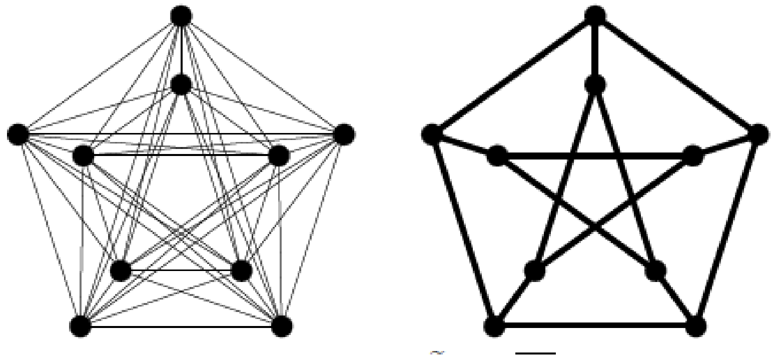
SPECTRAL OPTIMIZATION VIA MATROID INTERSECTION

Mohit Singh, Georgia Tech

SPECTRAL OPTIMIZATION

- Represent data using vectors and matrices.
- Study the spectral properties of natural matrices associated with combinations of data.
  - Determinant, maximum or minimum eigenvalue etc.

- Spectral Sparsifiers and Cut sparsifiers [Spielman, Teng'08].



- This talk will be about optimizing the spectral function: determinant.
  - Close relationship to matroid intersection.

OUTLINE

- Introduction
  - Problem Statement
  - Applications
- Results and Techniques
  - Combinatorial Methods: Matroid Intersection
- Open Questions

PROBLEM STATEMENT

- Given $v_1, \dots, v_n \in \mathbb{R}^d$, approximate

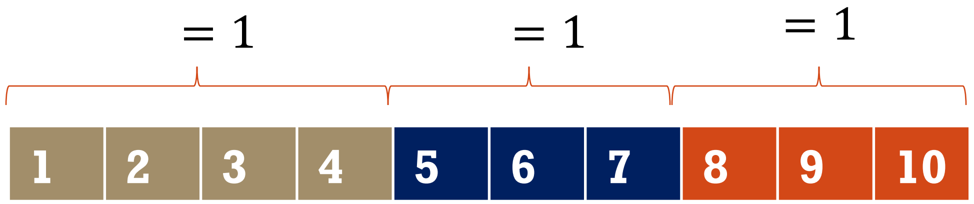
$$\mathrm{OPT}(v_1, \dots, v_n; \mathcal{U}_k) = \max\left\{\det\left(\sum_{i \in S} v_i v_i^T\right)^{\frac{1}{d}} : |S| \le k\right\}$$

- and more generally,

$$\mathrm{OPT}(v_1, \dots, v_n; \mathcal{B}) = \max\left\{\det\left(\sum_{i \in S} v_i v_i^T\right)^{\frac{1}{d}} : S \in \mathcal{B}\right\}$$

where $\mathcal{B}$ = bases of a matroid of rank $k \ge d$ over $[n]$

- Examples:
  - Uniform matroid: $\mathcal{B} = \{S \subseteq [n] : |S| = k\}$
  - Partition matroid: $\mathcal{B} = \{S \subseteq [n] : \forall i\ |S \cap P_i| = 1\}$, where $[n] = P_1 \uplus \dots \uplus P_k$



EXPERIMENTAL DESIGN

- Unknown $\theta^* \in \mathbb{R}^d$.
- Obtain linear measurements $y_i = v_i \cdot \theta^* + \eta_i$

  for some $v_i \in \{v_1, \ldots, v_n\}$ and noise $\eta_i = N(0, \sigma)$.
- Feasible sets of measurements: $S \in \mathcal{B}$, where $\mathcal{B} \subseteq \binom{[n]}{k}$
  - E.g., $\mathcal{B} = \{S : |S| = k\}$: can only make k measurements
- Goal: choose $S \in \mathcal{B}$ so that the MLE

$$\hat{\theta} = \arg\min \left\{ \sum_{i \in S} (y_i - v_i \cdot \theta)^2 : \theta \in \mathbb{R}^d \right\}$$

