Data-dependent Confidence Region of Singular Subspaces

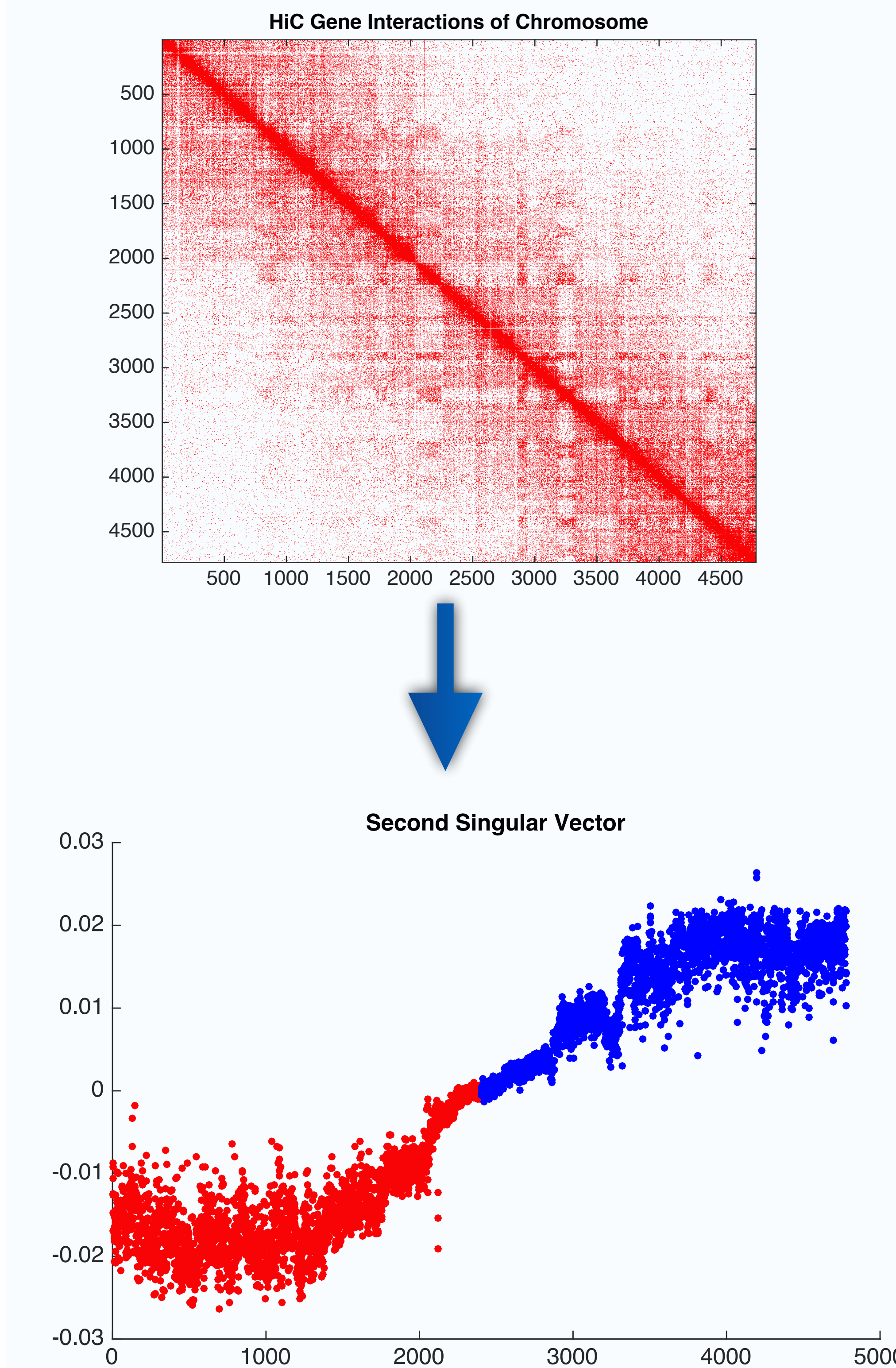
Dong Xia

Department of Mathematics
Hong Kong University of Science and Technology

Spectral Methods for Matrix Data Analysis

Example Applications

- Dimension Reduction
- Network Analysis
- Principal Component Analysis
- Classification and Clustering
- Matrix and Tensor Completion
- Low Rank Matrix Denoising



Dong Xia, HKUST

Preliminary

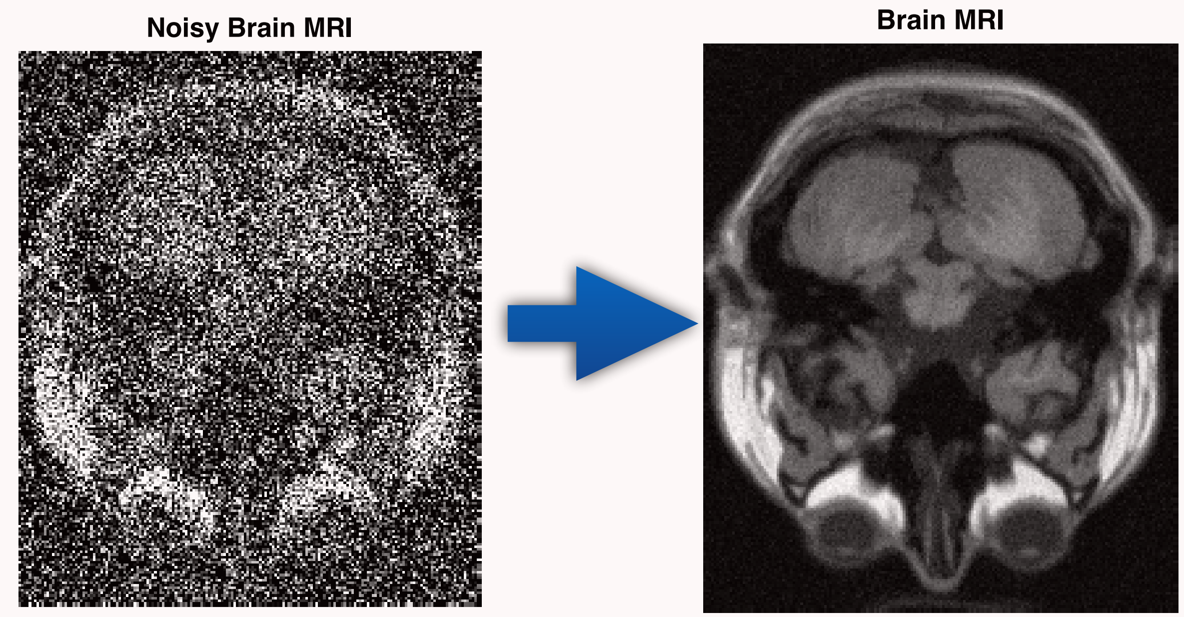
Spectral Methods for Matrix Data Analysis

Matrix Completion

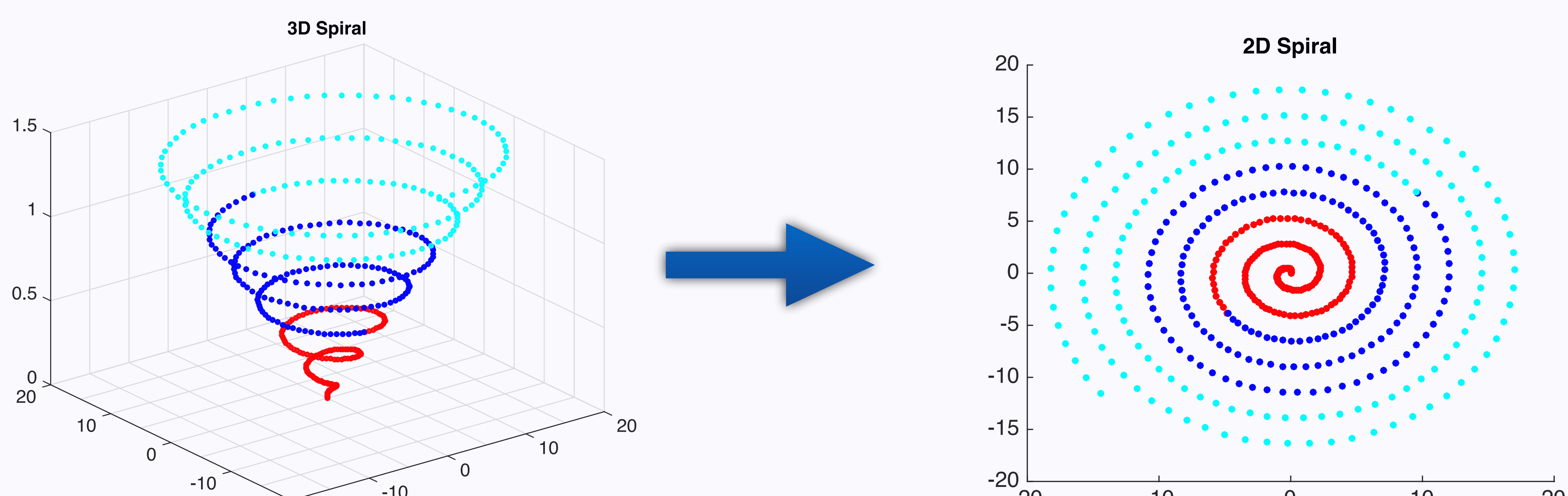
Movie

User					
	1		?	2	?
		?			4
	?	4.5		?	
		?	1.5		
	3.5				?

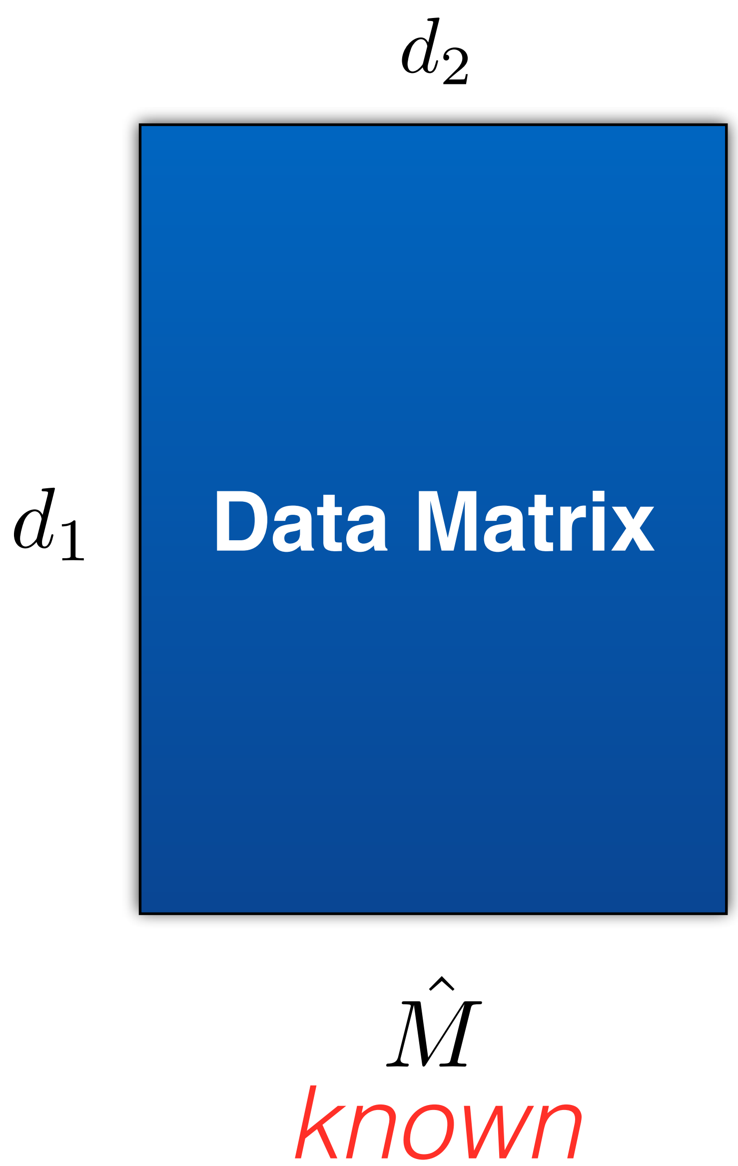
Image Denoising



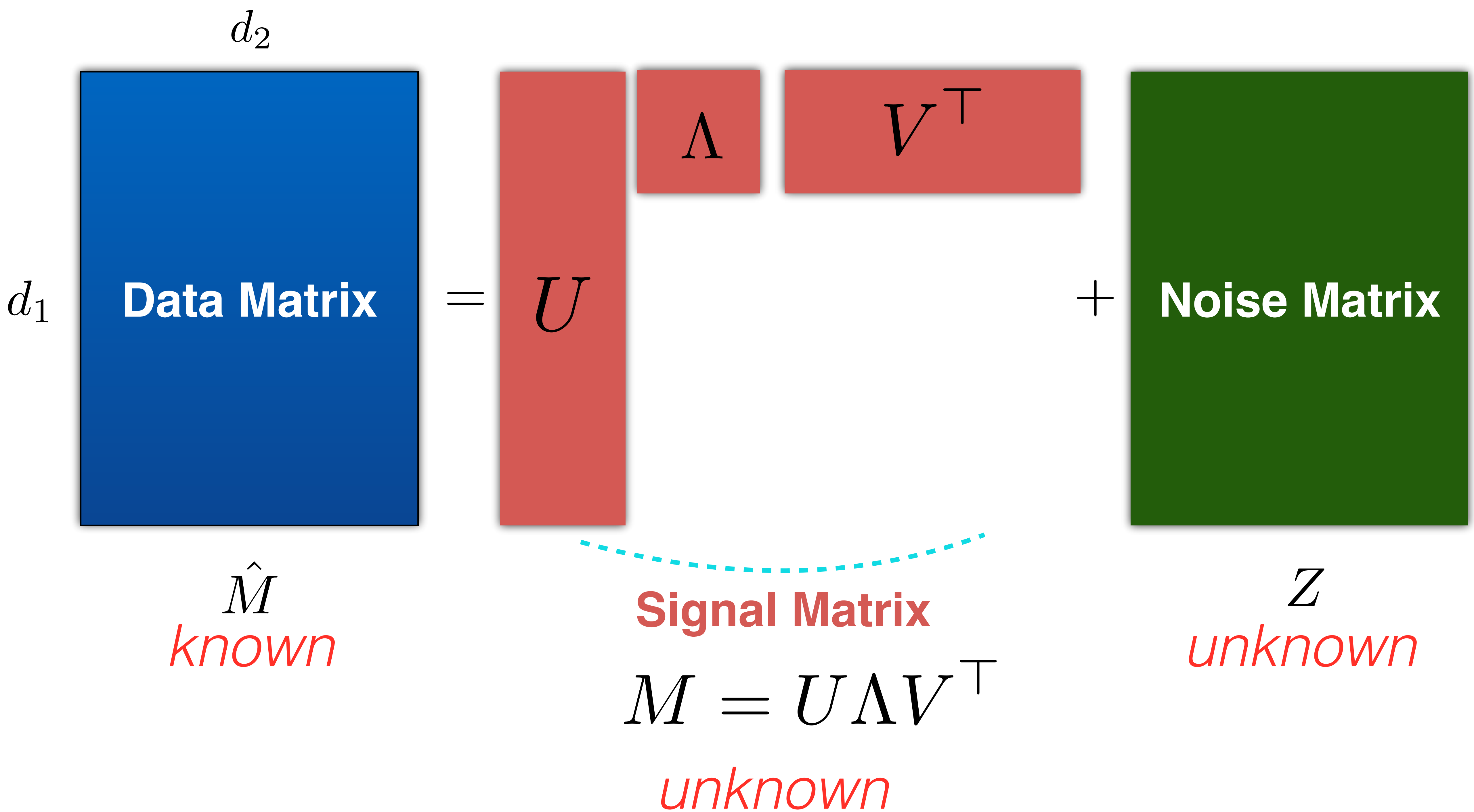
Dimension Reduction



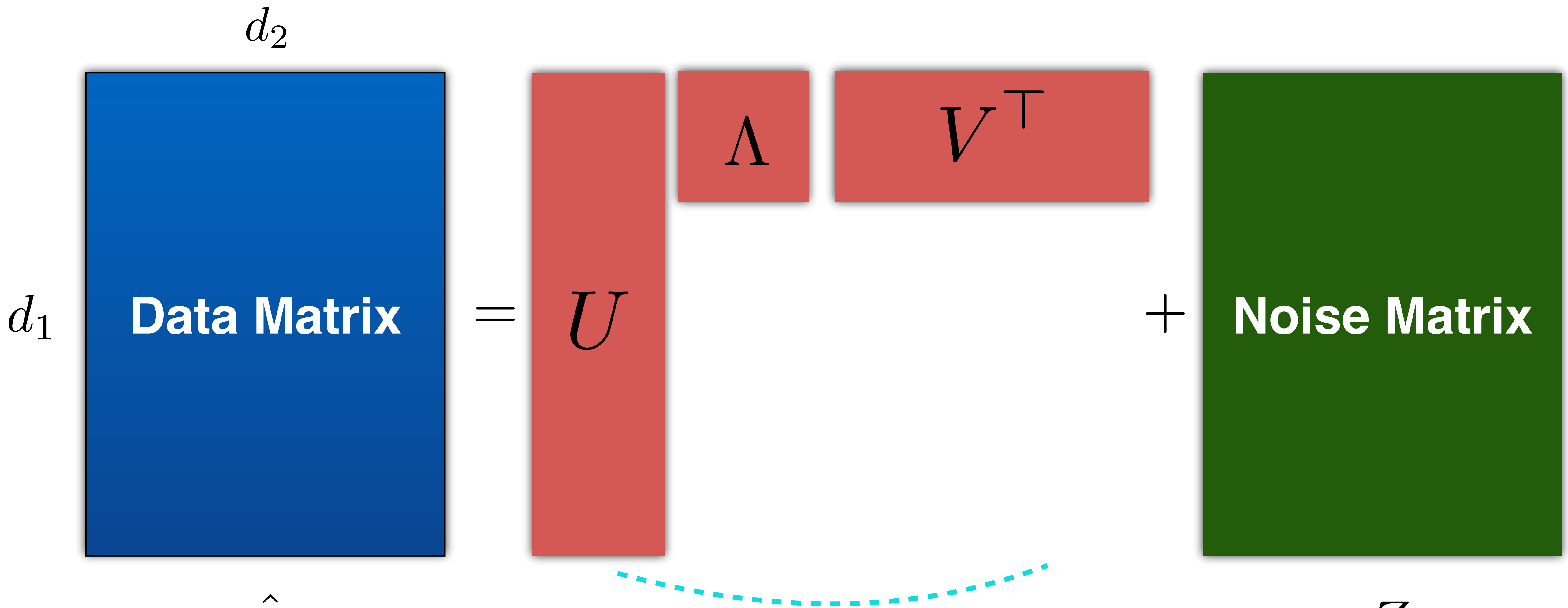
Matrix Perturbation with Noise



Matrix Perturbation with Noise



Matrix Perturbation with Noise



$\hat{M}$ known

Signal Matrix

$$M = U \Lambda V^\top$$

unknown

Z unknown

Singular Vectors: $U = (u_1, \cdots, u_r) \in \mathbb{R}^{d_1 \times r}$ and $V = (v_1, \cdots, v_r) \in \mathbb{R}^{d_2 \times r}$

Singular Values: $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_r)$

Deterministic Perturbations of SVD

Goal: given data matrix $\hat{\mathbf{M}}$, estimate $\mathbf{U}$ and $\mathbf{V}$

Deterministic Perturbations of SVD

Goal: given data matrix $\hat{\mathbf{M}}$, estimate $\mathbf{U}$ and $\mathbf{V}$

Plug-in Estimators: compute $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}$
top-r left and right singular vectors of $\hat{\mathbf{M}}$

Deterministic Perturbations of SVD

Goal: given data matrix $\hat{\mathbf{M}}$, estimate $\mathbf{U}$ and $\mathbf{V}$

Plug-in Estimators: compute $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}$

top-r left and right singular vectors of $\hat{\mathbf{M}}$

Davis - Kahan Theorem:

$$\max\left\{\|\hat{\mathbf{U}}\hat{\mathbf{U}}^\top - \mathbf{U}\mathbf{U}^\top\|_{\mathrm{F}}^2, \|\hat{\mathbf{V}}\hat{\mathbf{V}}^\top - \mathbf{V}\mathbf{V}^\top\|_{\mathrm{F}}^2\right\} \leq \frac{8r\|\mathbf{Z}\|^2}{\lambda_r^2}$$

if $\lambda_r > 2\|\mathbf{Z}\|$

Deterministic Perturbations of SVD

Goal: given data matrix $\hat{\mathbf{M}}$, estimate $\mathbf{U}$ and $\mathbf{V}$

Plug-in Estimators: compute $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}$
top-r left and right singular vectors of $\hat{\mathbf{M}}$

Davis - Kahan Theorem:

$$\max\left\{\|\hat{\mathbf{U}}\hat{\mathbf{U}}^\top - \mathbf{U}\mathbf{U}^\top\|_{\mathrm{F}}^2, \|\hat{\mathbf{V}}\hat{\mathbf{V}}^\top - \mathbf{V}\mathbf{V}^\top\|_{\mathrm{F}}^2\right\} \leq \boxed{\frac{8r\|\mathbf{Z}\|^2}{\lambda_r^2}}$$

if $\lambda_r > 2\|\mathbf{Z}\|$

deterministic bound

Dong Xia, HKUST

Preliminary

Deterministic Perturbations of SVD

Goal: given data matrix $\hat{\mathbf{M}}$, estimate $\mathbf{U}$ and $\mathbf{V}$

Plug-in Estimators: compute $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}$

top-r left and right singular vectors of $\hat{\mathbf{M}}$

Davis - Kahan Theorem:

$$\max\left\{\|\hat{\mathbf{U}}\hat{\mathbf{U}}^\top - \mathbf{U}\mathbf{U}^\top\|_{\mathrm{F}}^2, \|\hat{\mathbf{V}}\hat{\mathbf{V}}^\top - \mathbf{V}\mathbf{V}^\top\|_{\mathrm{F}}^2\right\} \le \boxed{\frac{8r\|\mathbf{Z}\|^2}{\lambda_r^2}}$$

operator norm

if $\lambda_r > 2\|\mathbf{Z}\|$

deterministic bound

Deterministic Perturbations of SVD

Goal: given data matrix $\hat{\mathbf{M}}$, estimate $\mathbf{U}$ and $\mathbf{V}$

Plug-in Estimators: compute $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}$
top-r left and right singular vectors of $\hat{\mathbf{M}}$

Davis - Kahan Theorem:

$$\max\left\{\|\hat{\mathbf{U}}\hat{\mathbf{U}}^\top - \mathbf{U}\mathbf{U}^\top\|_{\mathrm{F}}^2, \|\hat{\mathbf{V}}\hat{\mathbf{V}}^\top - \mathbf{V}\mathbf{V}^\top\|_{\mathrm{F}}^2\right\} \le \boxed{\frac{8r\|\mathbf{Z}\|^2}{\lambda_r^2}}$$

operator norm

$$\text{if } \lambda_r > 2\|\mathbf{Z}\|$$

deterministic bound

signal strength

Statistical Inference of Singular Subspaces

An deterministic upper bound for

$$\max\left\{\|\hat{\mathbf{U}}\hat{\mathbf{U}}^\top - \mathbf{U}\mathbf{U}^\top\|_{\mathrm{F}}^2, \|\hat{\mathbf{V}}\hat{\mathbf{V}}^\top - \mathbf{V}\mathbf{V}^\top\|_{\mathrm{F}}^2\right\}$$

is not enough for ***statisticians***!

