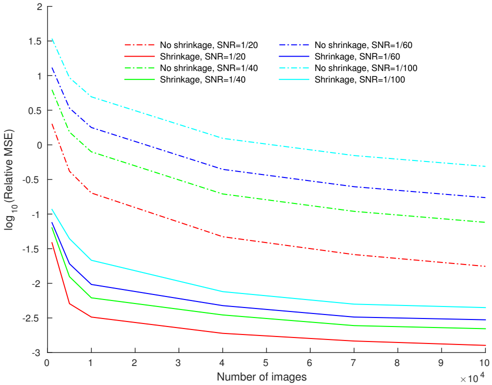
given clean projection $\mathbf{x}_k$.

- Using covariance estimation method, estimate $\mathrm{Cov}[\mathbf{x}_k]$.

- Construct generalized Wiener filter and invert T_k – estimate $\mathbf{x}_k$ from $\mathbf{y}_k$:

$$\Sigma_n T_k^T (T_k \Sigma_n T_k^T + \sigma^2 \mathrm{I}_q)^{-1} \mathbf{y}_k. \qquad (1)$$

Effect of shrinkage on simulated data

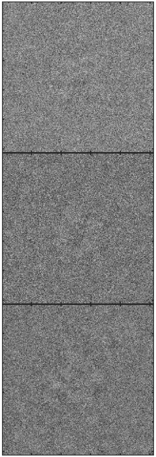


Applying shrinkage to right-hand side B_n in $L_n(\Sigma_n) = B_n$ results in significant increase in accuracy for the covariance estimation.

Experimental results: TRPV1 (EMPIAR-10005)

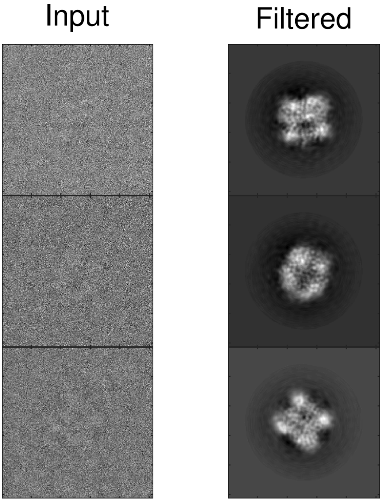
Dataset of 35645 projection images 256-by-256 pixels large.

Input



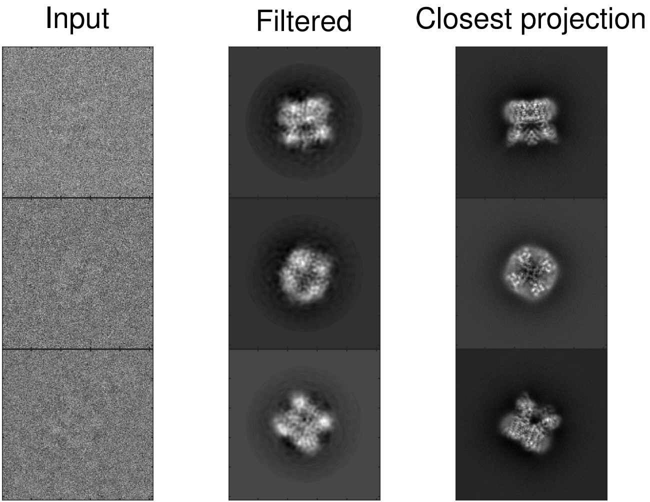
Experimental results: TRPV1 (EMPIAR-10005)

Dataset of 35645 projection images 256-by-256 pixels large.



Experimental results: TRPV1 (EMPIAR-10005)

Dataset of 35645 projection images 256-by-256 pixels large.



Heterogeneity problem

- Given images

$$\mathbf{y}_k = \overbrace{T_k P_k}^{A_k} \mathbf{x}_k + \mathbf{e}_k$$

classify $\mathbf{y}_k$ according to molecular state $\mathbf{x}_k$.

Heterogeneity problem

- Given images

$$\mathbf{y}_k = \overbrace{T_k P_k}^{A_k} \mathbf{x}_k + \mathbf{e}_k$$

classify $\mathbf{y}_k$ according to molecular state $\mathbf{x}_k$.

- We use proposed method to estimate covariance $\mathrm{Cov}[\mathbf{x}_k]$.

Heterogeneity problem

- Given images

$$\mathbf{y}_k = \overbrace{T_k P_k}^{A_k} \mathbf{x}_k + \mathbf{e}_k$$

  classify $\mathbf{y}_k$ according to molecular state $\mathbf{x}_k$.

- We use proposed method to estimate covariance $\mathrm{Cov}[\mathbf{x}_k]$.

- Reduce computational and memory costs by writing L_n as a convolution and preconditioning.

Heterogeneity problem

- Given images

$$\mathbf{y}_k = \overbrace{T_k P_k}^{A_k} \mathbf{x}_k + \mathbf{e}_k$$

classify $\mathbf{y}_k$ according to molecular state $\mathbf{x}_k$.

- We use proposed method to estimate covariance $\mathrm{Cov}[\mathbf{x}_k]$.

