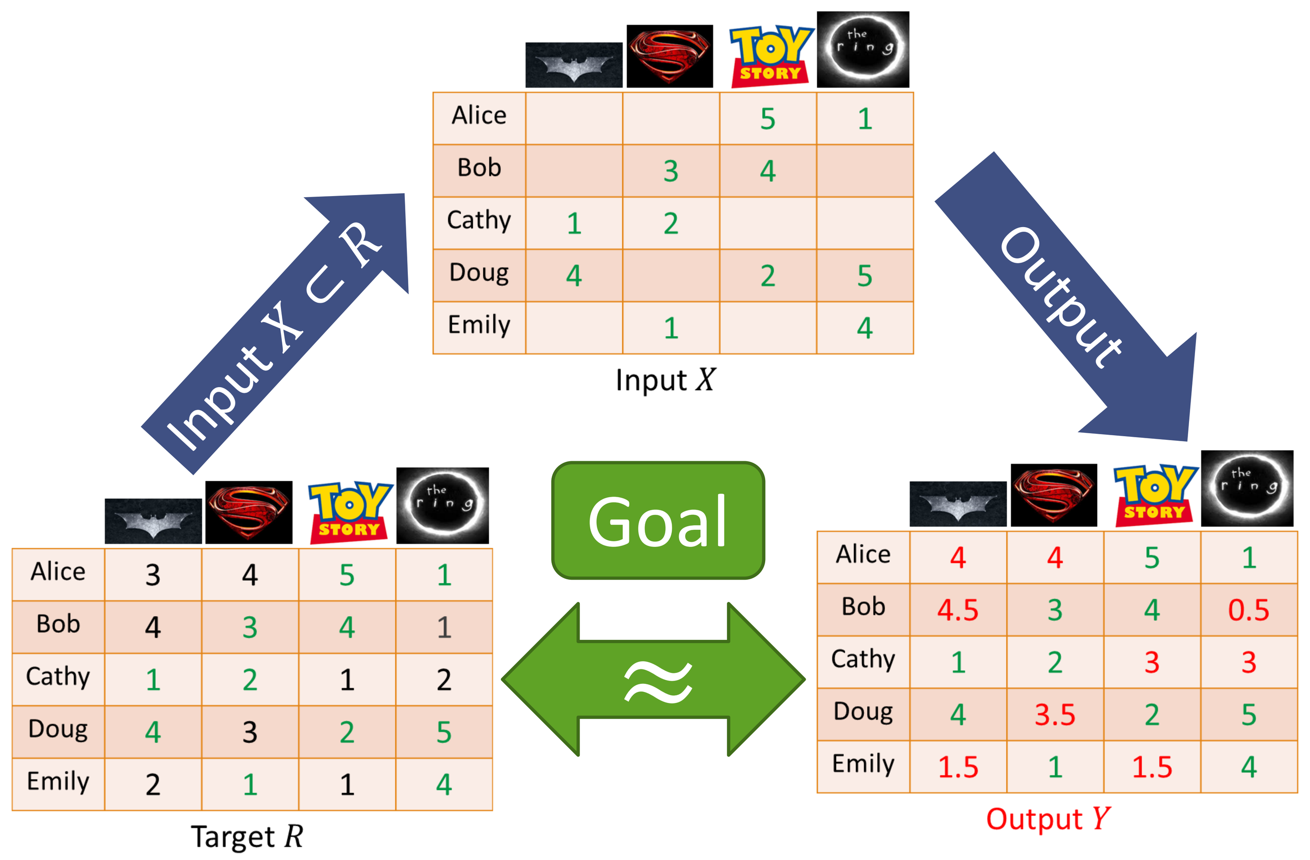
Differentially Private Matrix Completion Revisited

Prateek Jain[1], **Om Thakkar[2]**, Abhradeep Thakurta[3]

[1]Microsoft Research, [2]**Boston University**, [3]University of California-Santa Cruz

INTRODUCTION

Low-Rank Matrix Completion: Given an incomplete matrix $X \subset R$, s.t. R is low-rank, output Y, such that $Y \approx R$.