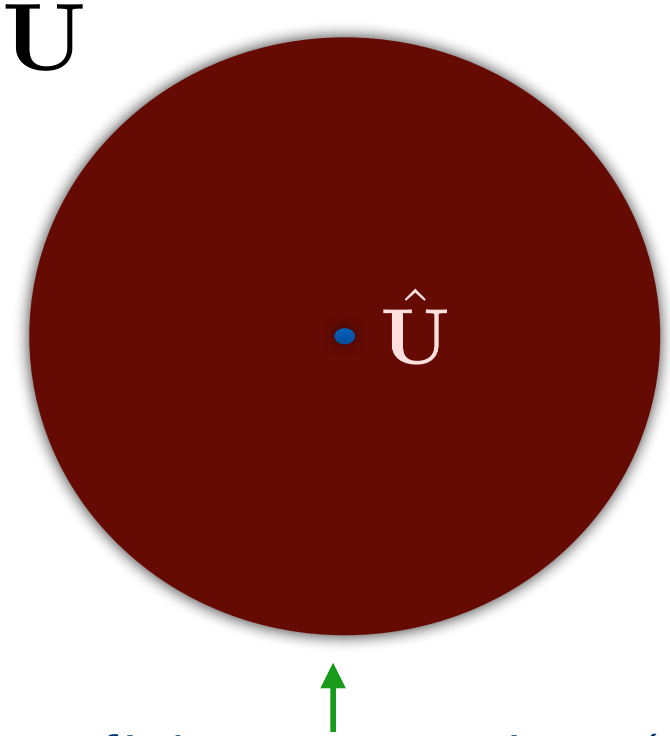
Statistical Inference of Singular Subspaces

An deterministic upper bound for

$$\max\left\{\|\hat{\mathbf{U}}\hat{\mathbf{U}}^\top - \mathbf{U}\mathbf{U}^\top\|_{\mathrm{F}}^2, \|\hat{\mathbf{V}}\hat{\mathbf{V}}^\top - \mathbf{V}\mathbf{V}^\top\|_{\mathrm{F}}^2\right\}$$

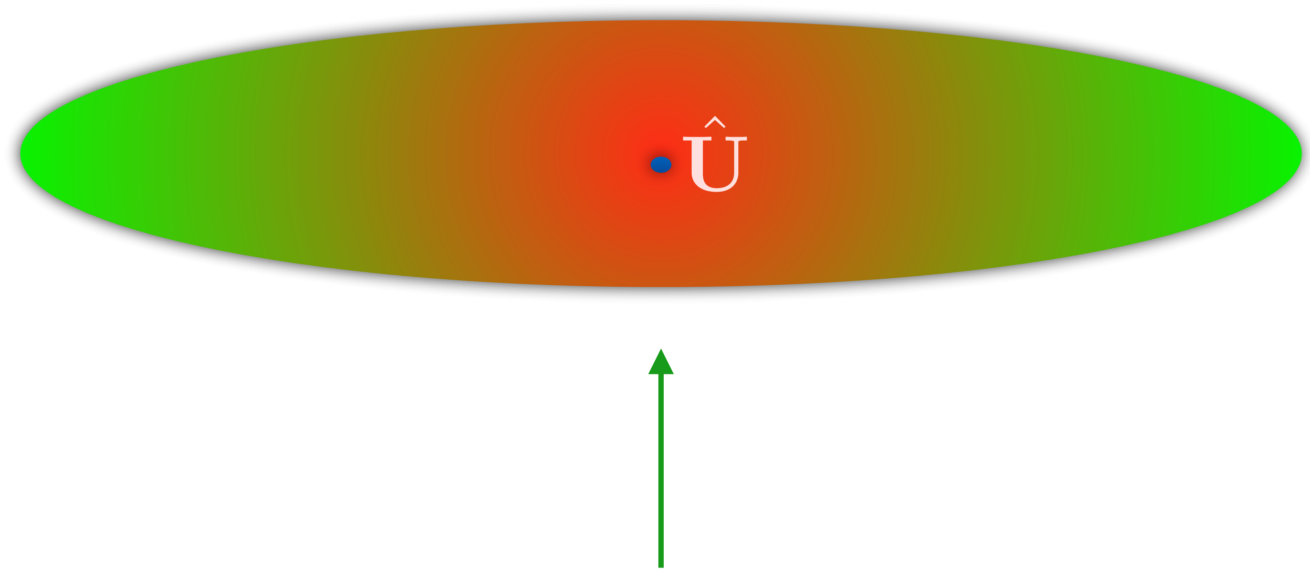
is not enough for ***statisticians****!*

CR of $\mathbf{U}$



confidence region (CR)
by deterministic bound

CR of $\mathbf{U}$



confidence region (CR)
statisticians *desired*

Statistical Model for Noise Matrix

Model: $\hat{M} = M + Z \text{ with } M = U\Lambda V^\top \in \mathbb{R}^{d_1 \times d_2}$

Statistical Model for Noise Matrix

Model: $\hat{M} = M + Z \text{ with } M = U\Lambda V^\top \in \mathbb{R}^{d_1 \times d_2}$

Assumption:

$$Z = (z_{ij})_{1\le i\le d_1}^{1\le j\le d_2}$$

$$\mathbf{z}_{ij} \text{ are i.i.d. and } \mathbf{z}_{ij} \sim \mathcal{N}(0,1)$$

Statistical Model for Noise Matrix

Model: $\hat{M} = M + Z \text{ with } M = U\Lambda V^\top \in \mathbb{R}^{d_1 \times d_2}$

Assumption: $Z = (z_{ij})_{1\le i\le d_1}^{1\le j\le d_2}$

$\mathbf{z}_{ij}$ are i.i.d. and $\mathbf{z}_{ij} \sim \mathcal{N}(0,1)$

Loss:

$$\text{dist}^2\left[(\hat{\mathbf{U}}, \hat{\mathbf{V}}), (\mathbf{U}, \mathbf{V})\right]$$

$$= \|\hat{\mathbf{U}}\hat{\mathbf{U}}^\top - \mathbf{U}\mathbf{U}^\top\|_{\mathrm{F}}^2 + \|\hat{\mathbf{V}}\hat{\mathbf{V}}^\top - \mathbf{V}\mathbf{V}^\top\|_{\mathrm{F}}^2$$

Dong Xia, HKUST

Random Perturbation of Singular Subspace

On $\|\mathbf{Z}\|$: $\|\mathbf{Z}\|$ is sub-gaussian with

$$\mathbb{E}\|Z\| = O(\bar{d}^{1/2}) \text{ with } \bar{d} = \max(d_1, d_2)$$

Random Perturbation of Singular Subspace

On $\|\mathbf{Z}\|$:

$$\|\mathbf{Z}\| \text{ is sub-gaussian with}$$
$$\mathbb{E}\|Z\| = O(\bar{d}^{1/2}) \text{ with } \bar{d} = \max(d_1, d_2)$$

By Davis - Kahan Bound:

$$\mathbb{E}\ \mathrm{dist}^2\big[(\hat{\mathbf{U}}, \hat{\mathbf{V}}), (\mathbf{U}, \mathbf{V})\big]$$
$$\lesssim \frac{r \cdot \mathbb{E}\|\mathbf{Z}\|^2}{\lambda_r^2}$$

Random Perturbation of Singular Subspace

On $\|\mathbf{Z}\|$:

$$\|\mathbf{Z}\| \text{ is sub-gaussian with}$$
$$\mathbb{E}\|Z\| = O(\bar{d}^{1/2}) \text{ with } \bar{d} = \max(d_1, d_2)$$

By Davis - Kahan Bound:

$$\mathbb{E}\ \mathrm{dist}^2\big[(\hat{\mathbf{U}}, \hat{\mathbf{V}}), (\mathbf{U}, \mathbf{V})\big]$$
$$\lesssim \frac{r \cdot \mathbb{E}\|\mathbf{Z}\|^2}{\lambda_r^2}$$
$$\lesssim \frac{r \cdot \bar{d}}{\lambda_r^2}$$

if $\lambda_r \gtrsim \bar{d}^{1/2}$

Random Perturbation of Singular Subspace

On $\|\mathbf{Z}\|$:

$$\|\mathbf{Z}\| \text{ is sub-gaussian with}$$
$$\mathbb{E}\|Z\| = O(\bar{d}^{1/2}) \text{ with } \bar{d} = \max(d_1, d_2)$$

By Davis - Kahan Bound:

$$\mathbb{E}\,\mathrm{dist}^2\big[(\hat{\mathbf{U}}, \hat{\mathbf{V}}), (\mathbf{U}, \mathbf{V})\big]$$
$$\lesssim \frac{r \cdot \mathbb{E}\|\mathbf{Z}\|^2}{\lambda_r^2}$$
$$\lesssim \frac{r \cdot \bar{d}}{\lambda_r^2}$$

$\boxed{\text{if } \lambda_r \gtrsim \bar{d}^{1/2}}$

Minimal requirement

Random Perturbation of Singular Subspace

On $\|\mathbf{Z}\|$:

$\|\mathbf{Z}\|$ is sub-gaussian with

$$\mathbb{E}\|Z\| = O(\bar{d}^{1/2}) \text{ with } \bar{d} = \max(d_1, d_2)$$

By Davis - Kahan Bound:

$$\mathbb{E}\ \mathrm{dist}^2\big[(\hat{\mathbf{U}}, \hat{\mathbf{V}}), (\mathbf{U}, \mathbf{V})\big]$$

$$\lesssim \frac{r \cdot \mathbb{E}\|\mathbf{Z}\|^2}{\lambda_r^2}$$

$$\boxed{\text{if } \lambda_r \gtrsim \bar{d}^{1/2}}$$

$$\lesssim \frac{r \cdot \bar{d}}{\lambda_r^2}$$

Minimal requirement

Random Perturbation of Singular Subspace

On $\|\mathbf{Z}\|$:

$$\|\mathbf{Z}\| \text{ is sub-gaussian with}$$
$$\mathbb{E}\|Z\| = O(\bar{d}^{1/2}) \text{ with } \bar{d} = \max(d_1, d_2)$$

By Davis - Kahan Bound:

$$\mathbb{E}\ \mathrm{dist}^2\big[(\hat{\mathbf{U}}, \hat{\mathbf{V}}), (\mathbf{U}, \mathbf{V})\big]$$
$$\lesssim \frac{r \cdot \mathbb{E}\|\mathbf{Z}\|^2}{\lambda_r^2}$$

$\boxed{\text{if } \lambda_r \gtrsim \bar{d}^{1/2}}$

$$\lesssim \frac{r \cdot \bar{d}}{\lambda_r^2}$$

Minimal requirement

does $\lambda_1, \cdots, \lambda_{r-1}$ matter?

Confidence Region of Singular Subspace

Outline

Representation of Spectral Projectors

Normal Approximation of Spectral Projectors

Data-dependent Confidence Regions

Representation of Spectral Projectors

Symmetric Matrices

d

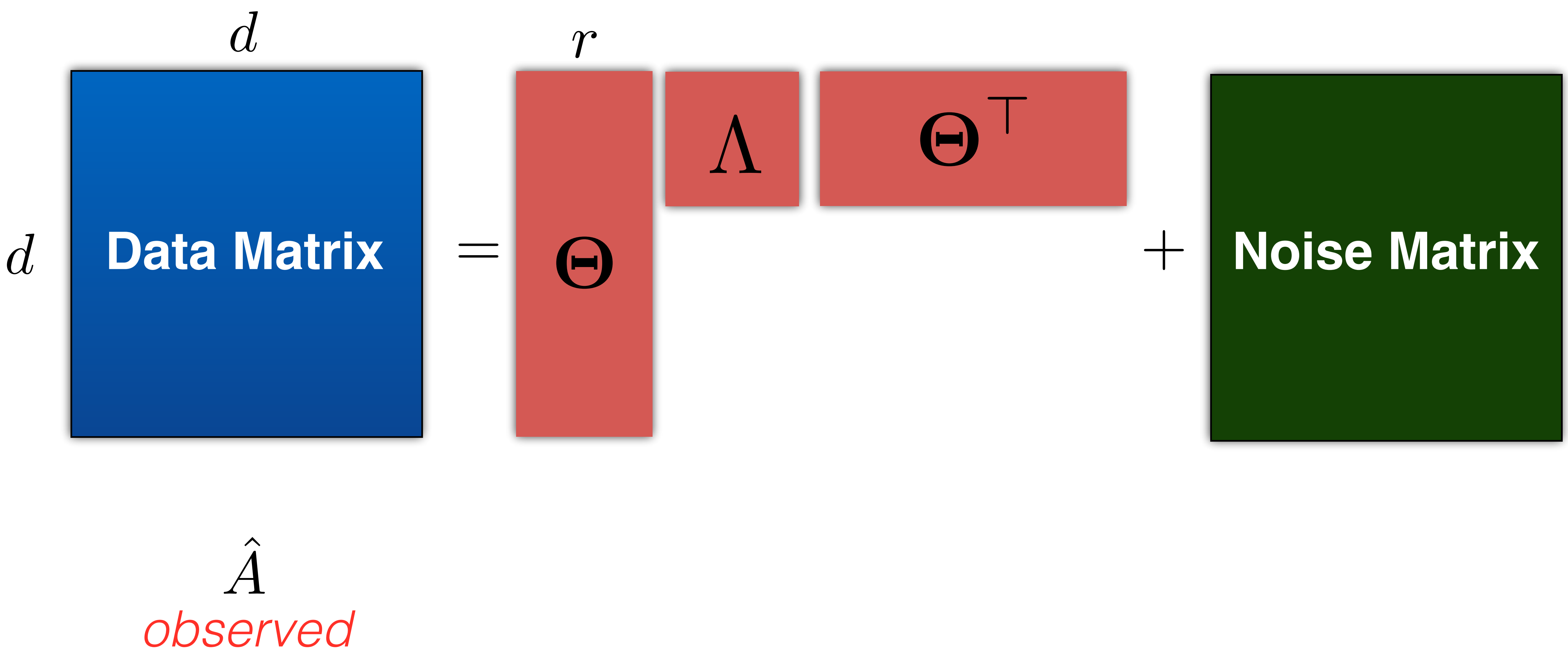
d

$\hat{A}$

observed

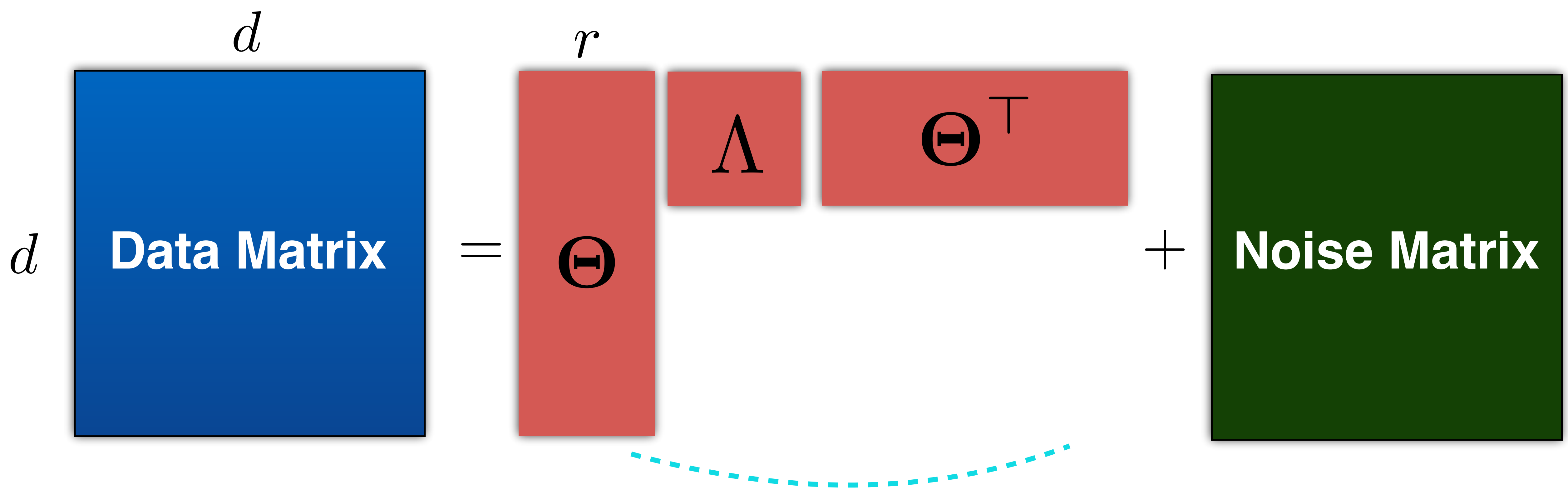
Representation of Spectral Projectors

Symmetric Matrices



Representation of Spectral Projectors

Symmetric Matrices



Signal Matrix

$\hat{A}$ *observed*

$A = \Theta\Lambda\Theta^\top$ *unknown*

X *unknown*

$$\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_r)$$
$$\Theta = (\theta_1, \cdots, \theta_r)$$

$\min_{1\leq i\leq r} |\lambda_i| > 0$

Representation of Spectral Projectors

Empirical Eigenvectors

d

d Data Matrix

eigen-decomposition

$$= \sum_{j=1}^{d} \hat{\lambda}_j \cdot \hat{\theta}_j \hat{\theta}_j^\top$$

$\hat{A}$

observed

Representation of Spectral Projectors

Empirical Eigenvectors

d

d **Data Matrix**

eigen-decomposition

$$= \sum_{j=1}^{d} \hat{\lambda}_j \cdot \hat{\theta}_j \hat{\theta}_j^\top$$

$$|\hat{\lambda}_1| \geq |\hat{\lambda}_2| \geq \cdots |\hat{\lambda}_d|$$

$\hat{A}$

observed

empirical eigenvectors

$$\hat{\Theta} = (\hat{\theta}_1, \cdots, \hat{\theta}_r)$$

Representation of Spectral Projectors

What is the exact relation between $\hat{\Theta}$ *and* Θ *?*

Representation of Spectral Projectors

🤔 *What is the exact relation between* $\hat{\Theta}$ *and* Θ *?*

Technical *Define*

$$\mathfrak{P}^{-s} = \Theta\Lambda^{-s}\Theta^{\top} \qquad \mathfrak{P}^{\perp} = \Theta_{\perp}\Theta_{\perp}^{\top}$$

Representation of Spectral Projectors

Representation of $\hat{\Theta}$

Lemma If $\|X\| < \min_{1\le i\le r} \frac{|\lambda_i|}{2}$, then

1st order term

$$\hat{\Theta}\hat{\Theta}^\top - \Theta\Theta^\top = \boxed{(\mathfrak{P}^{-1}X\mathfrak{P}^{\perp} + \mathfrak{P}^{\perp}X\mathfrak{P}^{-1})}$$

Representation of Spectral Projectors

Representation of $\hat{\Theta}$

Lemma If $\|X\| < \min_{1\le i\le r} \frac{|\lambda_i|}{2}$, then

1st order term

$$\hat{\Theta}\hat{\Theta}^\top - \Theta\Theta^\top = \boxed{(\mathfrak{P}^{-1}X\mathfrak{P}^\perp + \mathfrak{P}^\perp X\mathfrak{P}^{-1})}$$

$$\boxed{\begin{aligned}&+(\mathfrak{P}^{-2}X\mathfrak{P}^\perp X\mathfrak{P}^\perp + \mathfrak{P}^\perp X\mathfrak{P}^{-2}X\mathfrak{P}^\perp + \mathfrak{P}^\perp X\mathfrak{P}^\perp X\mathfrak{P}^{-2}\\ &-\mathfrak{P}^{-1}X\mathfrak{P}^{-1}X\mathfrak{P}^\perp - \mathfrak{P}^\perp X\mathfrak{P}^{-1}X\mathfrak{P}^{-1} - \mathfrak{P}^{-1}X\mathfrak{P}^\perp X\mathfrak{P}^{-1})\end{aligned}}$$

2nd order term

Representation of Spectral Projectors

Representation of $\hat{\Theta}$

Lemma

If $\|X\| < \min_{1\leq i\leq r} \frac{|\lambda_i|}{2}$, then

1st order approx

$$\hat{\Theta}\hat{\Theta}^\top - \Theta\Theta^\top = \boxed{(\mathfrak{P}^{-1}X\mathfrak{P}^{\perp} + \mathfrak{P}^{\perp}X\mathfrak{P}^{-1})}$$

$$\boxed{\begin{aligned} &+(\mathfrak{P}^{-2}X\mathfrak{P}^{\perp}X\mathfrak{P}^{\perp} + \mathfrak{P}^{\perp}X\mathfrak{P}^{-2}X\mathfrak{P}^{\perp} + \mathfrak{P}^{\perp}X\mathfrak{P}^{\perp}X\mathfrak{P}^{-2} \\ &-\mathfrak{P}^{-1}X\mathfrak{P}^{-1}X\mathfrak{P}^{\perp} - \mathfrak{P}^{\perp}X\mathfrak{P}^{-1}X\mathfrak{P}^{-1} - \mathfrak{P}^{-1}X\mathfrak{P}^{\perp}X\mathfrak{P}^{-1})\end{aligned}}$$

$$+ \cdots$$

2nd order term

Representation of Spectral Projectors

Representation of $\hat{\Theta}$

Lemma If $\|X\| < \min_{1\leq i\leq r} \frac{|\lambda_i|}{2}$, then for $s_1, \cdots, s_{k+1} \geq 0$

$$\hat{\Theta}\hat{\Theta}^\top - \Theta\Theta^\top$$

$$= \sum_{k\geq 1} \sum_{\mathbf{s}: s_1+\cdots+s_{k+1}=k} (-1)^{1+\tau(\mathbf{s})} \cdot \mathfrak{P}^{-s_1} X \mathfrak{P}^{-s_2} X \cdots X \mathfrak{P}^{-s_{k+1}}$$

where we denote $\mathfrak{P}^0 = \mathfrak{P}^\perp$ and

Representation of Spectral Projectors

Representation of $\hat{\Theta}$

Lemma If $\|X\| < \min_{1\leq i\leq r} \frac{|\lambda_i|}{2}$, then for $s_1, \cdots, s_{k+1} \geq 0$

$$\hat{\Theta}\hat{\Theta}^\top - \Theta\Theta^\top$$

$$= \sum_{k\geq 1} \sum_{\mathbf{s}: s_1+\cdots+s_{k+1}=k} \boxed{(-1)^{1+\tau(\mathbf{s})} \cdot \mathfrak{P}^{-s_1} X \mathfrak{P}^{-s_2} X \cdots X \mathfrak{P}^{-s_{k+1}}}$$

where we denote $\mathfrak{P}^0 = \mathfrak{P}^\perp$ and

k-th order term

Representation of Spectral Projectors

Representation of $\hat{\Theta}$

Lemma If $\|X\| < \min_{1\le i\le r} \frac{|\lambda_i|}{2}$, then for $s_1,\cdots,s_{k+1}\ge 0$

$$\hat{\Theta}\hat{\Theta}^\top - \Theta\Theta^\top = \sum_{k\ge 1}\sum_{\mathbf{s}:s_1+\cdots+s_{k+1}=k} \boxed{(-1)^{1+\tau(\mathbf{s})}\cdot \mathfrak{P}^{-s_1}X\mathfrak{P}^{-s_2}X\cdots X\mathfrak{P}^{-s_{k+1}}}$$

where we denote $\mathfrak{P}^0 = \mathfrak{P}^\perp$ and

$$\tau(\mathbf{s}) = \sum_{j=1}^{k+1}\mathbb{I}(s_j>0).$$

k-th order term

Representation of Spectral Projectors

Highlights

- *Deterministic representation of eigenvectors*
- *No eigen-gap conditions (except signal strength)*
- *The k-th order error terms:*

$$\Big\| \sum_{\mathbf{s}} (-1)^{1+\tau(\mathbf{s})} \cdot \mathfrak{P}^{-s_1} X \mathfrak{P}^{-s_2} X \cdots X \mathfrak{P}^{-s_{k+1}} \Big\| \leq \Big(\frac{4\|X\|}{\min_i |\lambda_i|} \Big)^k$$

Confidence Region of Singular Subspace

Outline

Representation of Spectral Projectors

Normal Approximation of Spectral Projectors

Data-dependent Confidence Regions

Normal Approximation of SVD

Model: $\hat{M} = M + Z \text{ with } M = U\Lambda V^\top \in \mathbb{R}^{d_1 \times d_2}$

Assumption: $Z = (z_{ij})_{1\le i\le d_1}^{1\le j\le d_2}$

$$\mathbf{z}_{ij} \text{ are i.i.d. and } \mathbf{z}_{ij} \sim \mathcal{N}(0,1)$$

Loss:

$$\text{dist}^2\big[(\hat{\mathbf{U}},\hat{\mathbf{V}}),(\mathbf{U},\mathbf{V})\big]$$

$$= \|\hat{\mathbf{U}}\hat{\mathbf{U}}^\top - \mathbf{U}\mathbf{U}^\top\|_{\mathrm{F}}^2 + \|\hat{\mathbf{V}}\hat{\mathbf{V}}^\top - \mathbf{V}\mathbf{V}^\top\|_{\mathrm{F}}^2$$

🤔 *Distribution of* $\text{dist}^2[(\hat{U},\hat{V}),(U,V)]$ *?*

Normal Approximation of SVD

Dilation

$$\hat{M} = M + Z$$

Normal Approximation of SVD

Dilation

$$\hat{M} = M + Z \longrightarrow \begin{pmatrix} 0 & \hat{M} \\ \hat{M}^\top & 0 \end{pmatrix} = \begin{pmatrix} 0 & M \\ M^\top & 0 \end{pmatrix} + \begin{pmatrix} 0 & Z \\ Z^\top & 0 \end{pmatrix}$$

Normal Approximation of SVD

Dilation

$$\hat{M} = M + Z \rightarrow \begin{pmatrix} 0 & \hat{M} \\ \hat{M}^\top & 0 \end{pmatrix} = \begin{pmatrix} 0 & M \\ M^\top & 0 \end{pmatrix} + \begin{pmatrix} 0 & Z \\ Z^\top & 0 \end{pmatrix}$$

$$\boxed{\hat{A} = A + X}$$

Symmetric !

Normal Approximation of SVD

Dilation

$$\hat{M} = M + Z \longrightarrow \begin{pmatrix} 0 & \hat{M} \\ \hat{M}^\top & 0 \end{pmatrix} = \begin{pmatrix} 0 & M \\ M^\top & 0 \end{pmatrix} + \begin{pmatrix} 0 & Z \\ Z^\top & 0 \end{pmatrix}$$

$$\boxed{\hat{A} = A + X}$$

Symmetric !

We can represent the empirical singular vectors !

Representation formula of singular vectors

$$A = \sum_{|j|\geq 1}^{r} \lambda_j \cdot \theta_j \theta_j^\top \longrightarrow \theta_j = \frac{1}{\sqrt{2}} \begin{pmatrix} u_j \\ v_j \end{pmatrix} \quad \theta_{-j} = \frac{1}{\sqrt{2}} \begin{pmatrix} u_j \\ -v_j \end{pmatrix}$$

$$\lambda_j = -\lambda_j$$

$$\Theta = (\theta_1, \cdots, \theta_r, \theta_{-r}, \cdots, \theta_{-1}) \in \mathbb{R}^{(d_1+d_2)\times(2r)}$$

Spectral Projector

$$\Theta\Theta^\top = \begin{pmatrix} UU^\top & 0 \\ 0 & VV^\top \end{pmatrix}$$

Representation formula of singular vectors

Empirical spectral projector $$\hat{\Theta}\hat{\Theta}^\top = \begin{pmatrix} \hat{U}\hat{U}^\top & 0 \\ 0 & \hat{V}\hat{V}^\top \end{pmatrix}$$

Technical *Define*

$$\mathfrak{P}^\perp = \begin{pmatrix} U_\perp U_\perp^\top & 0 \\ 0 & V_\perp V_\perp^\top \end{pmatrix}$$

$$\mathfrak{P}^{-k} = \begin{cases} \begin{pmatrix} U\Lambda^{-k}U^\top & 0 \\ 0 & V\Lambda^{-k}V^\top \end{pmatrix} & \text{if k is even} \\ \begin{pmatrix} 0 & U\Lambda^{-k}V^\top \\ V\Lambda^{-k}U^\top & 0 \end{pmatrix} & \text{if k is odd} \end{cases}$$

Representation formula of singular vectors

$$\begin{pmatrix} \hat U\hat U^\top - UU^\top & 0 \\ 0 & \hat V\hat V^\top - VV^\top \end{pmatrix} = \sum_{k\ge1} \boxed{\sum_{\mathbf{s}:s_1+\cdots+s_{k+1}=k} (-1)^{1+\tau(\mathbf{s})}\cdot \mathfrak{P}^{-s_1} X \mathfrak{P}^{-s_2} X \cdots X \mathfrak{P}^{-s_{k+1}}}$$

where $\mathfrak{P}^0 = \mathfrak{P}^\perp$.

