

Weighted Low Rank Approximations

Nathan Srebro and Tommi Jaakkola

Computer Science and Artificial Intelligence Laboratory

Massachusetts Institute of Technology

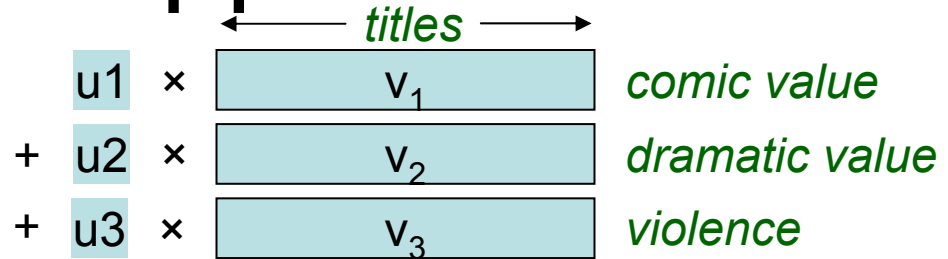
Weighted Low Rank Approximations

- What is a 'weighted low rank approximation'?
 - What is a 'low rank approximation'?
- Why weighted low rank approximations?
- How do we find a weighted low rank approximation?
- What can we do with weighted low rank approximations?

Low Rank Approximation

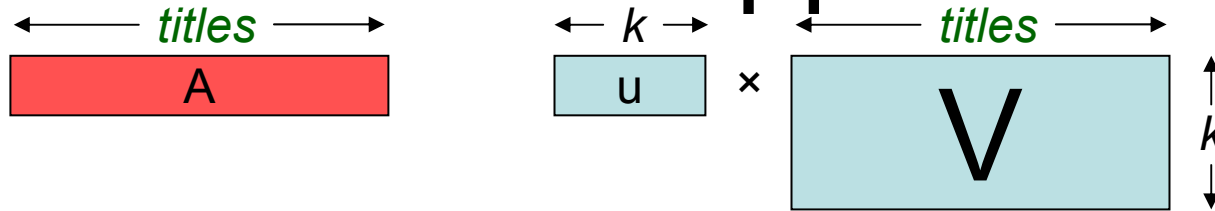


*preferences of a specific user
(real-valued preference level
for each title)*

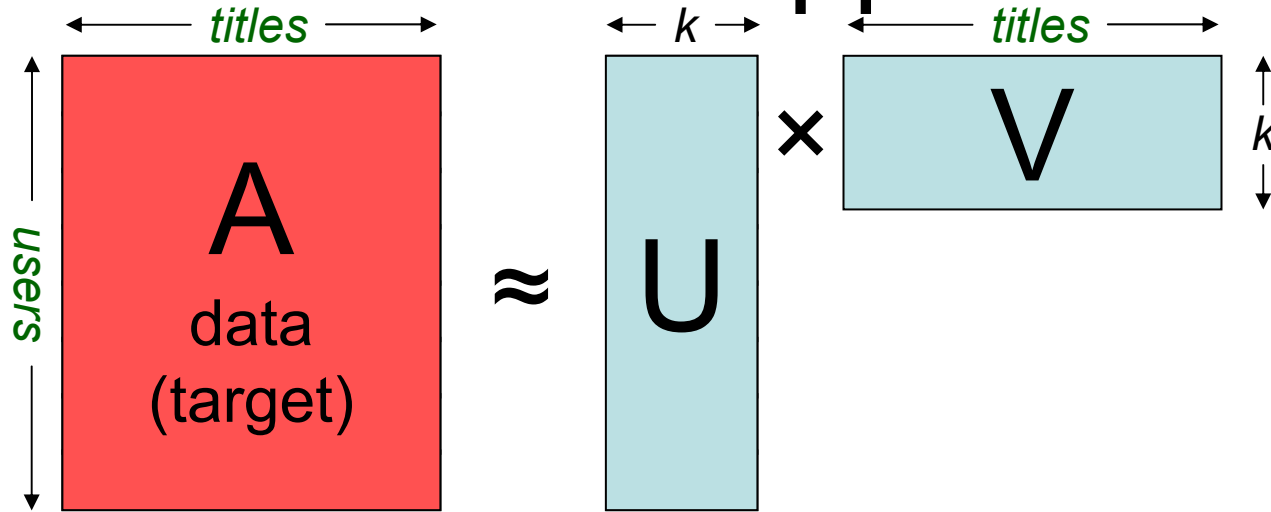


*characteristics
of the user*

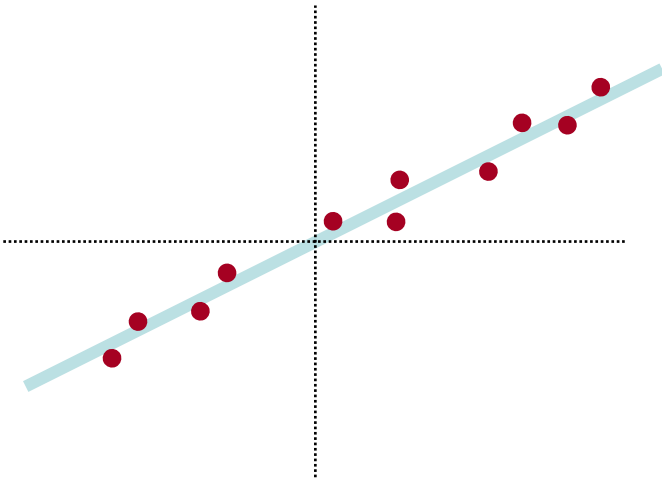
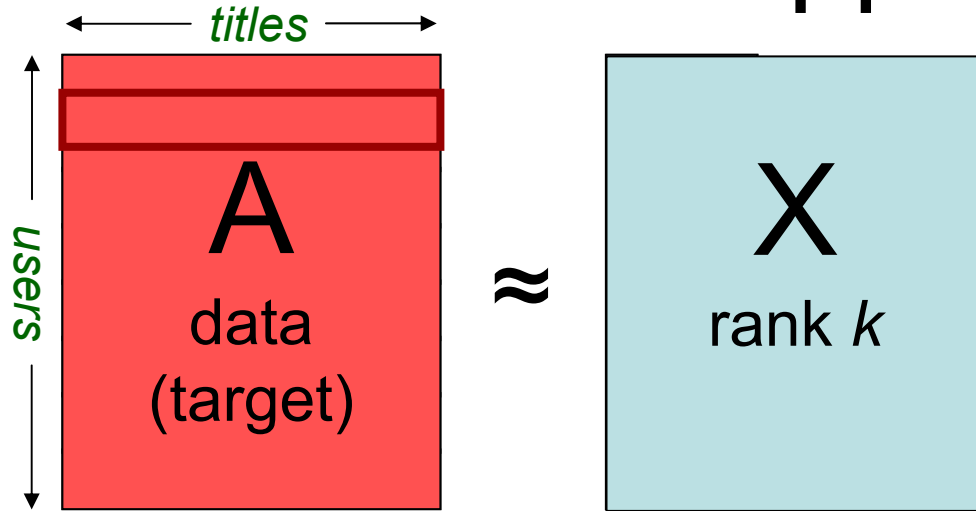
Low Rank Approximation



