Research Summary, 2022-2024

Alejandro Gomez-Leos

Single-Server Private Information Retrieval with Side Information Under Arbitrary Popularity Profiles

Alejandro Gomez-Leos
(UT Austin)

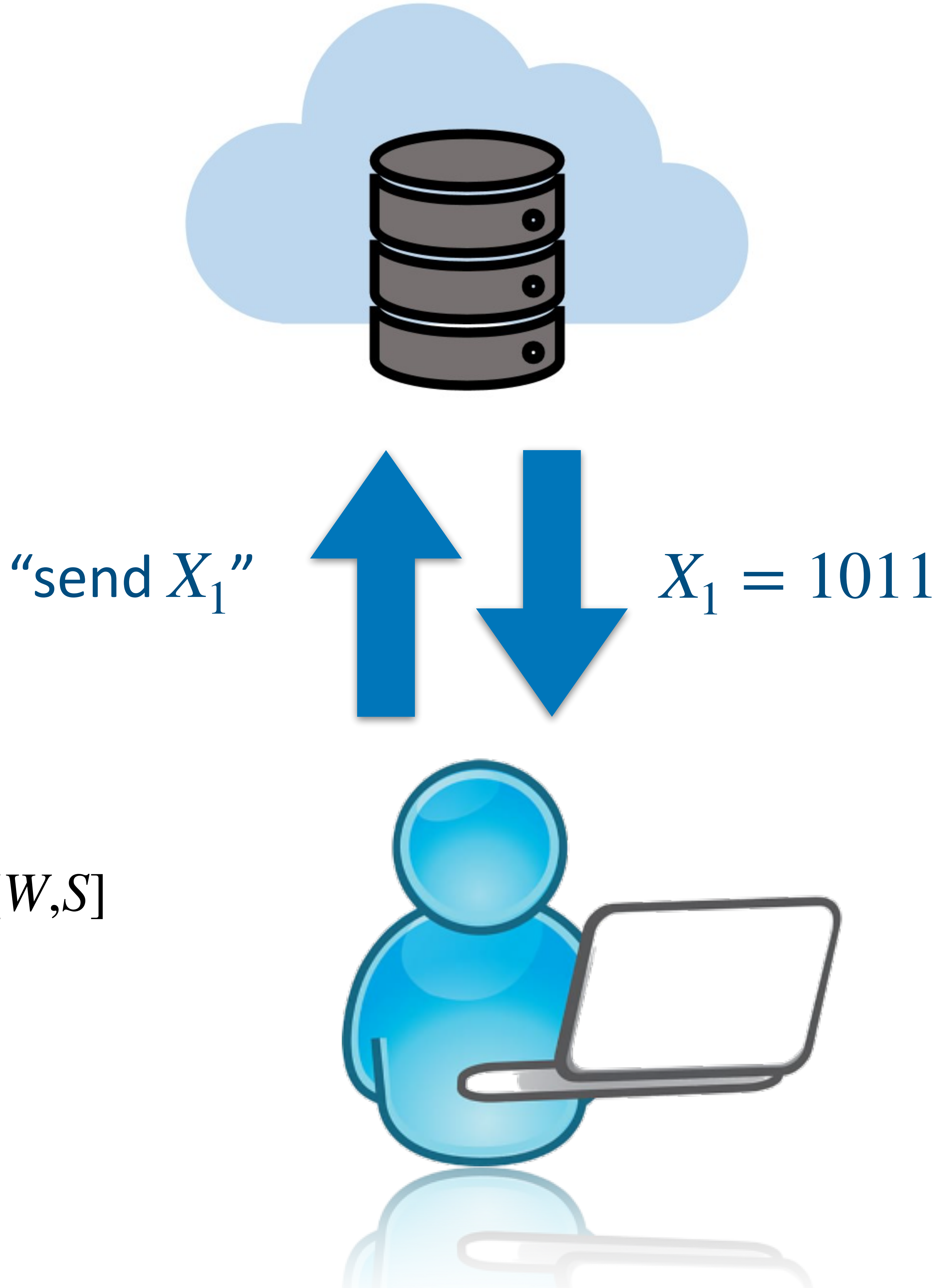
Anoosheh Heidarzadeh
(Santa Clara University)

This work was done while both authors were at Texas A&M University.

Presented at IEEE Information Theory Workshop 2022.

Private Information Retrieval with Side Information (PIR-SI)

- *Can two parties **efficiently** and **unilaterally** exchange information?*
- Two-party protocol client and server
  - Server holds data $\mathscr{D} := \{X_1, X_2, \ldots, X_k\}$ ($X_i \sim_{\mathrm{R}} \mathbb{F}_q^n$)
  - Client demands X_W and secretly knows $(X_i)_{i \in S}$, $m := |S|$
- A protocol prescribes query-answer procedure
  - Client sends bits $Q^{[W,S]}$ asking to compute $f(Q^{[W,S]}, \mathscr{D}) := A^{[W,S]}$
  - Server faithfully sends bits $A^{[W,S]}$



Private Information Retrieval with Side Information (PIR-SI)