$\mathcal{S}_{A,k}(X)$ ***k-th order term***

$$\hat\Theta\hat\Theta^\top - \Theta\Theta^\top = \sum_{k\ge1} \mathcal{S}_{A,k}(X)$$

$$\begin{aligned} \mathrm{dist}^2[(\hat U, \hat V), (U, V)] &= \|\hat\Theta\hat\Theta^\top - \Theta\Theta^\top\|_{\mathrm{F}}^2 \\ &= -2\sum_{k\ge2} \langle \Theta\Theta^\top, \mathcal{S}_{A,k}(X)\rangle \end{aligned}$$

Expectation of the Loss

First order approx. of $\mathbb{E}\,\mathrm{dist}^2\big[(\hat{\mathbf{U}},\hat{\mathbf{V}}),(\mathbf{U},\mathbf{V})\big]$

$$\mathbb{E}\,\mathrm{dist}^2\big[(\hat{\mathbf{U}},\hat{\mathbf{V}}),(\mathbf{U},\mathbf{V})\big] = -2\sum_{k\geq 2}\mathbb{E}\langle \Theta\Theta^\top, \mathcal{S}_{A,k}(X)\rangle$$

(the expectation is 0 whenever k is odd)

What if we just calculate k=2 and roughly bound remainders?

Theorem If $\lambda_r \gtrsim \sqrt{d_{\max}}$, then

$$\mathbb{E}\,\mathrm{dist}^2[(\hat{U},\hat{V}),(U,V)] = [2+o(1)](d_1+d_2-2r)\boxed{\|\Lambda^{-1}\|_{\mathrm{F}}^2}$$

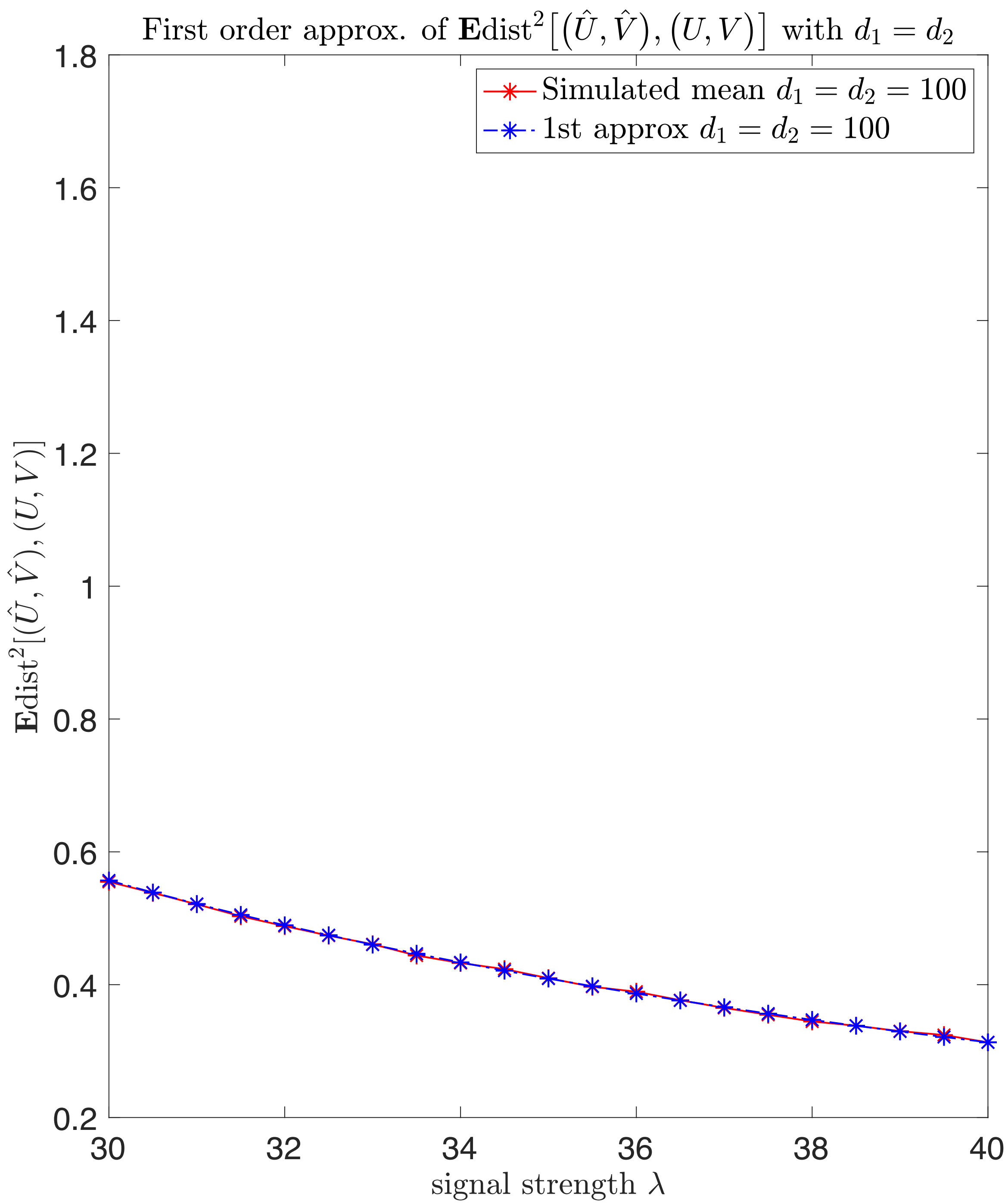
$$\frac{1}{\lambda_1^2}+\cdots+\frac{1}{\lambda_r^2}$$

Expectation of the Loss

Simulation of $\mathbb{E}\,\mathrm{dist}^2\big[(\hat{\mathbf{U}},\hat{\mathbf{V}}),(\mathbf{U},\mathbf{V})\big]$

1st approx

$$2(d_1+d_2-2r)\|\Lambda^{-1}\|_{\mathrm{F}}^2$$

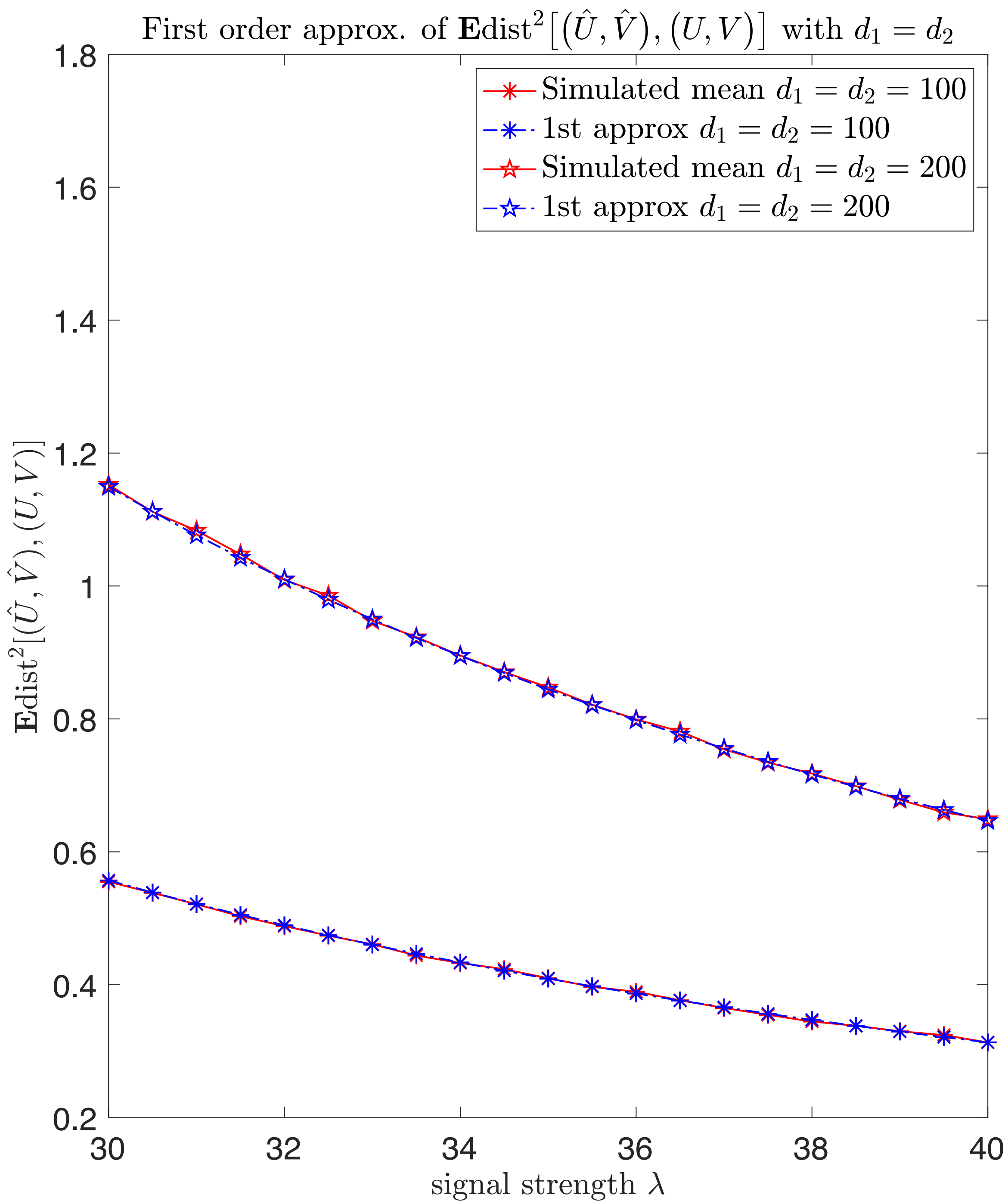


Expectation of the Loss

Simulation of $\mathbb{E}\,\mathrm{dist}^2\big[(\hat{\mathbf{U}},\hat{\mathbf{V}}),(\mathbf{U},\mathbf{V})\big]$

1st approx

$2(d_1+d_2-2r)\|\Lambda^{-1}\|_{\mathrm{F}}^2$

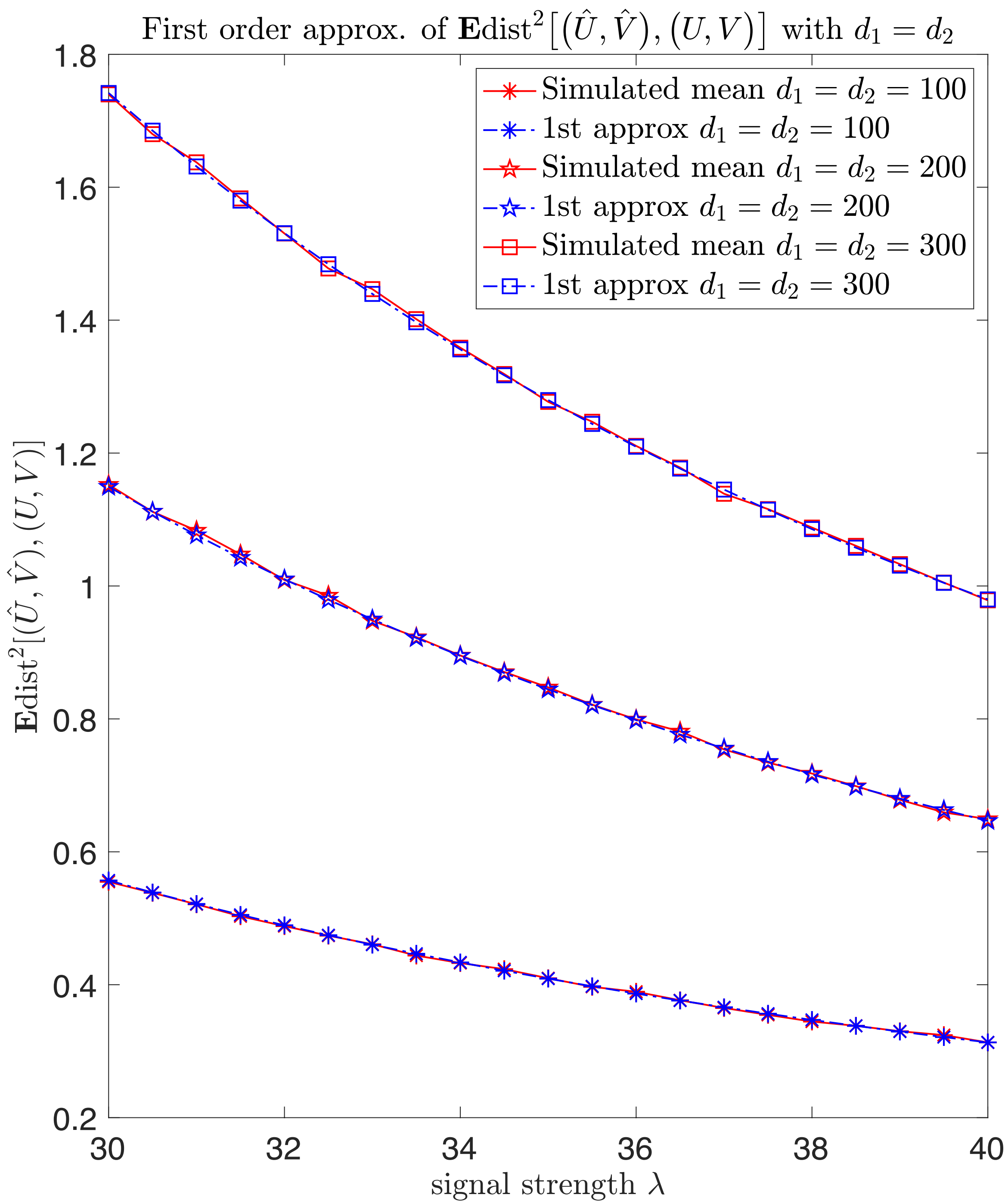


Expectation of the Loss

Simulation of $\mathbb{E}\,\mathrm{dist}^2\big[(\hat{\mathbf{U}},\hat{\mathbf{V}}),(\mathbf{U},\mathbf{V})\big]$

1st approx

$$2(d_1+d_2-2r)\|\Lambda^{-1}\|_{\mathrm{F}}^2$$



Very close!

Expectation of the Loss

Is first order approx. enough for $\mathbb{E}\,\mathrm{dist}^2\big[(\hat{\mathbf{U}},\hat{\mathbf{V}}),(\mathbf{U},\mathbf{V})\big]$?

1st approx

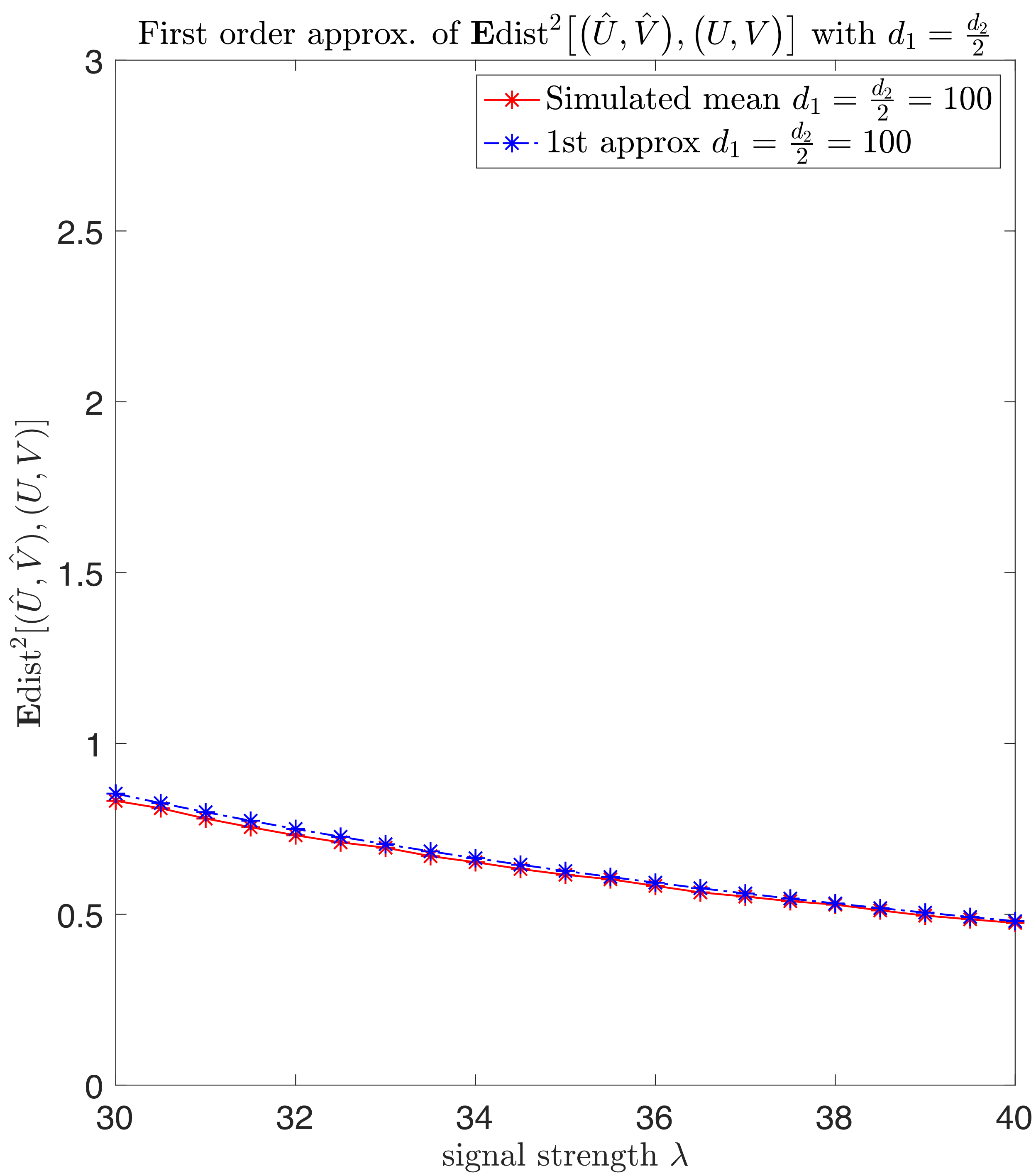
$$2(d_1+d_2-2r)\|\Lambda^{-1}\|_{\mathrm{F}}^2$$

Expectation of the Loss

Is first order approx. enough for $\mathbb{E}\,\mathrm{dist}^2\big[(\hat{\mathbf{U}},\hat{\mathbf{V}}),(\mathbf{U},\mathbf{V})\big]$?

1st approx

$$2(d_1+d_2-2r)\|\Lambda^{-1}\|_{\mathrm{F}}^2$$

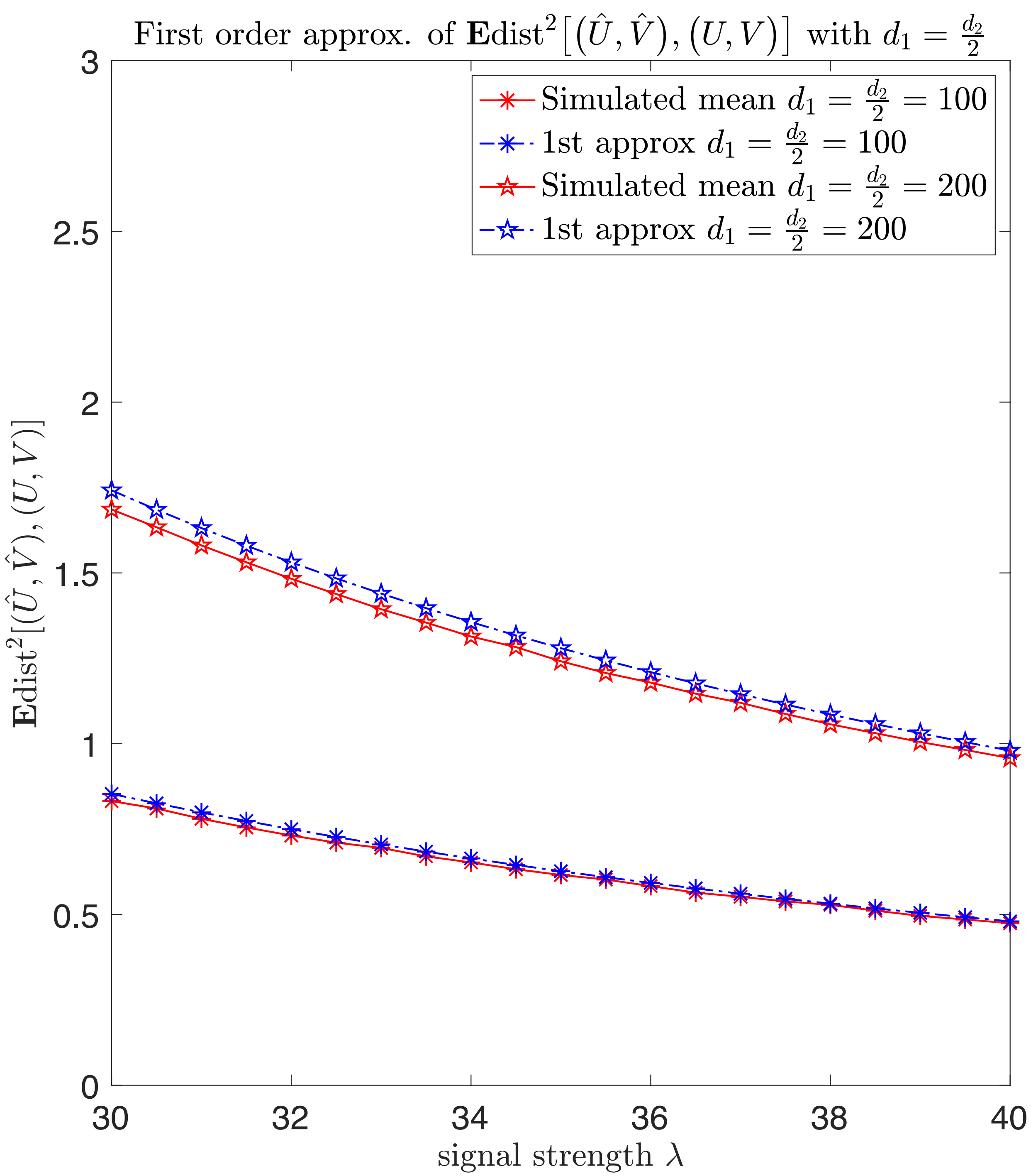


Expectation of the Loss

Is first order approx. enough for $\mathbb{E}\,\mathrm{dist}^2\big[(\hat{\mathbf{U}},\hat{\mathbf{V}}),(\mathbf{U},\mathbf{V})\big]$ **?**

1st approx

$$2(d_1+d_2-2r)\|\Lambda^{-1}\|_{\mathrm{F}}^2$$

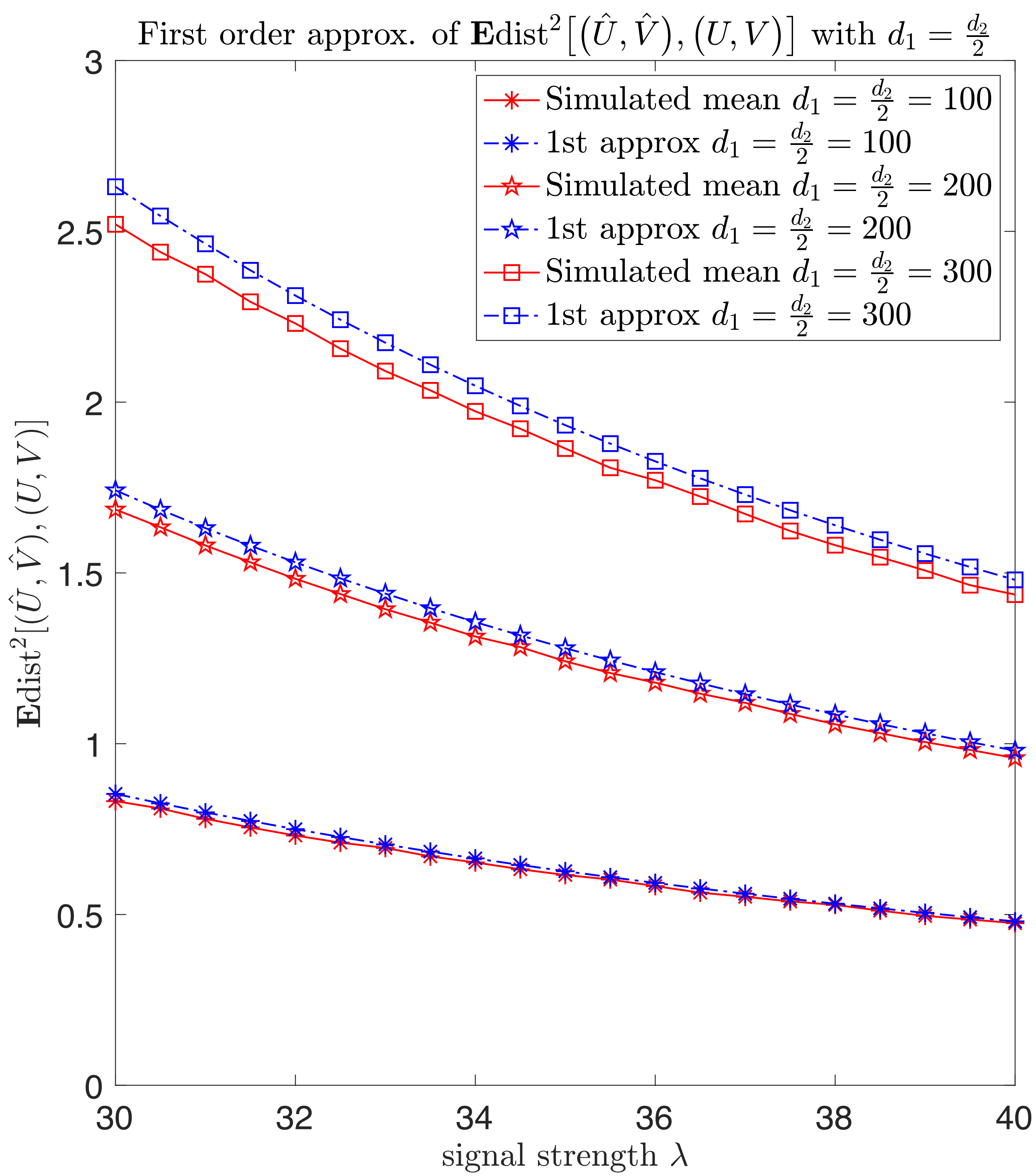


Expectation of the Loss

Is first order approx. enough for $\mathbb{E}\,\mathrm{dist}^2\big[(\hat{\mathbf{U}},\hat{\mathbf{V}}),(\mathbf{U},\mathbf{V})\big]$?