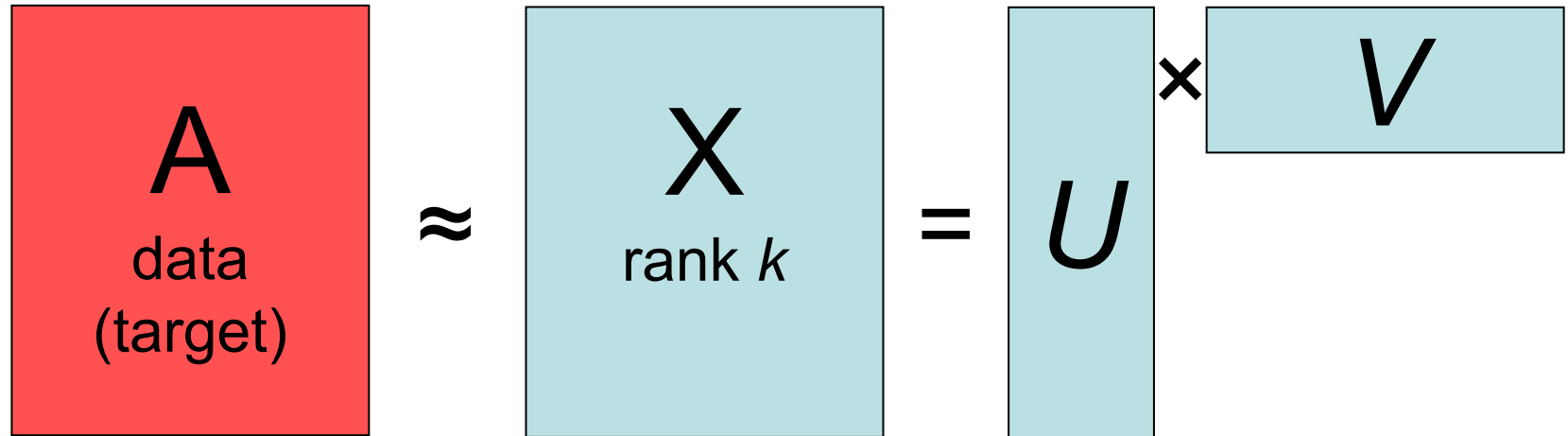
Low Rank Approximation



Low Rank Approximation

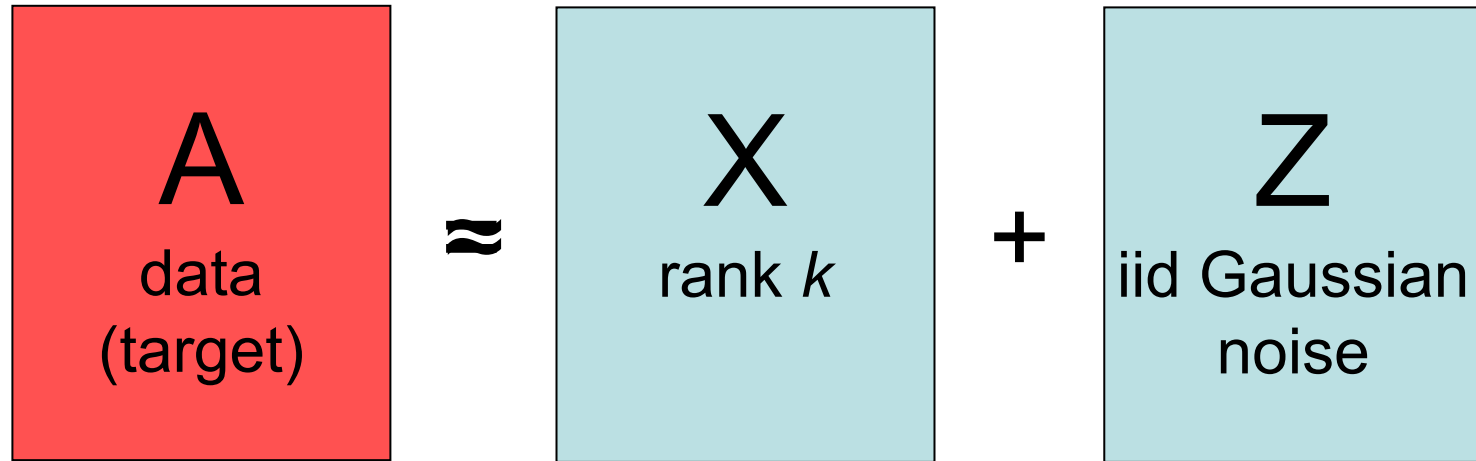


Low Rank Approximation



- Compression (mostly to reduce processing time)
- Prediction: collaborative filtering
- Reconstructing latent signal
 - biological processes through gene expression
- Capturing structure in a corpus
 - documents, images, etc
- Basic building block, e.g. for non-linear dimensionality reduction

