Denoising and Covariance Estimation of Cryo-EM Images
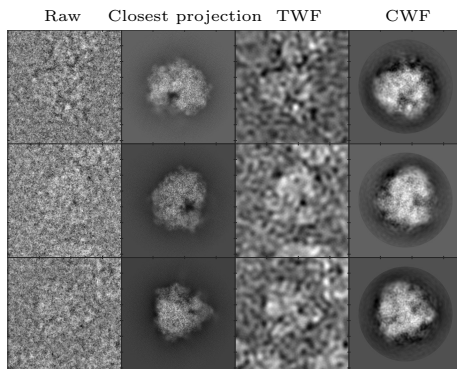
Tejal Bhamre

Princeton University
Joint work with Teng Zhang (UCF) and Amit Singer (Princeton)
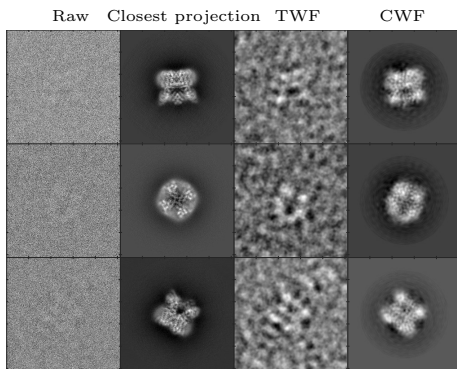
October 28, 2016

Experimental data - 80S ribosome

Raw   Closest projection   TWF   CWF

- FALCON II 4k×4k direct electron detector
- 105247 motion corrected, picked particle images of 360×360 pixels

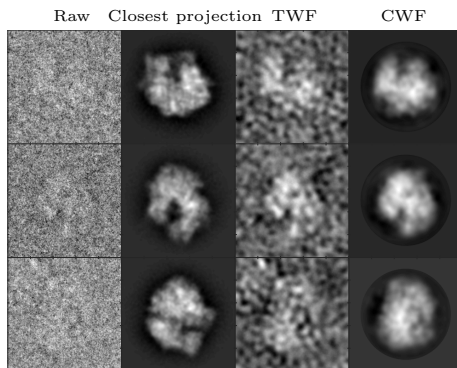# Experimental data - TRPV1

Raw   Closest projection   TWF   CWF

- K2 direct electron detector
- 35645 motion corrected, picked particle images of 256×256 pixels

Experimental data -IP$_3$R1

Raw Closest projection TWF CWF
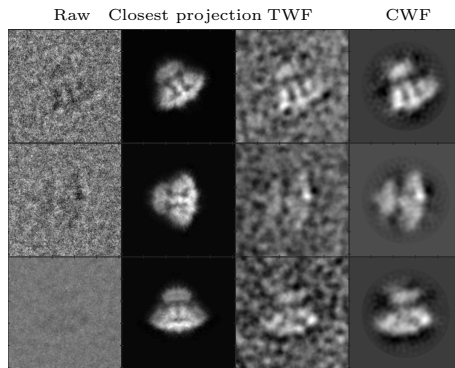
- Gatan 4k×4k CCD
- 37382 picked particle images of 256×256 pixels

Experimental data - 70S ribosome

- TVIPS TEMCAM-F415 (4k x 4k) CCD
- 216517 picked particle images of 250×250 pixels

Cryo-EM Pipeline

- Particle picking from micrographs

- Preprocessing

- 2D classification (Class averaging): inspect underlying particles, estimate viewing angles

- 3D classification

- 3D refinement

Cryo-EM Pipeline

- Particle picking from micrographs

- **Preprocessing**

- 2D classification (Class averaging): inspect underlying particles, estimate viewing angles

- 3D classification

- 3D refinement

Cryo-EM Pipeline

- Particle picking from micrographs
- **Preprocessing**
- **2D classification (Class averaging): inspect underlying particles, estimate viewing angles**
- 3D classification
- 3D refinement