- Reduce computational and memory costs by writing L_n as a convolution and preconditioning.

- Dominant eigenvectors of $\mathbf{\Sigma}_n$ define low-dimensional subspace where volumes $\mathbf{x}_k$ live. For C classes, $\mathrm{rank}\ \mathrm{Cov}[\mathbf{x}_k] = C - 1$.

Heterogeneity problem

- Given images

$$\mathbf{y}_k = \overbrace{T_k P_k}^{A_k} \mathbf{x}_k + \mathbf{e}_k$$

classify $\mathbf{y}_k$ according to molecular state $\mathbf{x}_k$.

- We use proposed method to estimate covariance $\mathrm{Cov}[\mathbf{x}_k]$.

- Reduce computational and memory costs by writing L_n as a convolution and preconditioning.

- Dominant eigenvectors of $\mathbf{\Sigma}_n$ define low-dimensional subspace where volumes $\mathbf{x}_k$ live. For C classes, $\mathrm{rank}\ \mathrm{Cov}[\mathbf{x}_k] = C - 1$.

- We can invert $T_k P_k$ to find coordinates in subspace and cluster.

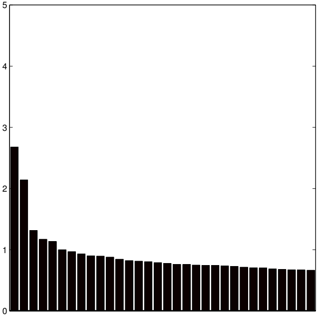
Experimental results: 70S ribosome

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

Experimental results: 70S ribosome

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

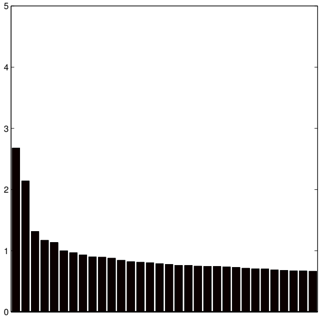
Largest 32 eigenvalues



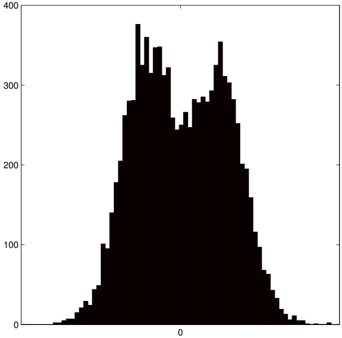
Experimental results: 70S ribosome

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

Largest 32 eigenvalues



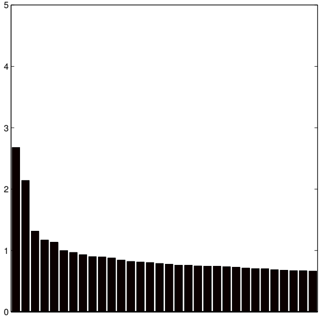
Coordinate histogram



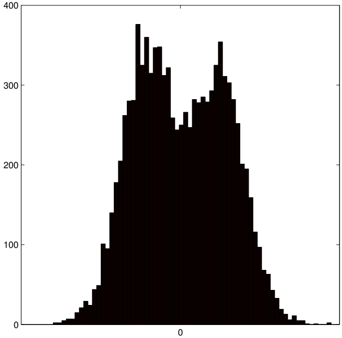
Experimental results: 70S ribosome

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

Largest 32 eigenvalues



Coordinate histogram

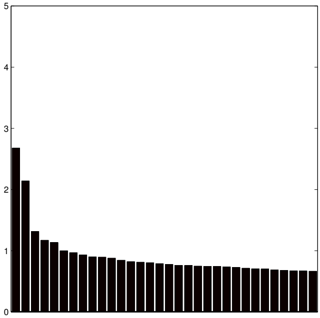


Clustering accuracy 89% with respect to dataset labeling.

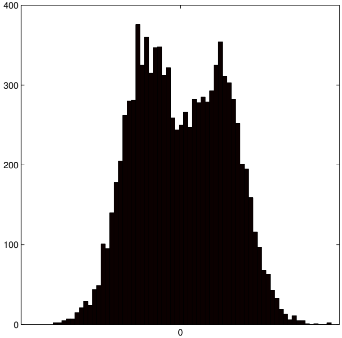
Experimental results: 70S ribosome

Dataset of 10000 images (130-by-130), 2 classes, courtesy Joachim Frank (Columbia University). Downsampled 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.

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.

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

- How to incorporate constraints on covariance Σ_n, such as sparsity.

Current and future work

- How to incorporate constraints on covariance Σ_n, such as sparsity.
- Understanding behavior of bulk noise distribution in Σ_n, not just in B_n.

Current and future work

- How to incorporate constraints on covariance Σ_n, such as sparsity.
- Understanding behavior of bulk noise distribution in Σ_n, not just in B_n.
- Application to other tasks, such as embryo development, magnetic resonance imaging, and matrix completion.

Bibliography

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.

T. Bhamre, T. Zhang, and A. Singer, “Denoising and covariance estimation of single particle cryo-em images,” *Journal of structural biology*, vol. 195, no. 1, pp. 72–81, 2016.

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.