Low Rank Approximation



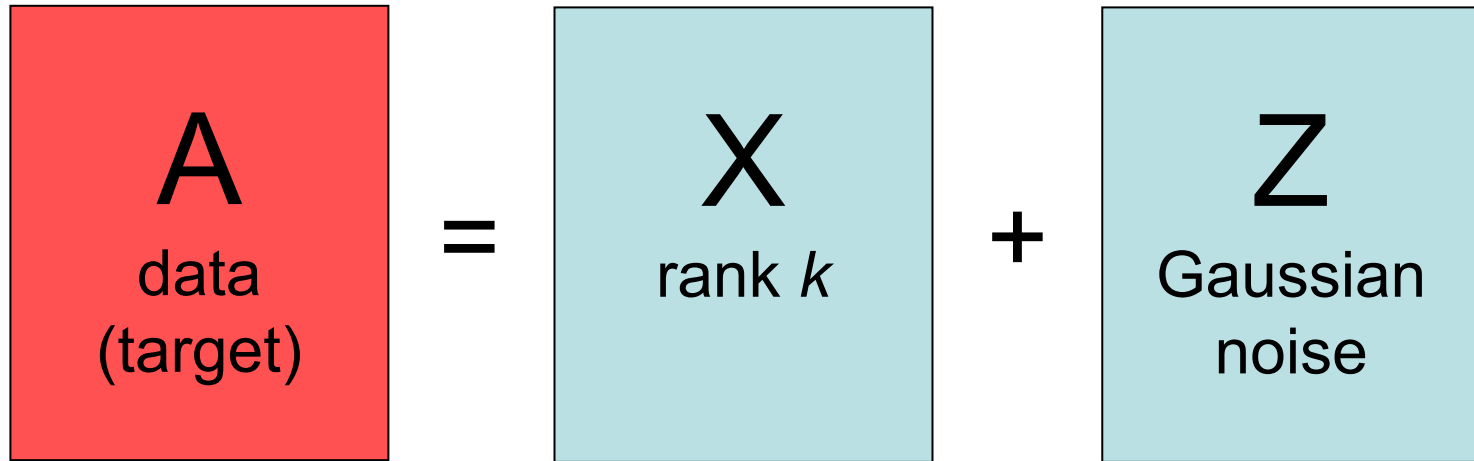
$$\log L(X; A) = \sum_{ij} \log P(A_{ij} | X_{ij}) = \frac{-1}{2\sigma^2} \sum_{ij} (A_{ij} - X_{ij})^2 + \text{const}$$

Max likelihood low-rank matrix given iid Gaussian noise

⇒ low-rank X minimizing sum-squared error

given explicitly in terms of SVD

Weighted Low Rank Approximation



$$Z_{ij} \sim N(0, \sigma_{ij}^2)$$

$$\log L(X; A) = \sum_{ij} \log P(A_{ij} | X_{ij}) = -\sum_{ij} w_{ij} (A_{ij} - X_{ij})^2 + \text{const}$$

$$w_{ij} = 1/\sigma_{ij}^2$$

Find low-rank matrix minimizing weighted sum-squared-error

[Young 1940]

Weighted Low Rank Approximation

low-rank matrix minimizing weighted sum-squared error

- External information about noise variance for each measurement
 - e.g. in gene expression analysis
- Missing data (0/1 weights)
 - collaborative filtering
- Different number of samples
 - e.g. separating style and content [Tenenbaum Freeman 00]
- Varying importance of different entries
 - e.g. 2D filters [Shpak 90, Lu et al 97]
- Subroutine for further generalizations

How?

Given A and W , find rank k matrix X minimizing the weighted sum-square difference

$$\sum_{ij} W_{ij} (A_{ij} - X_{ij})^2$$

(Unweighted) Low Rank Approximation

(Unweighted) Low Rank Approximation

$$J(X) = \left\| \begin{array}{c} d \\ \text{A} \\ \text{target} \end{array} - \begin{array}{c} \text{X} \\ \text{rank } k \end{array} \right\|_{Fro}^2$$

The diagram illustrates the objective function $J(X)$ for low rank approximation. It shows the Frobenius norm squared of the difference between a target matrix A and an approximation matrix X . Matrix A is represented by a red box with dimensions n (height) and d (width). Matrix X is represented by a light blue box with rank k . The Frobenius norm is indicated by the double vertical lines and the subscript Fro .

(Unweighted) Low Rank Approximation

$$J(UV') = \left\| \begin{array}{c} d \\ n \end{array} \begin{array}{c} \mathbf{A} \\ \text{target} \end{array} - \begin{array}{c} k \\ n \end{array} \begin{array}{c} \mathbf{U} \end{array} \times \begin{array}{c} d \\ k \end{array} \begin{array}{c} \mathbf{V}' \end{array} \right\|_{\text{Fro}}^2$$

$$\frac{\partial J}{\partial U} = 2(UV' - A)V = 0$$

$$UV'V = AV$$

$$\frac{\partial J}{\partial V} = 2(VU' - A')U = 0$$

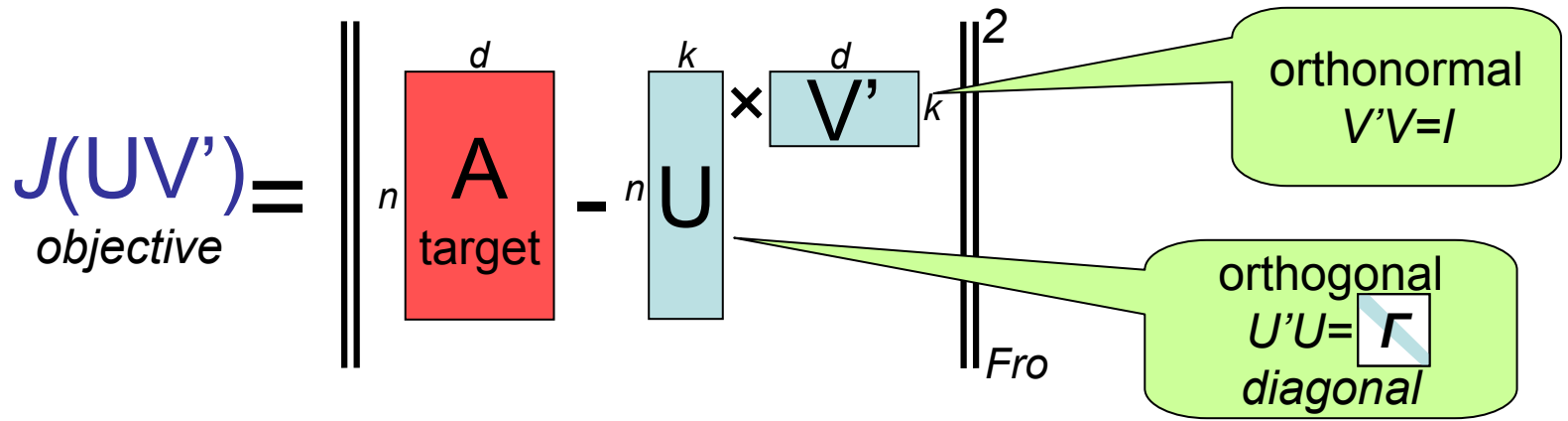
$$V = A'U$$

(Unweighted) Low Rank Approximation

$$\frac{\partial J(UV')}{\partial U, V} = 0$$



U, V are correspondingly spanned by eigenvectors of AA' and $A'A$



$$\frac{\partial J}{\partial U} = 2(UV^T - A)V = 0$$

$$\frac{\partial J}{\partial V} = 2(VU^T - A^T)U = 0$$

$$UV^T V = AV$$

$$V \Gamma = A^T AV$$

(Unweighted) Low Rank Approximation

$$\frac{\partial J(UV^T)}{\partial U, V} = 0$$



U, V are correspondingly spanned by eigenvectors of AA^T and $A^T A$

Global
Minimum



U, V are correspondingly spanned by **leading** eigenvectors

U, V spanned by eigenvectors of AA^T and AA^T

$$\implies J(UV^T) = \|A - UV^T\|^2 = \sum_{\text{unselected}} \text{eigenvalues}$$