Cryo-EM Pipeline

- Particle picking from micrographs
- **Preprocessing**
- **2D classification (Class averaging): inspect underlying particles, estimate viewing angles**
- 3D classification
- 3D refinement

Cryo-EM Pipeline

- Particle picking from micrographs
- **Preprocessing**
- **2D classification (Class averaging): inspect underlying particles, estimate viewing angles**
- 3D classification
- 3D refinement

Motivation

- Visualization of underlying particles without class averaging
- Image restoration (CTF correction and denoising) in a single step
- Outlier detection

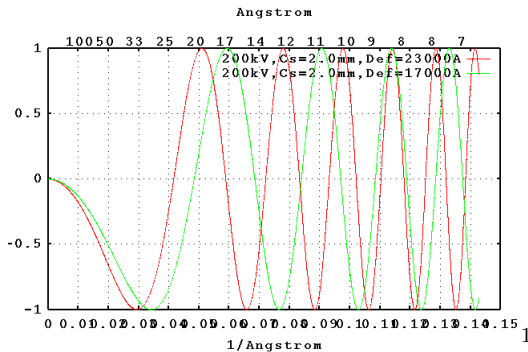# Motivation

- Visualization of underlying particles without class averaging
- Image restoration (CTF correction and denoising) in a single step
- Outlier detection

Motivation

- Visualization of underlying particles without class averaging
- Image restoration (CTF correction and denoising) in a single step
- Outlier detection

CTF Correction

- CTF suppresses information and inverts contrast
- Challenge: CTF cannot be trivially inverted (zero crossings)
- Information lost from one defocus group could be recovered from another group that has different zero crossings.

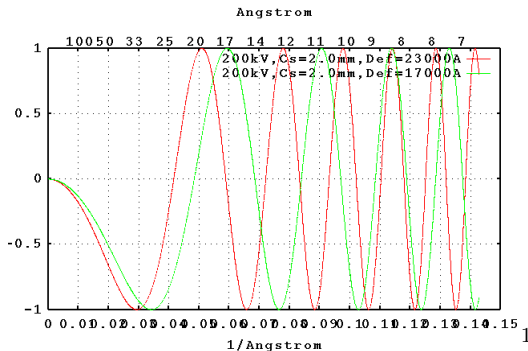[1] http://www.bio.brandeis.edu/~shaikh/lab/ctf.htm

CTF Correction

- CTF suppresses information and inverts contrast
- Challenge: CTF cannot be trivially inverted (zero crossings)
- Information lost from one defocus group could be recovered from another group that has different zero crossings.

[1] http://www.bio.brandeis.edu/~shaikh/lab/ctf.htm

CTF Correction

- CTF suppresses information and inverts contrast
- Challenge: CTF cannot be trivially inverted (zero crossings)
- Information lost from one defocus group could be recovered from another group that has different zero crossings.
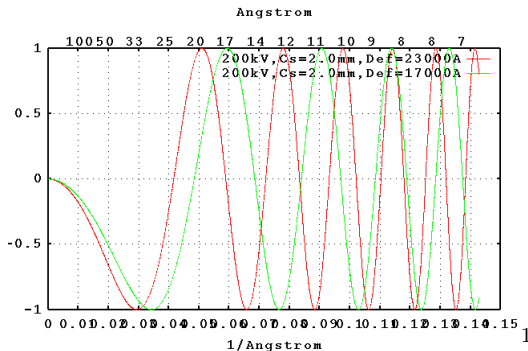
[1]http://www.bio.brandeis.edu/~shaikh/lab/ctf.htm