1st approx

$$2(d_1+d_2-2r)\|\Lambda^{-1}\|_{\mathrm{F}}^2$$



What happens

when $|d_1-d_2|\gg 0$

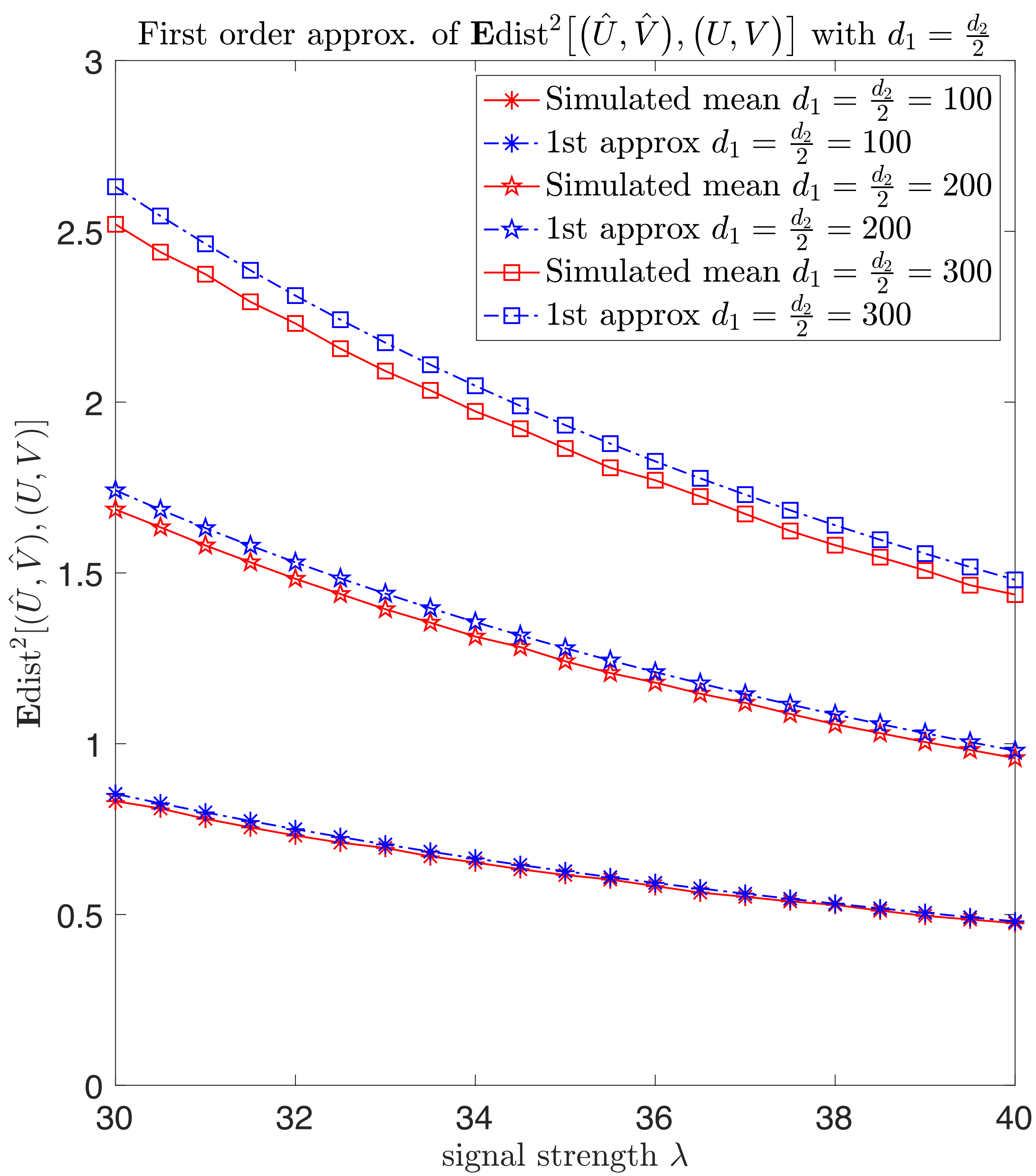
?

Expectation of the Loss

Is first order approx. enough for $\mathbb{E}\ \mathrm{dist}^2\big[(\hat{\mathbf{U}},\hat{\mathbf{V}}),(\mathbf{U},\mathbf{V})\big]$?

1st approx

$$2(d_1+d_2-2r)\|\Lambda^{-1}\|_{\mathrm{F}}^2$$



What happens

when $|d_1-d_2|\gg 0$

?

1st approx over-estimates !

Expectation of the Loss

Higher order approx. of $\mathbb{E}\,\mathrm{dist}^2\big[(\hat{\mathbf{U}},\hat{\mathbf{V}}),(\mathbf{U},\mathbf{V})\big]$

$$\mathbb{E}\,\mathrm{dist}^2\big[(\hat{\mathbf{U}},\hat{\mathbf{V}}),(\mathbf{U},\mathbf{V})\big] = -2\sum_{k\geq 2}\mathbb{E}\big\langle \Theta\Theta^\top, \mathcal{S}_{A,k}(X)\big\rangle$$

1st order term

$$-2\mathbb{E}\big\langle \Theta\Theta^\top, \mathcal{S}_{A,2}(X)\big\rangle = 2(d_1+d_2-2r)\|\Lambda^{-1}\|_{\mathrm{F}}^2$$

2nd order term $\quad -2\mathbb{E}\big\langle \Theta\Theta^\top, \mathcal{S}_{A,4}(X)\big\rangle \quad$ **?**

3rd order term $\quad -2\mathbb{E}\big\langle \Theta\Theta^\top, \mathcal{S}_{A,6}(X)\big\rangle \quad$ **?**

... $\quad$ **...** $\quad$ **...**

🤔

$$\mathcal{S}_{A,k}(X) = \sum_{\mathbf{s}}(-1)^{1+\tau(\mathbf{s})}\cdot \mathfrak{P}^{-s_1}X\mathfrak{P}^{-s_2}X\cdots X\mathfrak{P}^{-s_{k+1}}$$

Expectation of the Loss

Second order approx. of $\mathbb{E}\,\mathrm{dist}^2\big[(\hat{\mathbf{U}},\hat{\mathbf{V}}),(\mathbf{U},\mathbf{V})\big]$

Theorem If $\lambda_r \gtrsim \sqrt{d_{\max}}$, then

$$\Big|\mathbb{E}\big\langle \Theta\Theta^\top, \mathcal{S}_{A,4}(X)\big\rangle - (d_1-d_2)^2\|\Lambda^{-2}\|_{\mathrm{F}}^2\Big| = O\Big(\frac{r^2 d_{\max}}{\lambda_r^4}\Big)$$

and

$$\Big|\mathbb{E}\mathrm{dist}^2[(\hat{U},\hat{V}),(U,V)] - B_2\Big| = \boxed{O\Big(\frac{r d_{\max}^3}{\lambda_r^6}\Big)}$$

where

$$B_2 = 2(d_1+d_2-2r)\|\Lambda^{-1}\|_{\mathrm{F}}^2 - 2(d_1-d_2)^2\|\Lambda^{-2}\|_{\mathrm{F}}^2$$

2nd order approx

Expectation of the Loss

Third order approx. of $\mathbb{E}\,\mathrm{dist}^2\big[(\hat{\mathbf{U}},\hat{\mathbf{V}}),(\mathbf{U},\mathbf{V})\big]$

$d_{1-} = d_1 - r$
$d_{2-} = d_2 - r$

Theorem If $\lambda_r \gtrsim \sqrt{d_{\max}}$, then

$$\Big|\mathbb{E}\langle \Theta\Theta^\top, \mathcal{S}_{A,6}(X)\rangle + (d_{1-}^2 - d_{2-}^2)(d_1-d_2)\|\Lambda^{-3}\|_{\mathrm{F}}^2\Big| = O\Big(\frac{r^2}{\sqrt{d_{\max}}}\cdot\frac{d_{\max}^3}{\lambda_r^6}\Big)$$

and

$$\Big|\mathbb{E}\mathrm{dist}^2[(\hat{U},\hat{V}),(U,V)] - B_3\Big| = \boxed{O\Big(\frac{r d_{\max}^4}{\lambda_r^8}\Big)}$$

where

$$B_3 = B_2 + 2(d_{1-}^2 - d_{2-}^2)(d_1-d_2)\|\Lambda^{-3}\|_{\mathrm{F}}^2$$

3rd order approx

Expectation of the Loss

k-th order approx. of $\mathbb{E}\,\mathrm{dist}^2\big[(\hat{\mathbf{U}},\hat{\mathbf{V}}),(\mathbf{U},\mathbf{V})\big]$

$d_{1-} = d_1 - r$

$d_{2-} = d_2 - r$

Theorem If $\lambda_r \gtrsim \sqrt{d_{\max}}$, then

$$\Big|\mathbb{E}\langle \Theta\Theta^\top, \mathcal{S}_{A,2k}\rangle - (-1)^k(d_{1-}^{k-1} - d_{2-}^{k-1})(d_1 - d_2)\|\Lambda^{-k}\|_{\mathrm{F}}^2\Big| = O\Big(\frac{r^2}{\sqrt{d_{\max}}}\cdot\big(\frac{Cd_{\max}}{\lambda_r^2}\big)^k\Big)$$

and

$$\Big|\mathbb{E}\mathrm{dist}^2[(\hat U,\hat V),(U,V)] - B_k\Big| = \boxed{O\Big(\frac{r^2}{\sqrt{d_{\max}}}\cdot\frac{d_{\max}^3}{\lambda_r^6} + r\big(\frac{Cd_{\max}}{\lambda_r^2}\big)^{k+1}\Big)}$$

where

$$B_k = 2(d_{1-}+d_{2-})\|\Lambda^{-1}\|_{\mathrm{F}}^2 - 2\sum_{k_0=2}^{k}(-1)^{k_0}(d_{1-}^{k_0-1} - d_{2-}^{k_0-1})(d_1-d_2)\|\Lambda^{-k_0}\|_{\mathrm{F}}^2$$

k-th order approx

Expectation of the Loss

Remarks

If we choose $k = \lceil \log(d_1 + d_2) \rceil$, then

$$\left| \mathbb{E}\mathrm{dist}^2[(\hat{U}, \hat{V}), (U, V)] - B_{\lceil \log(d_1+d_2) \rceil} \right| = O\Big(\frac{r^2}{\sqrt{d_{\max}}} \cdot \frac{d_{\max}^3}{\lambda_r^6} \Big)$$

sum of $\lceil \log(d_1 + d_2) \rceil$ terms

If we choose $k = \infty$, then

$$B_\infty = 2 \sum_{j=1}^{r} \frac{1}{\lambda_j^2} \Big(d_{1-} \cdot \frac{\lambda_j^2 + d_{2-}}{\lambda_j^2 + d_{1-}} + d_{2-} \cdot \frac{\lambda_j^2 + d_{1-}}{\lambda_j^2 + d_{2-}} \Big)$$

Higher order approx. is unnecessary when $d_1 = d_2$.

Expectation of the Loss

Remarks

$$\left|\mathbb{E}\|\hat{U}\hat{U}^\top - UU^\top\|_{\rm F}^2 - 2\sum_{j=1}^r \frac{d_{1-}(\lambda_j^2+d_{2-})}{\lambda_j^2(\lambda_j^2+d_{1-})}\right| = O\Big(\frac{r^2}{\sqrt{d_{\max}}}\cdot\frac{d_{\max}^3}{\lambda_r^6}\Big)$$

$$\left|\mathbb{E}\|\hat{V}\hat{V}^\top - VV^\top\|_{\rm F}^2 - 2\sum_{j=1}^r \frac{d_{2-}(\lambda_j^2+d_{1-})}{\lambda_j^2(\lambda_j^2+d_{2-})}\right| = O\Big(\frac{r^2}{\sqrt{d_{\max}}}\cdot\frac{d_{\max}^3}{\lambda_r^6}\Big)$$

Contributions

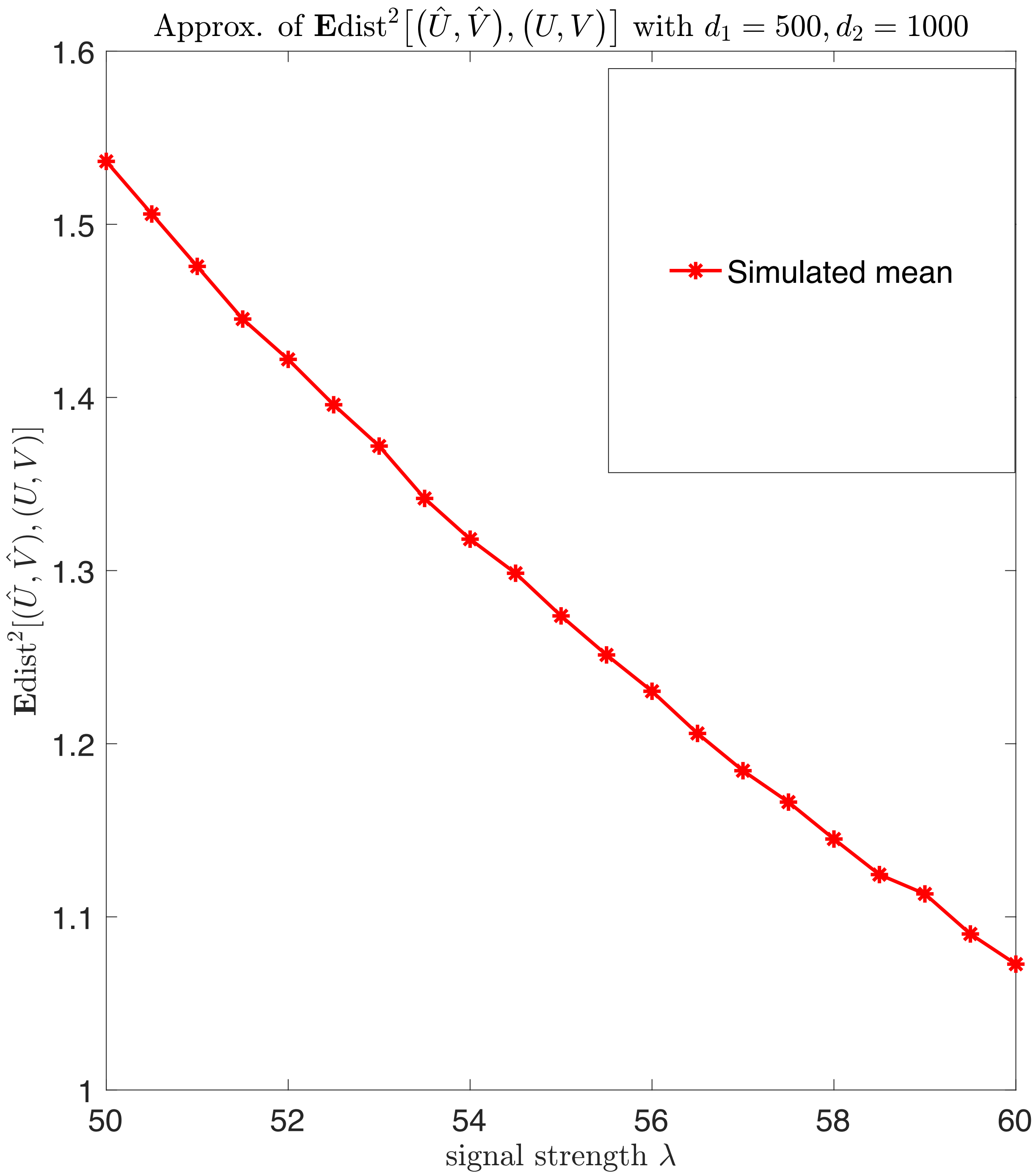
- Non-asymptotical results
- Rank r can diverge
- No eigen-gap conditions (except signal strength)