(Unweighted) Low Rank Approximation

$$\frac{\partial J(UV^T)}{\partial U, V} = 0$$



U, V are correspondingly spanned by eigenvectors of AA^T and $A^T A$

Global
Minimum



U, V are correspondingly spanned by **leading** eigenvectors

All other fixed points are saddle points
(no non-global local minima)

Weighted Low Rank Approximation

$$J(UV') = \sum_{ij} W_{ij} (A - UV')_{ij}^2$$

$$\frac{\partial J}{\partial U} = 2((UV' - A) \cdot W)V = 0$$

$$U_i = A_i \underline{W}_i V (V' \underline{W}_i V)^{-1}$$

If $\text{rank}(W)=1$, $(V' \underline{W}_i V)$ can be simultaneously diagonalized
 \Rightarrow eigen-methods apply [IraniAnandan 2000]

Otherwise, eigen-methods cannot be used,
solutions are not incremental

WLRA: Optimization

$$J(UV') = \sum_{ij} W_{ij} (A - UV')_{ij}^2$$

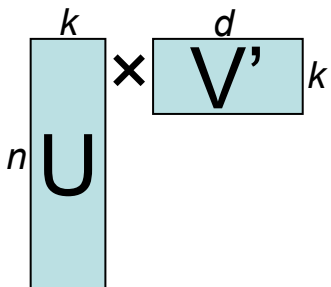
For fixed V , find optimal U

For fixed U , find optimal V

$$J^*(V) = \min_U J(UV')$$

$$\frac{\partial}{\partial V} J^*(V) = 2U^* ((U^* V' - Y) \otimes W)$$

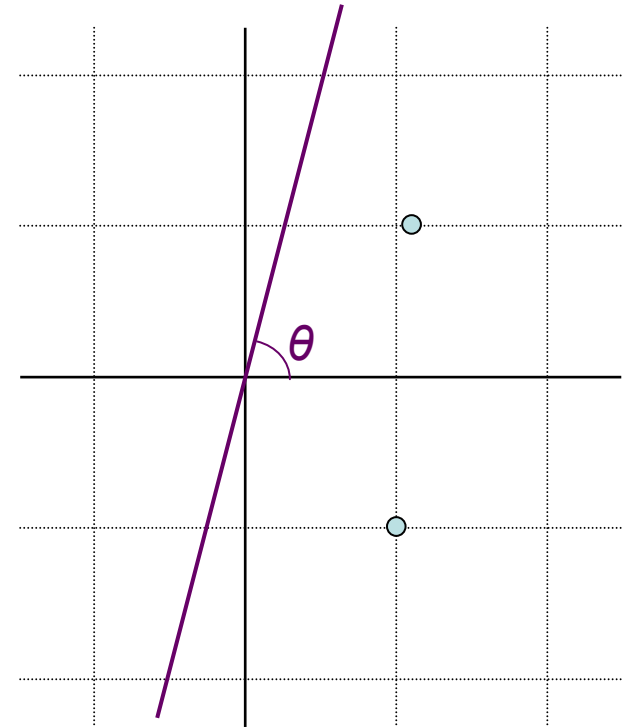
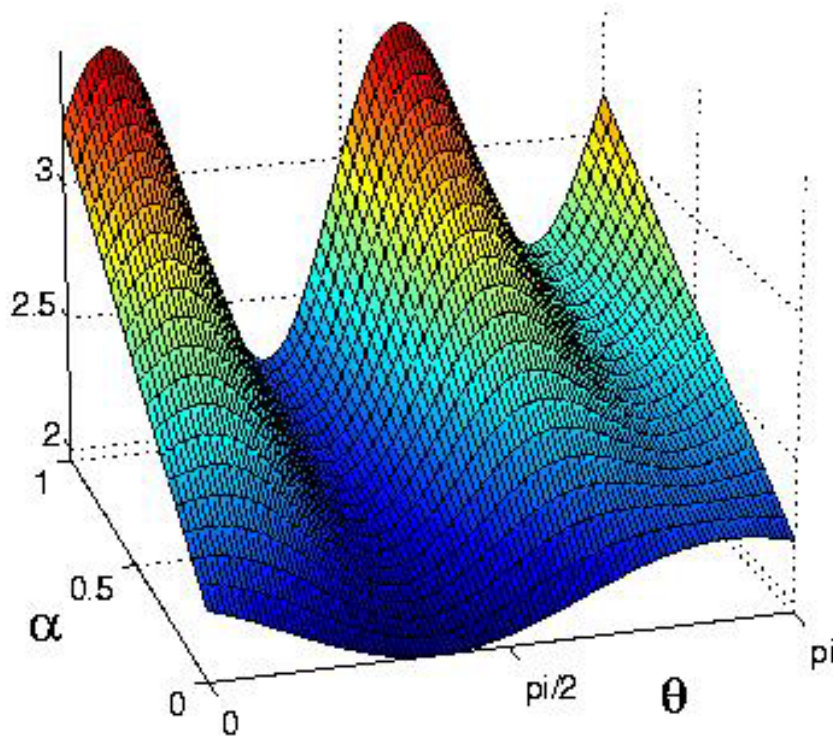
Conjugate gradient descent on J^*



Optimize kd parameters instead of $k(d+n)$

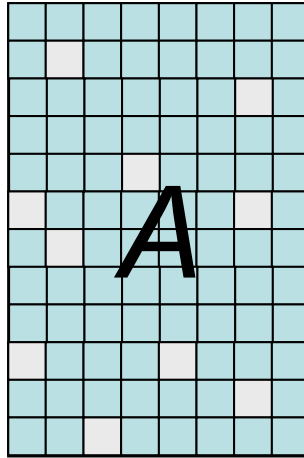
Local Minima in WLRA

$$A = \begin{bmatrix} 1 & 1.1 \\ 1 & -1 \end{bmatrix} \quad W = \begin{bmatrix} 1 + \alpha & 1 \\ 1 & 1 + \alpha \end{bmatrix}$$