Current Image Restoration Techniques

- **Phase flipping + steerable PCA (sPCA):**
  - Flip sign of the Fourier coefficients at frequencies for which the CTF is negative
  - Preserves noise statistics
  - Data adaptive basis: eigenvectors of the sample covariance matrix
  - Phase flipping corrects only phases
- **Traditional Wiener Filtering (TWF):**
  - Corrects both phases and amplitudes
  - Requires prior estimation of the spectral signal to noise ratio (SSNR)
  - Cannot restore information at zero crossings of the CTF
  - Not in a data adaptive basis (restricted to Fourier basis)

Current Image Restoration Techniques

- **Phase flipping + steerable PCA (sPCA):**
  - Flip sign of the Fourier coefficients at frequencies for which the CTF is negative
  - Preserves noise statistics
  - Data adaptive basis: eigenvectors of the sample covariance matrix
  - Phase flipping corrects only phases
- **Traditional Wiener Filtering (TWF):**
  - Corrects both phases and amplitudes
  - Requires prior estimation of the spectral signal to noise ratio (SSNR)
  - Cannot restore information at zero crossings of the CTF
  - Not in a data adaptive basis (restricted to Fourier basis)

Current Image Restoration Techniques

- **Phase flipping + steerable PCA (sPCA):**
    - Flip sign of the Fourier coefficients at frequencies for which the CTF is negative
    - Preserves noise statistics
    - Data adaptive basis: eigenvectors of the sample covariance matrix
    - Phase flipping corrects only phases
- Traditional Wiener Filtering (TWF):
    - Corrects both phases and amplitudes
    - Requires prior estimation of the spectral signal to noise ratio (SSNR)
    - Cannot restore information at zero crossings of the CTF
    - Not in a data adaptive basis (restricted to Fourier basis)

Current Image Restoration Techniques

- **Phase flipping + steerable PCA (sPCA)**:
  - Flip sign of the Fourier coefficients at frequencies for which the CTF is negative
  - Preserves noise statistics
  - Data adaptive basis: eigenvectors of the sample covariance matrix
  - Phase flipping corrects only phases
- Traditional Wiener Filtering (TWF):
  - Corrects both phases and amplitudes
  - Requires prior estimation of the spectral signal to noise ratio (SSNR)
  - Cannot restore information at zero crossings of the CTF
  - Not in a data adaptive basis (restricted to Fourier basis)

Current Image Restoration Techniques

- **Phase flipping + steerable PCA (sPCA):**
  - Flip sign of the Fourier coefficients at frequencies for which the CTF is negative
  - Preserves noise statistics
  - Data adaptive basis: eigenvectors of the sample covariance matrix
  - Phase flipping corrects only phases
- Traditional Wiener Filtering (TWF):
  - Corrects both phases and amplitudes
  - Requires prior estimation of the spectral signal to noise ratio (SSNR)
  - Cannot restore information at zero crossings of the CTF
  - Not in a data adaptive basis (restricted to Fourier basis)

Current Image Restoration Techniques

- **Phase flipping + steerable PCA (sPCA)**:
  - Flip sign of the Fourier coefficients at frequencies for which the CTF is negative
  - Preserves noise statistics
  - Data adaptive basis: eigenvectors of the sample covariance matrix
  - Phase flipping corrects only phases

- **Traditional Wiener Filtering (TWF)**:
  - Corrects both phases and amplitudes
  - Requires prior estimation of the spectral signal to noise ratio (SSNR)
  - Cannot restore information at zero crossings of the CTF
  - Not in a data adaptive basis (restricted to Fourier basis)

Current Image Restoration Techniques

- **Phase flipping + steerable PCA (sPCA)**:
  - Flip sign of the Fourier coefficients at frequencies for which the CTF is negative
  - Preserves noise statistics
  - Data adaptive basis: eigenvectors of the sample covariance matrix
  - Phase flipping corrects only phases

- **Traditional Wiener Filtering (TWF)**:
  - Corrects both phases and amplitudes
  - Requires prior estimation of the spectral signal to noise ratio (SSNR)
  - Cannot restore information at zero crossings of the CTF
  - Not in a data adaptive basis (restricted to Fourier basis)