Alice			5	1
Bob		3	4	
Cathy	1	2		
Doug	4		2	5
Emily		1		4

Input X

Alice	3	4	5	1
Bob	4	3	4	1
Cathy	1	2	1	2
Doug	4	3	2	5
Emily	2	1	1	4

Target R

Alice	4	4	5	1
Bob	4.5	3	4	0.5
Cathy	1	2	3	3
Doug	4	3.5	2	5
Emily	1.5	1	1.5	4

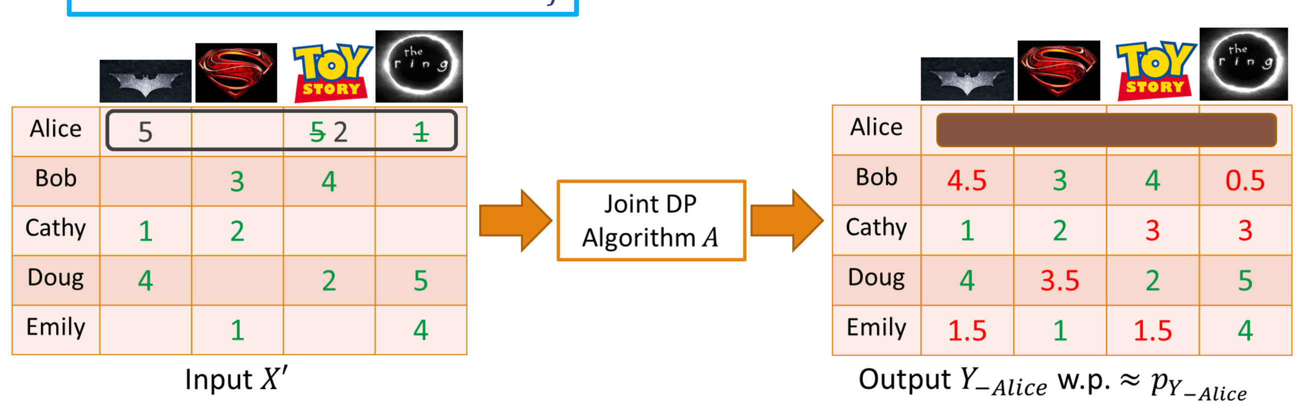
Output Y

Predictions can leak the privacy of user ratings!

Our Goal: Provide personalized recommendations while ensuring each user's privacy

JOINT DIFFERENTIAL PRIVACY [KPRU'14]

- Let Y_{-i} denote Y without user j's output.
- For randomized algorithm A, let $\boxed{\Pr\left(A(X)_{-j} = Y_{-j}\right) = p_{Y_{-j}}}$, $\forall j \in [n]$ and $Y_{-j} \in Range(A_{-j})$



Alice	5		5 2	1
Bob		3	4	
Cathy	1	2		
Doug	4		2	5
Emily		1		4

Input X'

Alice				
Bob	4.5	3	4	0.5
Cathy	1	2	3	3
Doug	4	3.5	2	5
Emily	1.5	1	1.5	4

Output Y_{-Alice} w.p. $\approx p_{Y_{-Alice}}$

(ϵ, δ) −Joint DP guarantee for each user $j \in [n]$:

- $\forall X'$ s.t. $d(X, X') = 1$ and user j's data changes, $j \in [n]$, we have $\boxed{\Pr\left(A(X')_{-j} = Y_{-j}\right) \le e^{\epsilon} \cdot p_{Y_{-j}} + \delta}$, $\forall Y_{-j} \in Range(A_{-j})$

Distinction from standard DP [DMNS'06]**:** Under Joint DP, A's output for user i can depend **arbitrarily** on i's input.

Consequence: Better personalized recommendations!

MAIN CONTRIBUTION

A Joint DP Matrix completion algorithm that provides

- the first *non-trivial* generalization error guarantee, and
- better empirical performance than state-of-the-art private algorithms