Expectation of the Loss

Simulations

higher order approx

when $|d_1 - d_2| \gg 0$

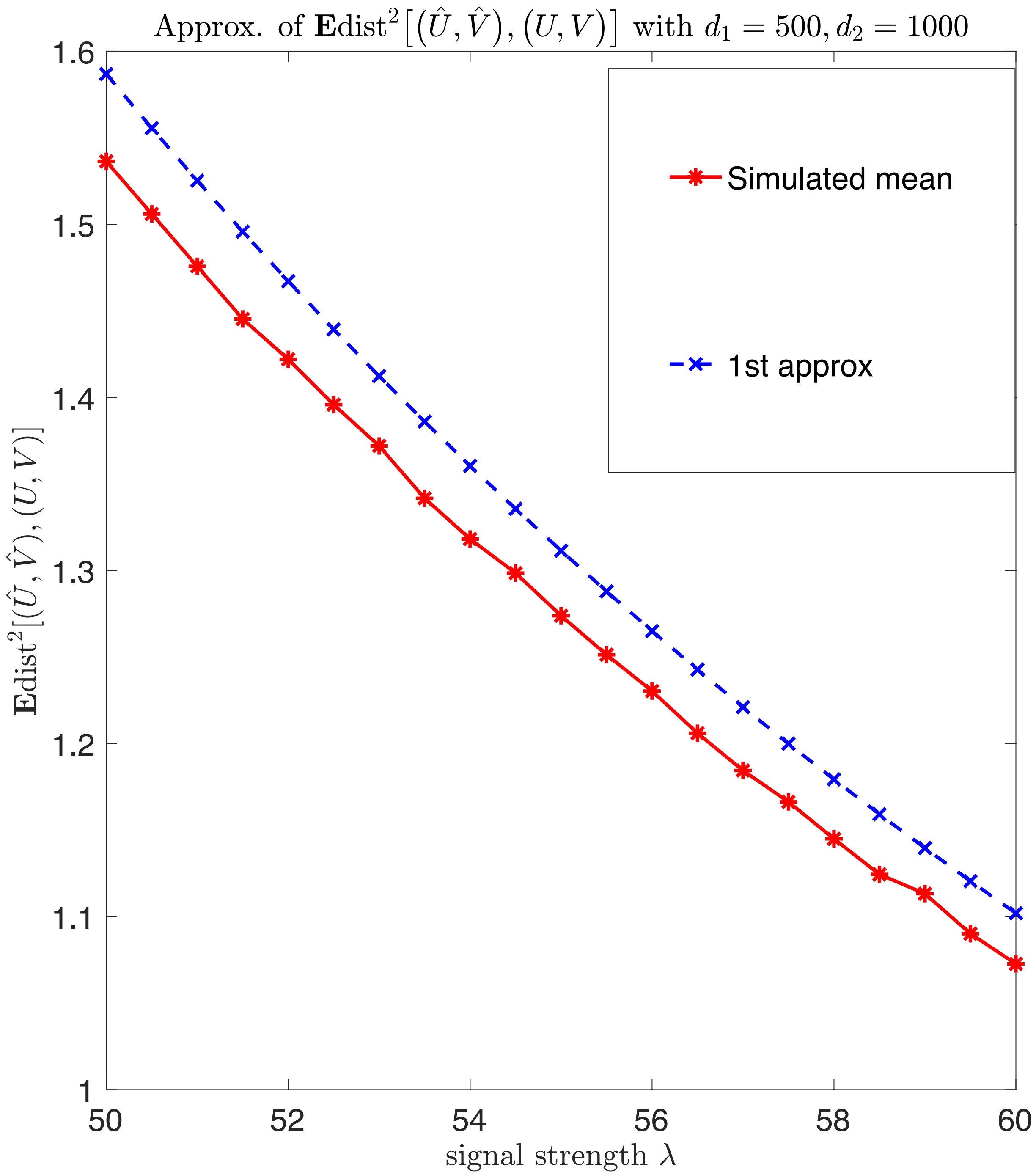


Expectation of the Loss

Simulations

higher order approx

when $|d_1 - d_2| \gg 0$

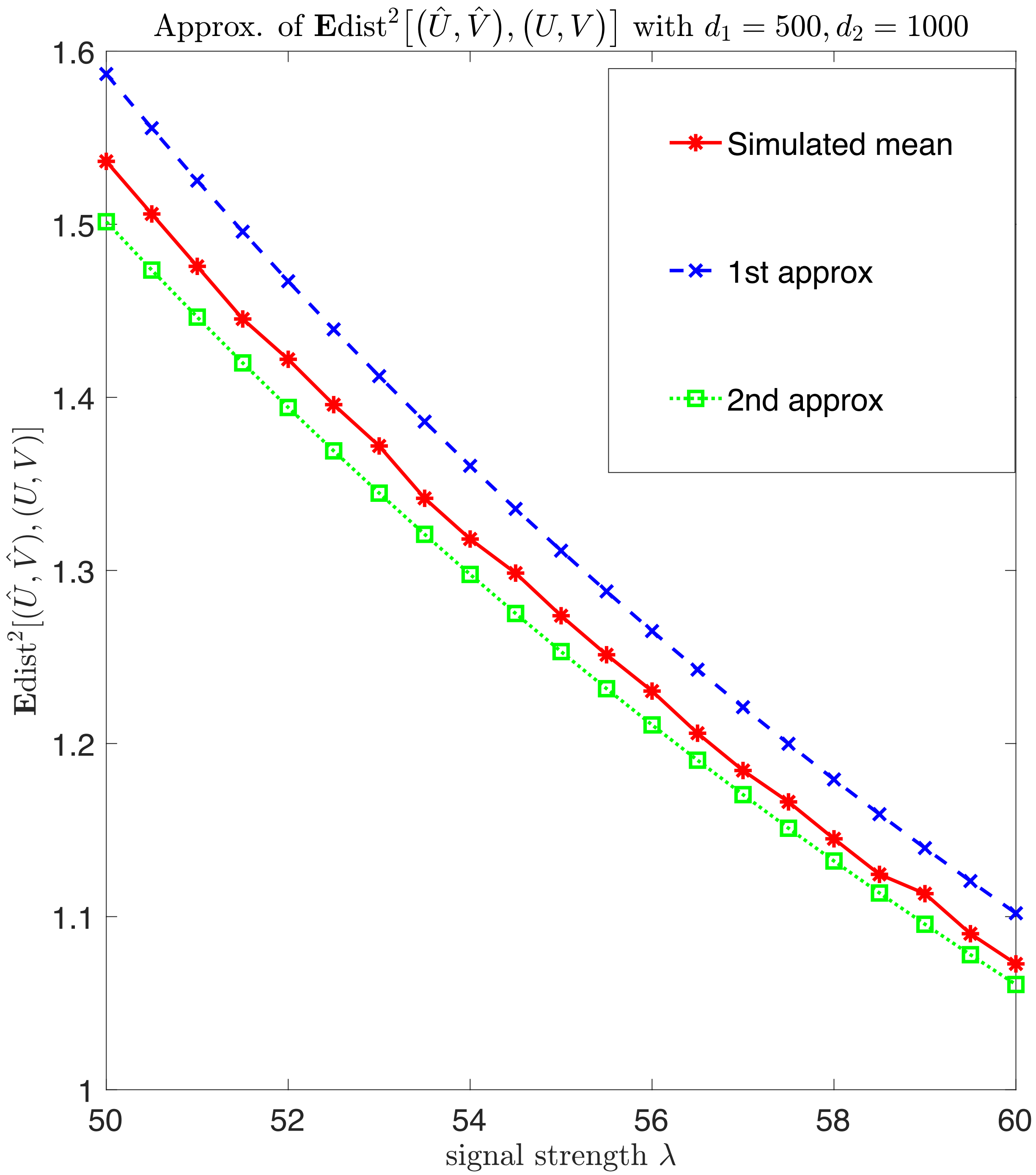


Expectation of the Loss

Simulations

higher order approx

when $|d_1 - d_2| \gg 0$

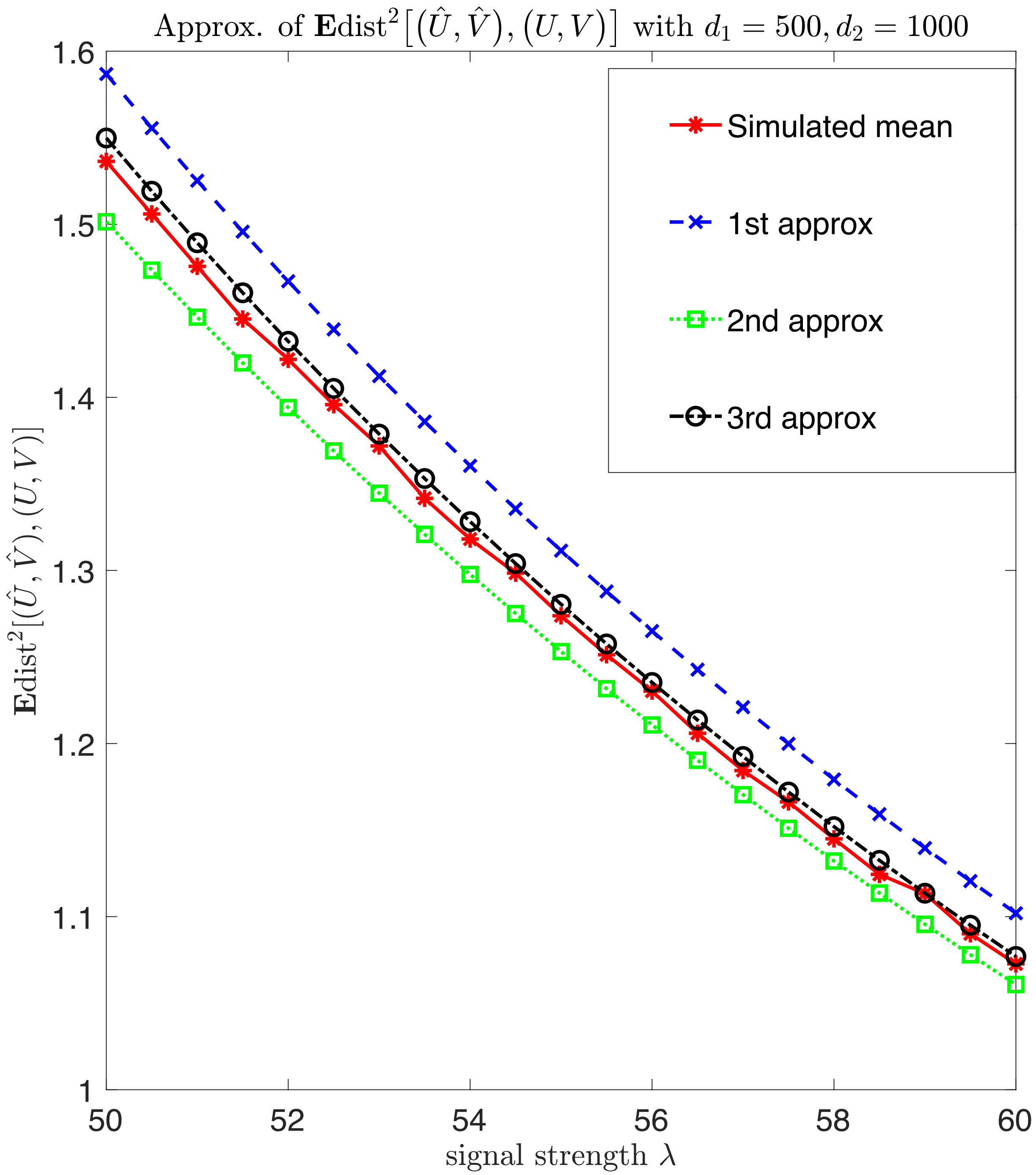


Expectation of the Loss

Simulations

higher order approx

when $|d_1 - d_2| \gg 0$

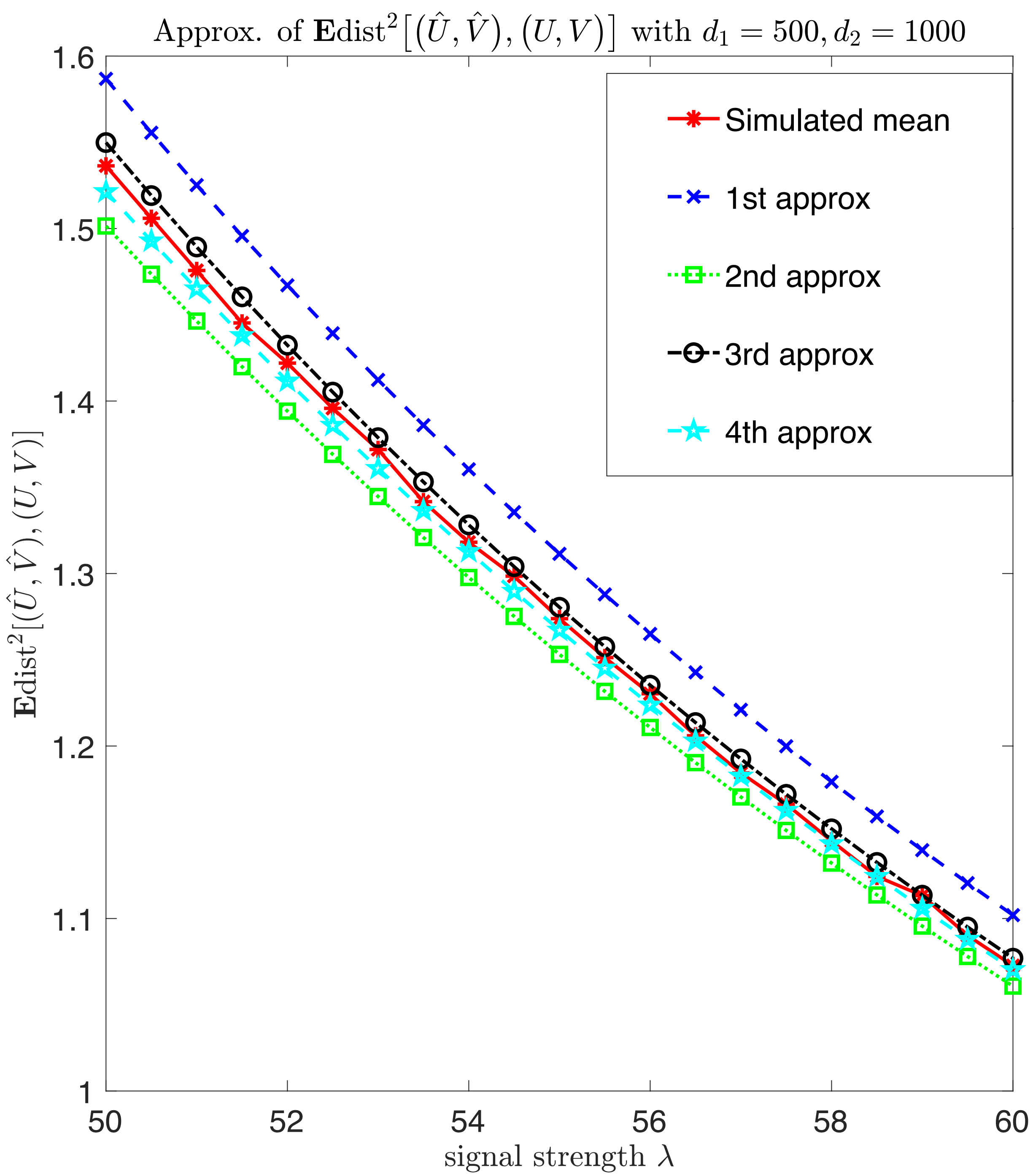


Expectation of the Loss

Simulations

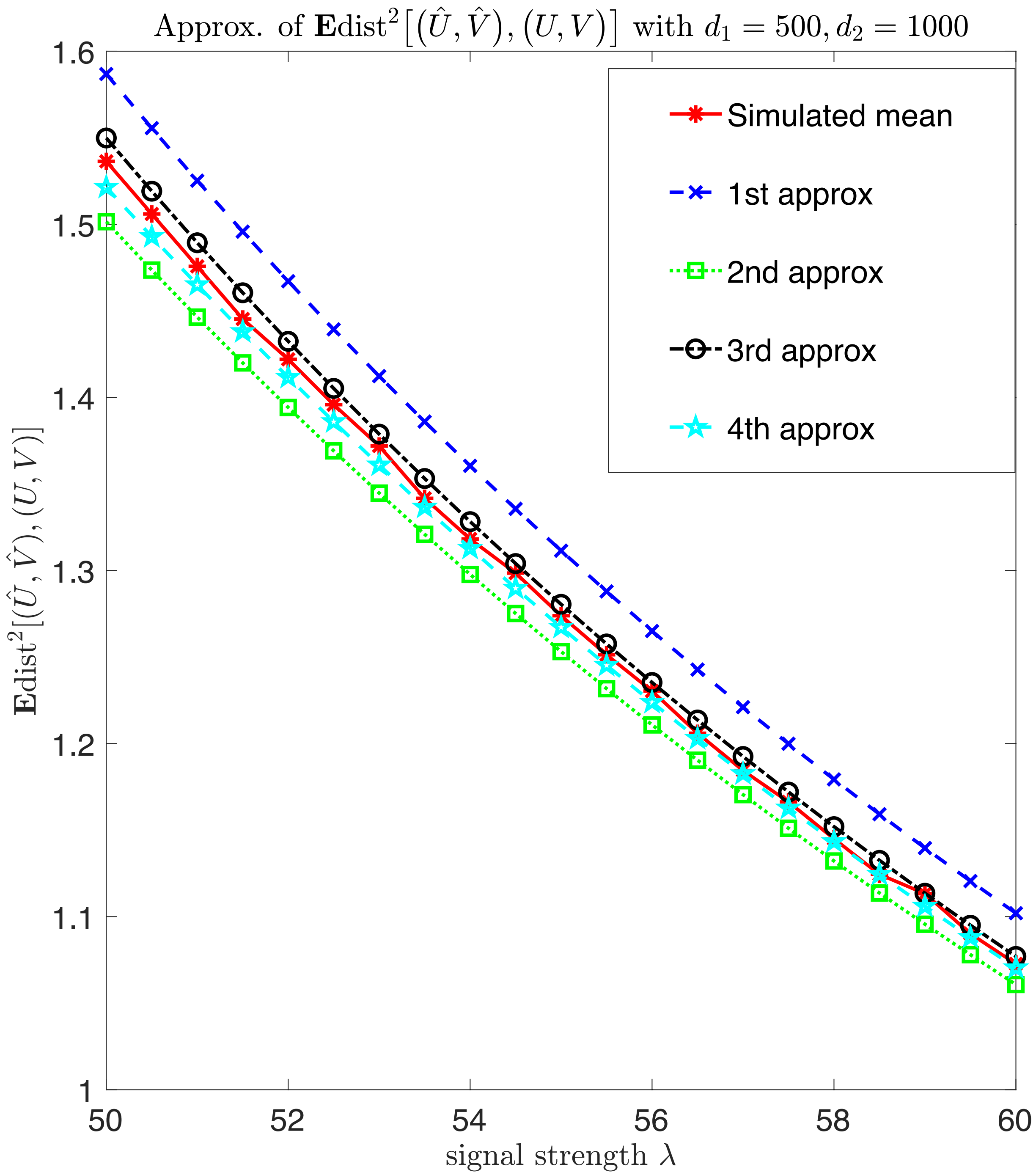
higher order approx

when $|d_1 - d_2| \gg 0$



Expectation of the Loss

Simulations



higher order approx

when $|d_1 - d_2| \gg 0$

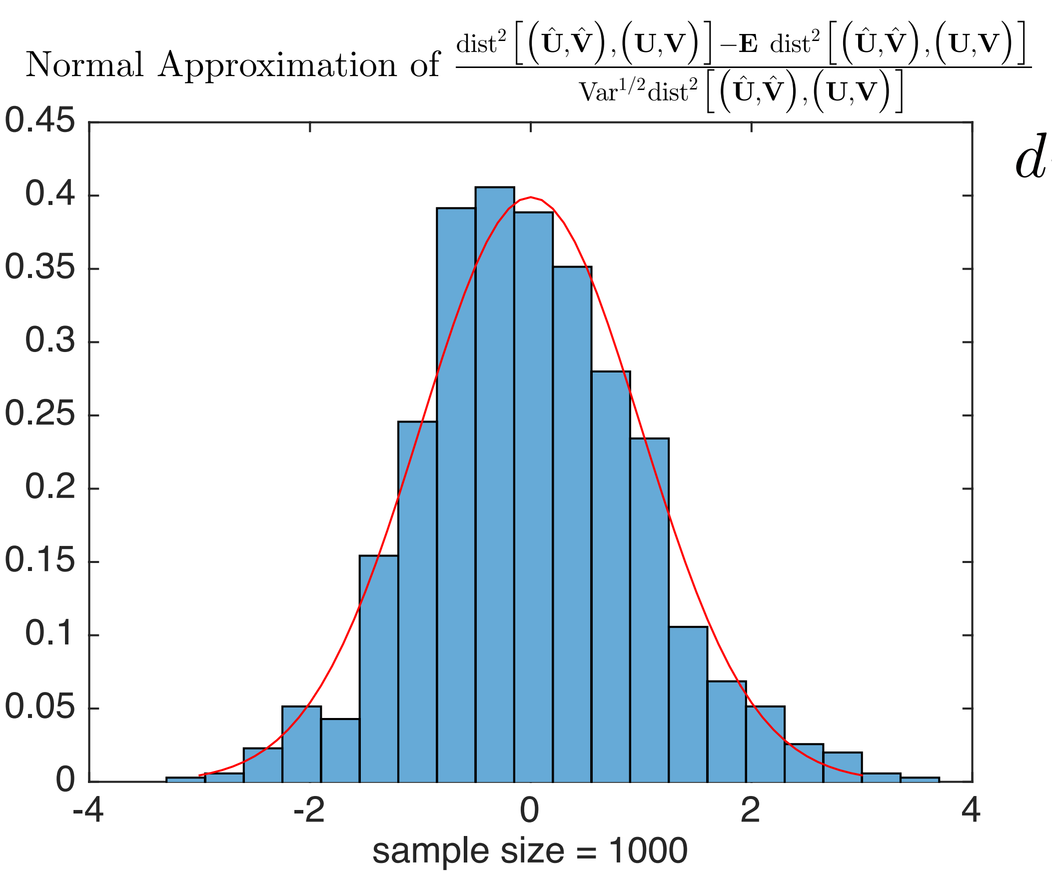
1st and 3rd over-estimate !

2nd and 4th under-estimate !

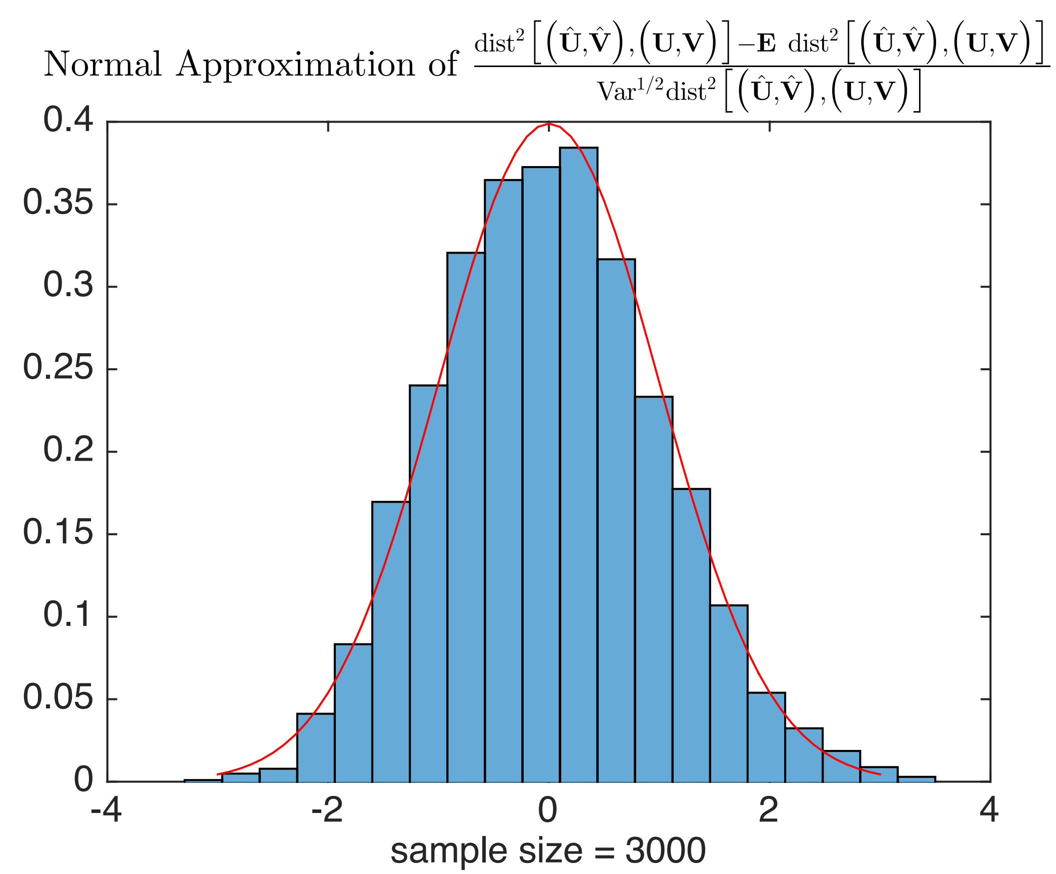
Normal Approximation of the Loss

Distribution of $\mathrm{dist}^2\big[(\hat{\mathbf{U}},\hat{\mathbf{V}}),(\mathbf{U},\mathbf{V})\big]$ **?**

Simulation study



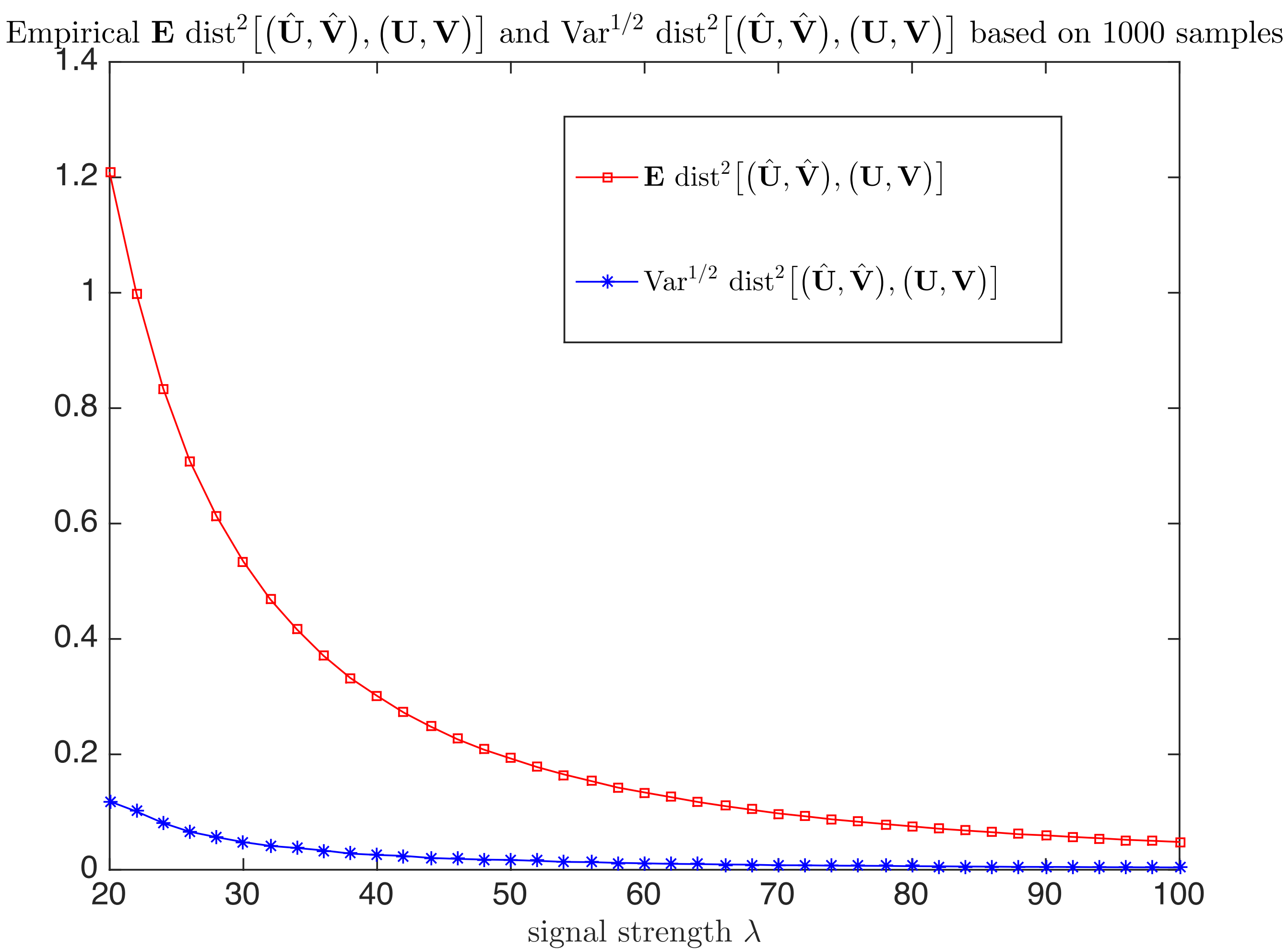
$$d_1 = d_2 = 100$$
$$r = 10$$
$$\lambda_k = 100$$

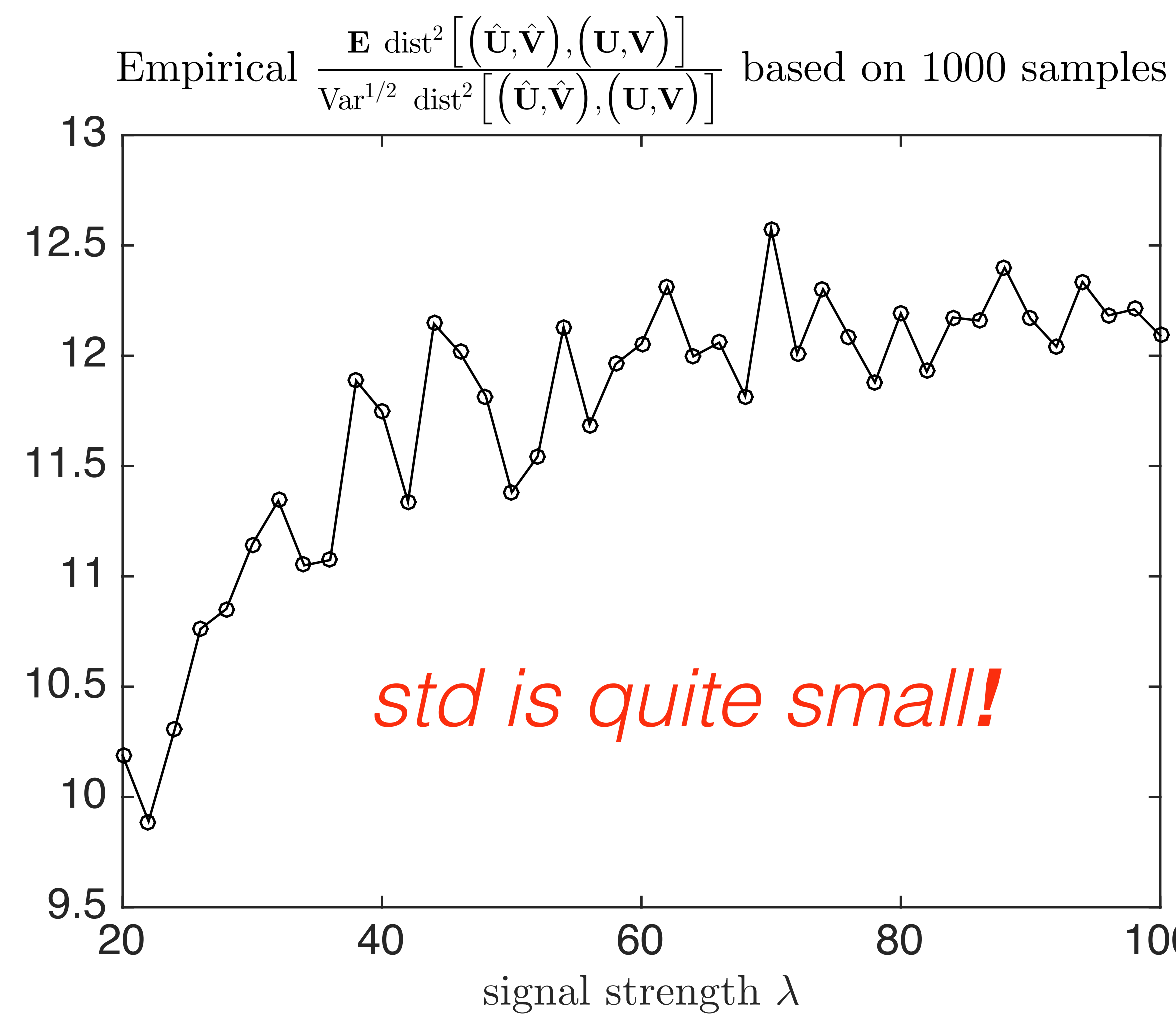


normal?