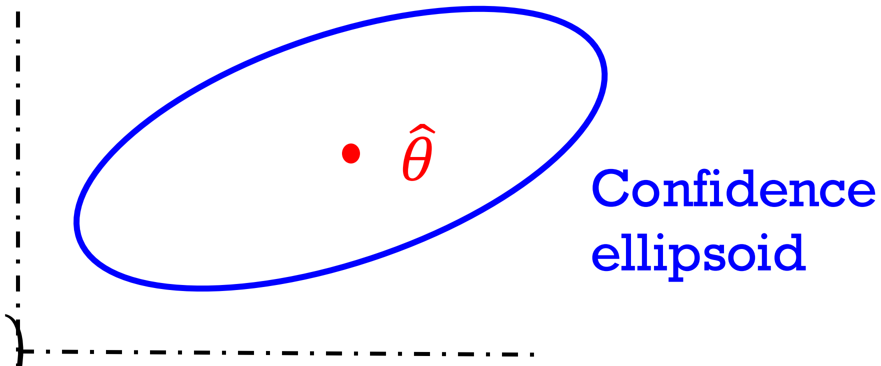
  is as accurate as possible

EXPERIMENTAL DESIGN: OBJECTIVES

- $\hat{\theta} = \arg\min\{\sum_{i\in S}(y_i - v_i \cdot \theta)^2 : \theta \in \mathbb{R}^d\}$
- $\hat{\theta} - \theta^* \sim N(0, \sigma \cdot \Sigma)$, where $\Sigma = \left(\sum_{i\in S} v_i v_i^T\right)^{-1}$
- Confidence ellipsoid : $E = \hat{\theta} + O(\sqrt{\mathrm{d}}\,\sigma) \cdot \Sigma^{1/2} \cdot B_2^d$
  - E contains θ^* with 95% probability
- D-optimal design:

$$\max\left\{\det\left(\sum_{i\in S} v_i v_i^T\right)^{\frac{1}{d}} : S \in \mathcal{B}\right\}$$

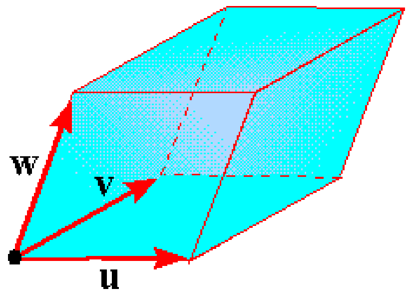
  - Equivalent to minimizing the volume of E
- Other objectives: minimize $tr(\Sigma)/\mathrm{d}$, $\|\Sigma\|_{op}$, …



[Kiefer '59, Fedorov '72, Kiefer '75, Atkinson and Donev '92, Miller and Nguyen '94, Pukelsheim '06, Avron and Boutsidis '13, Allen-Zhu, Li, Singh and Wang '17, Singh and Xie '18, Atkinson, Nikolov, Singh, and Tantipongpipat '19…..]

CONVEX GEOMETRY

- Given a set of n vectors $v_1, \ldots, v_n \in R^d$, pick (d+1) vectors to maximize the volume of the simplex formed by them.
- Given a set of n vectors $v_1, \ldots, v_n \in R^d$, pick d vectors to maximize the volume of the parallelopiped formed by them.
  - Volume of parallelopiped formed by $\{v_i : i \in S\}$ is exactly $\sqrt{\det(V_S^T V_S)}$.



- Can be reduced to Determinant maximization.
- Khachiyan[1995], Di Summa, Eisenbrand, Faenza, Moldenhauer[2014], Nikolov[2016].
- Closely related to Determinantal Point Processes [Kulesza,Taskar'2012] arriving in probability theory and machine learning.