Current Image Restoration Techniques

- **Phase flipping + steerable PCA (sPCA)**:
    - Flip sign of the Fourier coefficients at frequencies for which the CTF is negative
    - Preserves noise statistics
    - Data adaptive basis: eigenvectors of the sample covariance matrix
    - Phase flipping corrects only phases
- **Traditional Wiener Filtering (TWF)**:
    - Corrects both phases and amplitudes
    - Requires prior estimation of the spectral signal to noise ratio (SSNR)
    - Cannot restore information at zero crossings of the CTF
    - Not in a data adaptive basis (restricted to Fourier basis)

Current Image Restoration Techniques

- **Phase flipping + steerable PCA (sPCA)**:
  - Flip sign of the Fourier coefficients at frequencies for which the CTF is negative
  - Preserves noise statistics
  - Data adaptive basis: eigenvectors of the sample covariance matrix
  - Phase flipping corrects only phases
- **Traditional Wiener Filtering (TWF)**:
  - Corrects both phases and amplitudes
  - Requires prior estimation of the spectral signal to noise ratio (SSNR)
  - Cannot restore information at zero crossings of the CTF
  - Not in a data adaptive basis (restricted to Fourier basis)

Current Image Restoration Techniques

- **Phase flipping + steerable PCA (sPCA)**:
  - Flip sign of the Fourier coefficients at frequencies for which the CTF is negative
  - Preserves noise statistics
  - Data adaptive basis: eigenvectors of the sample covariance matrix
  - Phase flipping corrects only phases
- **Traditional Wiener Filtering (TWF)**:
  - Corrects both phases and amplitudes
  - Requires prior estimation of the spectral signal to noise ratio (SSNR)
  - Cannot restore information at zero crossings of the CTF
  - Not in a data adaptive basis (restricted to Fourier basis)

Covariance Wiener Filtering (CWF)

- Estimate the CTF-corrected covariance matrix of the underlying clean 2D projection images
- Wiener filtering to solve the image restoration deconvolution problem
- No averaging, act on each image separately
- CTF correction and denoising in a single step

Covariance Wiener Filtering (CWF)

- Estimate the CTF-corrected covariance matrix of the underlying clean 2D projection images
- Wiener filtering to solve the image restoration deconvolution problem
- No averaging, act on each image separately
- CTF correction and denoising in a single step

Covariance Wiener Filtering (CWF)

- Estimate the CTF-corrected covariance matrix of the underlying clean 2D projection images
- Wiener filtering to solve the image restoration deconvolution problem
- No averaging, act on each image separately
- CTF correction and denoising in a single step

Covariance Wiener Filtering (CWF)

- Estimate the CTF-corrected covariance matrix of the underlying clean 2D projection images
- Wiener filtering to solve the image restoration deconvolution problem
- No averaging, act on each image separately
- CTF correction and denoising in a single step

Covariance Wiener Filtering (CWF)

Table : Comparison of CTF Correction/Denoising Methods

Property	Phaseflip + sPCA	TWF	CWF
Applicable at preliminary stage	✓	✓	✓
Data dependent basis	✓	✗	✓
Correct both phases and amplitudes	✗	✓	✓
CTF corrected covariance estimate	✗	✗	✓

Computational Complexity: Timings

n images of size $L \times L$

UNIX environment with 60 cores, running at 2.3 GHz, with total RAM of 1.5TB

Table : Timing in seconds

Dataset	L	n	Basis coeffs	CWF
TRPV1	256	35645	312s	574s
80s	360	30000	731s	385s
IP3R1	256	37382	429s	589s
70s	250	99979	1174s	113s

Relative error of estimated clean images

(a)

