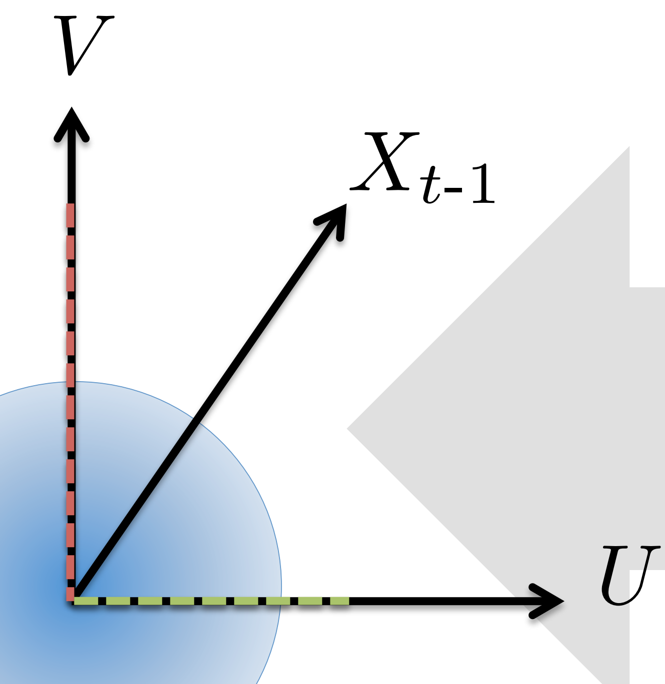
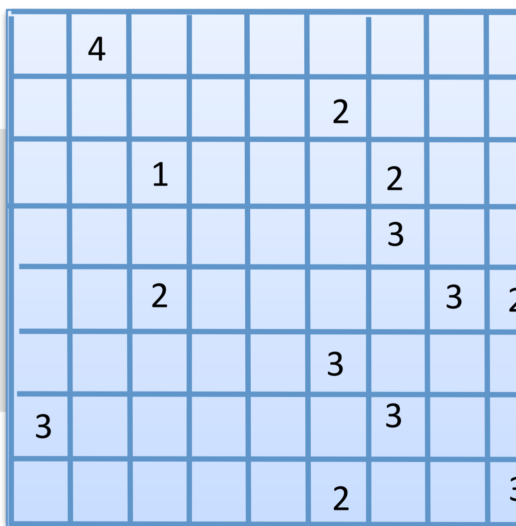
From Matrix Completion to Private Singular Vector Computation *and Back*



Moritz Hardt
IBM Almaden



Singular vector computation

Given matrix $\boldsymbol{A}$,

find u maximizing $|\boldsymbol{A}u|/|u|$

Top k singular vectors $u_1,\ldots,u_k$ defined recursively

Three Epochs

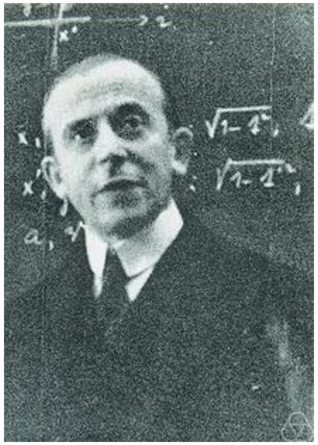
19th century

Cauchy initiates **spectral theory** of matrices [1829]

No efficient algorithms

20th century

Von Mises discovers **Power Method** [1929]



Efficient algorithms

Today

Focus on data:

- *noise*
- *missing values*
- *privacy concerns*

Efficient algorithms??