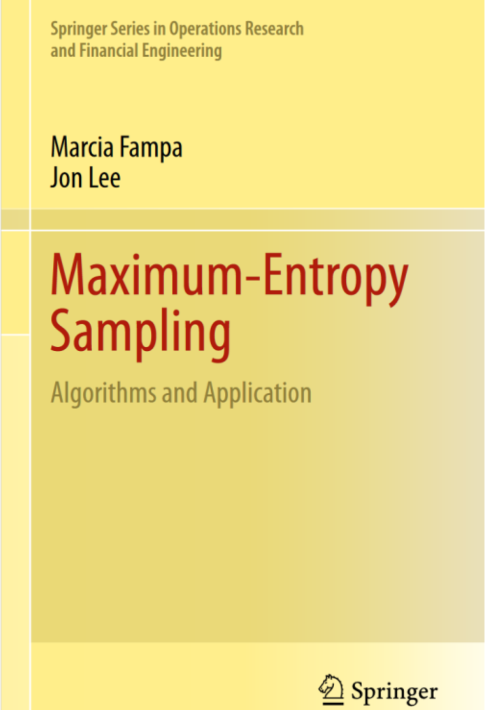
MAXIMUM ENTROPY SAMPLING

A closely related cousin of determinant maximization is the maximum entropy sampling problem [Fampa, Lee ‘23].

Given a $d \times d$ P.S.D. matrix C and integer $k \leq d$, find a principal submatrix B of C of maximum determinant.

-Work on stronger upper and lower bounds. [Ko, Lee, Wayne’98, Fampa, Lee’22]
-Branch and bound methods.[Ko, Lee, Queryanne’95][Lee’98]

Lots of this work also applies to determinant maximization.

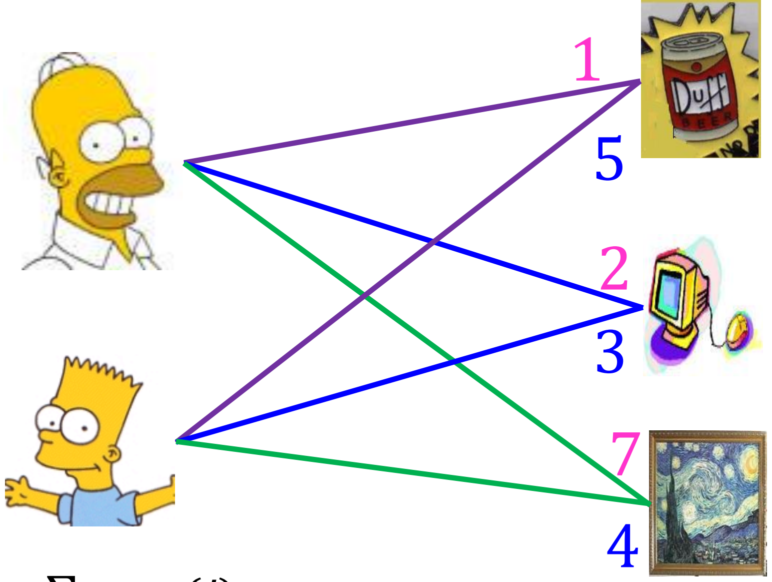


NETWORK DESIGN

- Given a graph $G = (V, E)$, find a spanning subgraph $H = (V, F)$ that is
  - *well connected.*
  - $H = (V, F)$ is chosen such that F satisfies certain combinatorial constraints, for example,
    - $|F| \leq k$, or
    - Given a coloring of E, pick at most 1 edge of each color in F.