JOINT DP FRANK-WOLFE

Notation: n = number of users, d = number of items

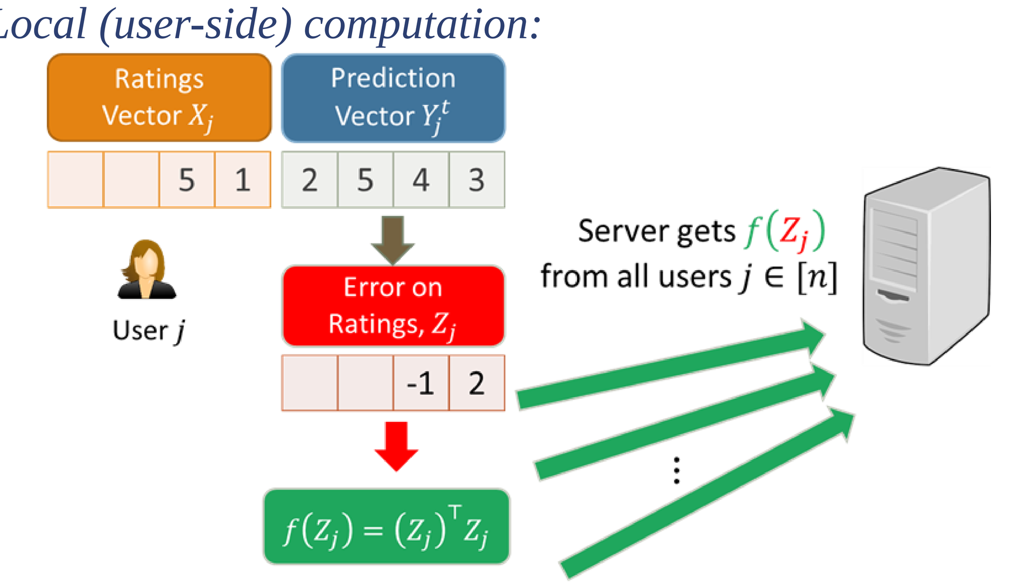
Input $X \subset R$, s.t. $\|R\|_{nuc} \le k$, and

X has non-zero entries only for positions in set Ω

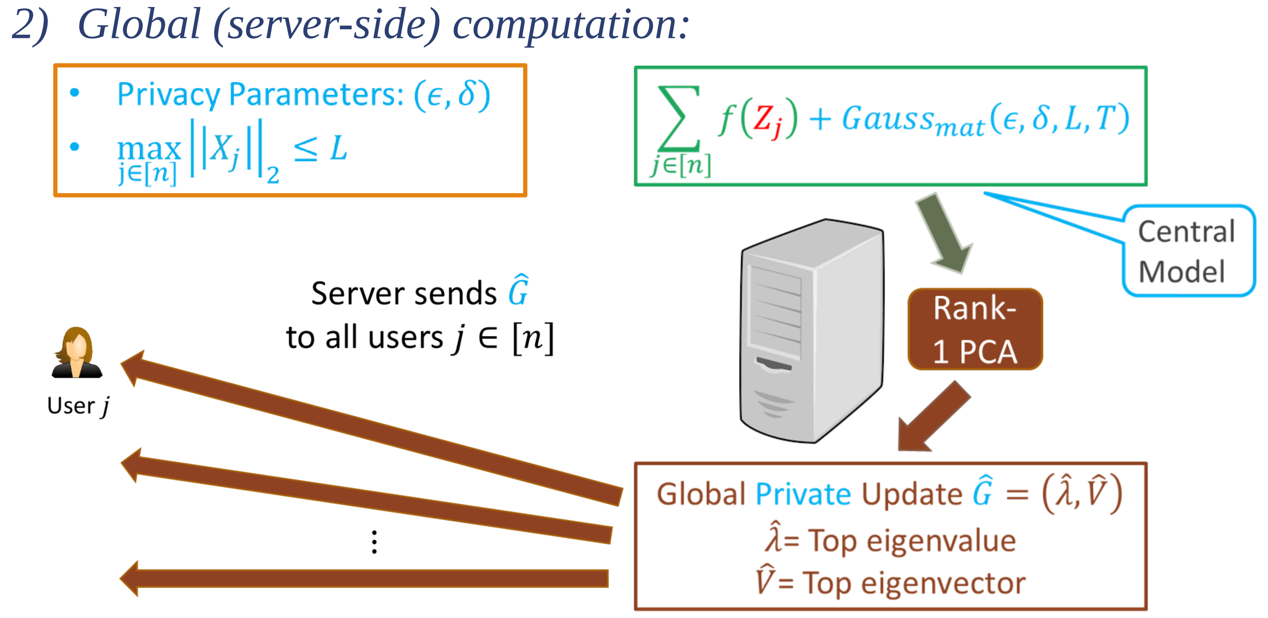
Objective: $$\min_{\|Y\|_{nuc} \le k} \frac{1}{|\Omega|} \sum_{i,j \in \Omega} \left(Y_{ij} - X_{ij}\right)^2$$