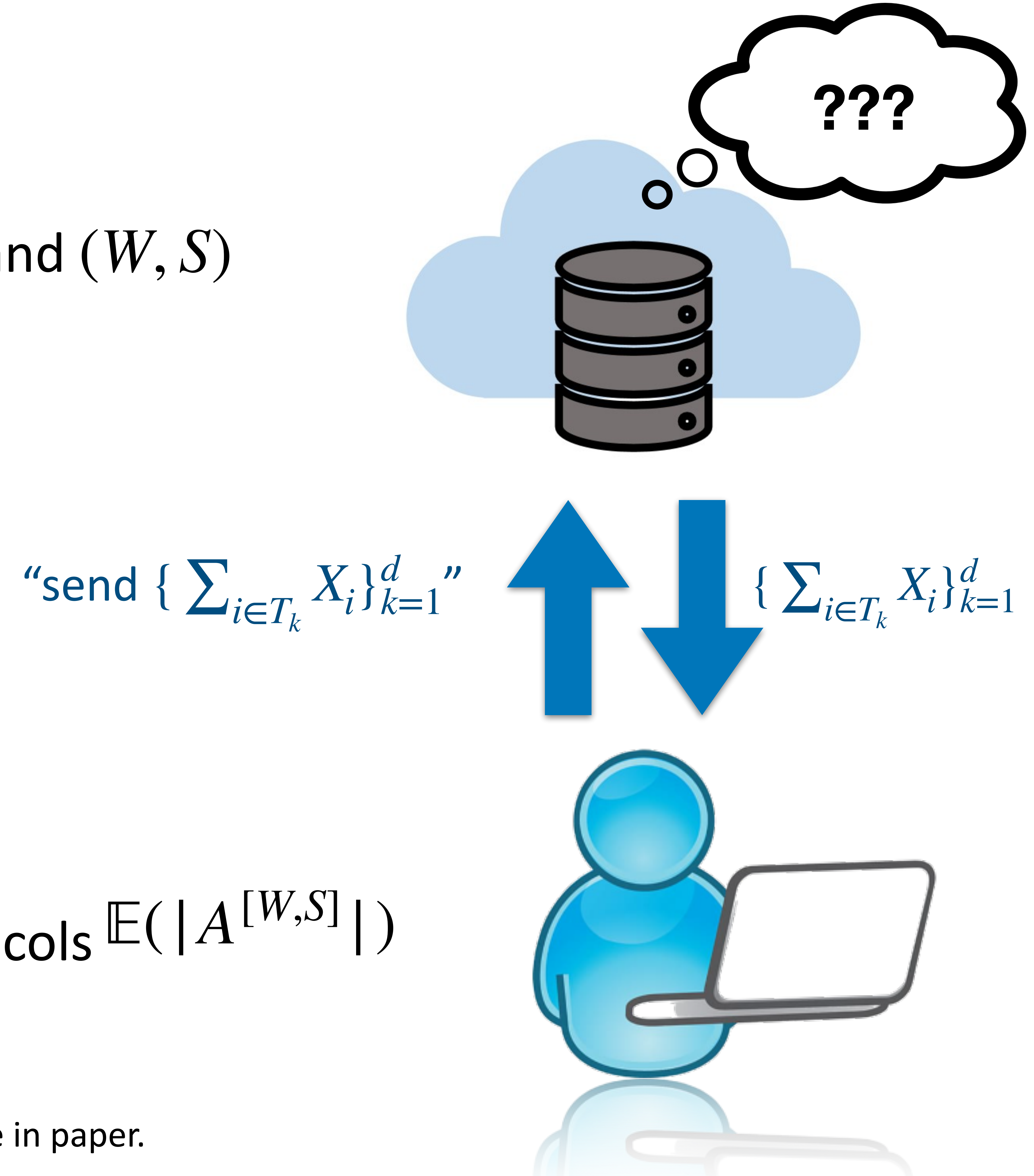
- Good PIR-SI protocols have extra desiderata
  - Privacy: statistical independence between query and (W, S)

$$\mathrm{I}(Q^{[W,S]}; (W, S)) = 0$$

  - Decodability: client can recover desired message

$$\mathrm{H}(X_W \mid A^{[W,S]}, (X_i)_{i \in S}) = 0$$

  - Efficiency: bit complexity* $L := \frac{1}{n \log q} \min_{\text{protocols}} \mathbb{E}(|A^{[W,S]}|)$



*Stated in alternative language in paper.

Related Work

- **Seminal result:** $\Omega(k)$ bits necessary if (i) info-theoretic privacy required (ii) "single database" (iii) $S = \varnothing$ [Chor-Goldreich-Kushilevitz-Sudan 1995, 1998]

- Scenarios admitting cheaper protocols?

  - Relaxing (i): server poly-bounded and exists one-way f [Chor-Gilboa 1997, Kushilevitz-Ostrovsky 1997, Cachin-Micali-Stadler 1999]

  - Relaxing (ii): multiple copies on non-colluding servers [Chor-Goldreich-Kushilevitz-Sudan 1995, 1998, Sun-Jafar 2016, Banawan-Ulukuus 2017, 2018]

  - Relaxing (iii): $S \neq \varnothing$ [Heidarzadeh et al. 2018, Li-Gastpar 2018, Kadhe et al. 2017, 2020, Heidarzadeh-Sprintson 2022]

Disclaimer: not state-of-the-art, many of these are studying slight variations (e.g. multi-message PIR)

Related Work (cont.)

	"Data popularity"	# Servers	Side Info.
Sun-Jafar '17	No	Multiple	No
Banawan-Ulukus '17, '18	No	Multiple	No
Kadhe *et al*. '20	No	Multiple	Yes
Kadhe *et al.* '17	No	Single	Yes
Heidarzadeh *et al*. '18	No	Single	Yes
Li-Gastpar '18	No	Single	Yes
Heidarzadeh-Sprintson '22	No	Single	Yes
Vithana-Banawan-Ulukus '20	Yes	Multiple	No
This work	**Yes**	**Single**	**Yes**

Prior work on PIR-SI assumes marginal distribution of demand is uniform.

We study the feasibility of PIR-SI after relaxing this.

Data Popularity Model

- Popularity profile: $\vec{p} := (p_1, p_2, \ldots, p_k) \in \mathbb{N}^k$. Relative weighting induces distribution:

$$\Pr[W = w \mid S = s] := \frac{p_w}{\sum_{i \notin S} p_i}, \Pr[S = s] := \binom{k}{m}^{-1}$$

Theorem [Kadhe et al. 2017]: Let $n, k, m \in \mathbb{N}$ where and $m + 1 \mid k$. If $\vec{p} = \mathbf{1}$, then

$$L(k, m, \vec{p}) = \frac{k}{m+1}.$$

- General case?