- Well connected=> maximize number of spanning trees in H
  [Li, Patterson, Yi, Zhang 19]
- Graph $G = (V, E)$. For an edge $e = ab \in E$, let $v_e = 1_a - 1_b \in R^V$
- Laplacian: $L_G = \sum_{e \in E} v_e v_e^T$
- [Kirchoff 1847] #{spanning trees of G} $= \det\left(L_G + \frac{1}{|V|^2} 1 1^T\right)$

ALLOCATION OF GOODS



- m goods, d agents; agent i has utility $u_i(j)$ for good j, and $u_i(S) = \sum_{j \in S} u_i(j)$
- Goal: find allocation $\sigma: [m] \to [d]$ to maximize welfare
- Nash Social Welfare: maximize geometric mean of agent utilities

$$\max\left\{\prod_{i=1}^{d}\left(\sum_{j:\sigma(j)=i} u_{i(j)}\right)^{\frac{1}{d}} : \sigma: [m] \to [d]\right\}$$

  - Interpolates between total utility and min utility

- Can be modeled as determinant maximization s.t. partition constraints [Anari, OveisGharan, Saberi, Singh 17]. e-approximation.
- Approximation for Nash Social Welfare: [Cole, Gkatzelis'16], [Barman, Krishnamurthy, Vaish'18] 1.45-approximation.
- Recent works generalize the models [Anari, Mai, Oveis-Gharan, Vazirani'18], [Garg, Vegh, Husic' 20], [Barman, Krishna, Kulkarni, Narang'21], [Barman, Verma' 21], [Li, Vondrak'21]

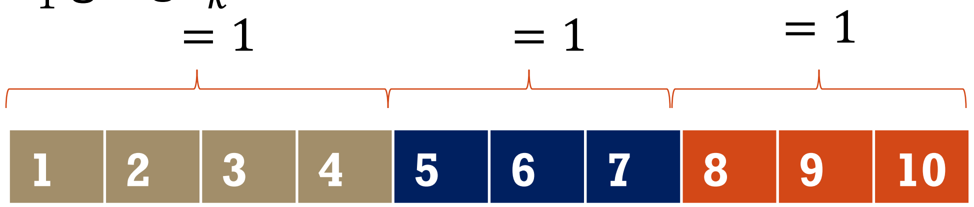
PROBLEM STATEMENT

- Given $v_1, \dots, v_n \in \mathbb{R}^d$, approximate

$$\mathrm{OPT}(v_1, \dots, v_n; \mathcal{B}) = \max\left\{\det\left(\sum_{i \in S} v_i v_i^T\right)^{\frac{1}{d}} : S \in \mathcal{B}\right\}$$

- $\mathcal{B}$ = bases of a matroid of rank $k \geq d$ over $[n]$

- Examples:
  - Uniform matroid: $\mathcal{B} = \{S \subseteq [n] : |S| = k\}$
  - Partition matroid: $\mathcal{B} = \{S \subseteq [n] : \forall i \; |S \cap P_i| = 1\}$, where $[n] = P_1 \cup \dots \cup P_k$



RESULTS AND TECHNIQUES

$$\mathrm{OPT}(v_1, \ldots, v_n; \mathcal{B}) = \max\left\{\det\left(\sum_{i\in S} v_i v_i^T\right)^{\frac{1}{d}} : S \in \mathcal{B}\right\}$$

	Convex Programming Based Methods	Combinatorial Methods
Cardinality Constraints (Pick k vectors)	Randomized Rounding: e-approximation [Nikolov'15, S' Xie'18] Spectral Sparsification: (1+ϵ)-approximation when $k \geq d/\epsilon^2$ [ALSW'17] Volume Sampling: (1+ϵ)-approximation when $k \geq O\left(\frac{d}{\epsilon}\right)$ [Nikolov, S',Tantipongpipat'18]	Local Search/Greedy: (1+ϵ)-approximation when $k \geq d/\epsilon$ [Madan, S',Tantipongpipat, Xie'19, Lau Zhou'22] (Widely implemented in SAS and other softwares)
General Matroid Constraint (Pick k vectors that form a basis)	Stable Polynomials and Strongly log-concave polynomials: e-approximate estimation and e^d-approximation [Nikolov, S' 16, Anari, Oveis-Gharan'17, Anari, Oveis-Gharan, Vinzant'19] (when k≤d) Sparsity of Convex Programs: d-approximate estimation and d^d-approximation [Madan, Nikolov, S', Tantipongpipat'19] (when k>d)	Matroid Intersection Based Methods: d-approximation algorithm [Brown, Laddha, Pittu, S', Tetali'22, '23]