(b)
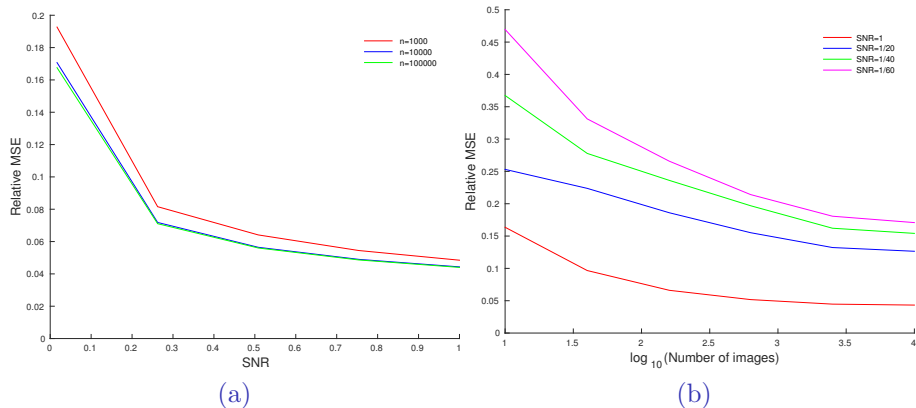
Figure : (a) Fixed number of images (b) Fixed SNR
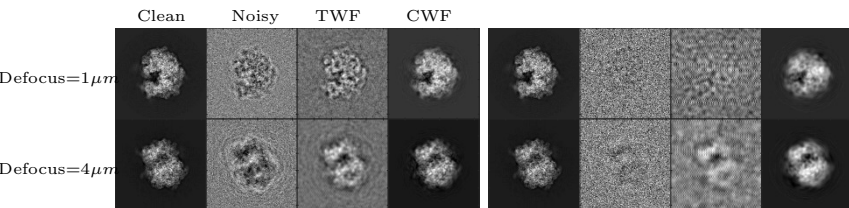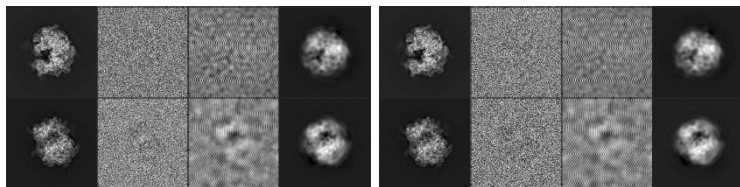
Relative error of estimated covariance
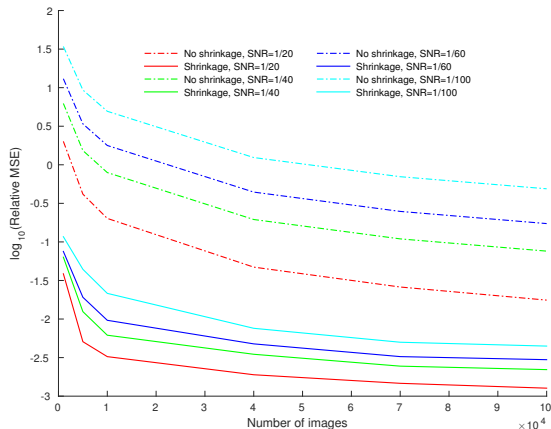
Figure : The estimator $\hat{\Sigma}$ can be shown to be consistent in the large sample limit $n \to \infty$

Simulations with white noise: 80S ribosome (EMDB-6454)

(a) SNR=1

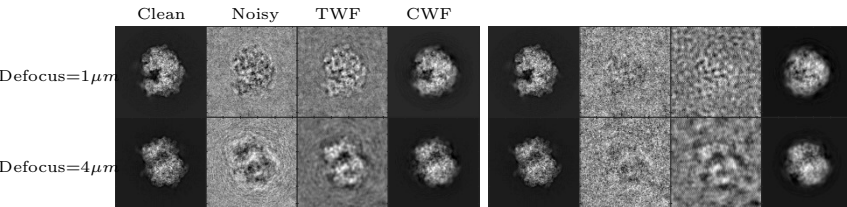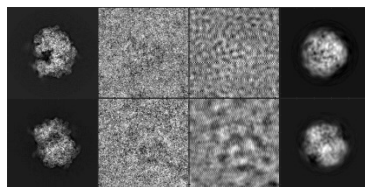(b) SNR=1/20