Main Result

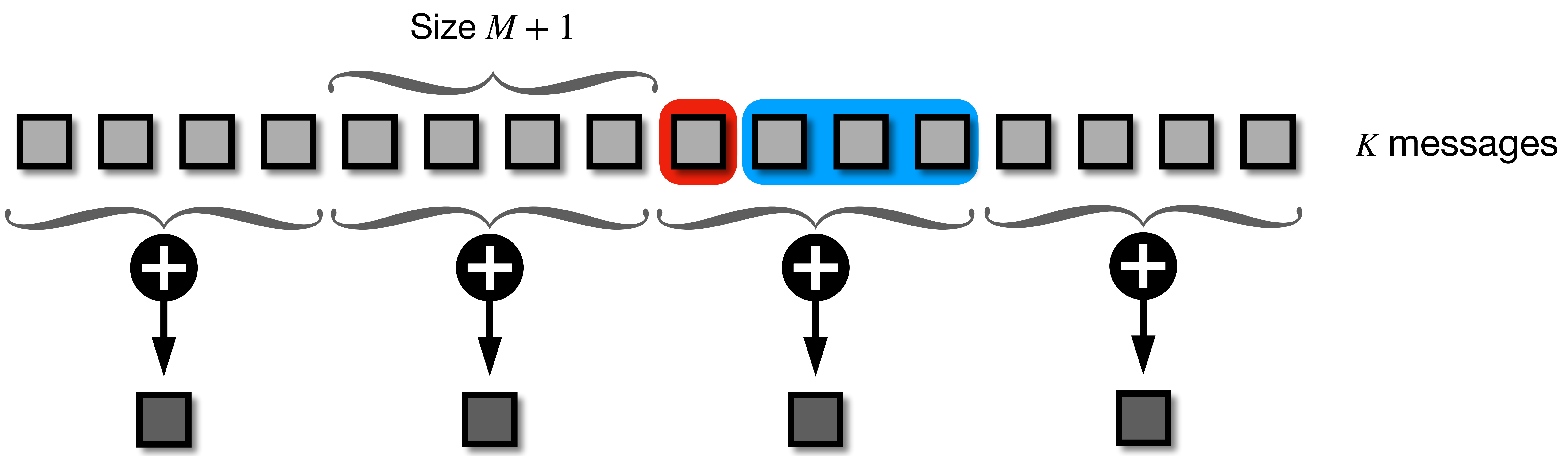
Theorem (this work): Let $k, m \in \mathbb{N}$ where $m+1 < \sqrt{k}$ and $m+1 \mid k$.

There's an $\delta := \delta(\vec{p})$ such that

$$\frac{k}{m+1} \leq L(k, m, \vec{p}) \leq (k-m) \cdot \delta + \frac{k}{m+1} \cdot (1-\delta)$$

- **Corollary**: $\max |p_i - p_j| = O(1) \Longrightarrow \delta = O(1/k)$ and RHS tight
- **Note:** protocol runs in time $k^{1+O(m)}$ or $O(m)$ if marginals pre-computed
- Based on optimal interpolation between two previously known protocols

Partition-and-Code Scheme [Kadhe et al. 2017]



- $P_0 := \{w\} \cup \{i : i \in S\}$
- Sample random $(m + 1)$-uniform partition of $[k] \backslash P_0$
- Query $k/(m + 1)$ sums over each

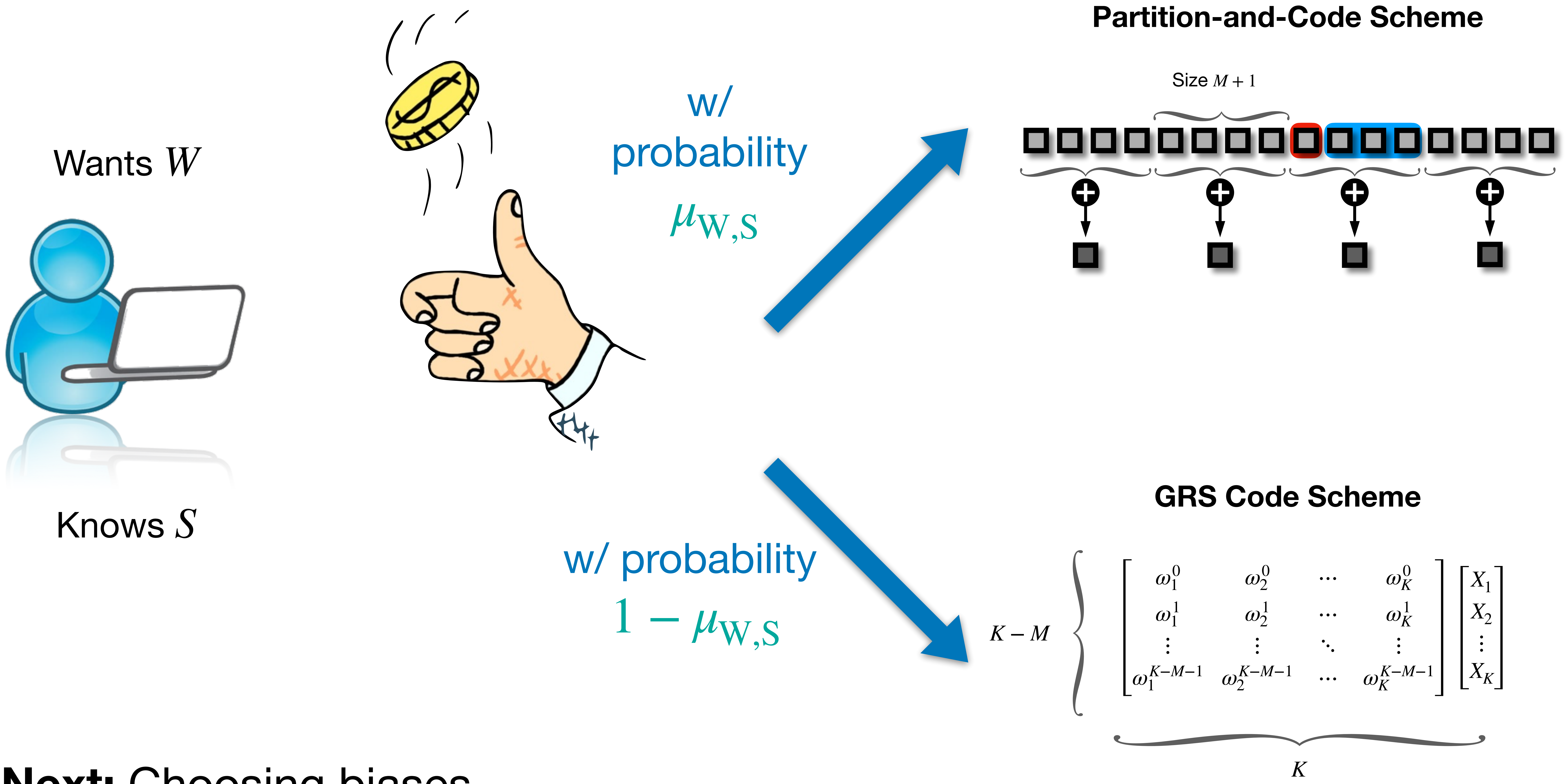
Generalized Reed-Solomon Scheme [Heidarzadeh et al. 2018]