MAIN RESULT FOR THE TALK

- Given $v_1, \dots, v_n \in \mathbb{R}^d$, approximate

$$\mathrm{OPT}(v_1, \dots, v_n; \mathcal{B}) = \max\left\{\det\left(\sum_{i\in S} v_i v_i^T\right)^{\frac{1}{d}} : S \in \mathcal{B}\right\}$$

- $\mathcal{B} =$ bases of a matroid of rank $k \geq d$ over $[n]$

- Theorem [Brown, Laddha, Pittu, S', Tetali '22,'23] : There exists a polynomial time algorithm for the determinant maximization problem under a matroid constraint that returns a solution S such that

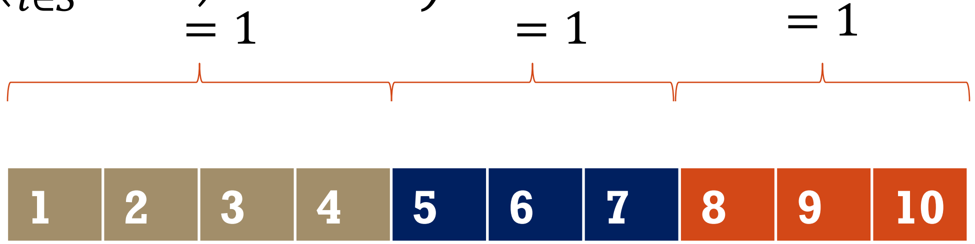
$$\det\left(V_S V_S^T\right) \geq \frac{1}{d^d} \det(V_T V_T^T)$$ where T is the optimal solution.

FEASIBILITY?

- Given $v_1, \dots, v_n \in \mathbb{R}^d$, approximate

$$\mathrm{OPT}(v_1, \dots, v_n; \mathcal{B}) = \max \left\{ \det \left(\sum_{i \in S} v_i v_i^T \right)^{\frac{1}{d}} : S \in \mathcal{B} \right\}$$

- $\mathcal{B}$ = bases of a matroid of rank $k \geq d$ over $[n]$



- Examples:
  - Partition matroid: $\mathcal{B} = \{S \subseteq [n] : \forall i \; |S \cap P_i| = 1\}$, where $[n] = P_1 \cup \dots \cup P_k$.

How do we even know there is a feasible solution with non-zero objective?

For ease of exposition, we assume k=d. Then $\det(V_S V_S^T) = \det(V_S)\det(V_S^T) = \det(V_S)^2 = vol(S)^2$.

MATROID INTERSECTION

Lemma: The objective of max determinant problem is non-zero iff there exists a feasible set S of vectors such that they form a basis R^d.

Matroid: $M = (V, I)$ is a matroid for some $I \subseteq 2^V$ if we have the following two axioms.

$A \in I$ and $B \subseteq A$ implies $B \in I$.

$A, B \in I$ s.t. $|A| > |B|$ then there exists $e \in A \setminus B$ such that $B \cup \{e\} \in I$.

Basis: size of the maximal set in I.

- Consider two matroid over the set of vectors.
  - Partition constraints $M_1 = (V, I_1)$: Pick at most 1 vector from each P_i, i.e. $A \in I_1$ if $|A \cap P_i| \leq 1$ for each i.
  - Linear Independence $M_2 = (V, I_2)$: : Independent set form a linearly independent set of vectors, i.e. $A \in I_2$ if vectors in A are linearly independent.

Lemma: The objective of max-determinant is non-zero iff there is a common basis of two matroid.

Theorem[Edmonds'71]: One can find a common basis of two matroids in polynomial time.