Joint DP Frank-Wolfe (FW) is an iterative algorithm in which each iteration $t \in [T]$ can be broadly divided into two parts:

1) *Local (user-side) computation:*



2) *Global (server-side) computation:*



Local update rule for $(t+1)^{st}$ iteration:

$$\boxed{Y_j^{t+1} = \left(1 - \frac{1}{T}\right) Y_j^t - \frac{1}{T}\left(h\left(Z_j, \hat{G}\right)\right)}$$

Projection free updates; Computationally light

PRIVATE OJA'S ALGORITHM

Let $Z = \sum_{j \in [n]} f(Z_j)$. Server update in Joint DP Frank-Wolfe:

$$\boxed{Z + Gauss_{mat}(\epsilon, \delta, L, T)}$$

Sparse $d \times d$ matrix

Dense $d \times d$ matrix

$\Omega(d^2)$ Space and Time
Solution: Private Oja's

Iteration $\tau \in [\Gamma]$ in Private Oja's:

- Update: $v_\tau = \hat{v}_{\tau-1} + c\left(\hat{v}_{\tau-1} \cdot Z + Gauss_{vec}(\epsilon, \delta, L, \Gamma)\right)$
- Normalize: $\hat{v}_\tau = \frac{v_\tau}{\|v_\tau\|}$

$1 \times d$ vector

Return $\hat{v}_\Gamma$, $\hat{\lambda}_\Gamma^2 = \|\hat{v}_\Gamma \cdot Z\|_2^2 + Gauss(\epsilon, \delta, L, \Gamma)$

$O(d)$ Space and Time

UTILITY OF JOINT DP FRANK-WOLFE

If Ω = set of non-zero indices in X, $\max_{j \in [n]} \|X_j\|_2 \le L$, $\|R\|_{nuc} \le k$,

and we run (ϵ, δ) Joint DP Frank-Wolfe for T iterations, then w.h.p.:

$$E_{emp} = \frac{1}{|\Omega|} \sum_{i,j \in \Omega} \left(Y_{ij} - X_{ij}\right)^2 = \tilde{O}\left(\frac{k^2}{|\Omega| T} + \frac{kL(dT)^{1/4}}{|\Omega|\sqrt{\epsilon}}\right)$$

Standard FW convergence error

Error due to Privacy

$$E_{gen} = \mathbb{E}_{i,j \sim u[n] \times [d]}\left[\left(Y_{ij} - R_{ij}\right)^2\right] = \tilde{O}\left(\left(\frac{k\sqrt{n+d}}{|\Omega|}\right)^{2/3} + \frac{k^{4/3} L d^{1/4}}{\sqrt{|\Omega|^{13/6} \epsilon (n+d)^{1/6}}}\right)$$

for $T = O\left(\left(\frac{k^4}{|\Omega|(n+d)}\right)^{1/3}\right)$,

and $\boxed{\Omega =_{unif} [n] \times [d]}$

Can be removed via [Shamir Shalev-Shwartz'11]

E.g.: Consider rank-one R s.t

- Elements in R are from a bounded range, e.g., $R_{ij} \in [-1, 1]$
- Each user provides $\approx \sqrt{d}$ ratings, i.e., $|\Omega| \approx n\sqrt{d}$
- The number of users is large w.r.t. items, i.e., $n = \omega(d^{5/4})$.

Hiding privacy parameters, this implies w.h.p.:

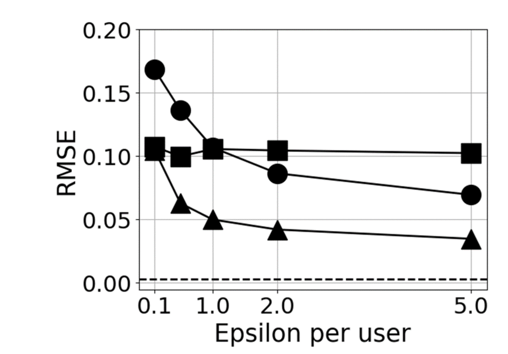
- $E_{emp} = \tilde{O}\left(\sqrt{d}/n^{2/5}\right)$ — $\Rightarrow E_{emp} = o(1)$
- $E_{gen} = o(1)$ — Non-privately, $E_{gen} = o(1)$ for $n = \omega(d)$ [SGS'11]