- Pick distinct $\omega_1, \ldots, \omega_k \in \mathbb{F}_q$.
- Query $k - m$ linear combinations of this form.

$$K-M \left\{ \begin{bmatrix} \omega_1^0 & \omega_2^0 & \cdots & \omega_K^0 \\ \omega_1^1 & \omega_2^1 & \cdots & \omega_K^1 \\ \vdots & \vdots & \ddots & \vdots \\ \omega_1^{K-M-1} & \omega_2^{K-M-1} & \cdots & \omega_K^{K-M-1} \end{bmatrix} \right. \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_K \end{bmatrix}$$

$$\underbrace{\hspace{10em}}_{K}$$

PC-GRS Scheme



Next: Choosing biases

Choosing Coin Biases

Minimize $$\sum_{(w,s)} \Pr[W = w, S = s] \cdot \left[\mu_{w,s}\left(\frac{K}{M+1}\right) + (1-\mu_{w,s})(K-M)\right]$$

s.t. $$\Pr[W = w \mid \text{query Q}] = \Pr[W = w] \quad \forall w \in [k]$$

- $\Omega((k/m)^m)$ size? **No!**
- $m + 1 < \sqrt{k}$ implies reduction to smaller program of size $O(m)$
  - Non-trivial pigeonholing argument (see Lemma 3)

Lower Bound

- **Claim:** $\Omega(k/(m+1))$ bits are necessary.

- **Proof sketch:** (use entropy chain-rule)

For query generated by protocol, consider any message X_i.

By decodability, $\exists\ m$ messages such that X_i recoverable from them (otherwise $p_i = 0$).

Repeat argument over remaining messages, yields $k/(m+1)$ pairs.

$\implies$ server's response must have at least $k/(m+1)$ linear combinations $\blacksquare$.

Summary

- Generalized PIR-SI to **nonuniform** demands
- Bounded the bit complexity of this problem
  - Optimal protocols when popularities pairwise within $O(1)$
  - $O(m)$ runtime for pre-computed marginals
- **Open problem:** Tight lower bound in other regimes?

Normal Bandits with Noisy Probes

Alejandro Gomez-Leos
(UT Austin)

Gustavo de Veciana
(UT Austin)

Sanjay Shakkottai
(UT Austin)

In-progress.

k-Armed Bandits

- Sequential decision-making*: how to compete with best player in hindsight?*
- Unknown distributions $D_1, \ldots, D_k$ with means $\mu_1, \ldots, \mu_k$
- Play for N rounds. On round $t \in [N]$:
  - Pick arm $a_t \in [k]$, obtain iid reward $x_t \sim D_{a_t}$
  - Suffer round regret $\text{REG}_t = \max \mu_i - \mathbb{E}(x_t)$
  - Use history to inform over next arm decision (follow seq. of policies $\pi_t : (a_1, x_1, a_2, x_2, \ldots, x_{t-1}) \mapsto [k]$)
- **Question:** Which policies admit asymptotically optimal (cumulative) regret?

The Classics

- UCB is the algorithmic workhorse for many variations:
  - **High level idea:** maintain time-averages and **u**pper-**c**onfidence **b**ounds for each. Pick arm with highest (empirical mean) + UCB.
  - Essentially optimal for σ^2-subgaussian inputs, yields regret $O(\sigma^2 k \log N)$ [Burnetas-Katehakis 1996, Auer-Bianchi-Fischer 2002]
  - *Why does it work?*

high regret $\implies$ chose too many suboptimal arms $\implies$ their bounds are small $\implies$ their empirical means are too high $\implies$ exp. small prob. event!

- **Intuition:** One should take risks for long-term wellbeing.