Normal Approximation of the Loss

$$\frac{\mathrm{dist}^2\big[(\hat{\mathbf{U}},\hat{\mathbf{V}}),(\mathbf{U},\mathbf{V})\big]-\mathbb{E}\,\mathrm{dist}^2\big[(\hat{\mathbf{U}},\hat{\mathbf{V}}),(\mathbf{U},\mathbf{V})\big]}{\mathrm{Var}^{1/2}\mathrm{dist}^2\big[(\hat{\mathbf{U}},\hat{\mathbf{V}}),(\mathbf{U},\mathbf{V})\big]}$$ *is approximately normal**?***





Variance of the Loss

Theorem If $\lambda_r \gg \sqrt{d_{\max}}$, then

$$\mathrm{Var}^{1/2}\mathrm{dist}^2[(\hat{U},\hat{V}),(U,V)] = [\sqrt{8}+o(1)](d_1+d_2-2r)^{1/2}\|\Lambda^{-2}\|_{\mathrm{F}}^2$$

Recall

$$\mathbb{E}\,\mathrm{dist}^2[(\hat{U},\hat{V}),(U,V)] = [2+o(1)](d_1+d_2-2r)\|\Lambda^{-1}\|_{\mathrm{F}}^2$$

🤔 $$\mathbb{E}\,\mathrm{dist}^2\big[(\hat{\mathbf{U}},\hat{\mathbf{V}}),(\mathbf{U},\mathbf{V})\big] \gg \mathrm{Var}^{1/2}\mathrm{dist}^2\big[(\hat{\mathbf{U}},\hat{\mathbf{V}}),(\mathbf{U},\mathbf{V})\big]$$

a factor of $\sqrt{d_{\max}}$

Normal Approximation of the Loss

Theorem If $\dfrac{\sqrt{r d_{\max}}}{\lambda_r} \to 0$ and $\dfrac{r^3}{d_{\max}} \to 0$, then

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left\{ \frac{\mathrm{dist}^2[(\hat{U},\hat{V}),(U,V)] - \mathbb{E}\mathrm{dist}^2[(\hat{U},\hat{V}),(U,V)]}{\sqrt{8d_\star}\|\Lambda^{-2}\|_{\mathrm{F}}} \le x \right\} - \boxed{\Phi(x)} \right| \to 0$$

where $d_\star = d_1 + d_2 - 2r$

c.d.f. of standard normal

convergence rate is of the order $\sqrt{\dfrac{(r d_{\max})^{1/2}}{\lambda_r}} + \sqrt{\dfrac{r^3}{d_{\max}}}$

Normal Approximation of the Loss

Remarks

- Non-asymptotical convergence rates established
- Rank r can diverge as fast as $(d_1 + d_2)^{1/3}$.
- No eigen-gap conditions (except signal strength)

Normal Approximation of the Loss

With explicit centering term

$$\frac{\mathrm{dist}^2[(\hat{U},\hat{V}),(U,V)] - \boxed{\mathbb{E}\mathrm{dist}^2[(\hat{U},\hat{V}),(U,V)]}}{\sqrt{8d_\star}\|\Lambda^{-2}\|_{\mathrm{F}}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1)$$

Replace the expected loss with k-th order approximation ?

$$\Big|\mathbb{E}\mathrm{dist}^2[(\hat{U},\hat{V}),(U,V)] - B_k\Big| = O\Big(\frac{r^2}{\sqrt{d_{\max}}}\cdot\frac{d_{\max}^3}{\lambda_r^6} + \boxed{r\big(\frac{Cd_{\max}}{\lambda_r^2}\big)^{k+1}\Big)}$$

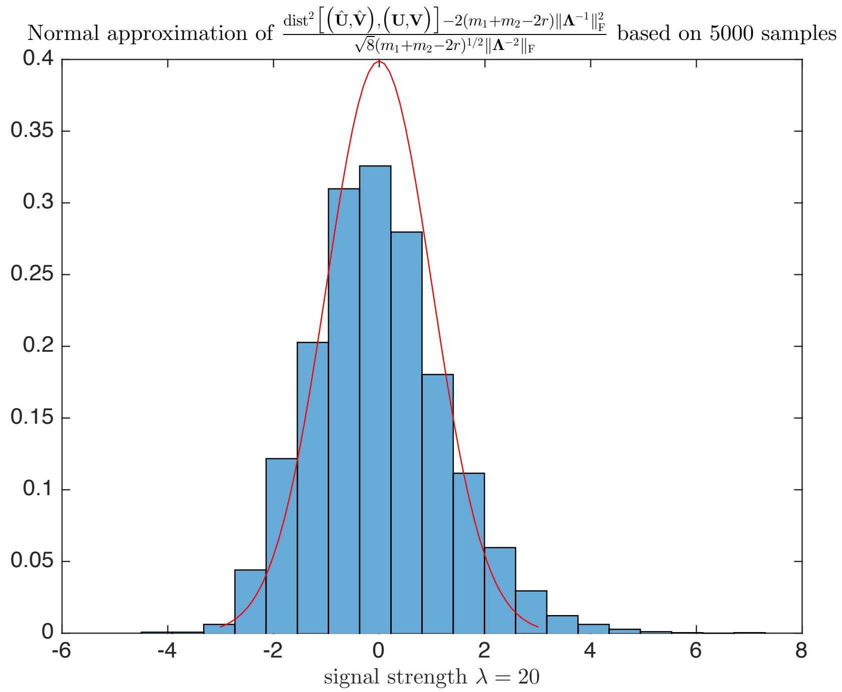
Standard deviation: $\boxed{O\Big(\sqrt{\frac{r}{d_{\max}}\cdot\frac{d_{\max}}{\lambda_r^2}}\Big)}$

need treatments

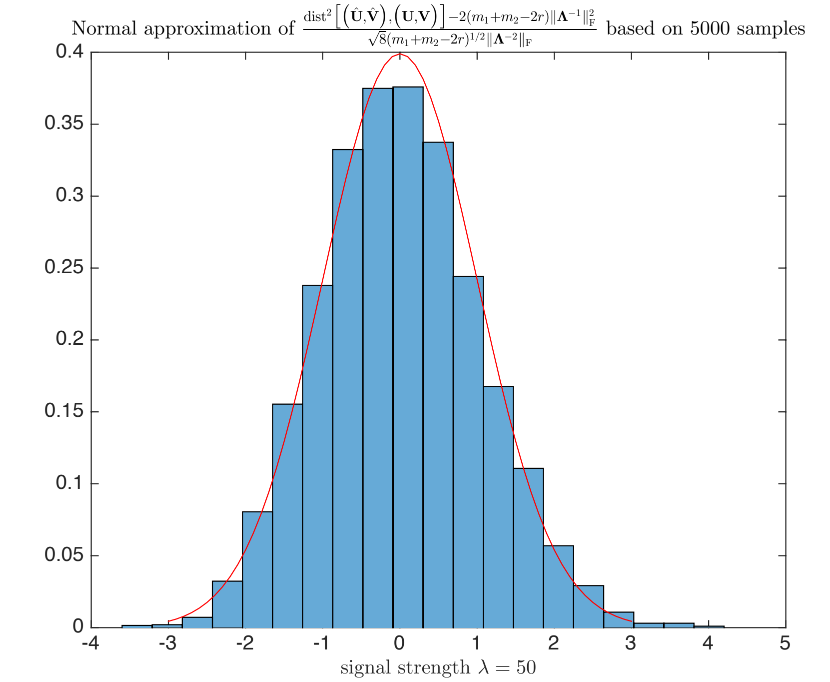
Normal Approximation of the Loss

Normal Approximation of $\frac{\mathrm{dist}^2[(\hat{U},\hat{V}),(U,V)] - B_1}{\sqrt{8d_\star}\|\Lambda^{-2}\|_{\mathrm{F}}}$

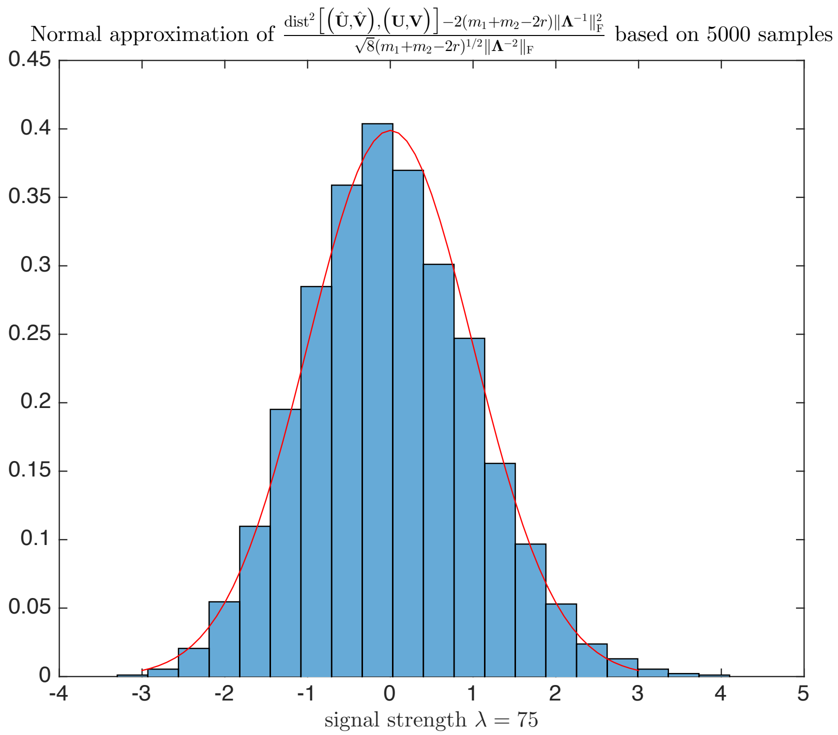
Simulation study

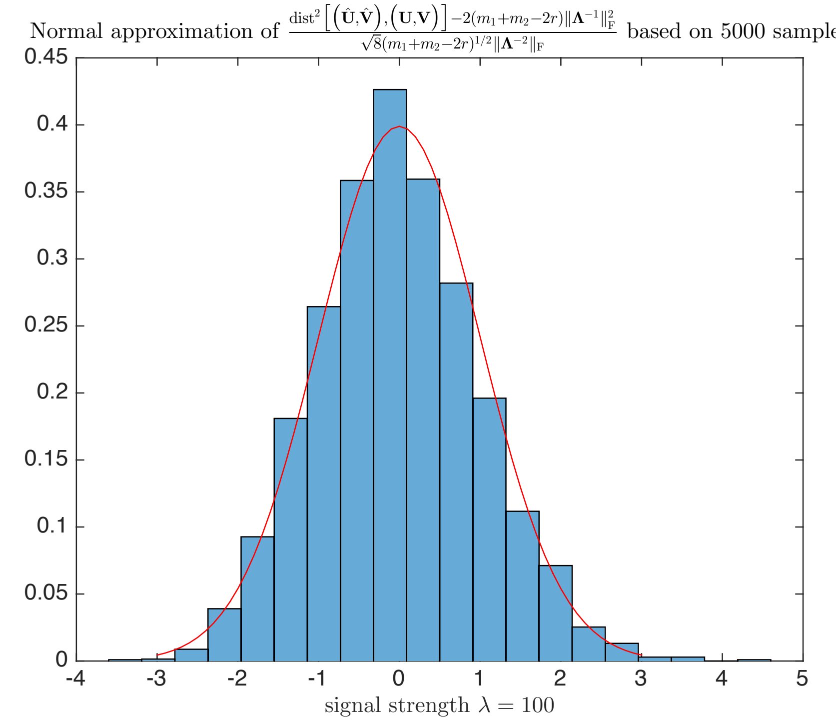


$$d_1 = d_2 = 100$$
$$r = 10$$
$$\lambda_k = \lambda \cdot 2^{r-k}$$



looks good!





Normal Approximation of the Loss

First order CLT

Theorem If $\boxed{\dfrac{\sqrt{r d_{\max}^{3/2}}}{\lambda_r} \to 0}$ and $\dfrac{r^3}{d_{\max}} \to 0$, then

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left\{ \frac{\mathrm{dist}^2[(\hat{U},\hat{V}),(U,V)] - 2d_\star \|\Lambda^{-1}\|_{\mathrm{F}}^2}{\sqrt{8d_\star}\|\Lambda^{-2}\|_{\mathrm{F}}} \le x \right\} - \Phi(x) \right| \to 0$$

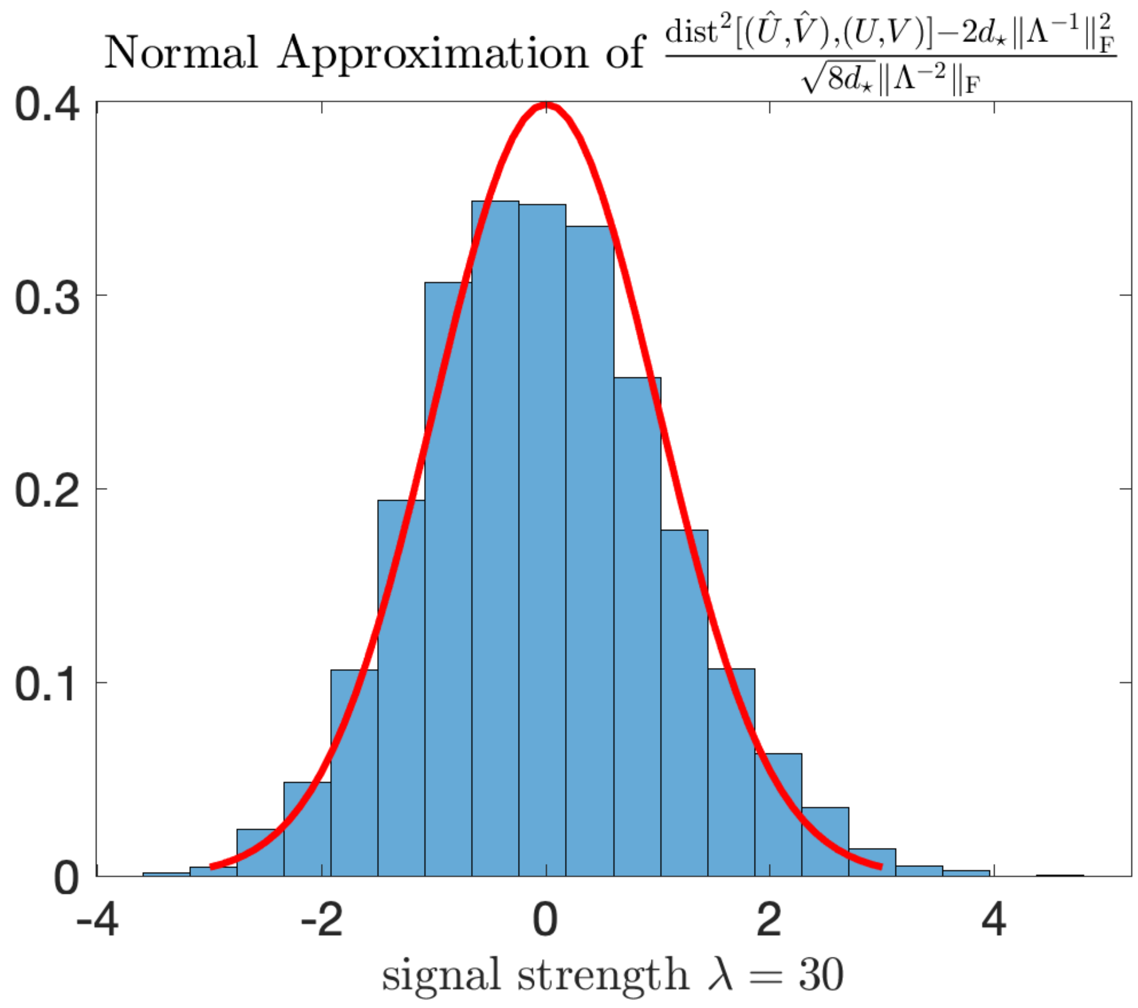
$\lambda_r \gg d_{\max}^{3/4}$ ***instead of*** $\lambda_r \gg \sqrt{d_{\max}}$

🤔 $$\mathbb{E}\ \mathrm{dist}^2[(\hat{U},\hat{V}),(U,V)] = 2d_\star\|\Lambda^{-1}\|_{\mathrm{F}}^2 + \boxed{O\Big(\frac{r\bar{d}^2}{\lambda_r^4}\Big)}$$

Normal Approximation of the Loss

First order CLT

Simulation study



$$d_1 = d_2 = 100$$
$$r = 10$$
$$\lambda_k = \lambda \cdot 2^{r-k}$$

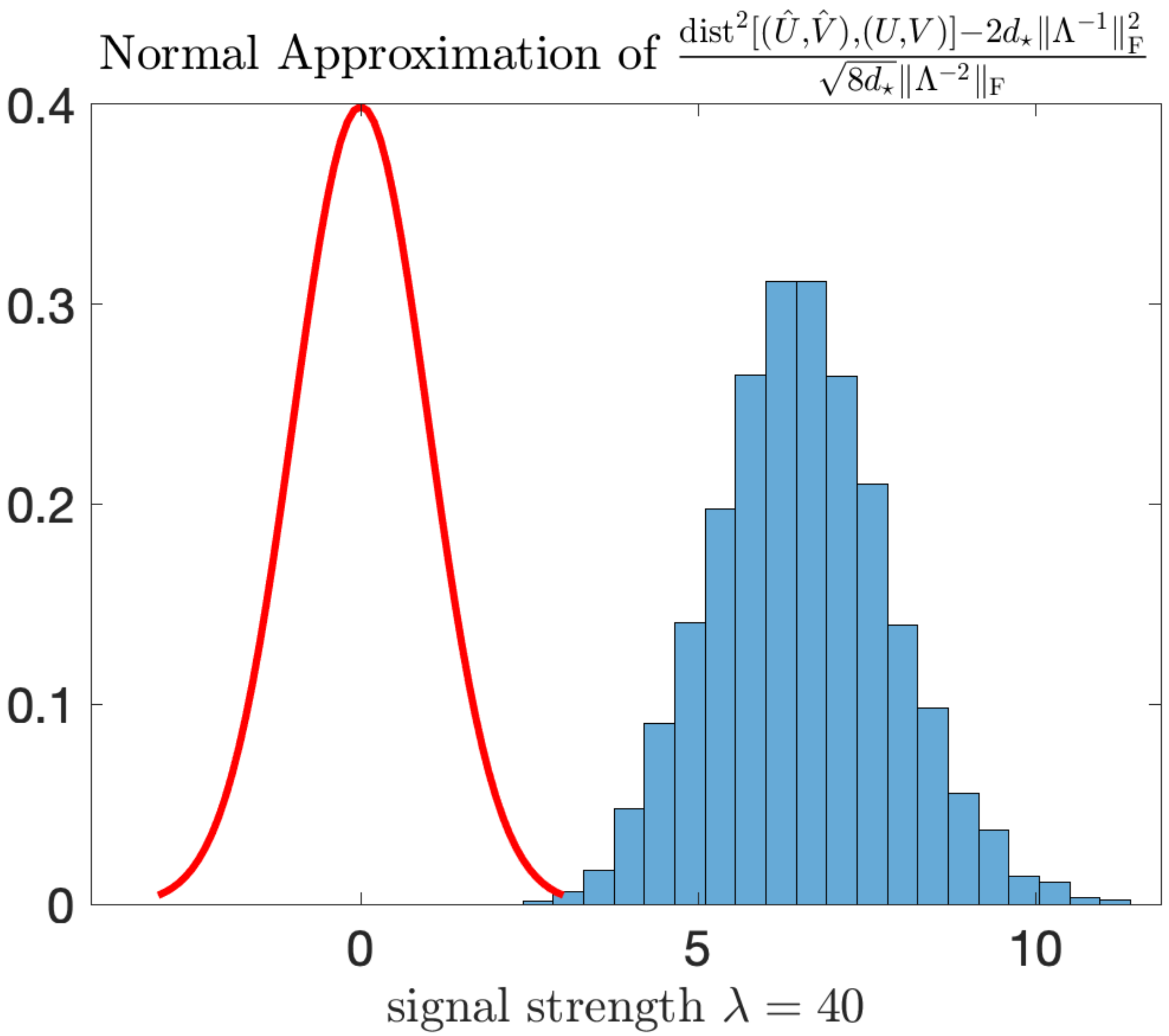
Looks good when

$\lambda_r = 30$ **!**

Normal Approximation of the Loss

First order CLT

Simulation study



$$d_1 = \frac{d_2}{2} = 100$$

$$r = 10$$

$$\lambda_k = \lambda \cdot 2^{r-k}$$

Not good at all !

Higher order approx. are necessary!

Normal Approximation of the Loss

Second order CLT

Theorem If $\boxed{\frac{\sqrt{r d_{\max}^{5/4}}}{\lambda_r} \to 0}$ and $\frac{r^3}{d_{\max}} \to 0$, then

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left\{\frac{\mathrm{dist}^2\big[(\hat{U},\hat{V}),(U,V)\big]-B_2}{\sqrt{8d_\star}\|\Lambda^{-2}\|_{\mathrm{F}}}\le x\right\}-\Phi(x)\right|\to 0$$

where $B_2 = 2(d_\star\|\Lambda^{-1}\|_{\mathrm{F}}^2 - \Delta_d^2\|\Lambda^{-2}\|_{\mathrm{F}}^2)$.

$\lambda_r \gg d_{\max}^{5/8}$ ***instead of*** $\lambda_r \gg \sqrt{d_{\max}}$

Normal Approximation of the Loss

K-th order CLT

Theorem If $\frac{\sqrt{rd_{\max}} + \sqrt{d_{\max}}(r^2 d_{\max})^{1/4k}}{\lambda_r} \to 0$ and $\frac{r^3}{d_{\max}} \to 0$, then

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left\{ \frac{\text{dist}^2\big[(\hat{U},\hat{V}),(U,V)\big] - B_k}{\sqrt{8d_\star}\|\Lambda^{-2}\|_{\mathrm{F}}} \le x \right\} - \Phi(x) \right| \to 0$$

with $B_k = 2(d_{1-} + d_{2-})\|\Lambda^{-1}\|_{\mathrm{F}}^2 - 2\sum_{k_0=2}^{k}(-1)^{k_0}(d_{1-}^{k_0-1} - d_{2-}^{k_0-1})(d_1 - d_2)\|\Lambda^{-k_0}\|_{\mathrm{F}}^2$

By choosing $k = \lceil \log(d_1 + d_2) \rceil$, we require $\lambda_r \gg \sqrt{d_{\max}}$

Normal Approximation of the Loss

∞ -th order CLT

Theorem If $\dfrac{\sqrt{rd_{\max}}}{\lambda_r} \to 0$ and $\dfrac{r^3}{d_{\max}} \to 0$, then

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left\{ \frac{\mathrm{dist}^2\big[(\hat{U},\hat{V}),(U,V)\big] - B_\infty}{\sqrt{8d_\star}\|\Lambda^{-2}\|_{\mathrm{F}}} \le x \right\} - \Phi(x) \right| \to 0$$

with $$B_\infty = 2\sum_{j=1}^{r} \frac{1}{\lambda_j^2}\left(d_{1-}\cdot \frac{\lambda_j^2 + d_{2-}}{\lambda_j^2 + d_{1-}} + d_{2-}\cdot \frac{\lambda_j^2 + d_{1-}}{\lambda_j^2 + d_{2-}} \right)$$

convergence rate is of the order $$\sqrt{\frac{(rd_{\max})^{1/2}}{\lambda_r}} + \sqrt{\frac{r^3}{d_{\max}}}$$

Confidence Region of Singular Subspace

Outline

- Representation of Spectral Projectors
- Normal Approximation of Spectral Projectors
- Data-dependent Confidence Regions

Confidence Region of Singular Subspaces

$\lceil \log(d_1+d_2) \rceil$-th order CR

Corollary

For any $\alpha \in (0,1)$, the $1-\alpha$ probability interval of $\text{dist}^2\big[(\hat{\mathbf{U}},\hat{\mathbf{V}}),(\mathbf{U},\mathbf{V})\big]$ is

$$\boxed{B_{\lceil \log(d_1+d_2) \rceil}} + [-z_{\alpha/2}, z_{\alpha/2}] * \sqrt{8d_\star}\boxed{\|\Lambda^{-2}\|_{\mathrm{F}}}$$

where $z_\alpha = \Phi^{-1}(1-\alpha)$.

$\lambda_1, \cdots, \lambda_r$ *are unknown*

$\hat{\mathbf{M}} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^\top + \mathbf{Z}$ *is known*

Confidence Region of Singular Subspaces

Plug-in estimates of singular values

$$\|\hat{\Lambda}^{-1}\|_{\mathrm{F}}^2 = \hat{\lambda}_1^{-2} + \cdots + \hat{\lambda}_r^{-2}$$

$$\|\hat{\Lambda}^{-2}\|_{\mathrm{F}}^2 = \hat{\lambda}_1^{-4} + \cdots + \hat{\lambda}_r^{-4}$$

$$\vdots$$

$$\|\hat{\Lambda}^{-k}\|_{\mathrm{F}}^2 = \hat{\lambda}_1^{-2k} + \cdots + \hat{\lambda}_r^{-2k}$$

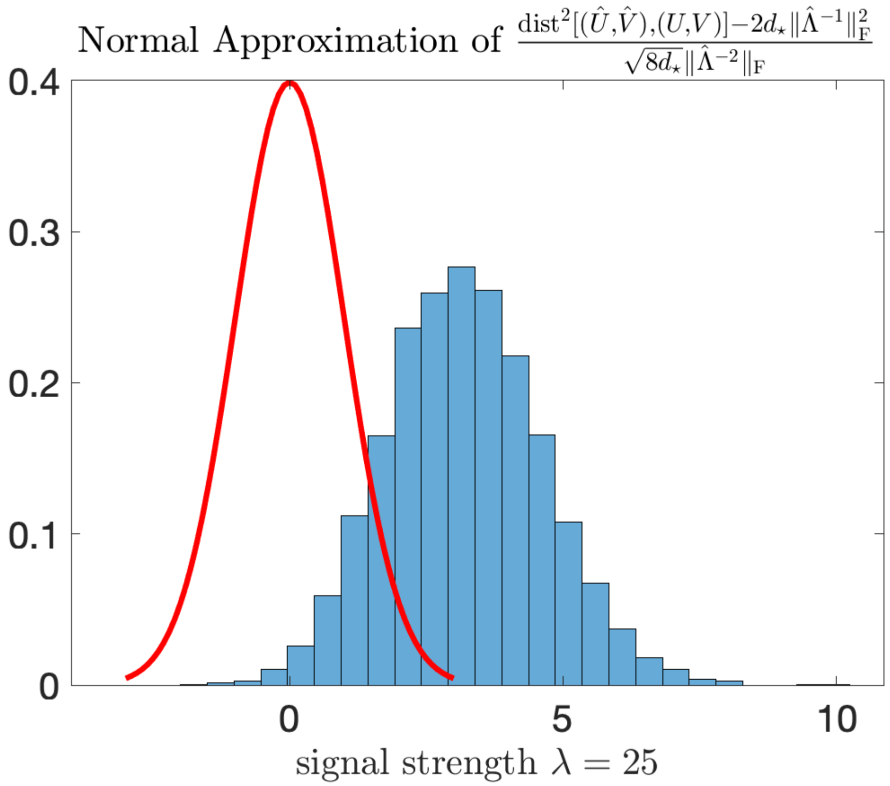
where $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_r$ are $\hat{\mathbf{M}}'s$ top-r singular values.