WLRA: An EM Approach

0/1 weights:



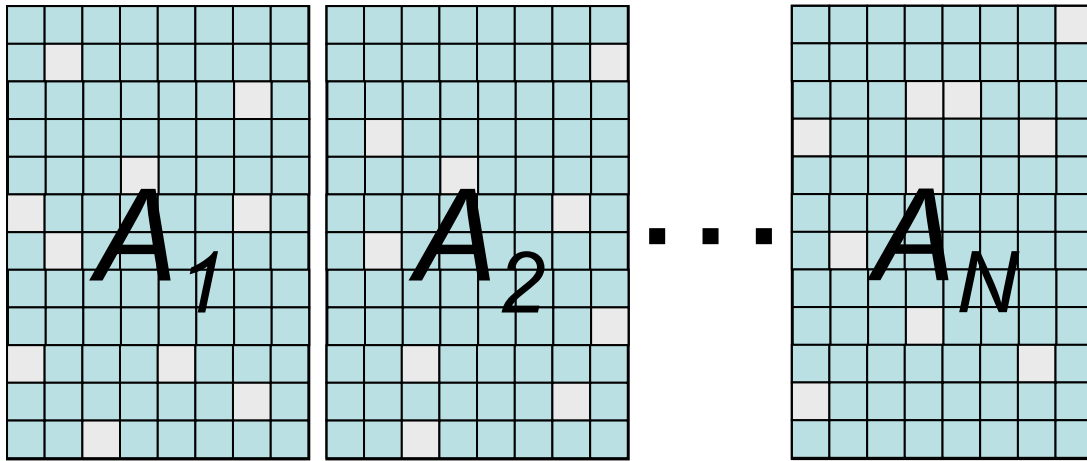
$$A = X + Z$$

observations

$N(0, 1)$ noise

parameters

WLRA: An EM Approach



$$A_r = X + Z_r$$

same low-rank X
for all targets A_r

independent noise Z_r
for each target A_r

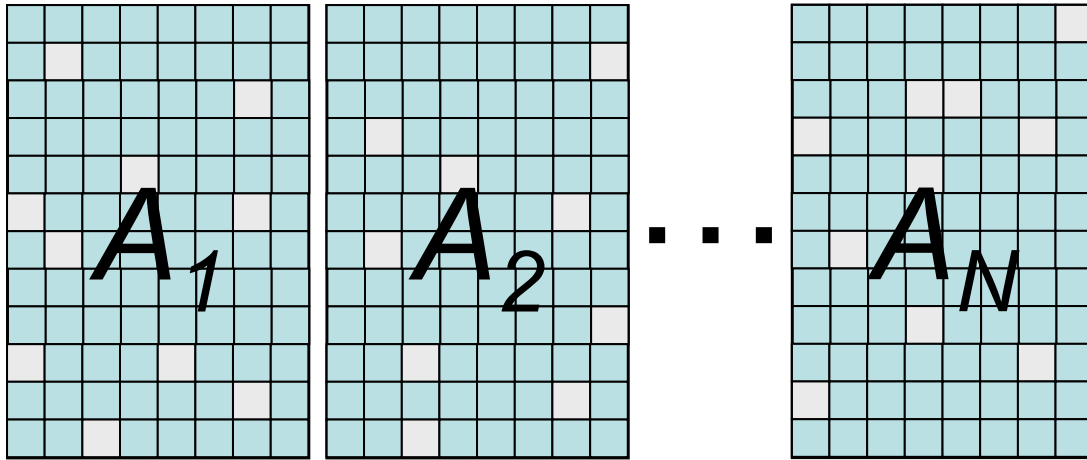
Expectation Step:

$$\text{missing } A_r[i,j] \leftarrow X[i,j]$$

Maximization Step:

$$X \leftarrow \text{LRA} \left(\frac{1}{N} \sum_r A_r \right)$$

WLRA: An EM Approach



$$A_r = X + Z_r$$

integer

WLRA(A, W) with $W[i, j] = w[i, j]/N$

$\Rightarrow A_r[i, j] = A[i, j]$, or missing if $w[i, j] < r$

$$X \leftarrow \text{LRA}(W \otimes A + (1 - W) \otimes X)$$

Expectation Step:

missing $A_r[i, j] \leftarrow X[i, j]$

Maximization Step:

$$X \leftarrow \text{LRA}\left(\frac{1}{N} \sum_r A_r\right)$$

WLRA: Optimization

$$J(UV') = \sum_{ij} W_{ij} (A - UV')_{ij}^2$$

For fixed V , find optimal U

For fixed U , find optimal V

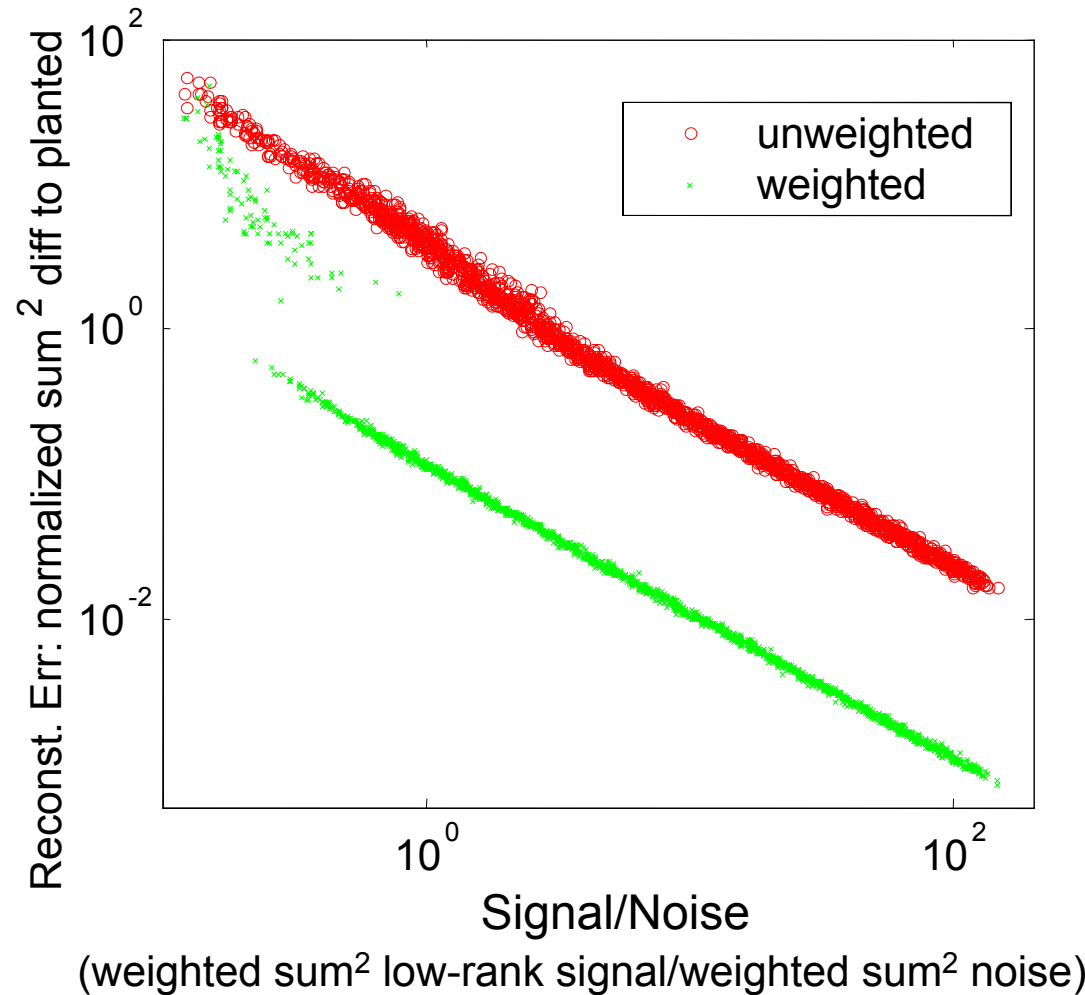
$$J^*(V) = \min_U J(UV')$$

$$\frac{\partial}{\partial V} J^*(V') = 2U^* ((U^* V' - A) \otimes W)$$

Conjugate gradient descent on J^*

$$X \leftarrow \text{LRA}(W \otimes A + (1 - W) \otimes X)$$

Estimations of a “planted” X with Non-Identical Noise Variances



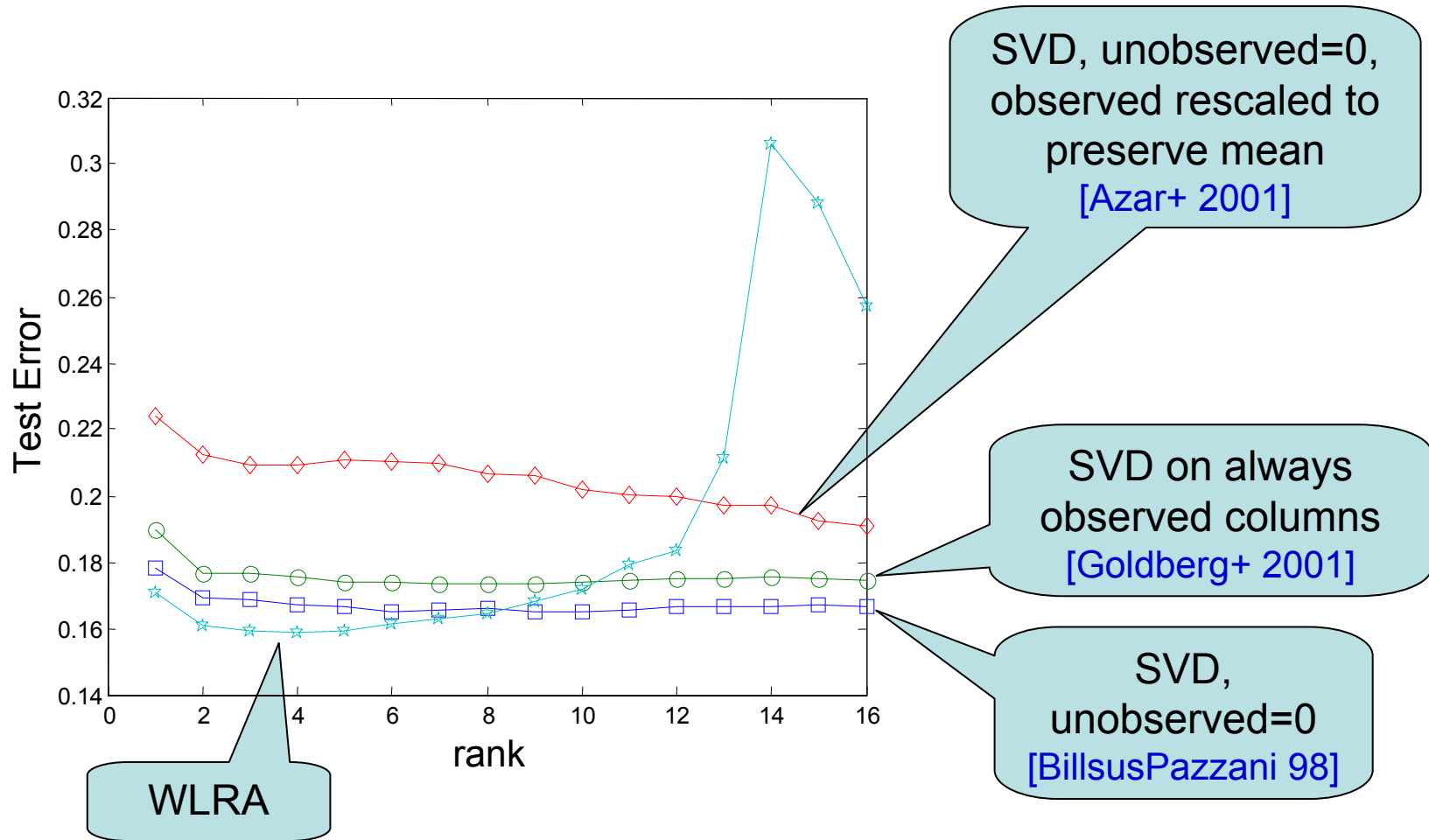
Collaborative Filtering of Joke Preferences (“Jester”)