k-Armed Bandits with a Twist

- $A \in \mathbb{N}$ actions, $P \in \mathbb{N}$ probes

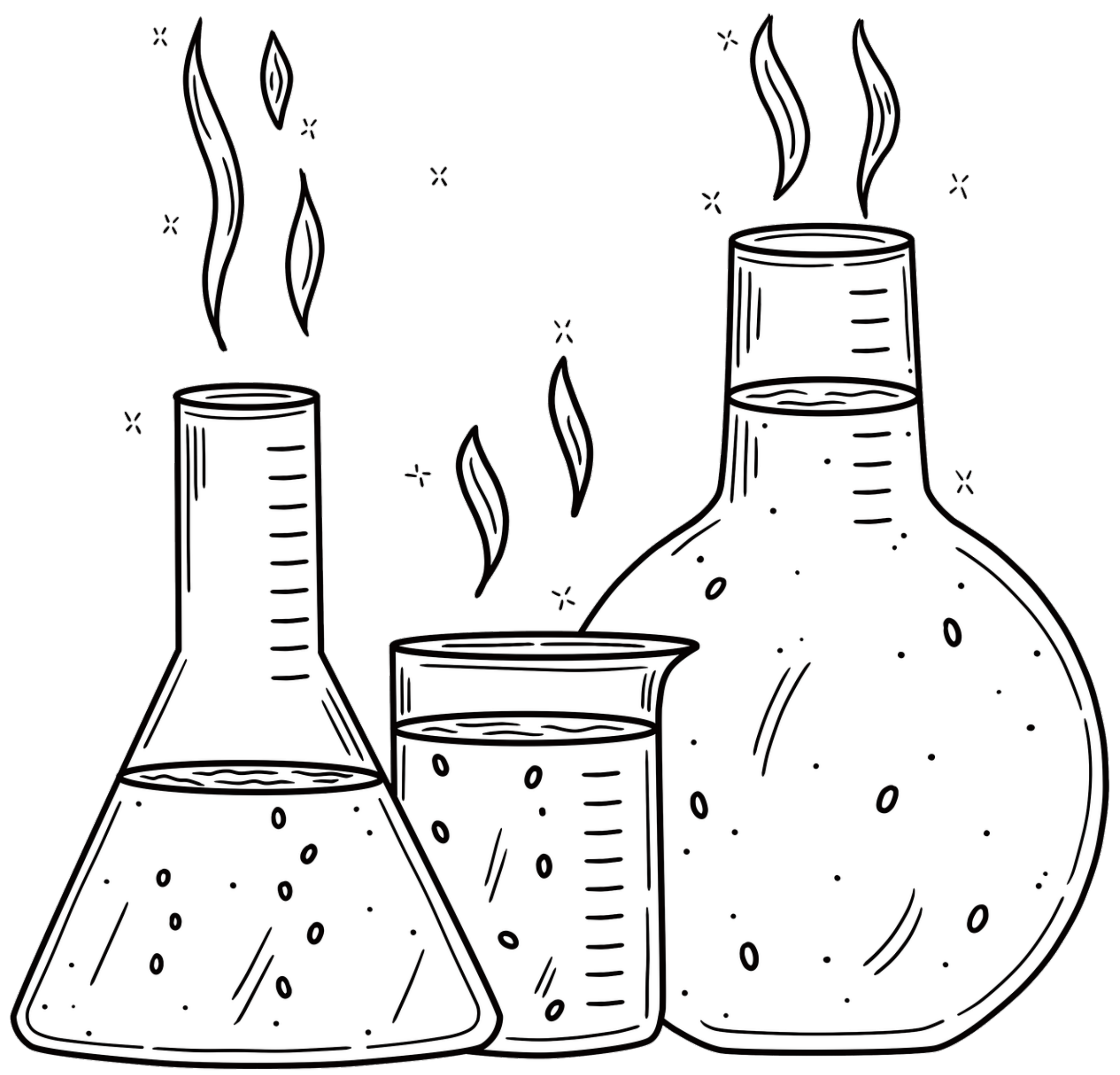
- Parameters $(\mu_a)_{a\in[A]} \subseteq \mathbb{R}$ and $(\nu_p)_{p\in[P]} \subseteq \mathbb{R}^+$ unknown beforehand

- On round $t \in [N]$, pick (action, probe) = $(a_t, p_t) \in [A] \times [P]$

    - Obtain iid $x_t \sim \mathcal{N}(\mu_{a_t}, \nu_{p_t})$

    - suffer Mean-Var* regret $\Delta^{\mathrm{MEAN}}_{a_t} + \Delta^{\mathrm{VAR}}_{p_t} = (\mu_{\mathrm{MAX}} - \mu_{a_t}) + (\nu_{p_t} - \nu_{\mathrm{MIN}})$

- **Motivation:** Decision-making in which measurement devices are part of action space. For example, industrial chemical manufacturing



*Canonical bandit problem variation

Main Result

- Trivial reduction to AP-armed bandit $\Longrightarrow O(AP \log N)$ regret policy
- Is $O((A+P) \cdot \log N)$ doable? **Yes!**

Theorem (this work): Suppose π suffers regret $R(\pi, N) \leq N^{0.99}$ on every instance. Then,

$$R(\pi, N) \gtrsim \sum_{a \in [A]: \Delta_a^\mu > 0} \frac{\nu_{MIN}^2 \log N}{\Delta_a^\mu} + \sum_{p \in [P]: \Delta_p^\nu > 0} \frac{\nu_p^2 \log N}{\Delta_p^\nu}$$

Moreover, [our policy] π^* suffers regret upper-bounded by a constant factor of the above.

- Ours is *consistency-optimal:* the only "better" policies sometimes suffer nearly-linear regret
- These are "silly" policies, such as "always-take-decision-X" which is trivially unbeatable in the environments where "X" is optimal

Upper Bound Main Ideas

- Our policy is inspired by alternating-optimization algorithms like *coordinate descent.*
  - **High-level idea:** maintain time-averages and <u>l</u>ower-<u>c</u>onfidence <u>b</u>ounds of variances. Each round, pick probe with lowest (empirical variance) - LCB. Then, pick action with highest (empirical mean) + UCB where this UCB score is specific to the chosen probe.
  - **Proof in a nutshell:** Separate regret from optimal and suboptimal probes. Former analysis similar to UCB. For the other part, carefully ensure LCB scores "concentrate quickly" and not too many suboptimal probes played.

Lower Bound Main Ideas

- **Lemma 1:** Let P, Q be probability measures on measurable space $(\Sigma, \mathscr{F})$. Let $A \in \mathscr{F}$ be any event. Then $P(A) + Q(\bar{A}) \geq (1/2) \cdot \exp(-D_{KL}(P \| Q))$ [Bretagnolle-Huber 1978, Tsybakov 2010].

- **Lemma 2 (informal):** The KL-divergence between two environments for the same policy is equivalent to the weighted sum of the associated decision distributions' KL-divergence [Tor-Szepesvari 2020].

$$D_{KL}(P_{\pi,E} || P_{\pi,E'}) = \sum_{a \in [A]} \sum_{p \in [P]} \mathbb{E}_{\pi,E}[N^T_{(a,p)}] \cdot D_{KL}(P^{(a,p)}_{\pi,E} || P^{(a,p)}_{\pi,E'})$$

- **Lower bound strategy:** We construct two instance families, each with small statistical diameter. Each forces policies to be sufficiently explorative, suffering appropriate regret. We take the max of both.