WEIGHTED MATROID INTERSECTION

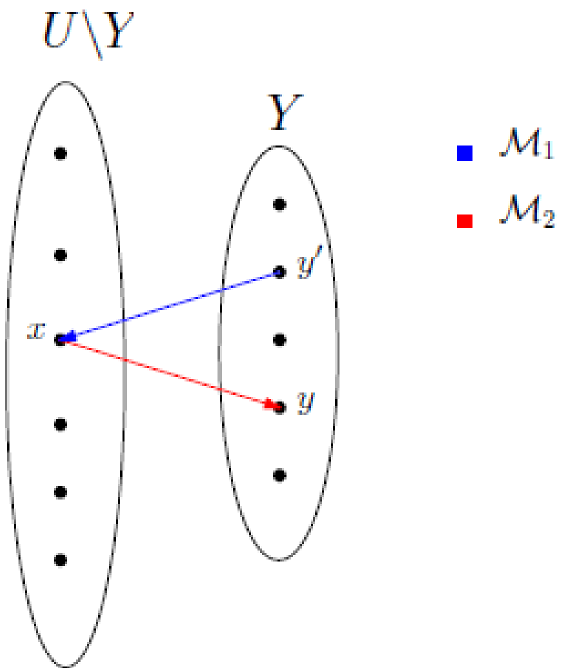
- We are interested in the weighted problem where if we pick a set S, the weight is given by $\det\left(\sum_{i\in S} v_i v_i^T\right)^{\frac{1}{d}}$.
- Classically additive weight functions are studied.
- Given weights w_i for each vector v_i. The weight of a subset S is $w(S) = \sum_{v_i \in S} w_i$.
- Problem: Given two matroids, find a common basis S of maximum weight $w(S)$.

- Theorem[Edmonds'71]: Maximum weight common basis problem is solvable in polynomial time.

The problem and algorithm generalizes maximum weight matching problem in bipartite graphs.

EXCHANGE GRAPH

- Given a feasible common basis S of two matroids, consider the directed bipartite graph D(S) with bipartition $(V \setminus S, S)$.
- Add arc (y, x) where y $\in S$ and $x \notin S$ if $S - y + x \in I_1$.
- Add arc (x, y) where $v \in S$ and $x \notin S$ if $S - y + x \in I_2$.
- Let $\mathrm{l}(y) = w(\mathrm{y})$ for each $y \in \mathrm{S}$ and $\mathrm{l(x)} = -w(x)$ for each $x \notin S$.



- Theorem: S is maximum weight basis iff there is a no negative

length cycle C in S.

- Algorithm for maximum weight common basis:
  - Initialize with any common basis S.
  - Define the exchange graph D(S).
  - If there is no negative length cycle in D(S), declare optimal.
  - Else, find a shortest hop negative length cycle C. Let $S \leftarrow S \Delta C$.

EXCHANGE GRAPH: UPDATED WEIGHTS

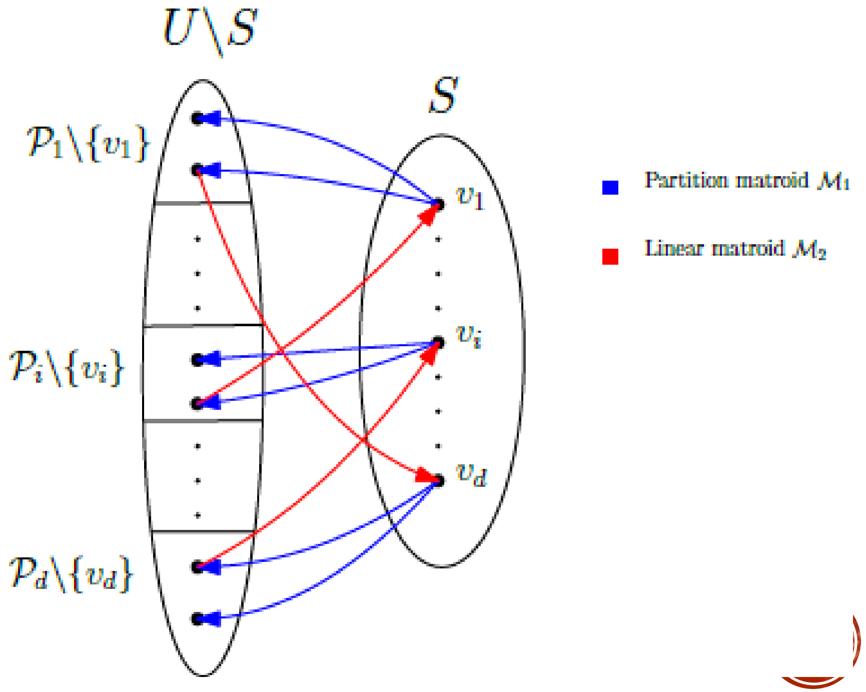
- Challenge: Recall our weight function is not additive.