First *non-trivial* generalization error guarantees with Joint DP

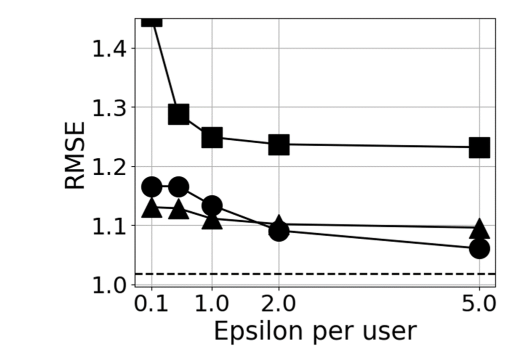
- FW rank-1 updates ⇒ Non-trivial utility via DP

EXPERIMENTAL RESULTS

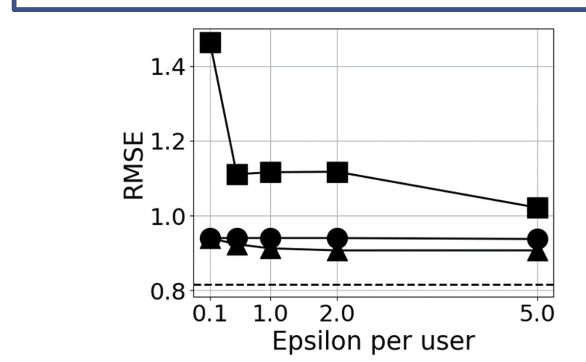
- n = number of users, d = number of items, $\delta = 10^{-6}$
- Unless specified, each rating $\in [0,5]$, sample 80 ratings per user
- (Top 400) – Selecting the 400 most rated items
- Algorithms: DP Frank-Wolfe, DP SVD after cleansing [MM'09], DP Projected Gradient Descent (PGD) [CCS'10, BST'14, ACG+'16]



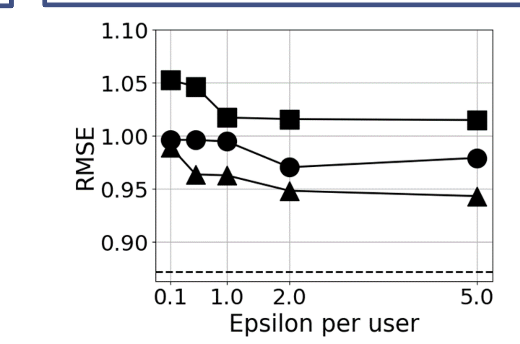
Synthetic dataset
$Y = uv$, $u = [0,1]^{n \times 1}$, $v = [0,1]^{1 \times d}$
$n = 500k$, $d = 400$, ratings $\in [0,1]$



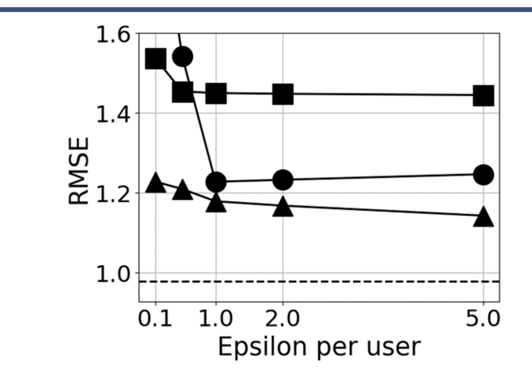
Jester dataset
$n \approx 73k$, $d = 100$ jokes,
No sampling of ratings



MovieLens10M (Top 400)
$n \approx 70k$



Netflix (Top 400)
$n \approx 474k$



Yahoo (Top 400)
$m \approx 995k$

-----	Non-private Baseline
▲	FW Private run
■	[MM09] Private run
●	PGD Private run

Legend for all the plots

REFERENCES

[ACG+'16] Abadi Chu Goodfellow McMahan Mironov Talwar Zhang, *CCS*'16.

[BST'14] Bassily Smith Thakurta, *FOCS*'14.

[CCS'10] Cai Candès Shen, *SIOPT*'10.

[DMNS'06] Dwork McSherry Nissim Smith, *TCC*'06.

[KPRU'14] Kearns Pai Roth Ullman, *ITCS*'14.

[MM'09] McSherry Mironov, *KDD*'09.