- **Lemma 1:** Let P, Q be probability measures on measurable space $(\Sigma, \mathscr{F})$. Let $A \in \mathscr{F}$ be any event. Then $P(A) + Q(\overline{A}) \geq (1/2) \cdot \exp(-D_{KL}(P \| Q))$ [Bretagnolle-Huber 1978, Tsybakov 2010].

- **Lemma 2 (informal):** The KL-divergence between two environments for the same policy is equivalent to the weighted sum of the associated decision distributions [Tor-Szepesvari 2020].

$$D_{KL}(P_{\pi,E} \,||\, P_{\pi,E'}) = \sum_{a \in [A]} \sum_{p \in [P]} \mathbb{E}_{\pi,E}[T^{(a,p)}(N)] \cdot D_{KL}(P^{(a,p)}_{\pi,E} \,||\, P^{(a,p)}_{\pi,E'})$$

- **Main idea of lower bound:** Consider environments E and E' in which optimal probe is switched (say probe p and p').

$$2N^{0.99} \geq R(\pi, N, E) + R(\pi, N, E') \quad \text{"consistency"}$$

$$\gtrsim N/2 \left(P_{\pi,E}(\text{play p' N/2 times}) + P_{\pi,E'}(\text{play p N/2 times}) \right)$$

$$\gtrsim (N/2) \cdot \exp(-D_{KL}(P_{\pi,E} \| \| P_{\pi,E})) \quad \text{"lemma 1"}$$

$$\gtrsim (N/2) \cdot \exp(-\mathbb{E}_{\pi,E}[T^{(a,p)}(N)]) \quad \text{"lemma 2"}$$

$$\therefore \mathbb{E}_{\pi,E}[T^{(a,p)}(N)] \gtrsim \log N$$

Summary

- **Main takeaways**:
  - UCB works very well when we know *how much* we ought to boost estimates, which requires knowledge of variance or estimates [Audibert-Munos-Szepesvari 2009, Wesley-Honda-Katehakis 2017].
  - If samples are correlated in our sense, then those variance estimates can be adaptively and optimally be controlled via another layer of optimism.
- **Open problem:** Extensions, high-probability regret guarantees, weighted mean-var regret …

Simple and Nearly-Optimal Sampling for Rank-1 Tensor Completion via Gauss-Jordan

Alejandro Gomez-Leos
(UT Austin)

Oscar F. Lopez
(Florida Atlantic University)

Preprint.

Introduction

$$\begin{bmatrix} 1 & ? & ? & ? \\ 5 & ? & 7 & ? \\ ? & 10 & -7 & 4 \\ 13 & ? & ? & ? \end{bmatrix}$$

- Assume sample access to a **low-rank** matrix $M \in \mathbb{R}^{d \times d}$

- Matrix completion: *how many samples required fill in M?*

- Tensor completion: generalization to low-rank multilinear forms in $\bigotimes_{i=1}^{N} \mathbb{R}^d$

- Theory motivated by practical success in industrial and scientific computing

- **Def:** A tensor $\mathscr{U} \in \bigotimes_{i=1}^{N} \mathbb{R}^d$ is rank-1 if $\exists \, \{\mathbf{u}_1, \ldots, \mathbf{u}_N\} \subseteq \mathbb{R}^d$ s.t. $\mathscr{U}_{(i_1, i_2, \ldots, i_N)} = \prod_{k=1}^{N} (\mathbf{u}_k)_{i_k}$

Introduction (cont.)

- **Problem:** Given uniformly drawn entries $\mathscr{U}$, output $\hat{\mathscr{U}}$ where $\hat{\mathscr{U}} = \mathscr{U}$ w.p. $\geq 2/3$
- Why study?
  - Special case of well-studied generalizations (results up next)
  - Independent interest, particularly from geometric perspective [Kahle et al. 2017, Jaramillo 2018, Singh-Shapiro-Zhang 2020, Zhou-Ne-Peng-Zhou 2024]
- **This work:** a simple linear algebraic characterization, and application to problem above
- Assume for simplicity all components are nonzero

Main Result

Theorem (this work): Let $d, N, q \in \mathbb{N}$. If $\mathscr{U} \in \bigotimes_{i=1}^{N} \mathbb{R}^d_{\neq 0}$ is a rank-1 tensor, then

1. $m = O((dN)^2 \cdot \log d)$ samples suffice to recover q entries of $\mathscr{U}$ in time $O(qN + md^2)$.
2. Moreover, $\Omega(d \cdot \log(dN))$ samples are necessary.

- Sampling complexity upper-bounds usually dependent on *incoherence** μ ($= \Omega(d)$ in worst-case).
  - $N = 2$: $d\mu \log^{O(1)} d$ entries suffice [Candes-Tao 2010, Recht 2011, Candes-Recht 2012, Chen 2015]
  - $N = 3$: $d^{3/2}\mu^{O(1)} \log^{O(1)} d$ entries suffice [Jain-Oh 2014, Xia-Yuan 2019, Liu-Moitra 2020]
  - $N \geq 4$: $d^{N/2}(\mu N)^{O(N)} \log^{O(1)} d$ entries suffice [Krishnamurthy-Singh 2013, Montanari-Sun 2018, Stephan-Zhu 2024, Haselby et al. 2024]

*Informally measures how well components discorrelate with fixed basis.

Main Result

Theorem (this work): Let $d, N, q \in \mathbb{N}$. If $\mathscr{U} \in \bigotimes_{i=1}^{N} \mathbb{R}^{d}_{\neq 0}$ is a rank-1 tensor, then

1. $m = O((dN)^2 \cdot \log d)$ samples suffice to recover q entries of $\mathscr{U}$ in time $O(qN + md^2)$.
2. Moreover, $\Omega(d \cdot \log(dN))$ samples are necessary.