Are they good?

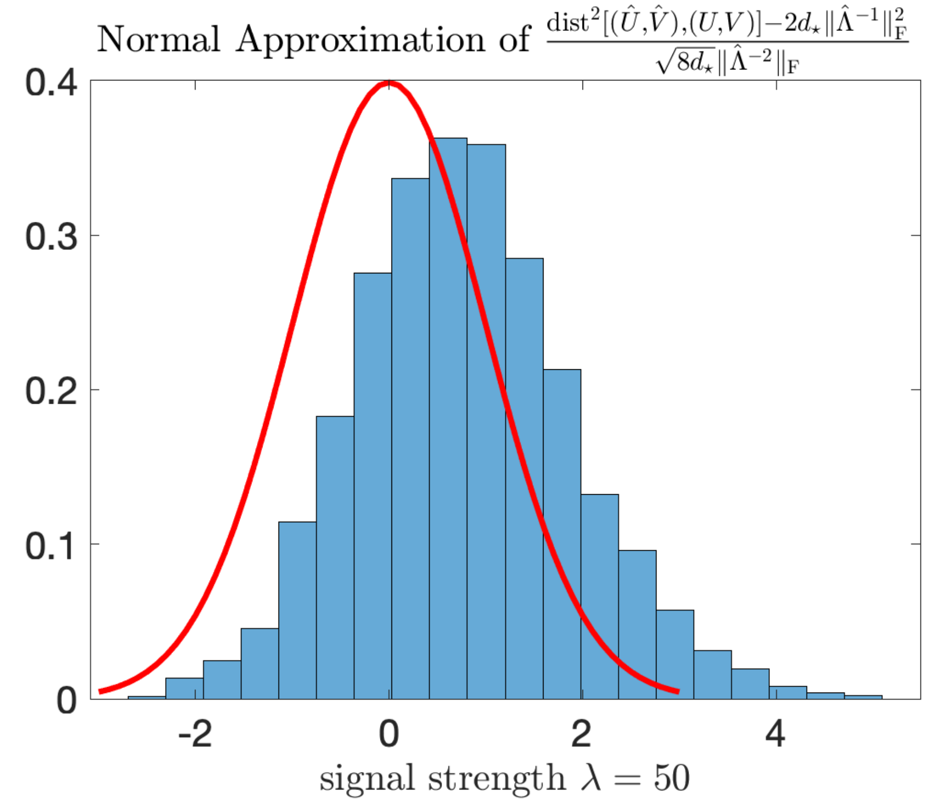
Confidence Region of Singular Subspaces

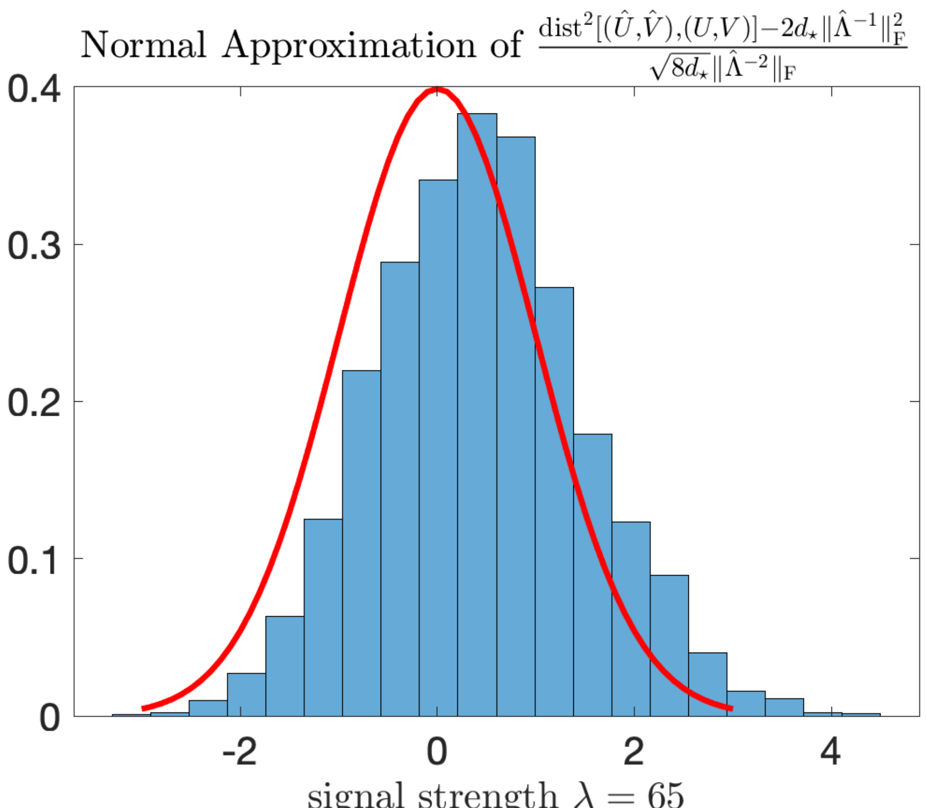
Simulation

Normal Approximation of $\frac{\mathrm{dist}^2[(\hat{U},\hat{V}),(U,V)] - 2d_\star\|\hat{\Lambda}^{-1}\|_{\mathrm{F}}^2}{\sqrt{8d_\star}\|\hat{\Lambda}^{-2}\|_{\mathrm{F}}}$

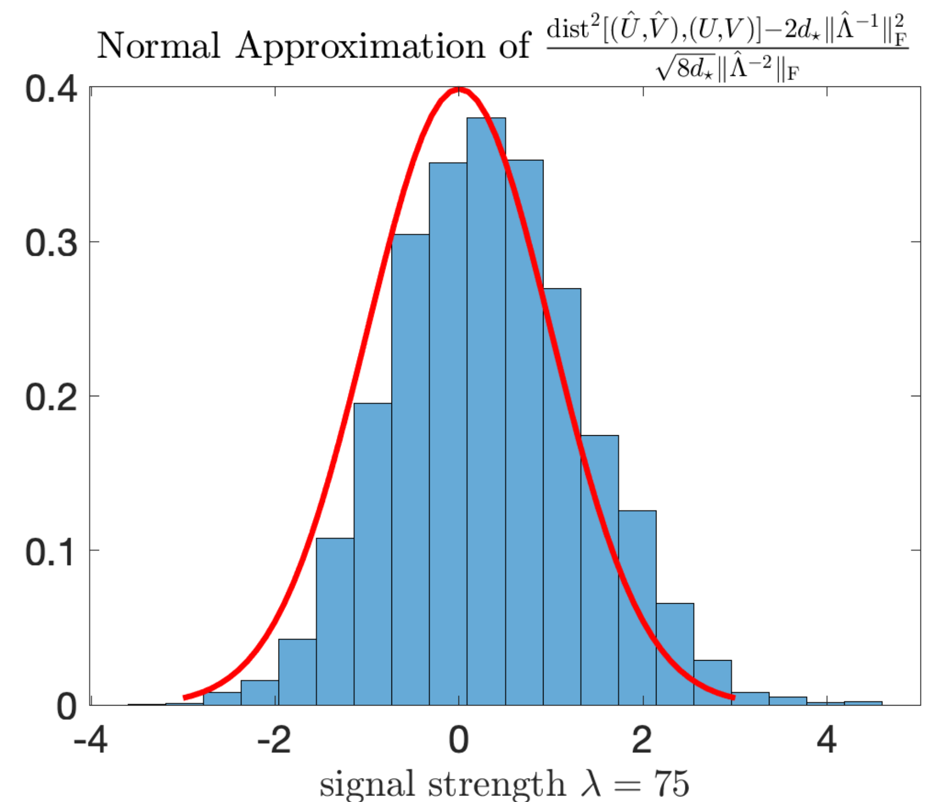


$$d_1 = d_2 = 100$$
$$r = 6$$
$$\lambda_k = \lambda \cdot 2^{r-k}$$





Not good!



Shrinkage Estimates of Singular Values

By random matrix theory (Ding, 2017, arxiv:1702.06975)

$$\left|\hat{\lambda}_j^2 - \left(\lambda_j^2 + (d_1 + d_2) + \frac{d_1 d_2}{\lambda_j^2}\right)\right| = O_P\left(d_{\max}^{1/4}\sqrt{\lambda_j}\right)$$

Define shrinkage estimators

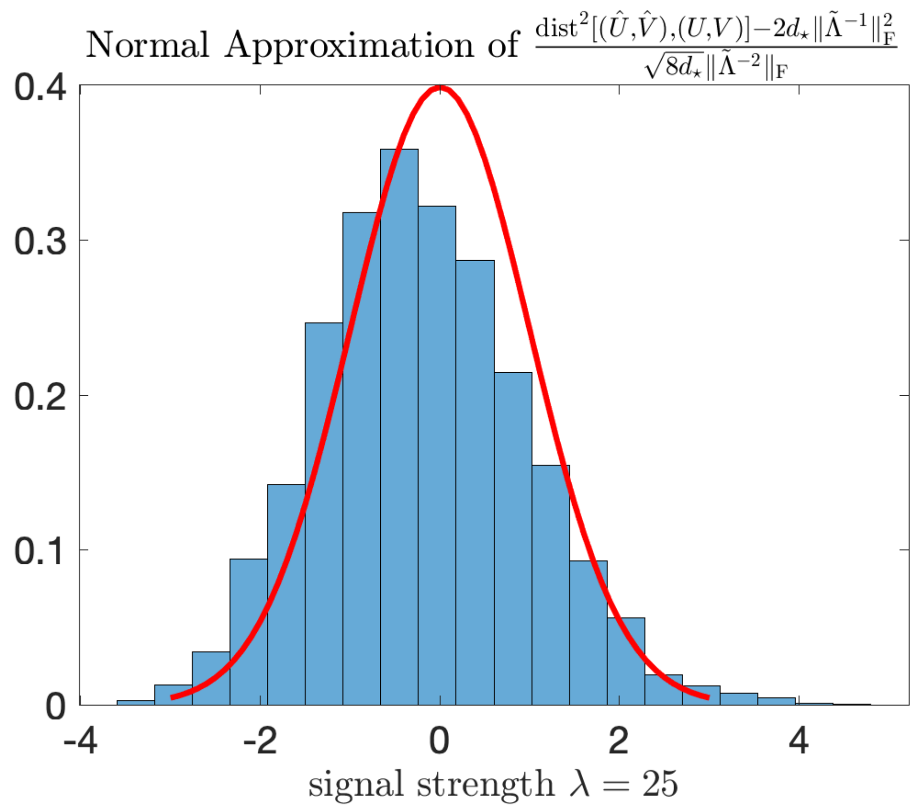
$$\tilde{\lambda}_j^2 = \frac{\hat{\lambda}_j^2 - d_1 - d_2}{2} + \frac{\sqrt{(\hat{\lambda}_j^2 - d_1 - d_2)^2 - 4d_1 d_2}}{2}$$

Denote

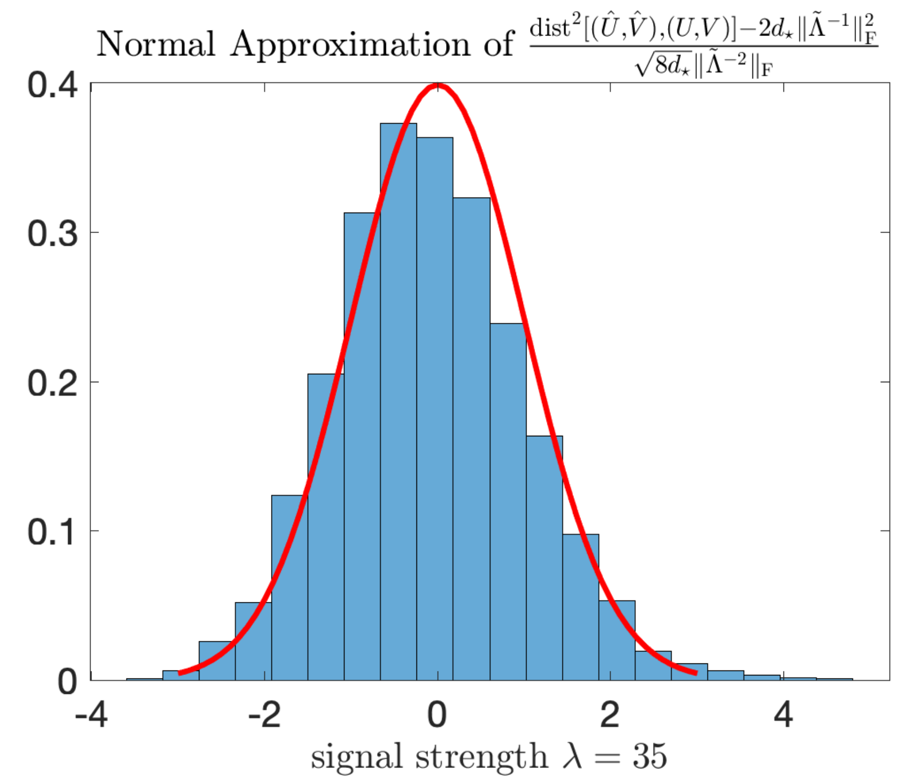
$$\boxed{\tilde{\Lambda} = \mathrm{diag}(\tilde{\lambda}_1, \cdots, \tilde{\lambda}_r)}$$

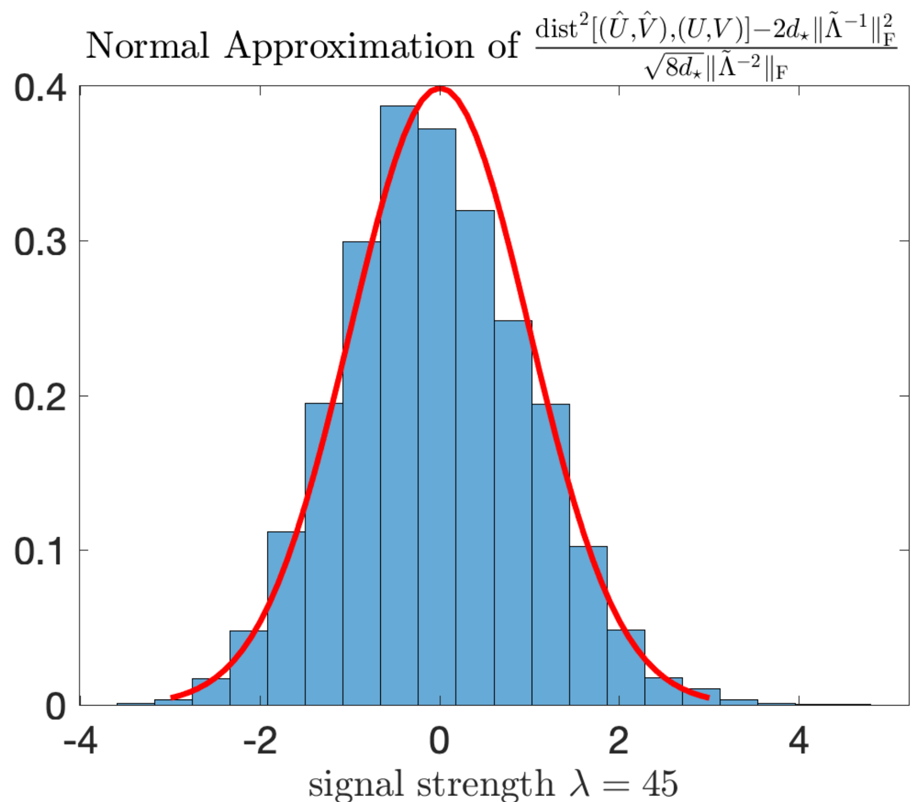
Confidence Region of Singular Subspaces

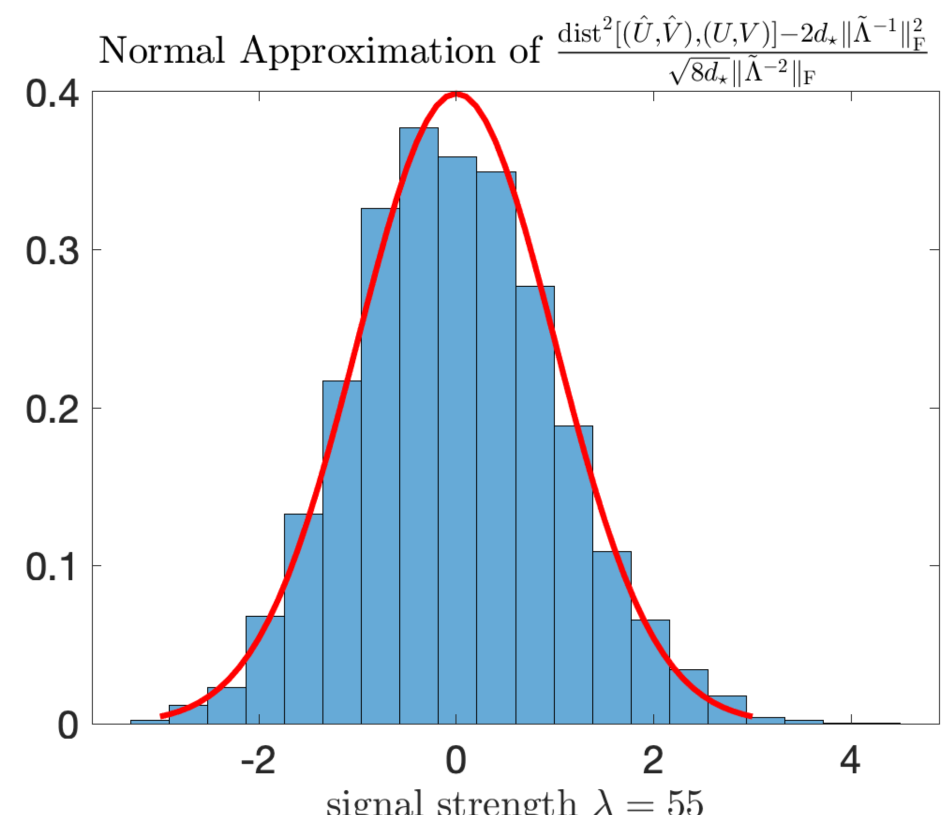
Simulation **Normal Approximation of** $\frac{\mathrm{dist}^2[(\hat{U},\hat{V}),(U,V)] - 2d_\star\|\tilde{\Lambda}^{-1}\|_F^2}{\sqrt{8d_\star}\|\tilde{\Lambda}^{-2}\|_F}$



$$d_1 = d_2 = 100$$
$$r = 6$$
$$\lambda_k = \lambda \cdot 2^{r-k}$$



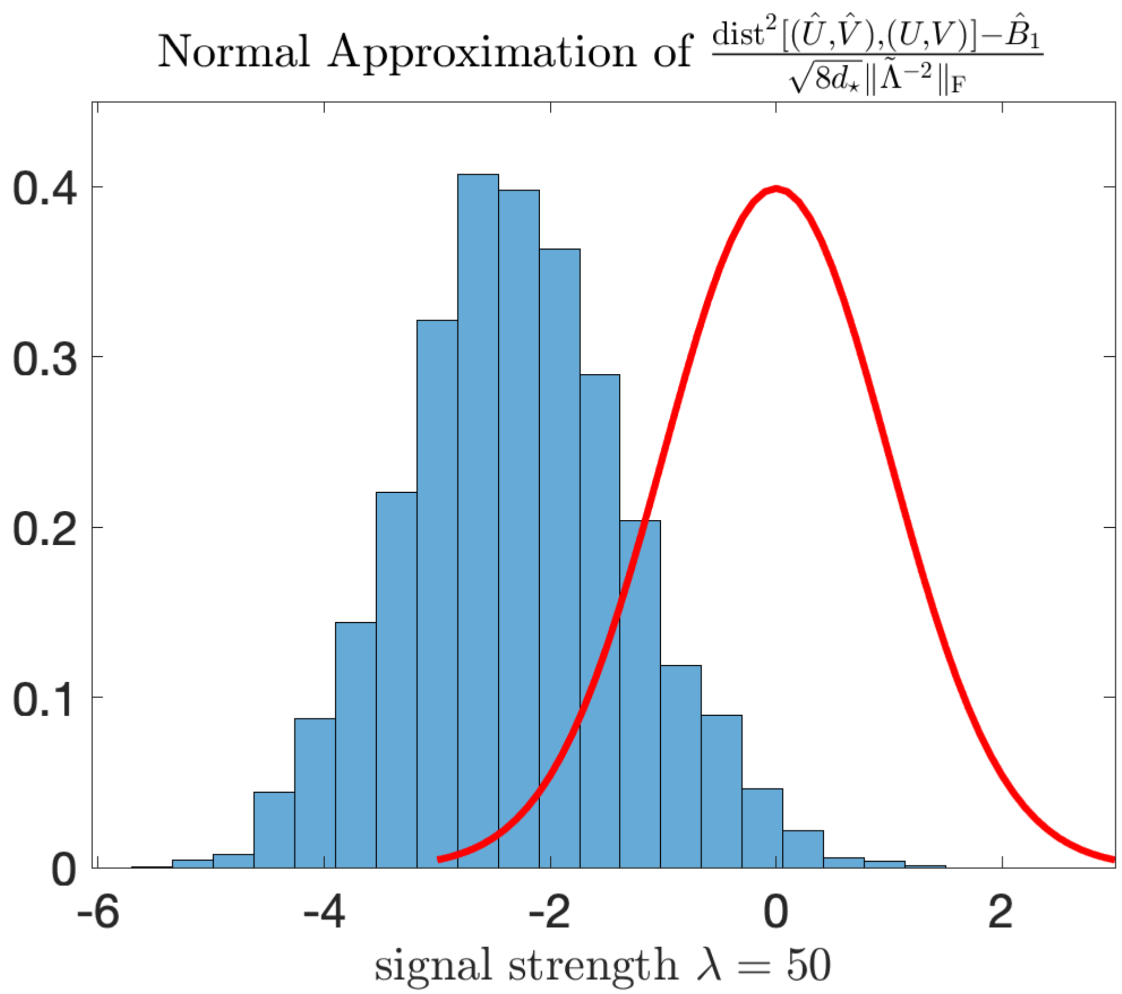




Confidence Region of Singular Subspaces

Normal Approximation of $\dfrac{\mathrm{dist}^2[(\hat{U},\hat{V}),(U,V)] - 2d_\star\|\tilde{\Lambda}^{-1}\|_{\mathrm{F}}^2}{\sqrt{8d_\star}\|\tilde{\Lambda}^{-2}\|_{\mathrm{F}}}$

what if $|d_1 - d_2| \gg 0$ **?**



$$d_1 = \frac{d_2}{6} = 100$$

$$r = 6$$

$$\lambda_k = \lambda \cdot 2^{r-k}$$

1st order approx $\hat{B}_1 = 2d_\star\|\tilde{\Lambda}^{-1}\|_{\mathrm{F}}^2$

Try higher order bias corrections !