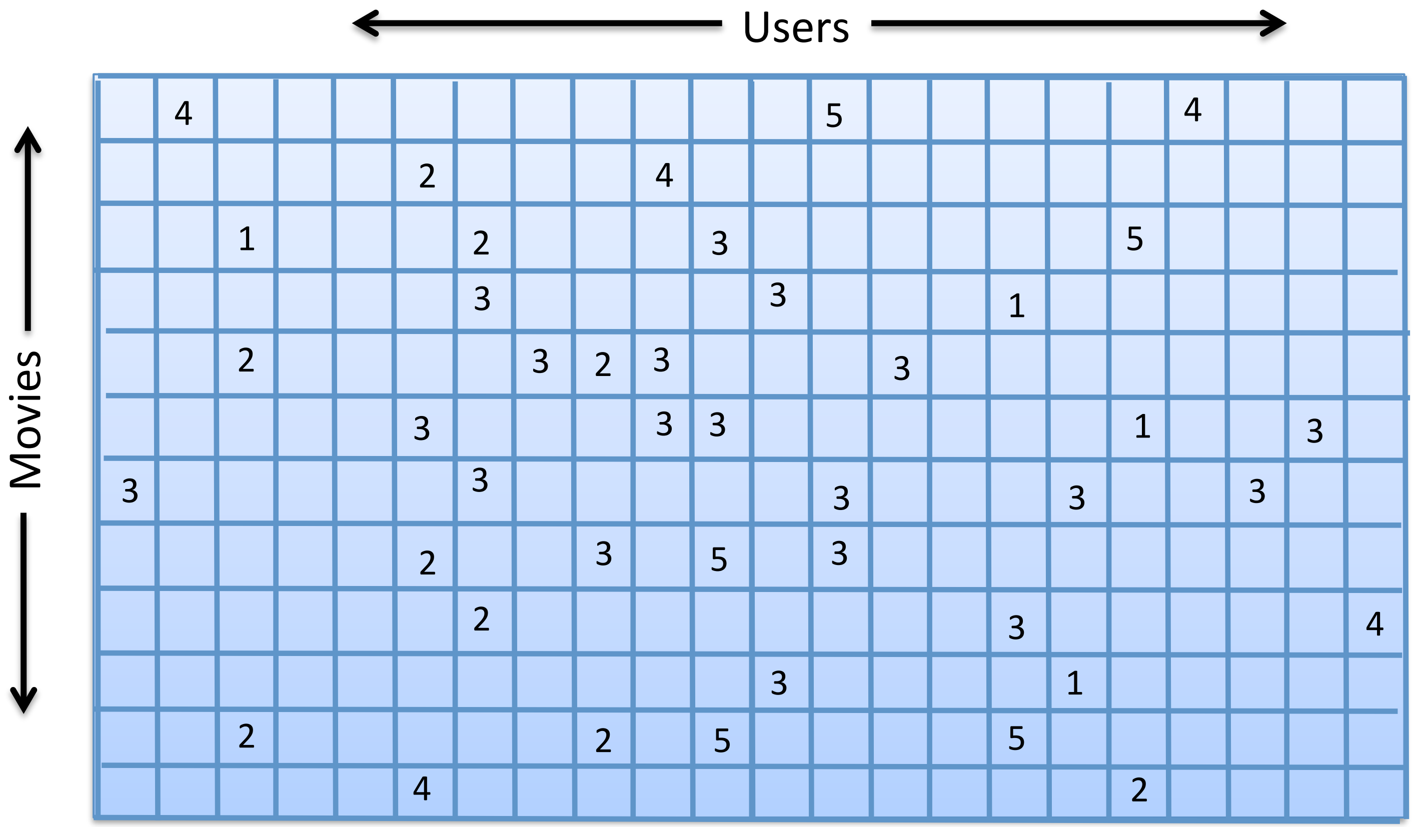
Matrix Completion

	4											5						4			
					2				4												
		1				2				3							5				
						3					3				1						
		2					3	2	3				3								
					3				3	3							1			3	
3						3						3				3			3		
					2			3		5		3									
						2									3						4
											3					1					
		2						2		5					5						
					4												2				

Goal: Reconstruct missing entries

I.e: Approximate dominant SVs of *unknown* matrix

Matrix Completion aka Netflix Problem



Goal: Reconstruct missing entries

I.e: Approximate dominant SVs of *unknown* matrix

Feasibility Assumptions

[CT,CR,R,KMO,........(long line of work)....]

- Matrix A approximately rank $k << n$
- Top singular space U is *incoherent*

  for all e_i : $|e_i^T U|$ small
- Subsample uniformly *random*
- ~~It *never rains* in San Francisco~~

Twin assumptions

Strong, but lead to informative theory!

State of the Art

Algorithm	Space/ Running time	(Provable) sample bounds
Nuclear Norm	Slow	Nearly optimal
Alternating Minimization	Nearly linear	High

Alternating Minimization method of choice in practice!
First bounds due to Keshavan12, Jain-Netrapalli-Sanghavi13

[H'13]	Nearly linear	Not too far from optimal

Based on Alternating Minimization
See arXiv:1312.0925 for details.