← jokes →

	+1	-2	+3			+6			+5			-2		
	+6	+4	-7				-2		+4			-3		
	-5	+7	-2	+6	+2				-1	-7				
	-1	+2	+4		+1			+3		-3				+7
	+2	-3	-6			+4								
	0	-2	-1				+5	-2		+6				+1
	+5	+2	-7		+3					-2	-3	-1		
	+7	-3	-2	-6		+2	+2		-4		-7			+3
	-4	+2	+2				+1					+2	-5	
	-3	-6	-1							+6	+5			-7
	+1	-6	+7	-3	-2			+6	+5					
	+9	+4	+5									-2		-6
	+7	-5	+2			+4		-2				-3		
	+1	+6	-2						+2	+4		-2		+3
	-3	+3	-6		-4	-9		+7				0		+2
	-4	+3	+2	-2			0						-4	

↑ users ↓

Collaborative Filtering of Joke Preferences (“Jester”)



WLRA as a Subroutine for Other Cost Functions

$$\sum_{ij} (A_{ij} - X_{ij})^2$$

⇒ Low Rank Approximation (PCA)

$$\sum_{ij} W_{ij} (A_{ij} - X_{ij})^2$$

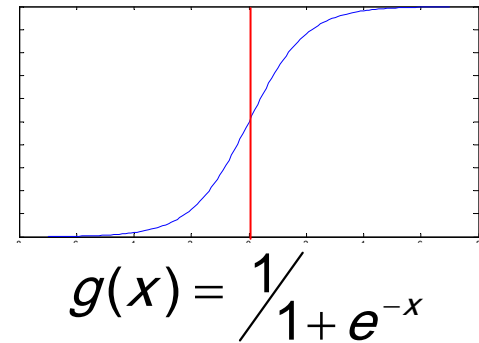
⇒ Weighted Low Rank Approximation

$$\sum_{ij} \text{Loss}(A_{ij}, X_{ij})$$

Logistic Low Rank Regression

[Collins+ 2002, Gordon 2003, Schein+ 2003]

$$\Pr \left(\begin{array}{c|c} \boxed{Y} & \boxed{X} \\ \text{observed} & \text{rank } k \end{array} \right) = g \left(\boxed{X} \right)$$



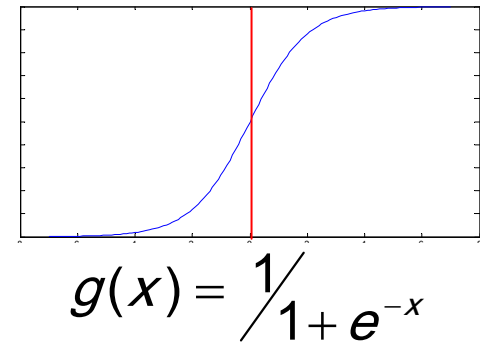
$$Loss(Y, X) = \sum_{ij} \log g(Y_{ij}, X_{ij})$$

$$\approx - \sum_{ij} \frac{g(Y_{ij} X_{ij}^{(t-1)}) g(-Y_{ij} X_{ij}^{(t-1)})}{2} \left(X_{ij}^{(t)} - \left(X_{ij}^{(t-1)} + \frac{Y_{ij}}{g(Y_{ij} X_{ij}^{(t-1)})} \right) \right)^2 + Const$$

Logistic Low Rank Regression

[Collins+ 2002, Gordon 2003, Schein+ 2003]

$$\Pr \left(\begin{array}{c|c} \boxed{Y} & \boxed{X} \\ \text{observed} & \text{rank } k \end{array} \right) = g \left(\begin{array}{c} \boxed{X} \\ \text{rank } k \end{array} \right)$$



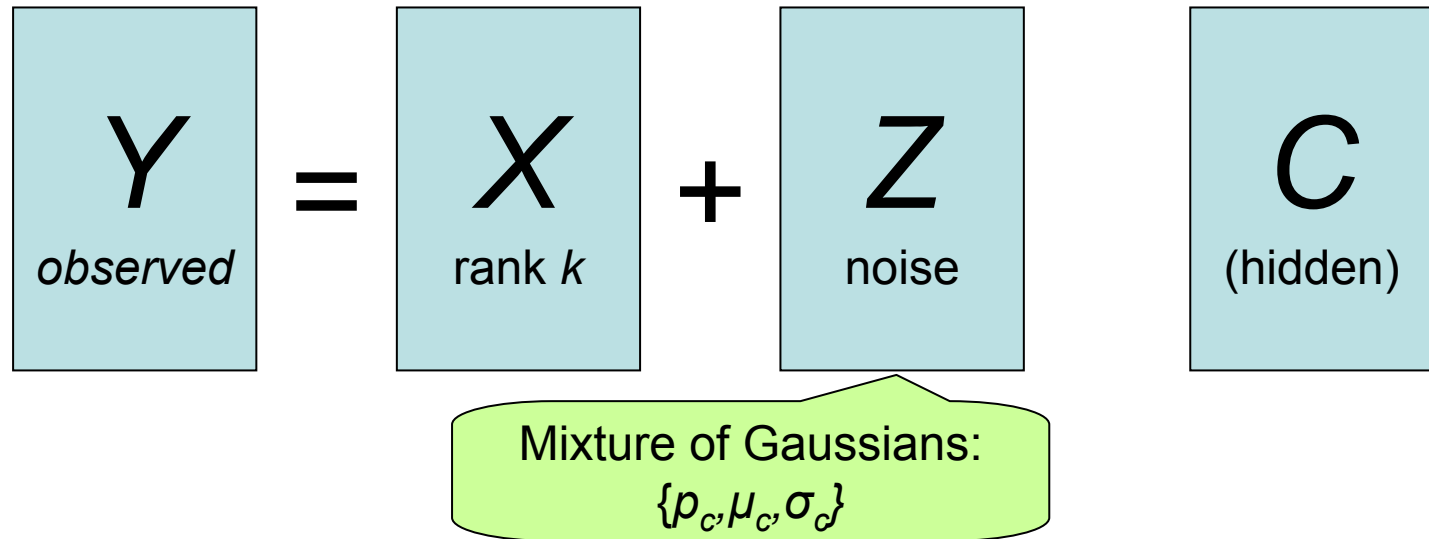
$$Loss(Y, X) = \sum_{ij} \log g(Y_{ij}, X_{ij})$$

$$\approx - \sum_{ij} W_{ij}^{(t)} \left(X_{ij}^{(t)} - \left(A_{ij}^{(t)} \right) \right)^2 + Const$$

$$W_{ij}^{(t)} = \frac{g(Y_{ij} X_{ij}^{(t-1)}) g(-Y_{ij} X_{ij}^{(t-1)})}{2}$$

$$A_{ij}^{(t)} = X_{ij}^{(t-1)} + \frac{Y_{ij}}{g(Y_{ij} X_{ij}^{(t-1)})}$$

Maximum Likelihood Estimation with Gaussian Mixture Noise



E step: calculate posteriors of C

M step: WLRA with

$$W_{ij} = \sum_c \frac{\Pr(C_{ij} = c)}{\sigma_c^2}$$

$$A_{ij} = Y_{ij} + \sum_c \frac{\Pr(C_{ij} = c) \mu_c}{\sigma_c^2} / W_{ij}$$

Weighted Low Rank Approximations

Nathan Srebro

Tommi Jaakkola

- Weights often appropriate in low rank approximation
- WLRA more complicated than unweighted LRA
 - Eigenmethods (SVD) do not apply
 - Non-incremental
 - Local minima
- Optimization approaches:
 - Alternate optimization
 - Gradient methods on $J^*(V)$
 - EM method: $X \leftarrow \text{LRA}(W \otimes A + (1-W) \otimes X)$
- WLRA useful as subroutine for more general loss functions