- **Notes:**
  - Tight up to a factor of d when $N = \Theta(1)$ ($d \gg N$ in practice)
  - $\forall \rho > 0, \quad \exists$ hard instance family where few samples $\Longrightarrow \|\mathscr{U} - \hat{\mathscr{U}}\|_F \geq \rho d^{(N-1)/2}$ with large prob.
  - No dependence on μ

Main Ideas

Lemma: There exists a unique matrix $\mathbf{A}$ and bijections $f, \tilde{f}$ with the following property.

Any nonzero tensors $\mathscr{U}, \mathscr{T}$ induce the linear systems

1. $\mathbf{A}x = f(\mathscr{U})$ over $\mathbb{F}_2$ and $\mathbf{A}x = \tilde{f}(\mathscr{U})$ over $\mathbb{R}$,

2. $\mathbf{A}x = f(\mathscr{T})$ over $\mathbb{F}_2$ and $\mathbf{A}x = \tilde{f}(\mathscr{T})$ over $\mathbb{R}$,

where (i) $\mathscr{U}$ **is rank-1 iff** (1) **is consistent**, and (ii) $\mathscr{U} = \mathscr{T}$ **and rank-1 iff (1) and (2) have same solution sets.**

- **Proof sketch:**

$$\mathscr{U}_{(i_1,i_2,\ldots,i_N)} = \textbf{sign}\left(\prod_{\ell=1}^{N}(\mathbf{u}_\ell)_{i_\ell}\right)\left|\prod_{\ell=1}^{N}(\mathbf{u}_\ell)_{i_\ell}\right| = \left(\prod_{\ell=1}^{N}\textbf{sign}\left((\mathbf{u}_\ell)_{i_\ell}\right)\right)\left(\exp\left(\sum_{\ell=1}^{N}\log\left|(\mathbf{u}_\ell)_{i_\ell}\right|\right)\right) := \mathscr{U}'_{(i_1,i_2,\ldots,i_N)}\exp\left(\mathscr{U}''_{(i_1,i_2,\ldots,i_N)}\right)$$

$$\varphi\left(\mathscr{U}'_{(i_1,i_2,\ldots,i_N)}\right) = \sum_{\ell}\varphi\left(\textbf{sign}\,(\mathbf{u}_\ell)_{i_\ell}\right) \iff \mathbf{A}x = \varphi\left(\textbf{sign}(\textbf{vec}\mathscr{U})\right)$$

$$\mathscr{U}''_{(i_1,i_2,\ldots,i_N)} = \sum_{\ell}\log\left|(\mathbf{u}_\ell)_{i_\ell}\right| \iff \mathbf{A}x = \log|\textbf{vec}\mathscr{U}|$$

Main Ideas (cont.)

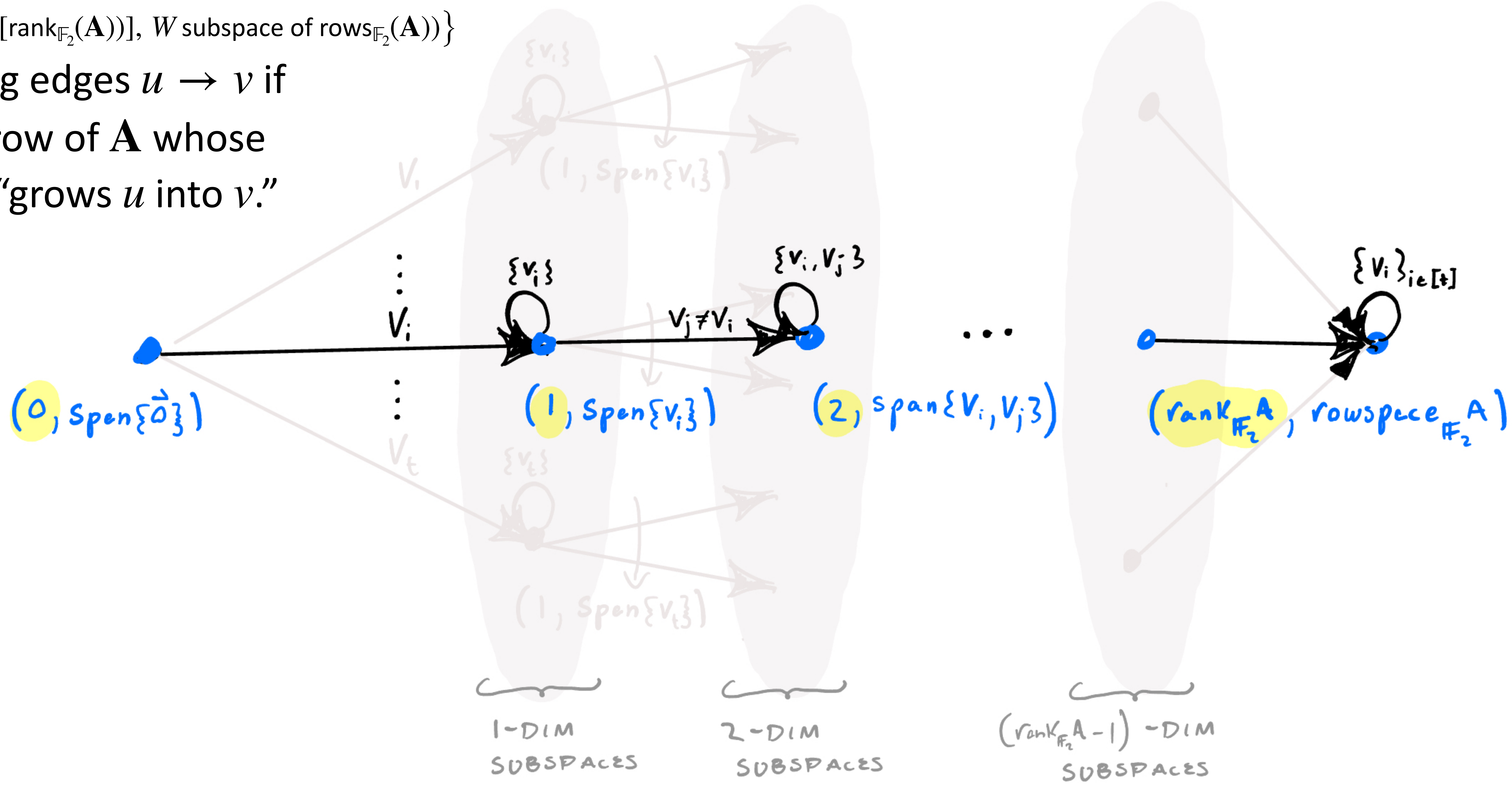
$$\begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