Privacy: The *other* Netflix Problem

Dramatic reference to [Narayanan-Shmatikov'08]

Basic Question: Given matrix A, approximate top k singular vectors subject to differential privacy

Lots of work, e.g., BDMN05,MM09,CSS12, HR12,BBDS12,KT13, HR13,DTTZ

Results [H'13]

Nearly linear time algorithm with following guarantees:

Entry-level privacy

Nearly optimal error in *k* and coherence

- Only logarithmic in *n*
- (ε,δ)-diff priv
- resolved main question of *H*-Roth (2013)

Unit spectral norm

Nearly optimal error in *k* and *n*

- for both (ε,0) and (ε,δ)-dp
- running time down from > n^3 [Kapralov-Talwar 13]

See arXiv:1311.2495 for details.

Main message

Noisy Power Method

solves both AltMin and Private SVD

Incoherence

controls error rates in both problems

Techniques

transfer from one problem to the other

Noisy Power Method

Input: n x n matrix A *symmetric*, target rank k

X_0 random orthonormal matrix

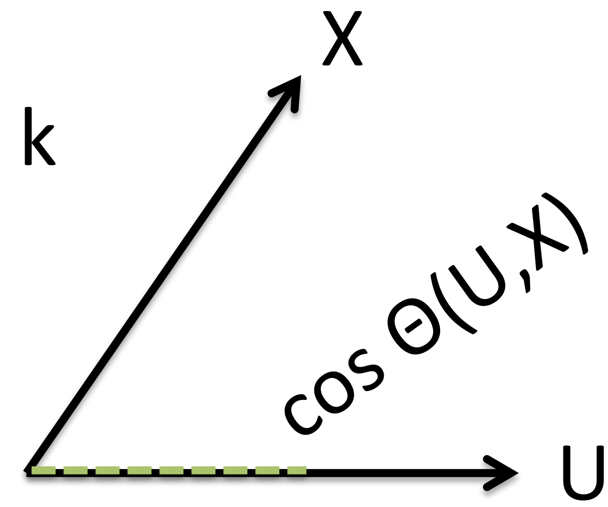
For $t = 1$ to T:

- Nature chooses perturbation G_t
- We observe $Y_t = AX_{t-1} + G_t$
- X_t = Orthonormalize(Y_t)

Output X_T *(approx top k singular vectors)*

Principle Angle Between Subspaces

Let U, X subspaces of dimension k



k=1 $\quad \cos \Theta(U,X) = |U^T X|$

In general $\quad \cos \Theta(U,X) = \sigma_{min}(U^T X)$

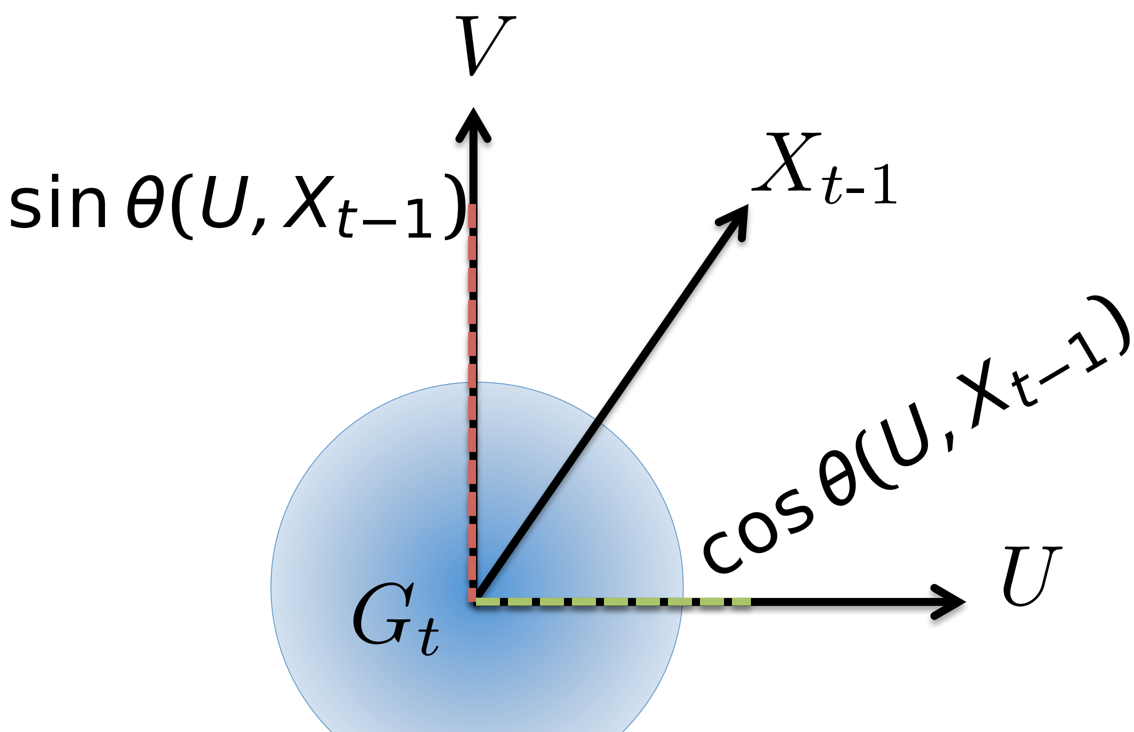
$\sin \Theta(U,X) = \sigma_{max}(V^T X)$
where V orthog. complement of U