(c) SNR=1/40

(d) SNR=1/60

Simulations with colored noise: 80S ribosome (EMDB-6454)

Clean   Noisy   TWF   CWF

Defocus=$1\mu m$

Defocus=$4\mu m$

(e) SNR=1

(f) SNR=1/10

(g) SNR=1/20

Outlier Detection

- Significant amount of time is spent on discarding outliers by visual inspection after the particle picking

- CWF: automatic way to classify picked particles

- Specimen particles at various depths in the ice layer: acquired projection images can have different contrasts

- Outlier images typically have low contrast after denoising: linear classifier after CWF

Outlier Detection

- Significant amount of time is spent on discarding outliers by visual inspection after the particle picking

- CWF: automatic way to classify picked particles

- Specimen particles at various depths in the ice layer: acquired projection images can have different contrasts

- Outlier images typically have low contrast after denoising: linear classifier after CWF

Outlier Detection

- Significant amount of time is spent on discarding outliers by visual inspection after the particle picking
- CWF: automatic way to classify picked particles
- Specimen particles at various depths in the ice layer: acquired projection images can have different contrasts
- Outlier images typically have low contrast after denoising: linear classifier after CWF

Outlier Detection

- Significant amount of time is spent on discarding outliers by visual inspection after the particle picking
- CWF: automatic way to classify picked particles
- Specimen particles at various depths in the ice layer: acquired projection images can have different contrasts
- Outlier images typically have low contrast after denoising: linear classifier after CWF

Outlier Detection: 80S ribosome (EMDB-6454)

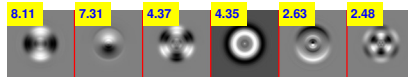SNR=1/20 $\alpha \in [0.75, 1.5]$ 10% images are pure noise
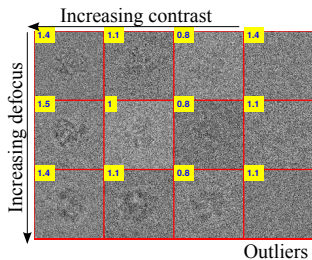
(h)
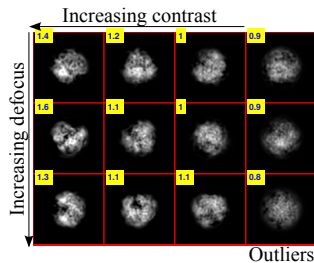
(i)

(j)

(k)

Applications

- Better class averages

- Covariance matrix: Orthogonal Replacement using Kam's method

- 3D reconstruction from denoised images, without class averaging

Applications

- Better class averages
- Covariance matrix: Orthogonal Replacement using Kam's method
- 3D reconstruction from denoised images, without class averaging

Applications

- Better class averages
- Covariance matrix: Orthogonal Replacement using Kam's method
- 3D reconstruction from denoised images, without class averaging

Resources

- Code: `https://github.com/PrincetonUniversity/cwf_denoise`
- Paper: Journal of Structural Biology:
  `10.1016/j.jsb.2016.04.013`

Pipeline

- Preprocessing
- Mean estimation
- Covariance Estimation
- Deconvolution

The Model: Real space

Linear, weak phase approximation

$$y_i = a_i * x_i + \epsilon_i, \quad i = 1, 2, \ldots, n \tag{1}$$

n: number of images
$*$: convolution operation
y_i: noisy, CTF filtered i'th image in real space
x_i: underlying clean projection image in real space
a_i: the point spread function of the microscope
ϵ_i: additive Gaussian noise that corrupts the image