Confidence Region of Singular Subspaces

Higher order empirical bias corrections

$$\Delta_d = d_1 - d_2$$

1-st $$\hat{B}_1 = 2d_\star \|\tilde{\Lambda}^{-1}\|_{\mathrm{F}}^2$$

2-nd $$\hat{B}_2 = 2\big(d_\star \|\tilde{\Lambda}^{-1}\|_{\mathrm{F}}^2 - \Delta_d^2 \|\tilde{\Lambda}^{-2}\|_{\mathrm{F}}^2\big)$$

$$\vdots$$

k-th $$\hat{B}_k = 2(d_{1-} + d_{2-})\|\tilde{\Lambda}^{-1}\|_{\mathrm{F}}^2 - 2\sum_{k_0=2}^{k}(-1)^{k_0}(d_{1-}^{k_0-1} - d_{2-}^{k_0-1})(d_1 - d_2)\|\tilde{\Lambda}^{-k_0}\|_{\mathrm{F}}^2$$

Confidence Region of Singular Subspaces

Simulation

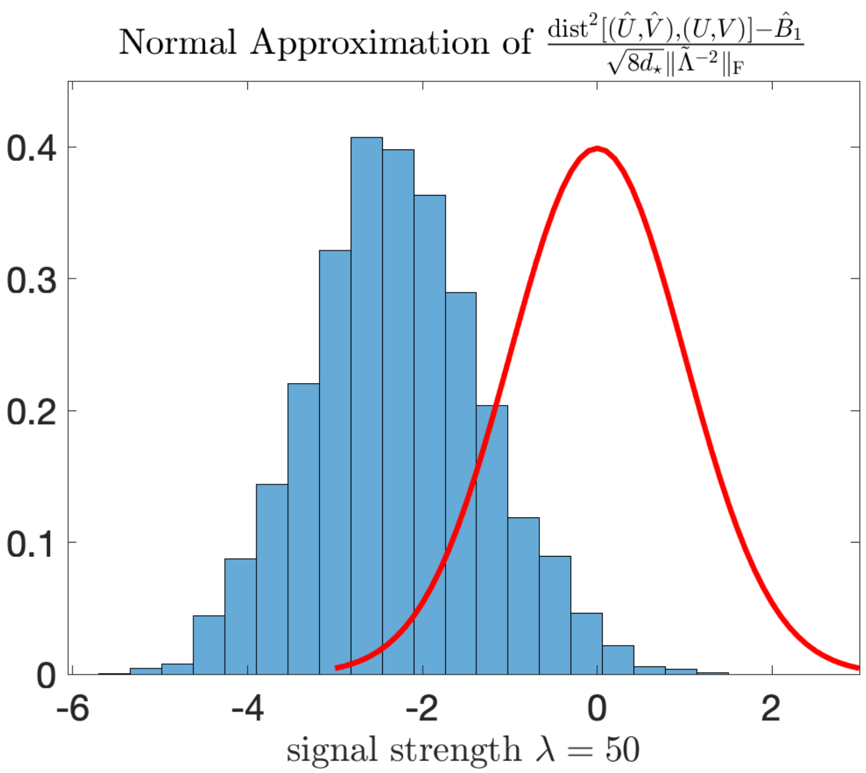
$$d_1 = \frac{d_2}{6} = 100$$

$$r = 6$$

$$\lambda_k = \lambda \cdot 2^{r-k}$$

Confidence Region of Singular Subspaces

Simulation



1st

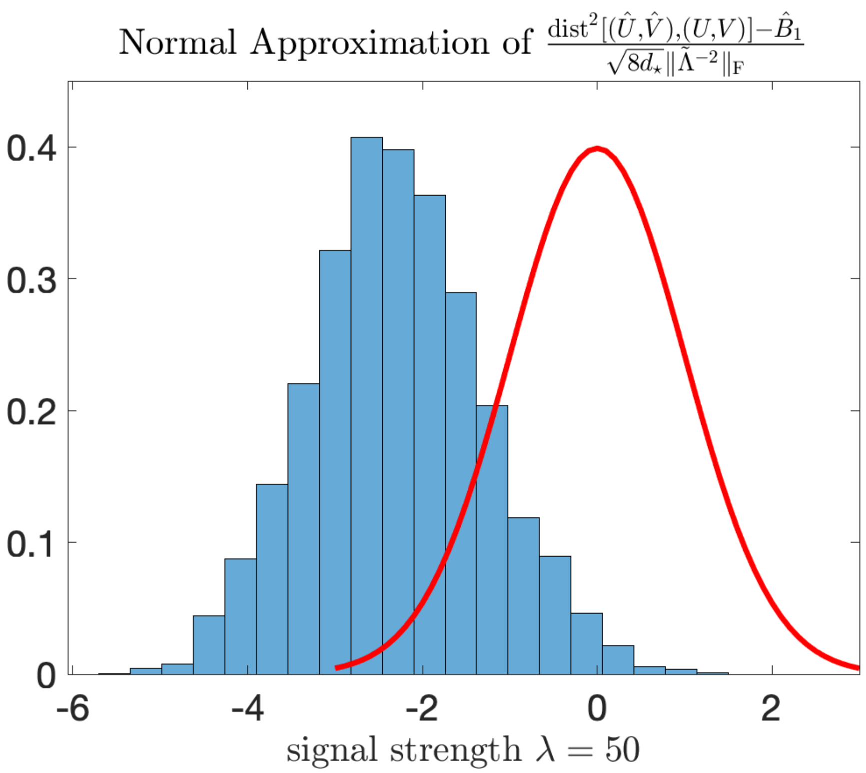
$$d_1 = \frac{d_2}{6} = 100$$

$$r = 6$$

$$\lambda_k = \lambda \cdot 2^{r-k}$$

Confidence Region of Singular Subspaces

Simulation



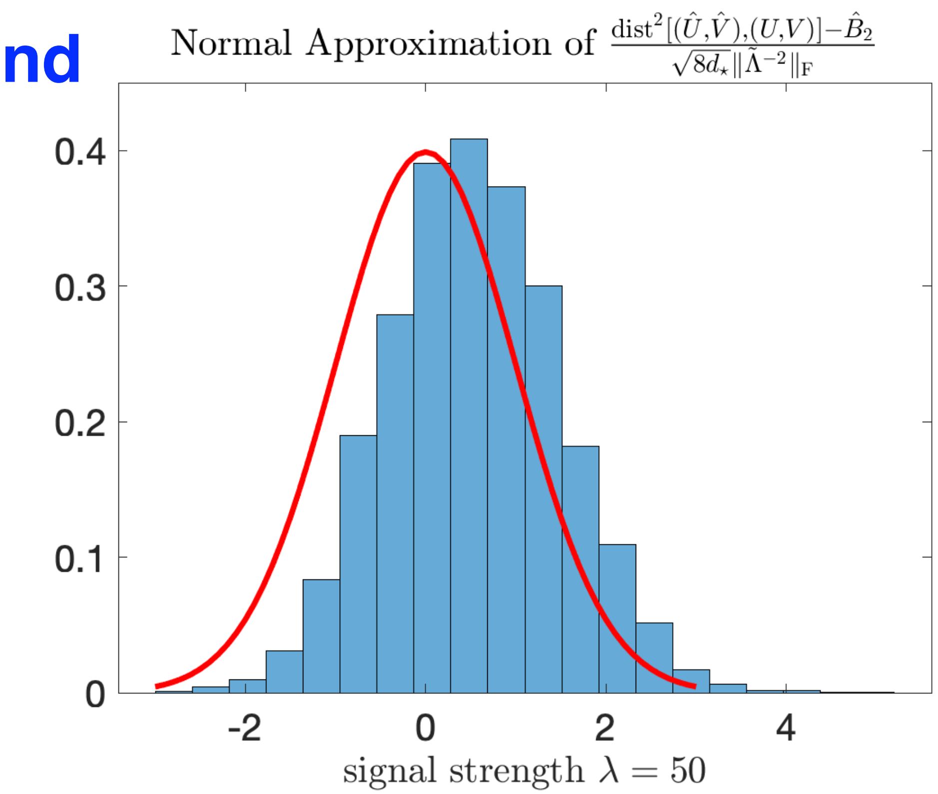
1st

$$d_1 = \frac{d_2}{6} = 100$$

$$r = 6$$

$$\lambda_k = \lambda \cdot 2^{r-k}$$

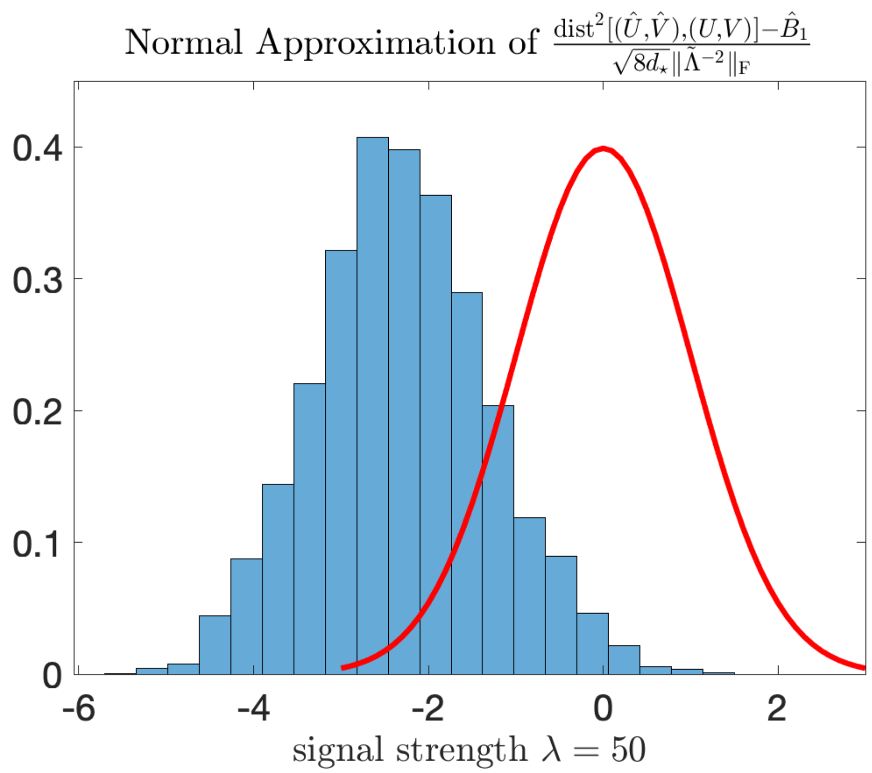
2nd



Confidence Region of Singular Subspaces

Simulation

1st

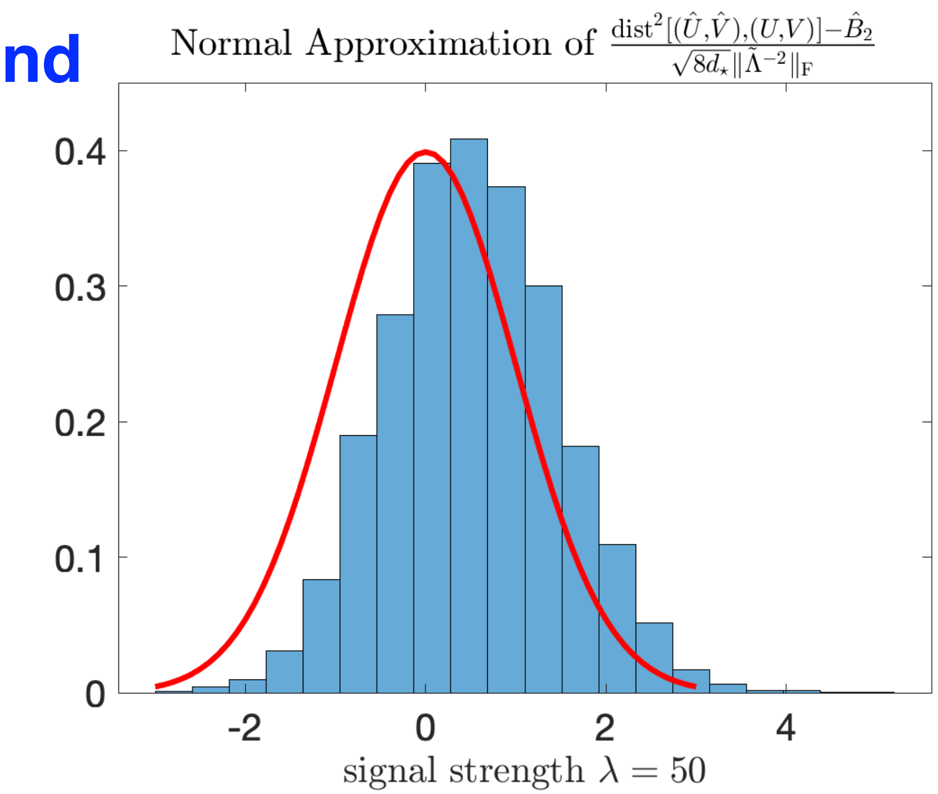


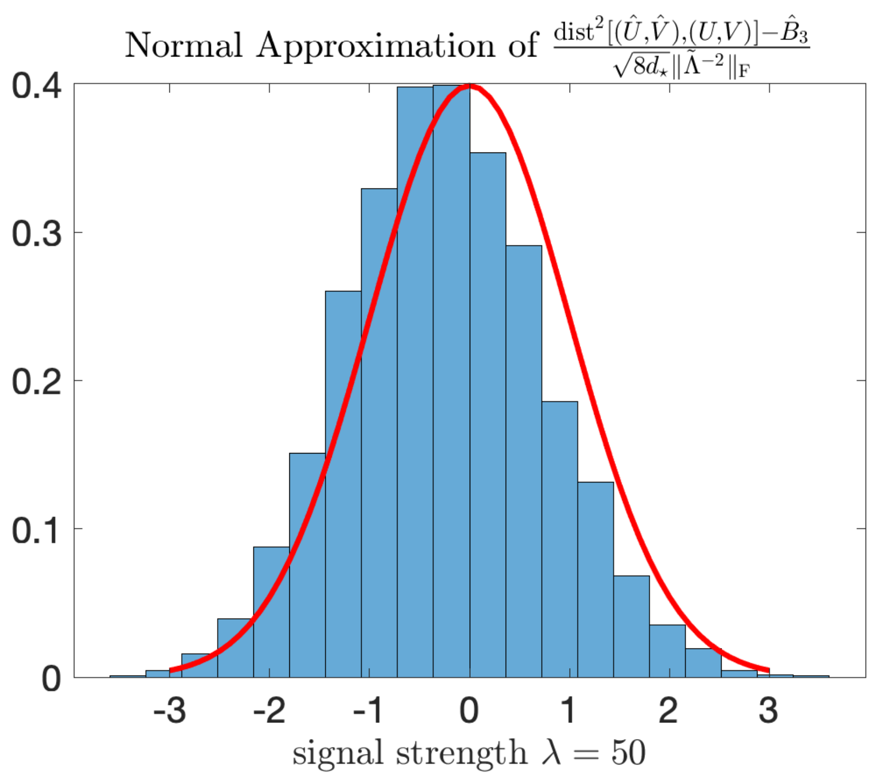
$$d_1 = \frac{d_2}{6} = 100$$

$$r = 6$$

$$\lambda_k = \lambda \cdot 2^{r-k}$$

2nd



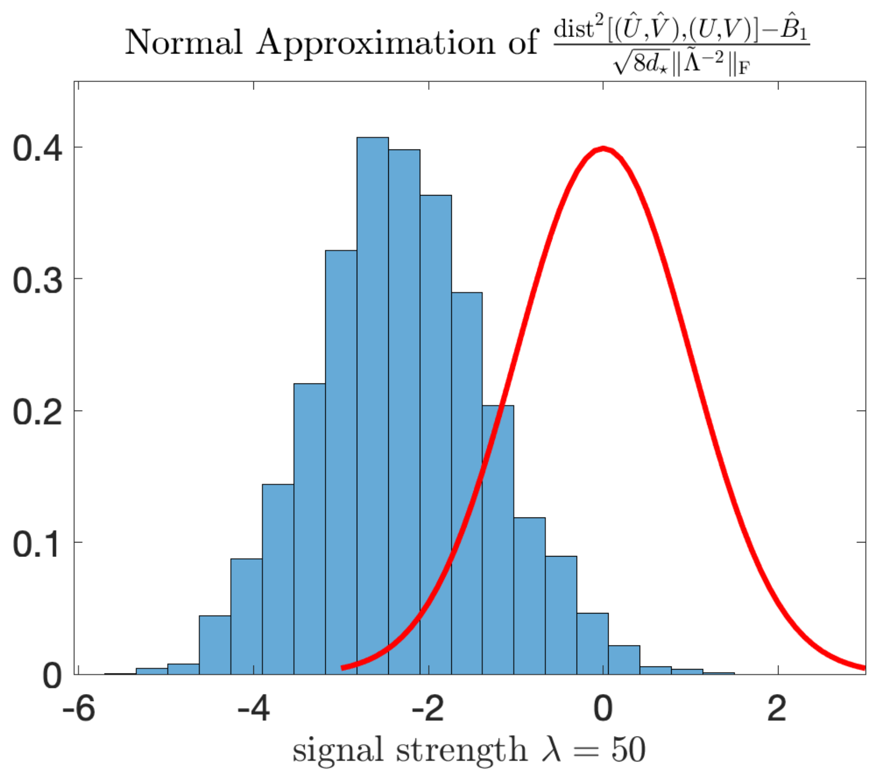


3rd

Confidence Region of Singular Subspaces

Simulation

1st

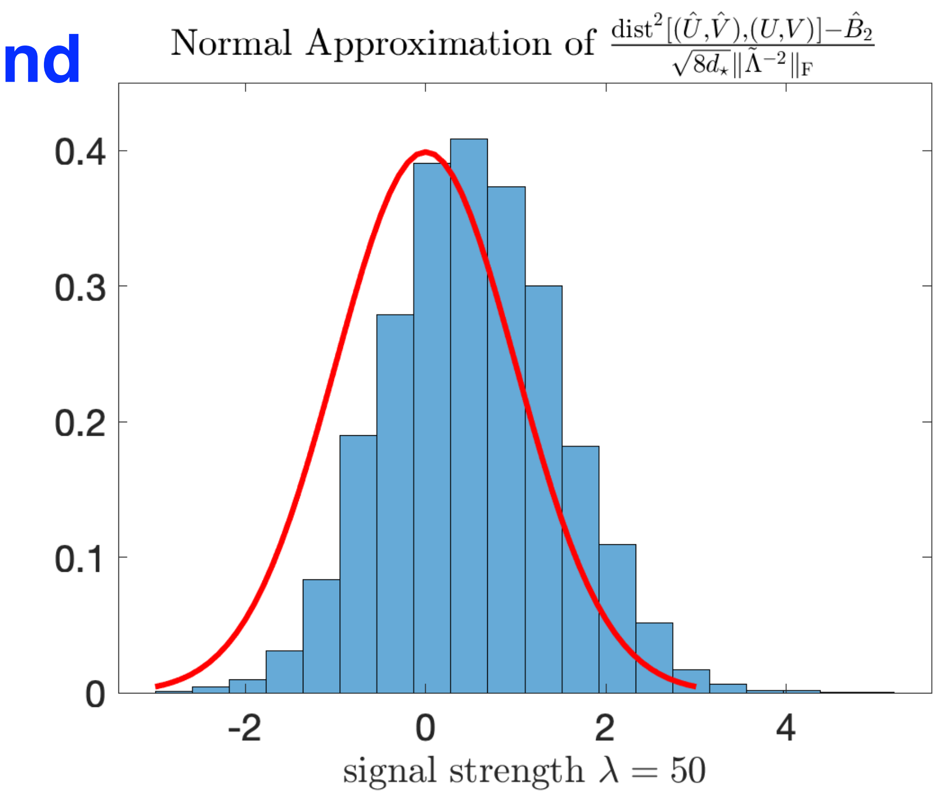


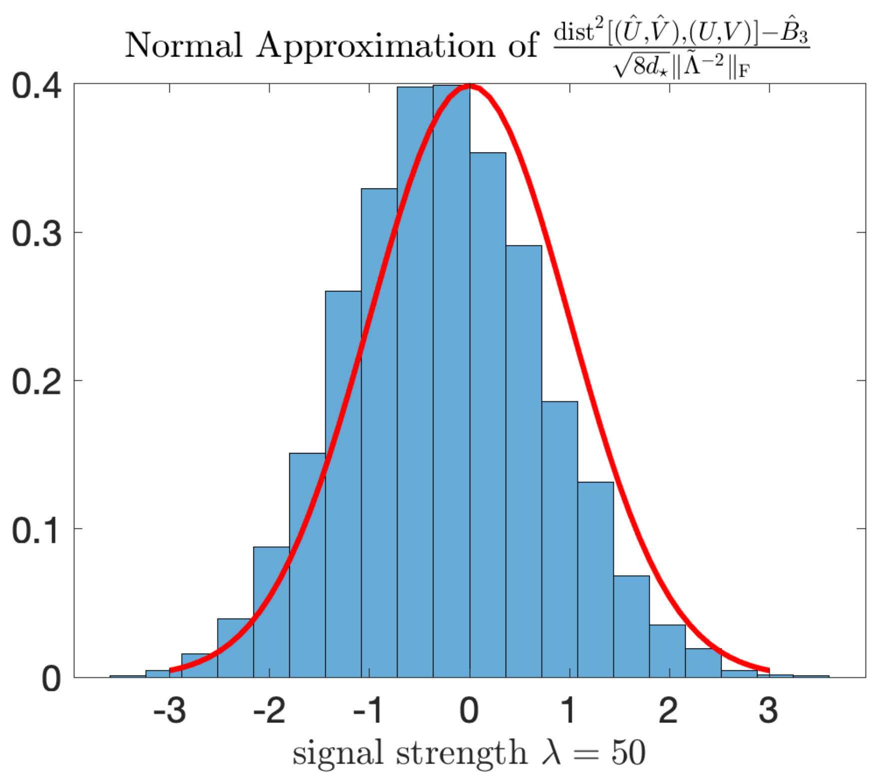
$$d_1 = \frac{d_2}{6} = 100$$

$$r = 6$$

$$\lambda_k = \lambda \cdot 2^{r-k}$$

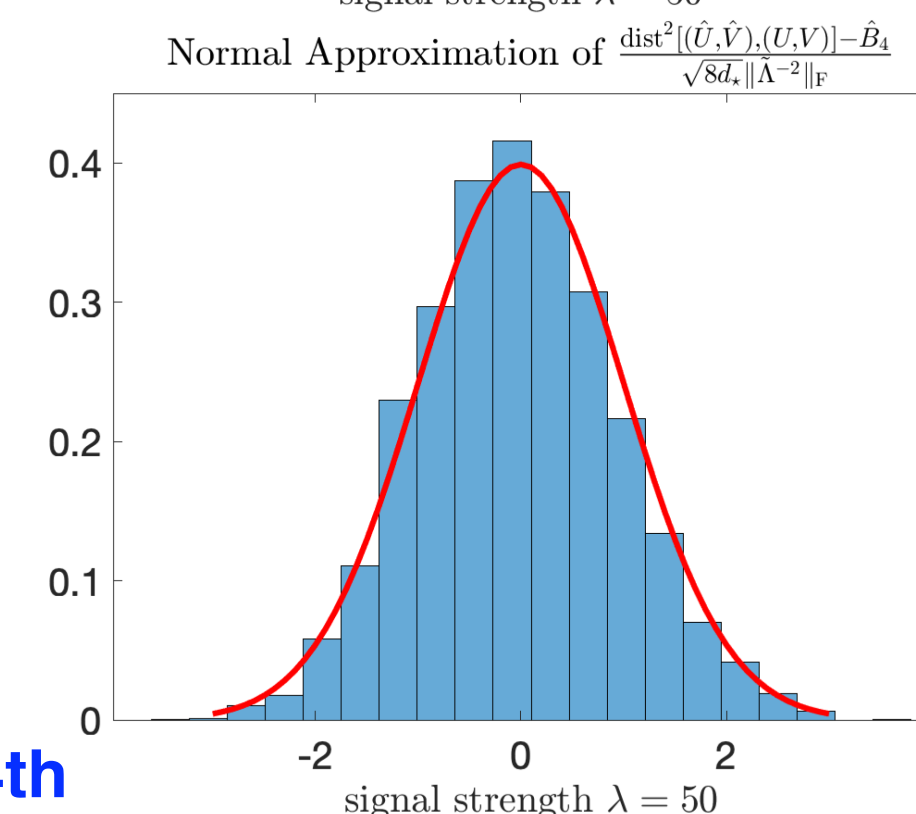
2nd





3rd

Higher order correction works!



4th

Confidence Region of Singular Subspaces

Data-dependent normal approximation

Theorem If $\frac{\sqrt{r d_{\max}}}{\lambda_r} \to 0$ and $\frac{r^3}{d_{\max}} \to 0$, then

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left\{\frac{\mathrm{dist}^2\big[(\hat{U},\hat{V}),(U,V)\big]-\hat{B}_k}{\sqrt{8d_\star}\|\tilde{\Lambda}^{-2}\|_{\mathrm{F}}}\le x\right\}-\Phi(x)\right|\to 0.$$

where

$$\hat{B}_k = 2(d_{1-}+d_{2-})\|\tilde{\Lambda}^{-1}\|_{\mathrm{F}}^2 - 2\sum_{k_0=2}^{k}(-1)^{k_0}(d_{1-}^{k_0-1}-d_{2-}^{k_0-1})(d_1-d_2)\|\tilde{\Lambda}^{-k_0}\|_{\mathrm{F}}^2$$

as long as $k \ge \lceil \log(d_1+d_2)\rceil$.

Confidence Region of Singular Subspaces

Data-dependent confidence region

For any $\alpha \in (0,1)$, the $1-\alpha$ confidence region of $\mathbf{U}$ and $\mathbf{V}$ is

$$\boxed{\begin{aligned}\mathcal{M}_\alpha(\hat{U},\hat{V}) = \Big\{&(L,R): L \in \mathbb{R}^{d_1\times r}, R \in \mathbb{R}^{d_2\times r}, L^\top L = R^\top R = I_r \\ &, \big|\mathrm{dist}^2[(L,R),(\hat{U},\hat{V})] - \hat{B}_k\big| \le \sqrt{8d_\star} z_{\alpha/2}\|\tilde{\Lambda}^{-2}\|_{\mathrm{F}}\Big\}\end{aligned}}$$

If $\dfrac{\sqrt{r d_{\max}}}{\lambda_r} \to 0$ and $\dfrac{r^3}{d_{\max}} \to 0$, then as long as $k \ge \lceil \log(d_1+d_2) \rceil$.

$$\lim_{\bar{d}\to\infty} \mathbb{P}\big((U,V) \in \mathcal{M}_\alpha(\hat{U},\hat{V})\big) = \alpha$$

Confidence Region of Singular Subspaces

Conclusions

- Deterministic representation of singular vectors
- Non-asymptotical normal approximation with convergence rates
- Rank r can diverge as fast as $(d_1 + d_2)^{1/3}$.
- No eigen-gap conditions (except signal strength)
- Data-dependent confidence regions


Thank
You