Example $\mathbf{A}$ when $(d, N) = (2,3)$

- Think of linear systems represented by their augmented matrices
  - $\implies$ "observed entries are isomorphic to partial linear systems"
  - $\implies$ rank-1 TC $\equiv$ sketching $\mathbf{A}$!
  - Over $\mathbb{R}$, leverage-score sampling says $O((dN) \cdot \log(dN))$ samples suffice [Cohen-Lee-Musco-Musco-Peng-Sidford 2014]
- **Challenge:** Working over $\mathbb{F}_2$ as well (other machinery requires matrix-Chernoff bounds)
- **Fix:** Express sample complexity as hitting time of random walk on subspace graph

Main Ideas (cont.)

- Consider digraph on vertices $\{(\alpha, W) \mid \alpha \in [\text{rank}_{\mathbb{F}_2}(\mathbf{A})],\ W \text{ subspace of rows}_{\mathbb{F}_2}(\mathbf{A}))\}$ connecting edges $u \to v$ if there's a row of $\mathbf{A}$ whose inclusion "grows u into v."

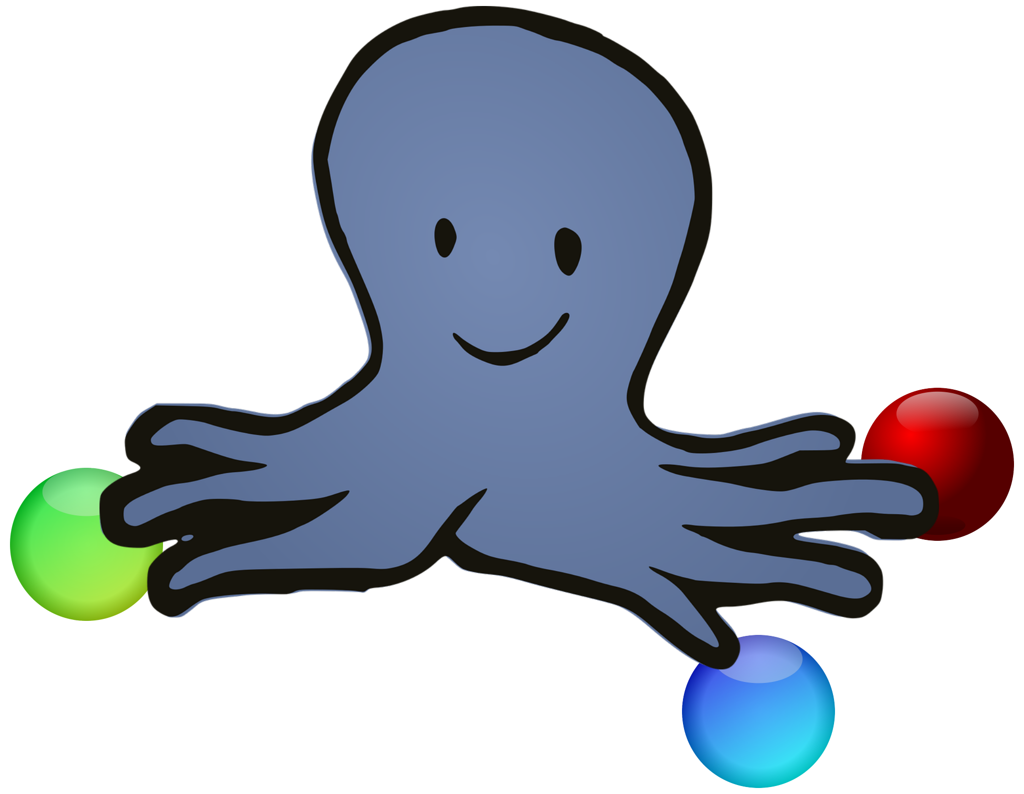


Main Ideas (cont.)

- **Lemma:** A random walk on this graph starting at node $(0,\mathbf{0})$ hits the absorbing state $(\text{rank}_{\mathbb{F}_2}(\mathbf{A}), \text{rowspace}_{\mathbb{F}_2}(\mathbf{A}))$ w.p. $\geq 2/3$ after $\lesssim d^2N$ steps.

- **Proof sketch:** Chain self-loops w.p. $\geq 1/d$. In expectation d steps before transitioning. Cannot transition more than $\text{rank}_{\mathbb{F}_2}(A) = \Theta(dN)$ steps. Entire trajectory takes d^2N steps. Claim follows by Markov's inequality.

Lower Bound



- **Lemma (informal):** Consider coupon collector variant: N urns, each with d unique balls. Each round draw in parallel a ball from each (N per round). $\Omega(d \log dN)$ draws necessary.

- **Proof sketch:** Track martingale generated by "did-we-observe it" indicator variables. Apply Hoeffding's lemma in manner similar to Azuma-Hoeffding proof.

- **Rough sketch of lower bound**: Pick $\mathbf{u}_i's \sim_R \{\pm 1\}^d$, let $\mathscr{U} = \rho(\mathbf{u}_1 \otimes \ldots \otimes \mathbf{u}_N)$.

Correspond balls to component coordinates and correspond draws to observations.

One can show $\mathbb{E}[\|\hat{\mathscr{U}} - \mathscr{U}\|_F^2] \geq \rho d^{N-1}$, and then apply reversed Markov inequality.

Summary

- Simplified pre-existing understanding of rank-1 tensor completion.
- Problem difficulty doesn't depend on incoherence, problem reduces to matrix sketching problem.
- **Open problem:** Improve upper bound to match lower bound.

Thanks!