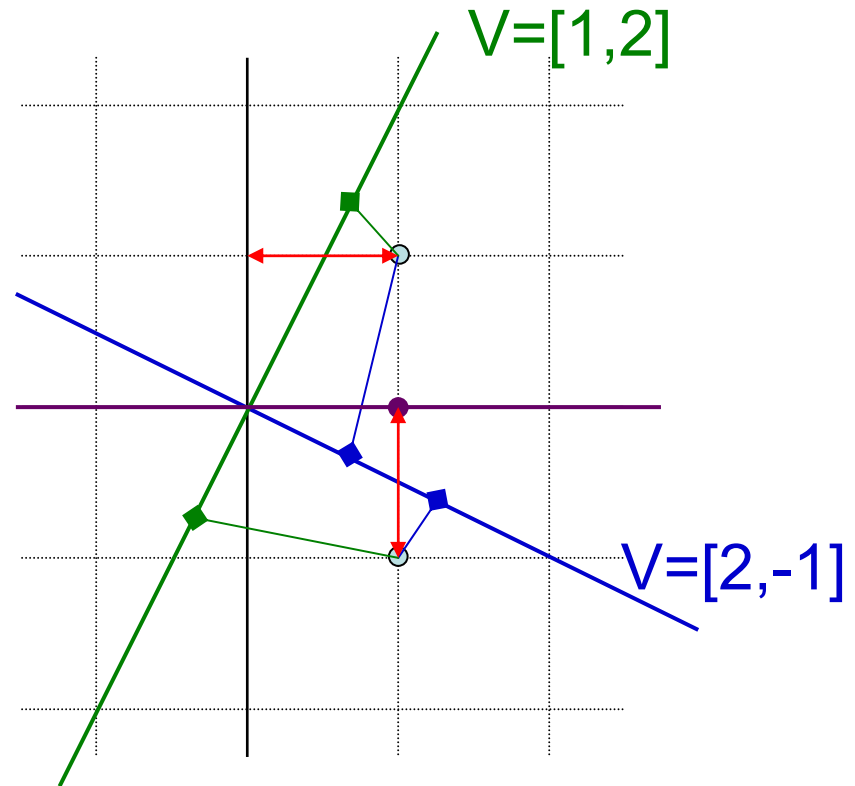
END

Local Minima in WLRA

$$A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad W = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

$$X = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \\ 1\frac{1}{3} & -\frac{2}{3} \end{bmatrix} \quad J=3$$

$$X = \begin{bmatrix} \frac{2}{3} & 1\frac{1}{3} \\ -\frac{1}{3} & -\frac{2}{3} \end{bmatrix} \quad J=2\frac{1}{3}$$

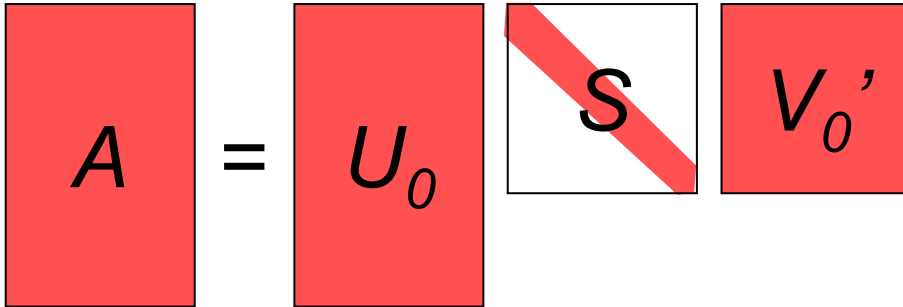


(Unweighted) Low Rank Approximation

$$\frac{\partial J(UV')}{\partial U, V} = 0$$



U, V are correspondingly spanned
by eigenvectors of AA' and $A'A$



The diagram illustrates the matrix decomposition $A = U_0 S V_0'$. It consists of four red rectangular boxes arranged horizontally. The first box contains the letter A . To its right is an equals sign. The second box contains U_0 . To its right is a white square box with a red diagonal line from the top-left to the bottom-right, and the letter S centered inside. To the right of this box is the third red box containing V_0' .

(Unweighted) Low Rank Approximation

$$\frac{\partial J(UV')}{\partial U, V} = 0$$



U, V are correspondingly spanned by eigenvectors of AA' and $A'A$

$$A = U_0 S V_0'$$

$$Q_U' Q_V = I$$

$$U = U_0 S Q_U$$

$$V = V_0 Q_V$$

(Unweighted) Low Rank Approximation

$$\frac{\partial J(UV^T)}{\partial U, V} = 0 \iff U, V \text{ are correspondingly spanned by eigenvectors of } AA^T \text{ and } A^T A$$

$$A = U_0 \begin{matrix} s_1 & & & \\ & s_2 & & \\ & & s_3 & \dots \\ & & & \dots & s_m \end{matrix} V_0^T$$

$$\|A\|^2 = \sum_{\alpha} s_{\alpha}^2$$

$$\|UV^T\|^2 = \sum_{\text{selected } \alpha} s_{\alpha}^2$$

$$\|A - UV^T\|^2 = \sum_{\text{unselected } \alpha} s_{\alpha}^2$$

$$U^* = U_0 S \quad V^* = V_0$$

0/1 "permuted diagonal" matrix

(Unweighted) Low Rank Approximation

$$\frac{\partial J(UV^T)}{\partial U, V} = 0$$



U, V are correspondingly spanned by eigenvectors of AA^T and $A^T A$

Global Minimum



U, V are correspondingly spanned by **leading** eigenvectors

$$U = U_0 S$$

$$V = V_0$$