Updated weights: We place weights w_{uv} on edges of $D(S)$ and not on vertices.

Indeed if $(u, v) \in D(S)$ where $u \notin S$ and $v \in S$ we place a weight of $-\log \frac{vol(S-v+u)}{vol(S)}$, i.e., change in volume when replacing v by u.

The backward edges $(v, u) \in D(S)$ where $u \notin S$ and $v \in S$ get weight 0.

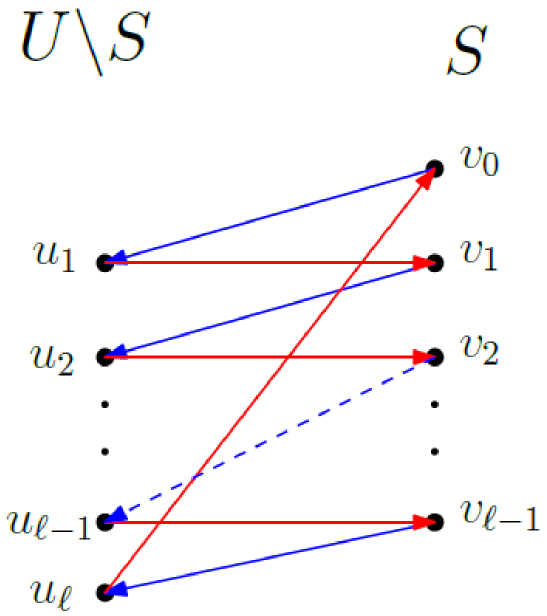
Let S be a current solution. It is a basis, so every u $\notin S$,
u $= \sum_{w \in S} a_{uw}\, w$ for some reals a_{uw}.
Claim: For any u $\notin S$ *and* $v \in S$, *we have* $\frac{vol(S-v+u)}{vol(S)} = |a_{uv}|$.
Proof: $vol_d(S) = vol_{d-1}(S - v) \cdot |v^{\perp}|$ where $v^{\perp}$ is the component of v orthogonal to span $S - v$.
$vol_d(S - v + u) = vol_{d-1}(S - v) \cdot |u^{\perp}|$ where where $u^{\perp}$ is the component of u orthogonal to span $S - v$.
Since u$= a_{uv}\, v + \sum_{\{w \in S - v\}} a_{uw} w$, we have $u^{\perp} = a_{uv} v^{\perp}$



CYCLES AND DETERMINANTS

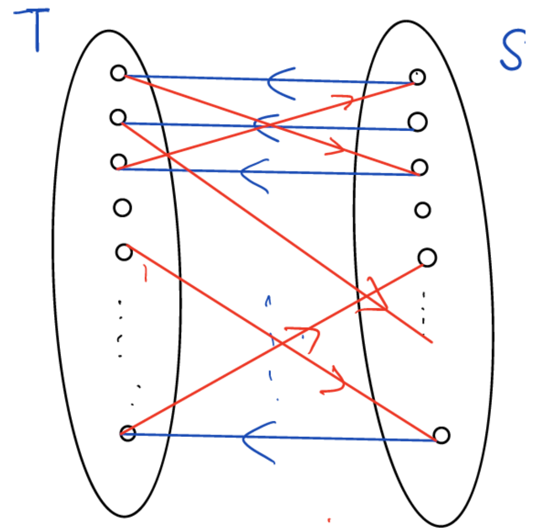
- **Lemma[Determinant to Cycle]:** Let T denote the set of vectors in the optimal solution. If $\frac{vol(T)}{vol(S)} > d^{O(d)}$ then there exists a cycle C in D(S) such that $\sum_{e \in C} w_e \leq -5|C| \log |C|$.