The Model: Fourier space

$$Y_i = A_i X_i + \xi_i, \quad i = 1, 2, \ldots, n \tag{2}$$

A_i: diagonal operator, whose diagonal consists of the Fourier transform of the point spread function

$X_1, \ldots, X_n$: vectors in $\mathbb{C}^p$, where p is the number of pixels
i.i.d. samples from a distribution with mean $\mathbb{E}[\mathbf{X}] = \mu$ and covariance
$\mathbb{E}[(\mathbf{X} - \mu)(\mathbf{X} - \mu)^T] = \Sigma$

"All models are wrong but some are useful" - George Box

The Model

$$\mathbb{E}[\mathbf{Y}_i] = A_i \mathbb{E}[\mathbf{X}_i], \quad i = 1, 2, \ldots, n. \tag{3}$$

$$\mathbb{E}[(\mathbf{Y}_i - \mathbb{E}[\mathbf{Y}_i])(\mathbf{Y}_i - \mathbb{E}[\mathbf{Y}_i])^T] = \mathbb{E}[A_i(\mathbf{X}_i - \mu)(\mathbf{X}_i - \mu)^T A_i^T] + \sigma^2 I$$
$$= A_i \Sigma A_i^T + \sigma^2 I. \tag{4}$$

Relates the second order statistics of the noisy images with the population covariance Σ of the clean images

Mean Estimation

$$\hat{\mu} = \arg\min_{\mu} \sum_{i=1}^{n} ||(Y_i - A_i\mu)||_2^2 + \lambda||\mu||_2^2 \tag{5}$$

$$\hat{\mu} = (\sum_{i=1}^{n} A_i^T A_i + \lambda I)^{-1}(\sum_{i=1}^{n} A_i^T Y_i). \tag{6}$$

Covariance Estimation

$$\hat{\Sigma} = \arg\min_{\Sigma} \sum_{i=1}^{n} ||(Y_i - \mathbb{E}[\mathbf{Y}_i])(Y_i - \mathbb{E}[\mathbf{Y}_i])^T - (A_i \Sigma A_i^T + \sigma^2 I)||_F^2$$

$$= \arg\min_{\Sigma} \sum_{i=1}^{n} ||A_i \Sigma A_i^T + \sigma^2 I - C_i||_F^2$$

(7)

where $C_i = (Y_i - A_i \mu)(Y_i - A_i \mu)^T$ and $||.||_F$ is the Frobenius matrix norm.

Solving using Conjugate Gradient

System of linear equations for the elements of the matrix $\hat{\Sigma}$

$$\sum_{i=1}^{n} A_i^T A_i \hat{\Sigma} A_i^T A_i = \sum_{i=1}^{n} A_i^T C_i A_i - \sum_{i=1}^{n} \sigma^2 A_i^T A_i \tag{8}$$

$$L(\hat{\Sigma}) = B \tag{9}$$

where $L : \mathbb{R}^{p \times p} \to \mathbb{R}^{p \times p}$ is the linear operator acting on $\hat{\Sigma}$ defined by the left hand side of eqn. 8, and B is the right hand side.

- Direct inversion of this linear system is slow for large image sizes
- Applying L only involves matrix multiplications: fast!
- Conjugate gradient

Wiener Filtering

- White noise: estimate X_i as

$$\hat{X}_i = (I - H_i A_i)\hat{\mu} + H_i Y_i \qquad (10)$$

  where $H_i = \hat{\Sigma} A_i^T (A_i \hat{\Sigma} A_i^T + \sigma^2 I)^{-1}$ is the linear Wiener filter

- Colored noise: estimate X_i as

$$\hat{X}_i = (I - H_i W A_i)\hat{\mu} + H_i Y_i \qquad (11)$$

  with $H_i = \hat{\Sigma} A_i^T W^T (W A_i \hat{\Sigma} A_i^T W^T + \sigma^2 I)^{-1}$

Acknowledgement