$\tan \Theta(U,X) = \sin \Theta(U,X) / \cos \Theta(U,X)$

Main Convergence Lemma

$$\tan\theta(U, X_t) \leq \frac{\sigma_{k+1}\sin\theta(U, X_{t-1})}{\sigma_k \cos\theta(U, X_{t-1})}$$

If G_t=0



So, what can we do with this?

Alternating Minimization

Input: Subsample P_Ω A of unknown matrix A

Pick X_0 uniformly at random

For $t = 1$ **to** L:

$$Y_t = \arg\min_Y \|P_\Omega(A - X_{t-1}Y^T)\|_F^2$$

where P_Ω is projection on subsample

X_t = Orthonormalize(Y_t)

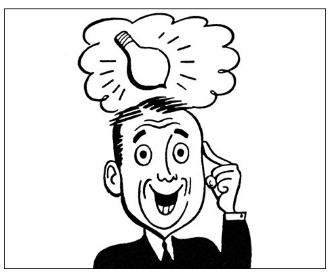
Output: $B = X_{L-1}Y_L^T$

AltMin as Noisy Power Method

- Update step in AltMin

$$Y_t = \arg\min \|P_\Omega(A - X_{t-1}Y^T)\|_F^2$$

where P_Ω is projection on subsample



With full information, $Y_t = AX_{t-1}$

Observation. We can write $Y_t = AX_{t-1} + G_t$ where G_t captures "sampling error"

Main hurdle

Observation. We can write $Y_t = AX_{t-1} + G_t$ where G_t captures "sampling error"

Norm of G_t is controlled by coherence of X_{t-1}

Def: Coherence $\mu(X) = (n/k) \max_i |e_i^T X|^2$

Reasoning about Coherence

Coherence propagation:

If A is incoherent, then so is every iterate X_t

AltMin:

Incoherent X_t ensures low sample complexity

Argue via Smooth Orthonormalization [H-Roth12]

Private SVD:

Incoherent X_t ensures small Gaussian noise

- Argue via rotational invariance of Gaussians

Recap

Noisy Power Method

solves both AltMin and Private SVD

Incoherence

controls error rates in both problems

Techniques

transfer from one problem to the other

Conclusion and Open Problems

- Robustness is the common denominator between privacy and machine learning
  - Focus on finding the "right" robust analysis of algorithms
- Lots of technical problems:
  - Weaker coherence notion sufficient for privacy?
  - Tight sample complexity bounds for AltMin?
  - Privacy-preserving AltMin?
  - Robustness of gradient descent?

Thank you.

Results

Recall $\alpha = |(I - UU^T)X|$

Entry-level privacy:

Tight dependence on k and μ(A)

H'13 $$\alpha = \tilde{O}_{\epsilon,\delta}\left(\frac{1}{\sigma_k}\sqrt{k\mu(A)}\right)$$

H-Roth13: $$\alpha = \tilde{O}_{\epsilon,\delta}\left(\frac{1}{\sigma_k}k\sqrt{\mathrm{rk}(A)\mu(A)}\right)$$

Settings of [Kapralov-Talwar 13] and [Chaudhuri-Sarwate-Sinha 12]

This work: Tight dependence on *k*, *n*.

Applies to (ε,δ)-dp as well.

Worst-case running time linear in *n*. Down from > n^3.