Lemma [Cycle to Determinant]: If C is the shortest hop length cycle with $\sum_{e \in C} w_e \leq -5|C| \log |C|$ then $\frac{\det(S \Delta C)}{\det(S)} \geq 2$. Moreover, $S \Delta C$ is a basis of constraint matroid.



EXISTENCE OF CYCLES

- Lemma[Determinant to Cycle]: Let T denote the set of vectors in the optimal solution. If $\frac{vol(T)}{vol(S)} > d^{O(d)}$ then there exists a cycle C in D(S) such that $\sum_{e\in C} w_e \leq -5|C|\log|C|$.
- Proof: Order vectors in T and S appropriately.



Write each vector in T as a linear combination of vectors in S.

$V_T = V_S A$ where A is a $d \times d$ matrix.

The entries of A are also the weight on the edges.

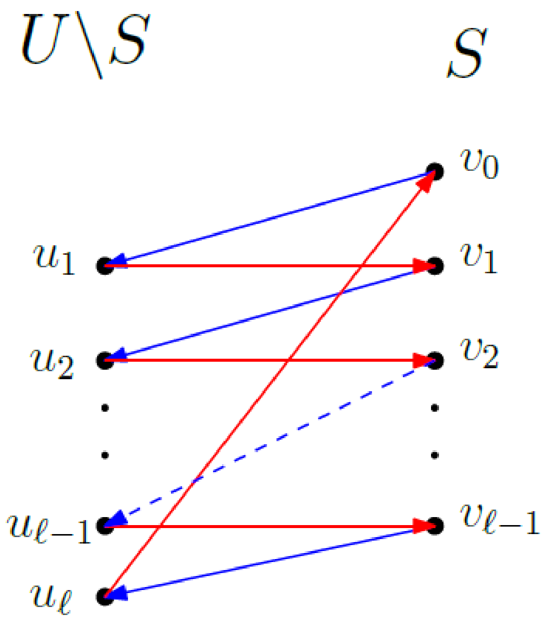
Now, $\frac{vol(T)}{vol(S)} = \frac{\det(V_T)}{\det(V_S)} = \det(A)$. Thus $\det(A) > d^{O(d)}$.

Thus there exists a permutation $\pi: T \to S$ such that

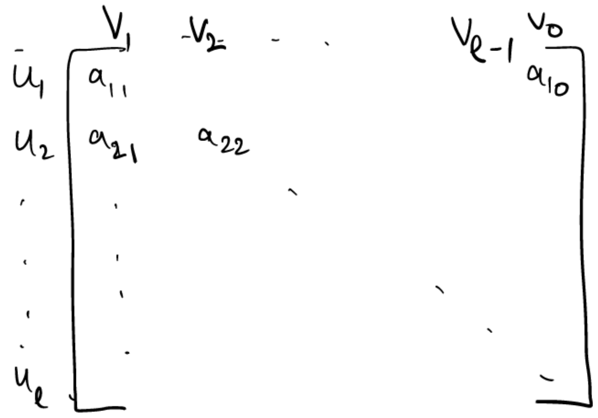
$\prod_{u\in T} a_{u\pi(u)} \geq \frac{d^{O(d)}}{d!} = d^{O(d)}$.

But this permutation corresponds to a collection of cycles and therefore one of the cycles must have really negative length.

VOLUME CHANGE IN CYCLE EXCHANGE



Lemma [Cycle to Determinant]: If C is the shortest hop length cycle with $\sum_{e \in C} w_e \leq -5|C| \log |C|$ then $\frac{\det(S\Delta C)}{\det(S)} \geq 2$. Moreover, $S\Delta C$ is a basis of constraint matroid.

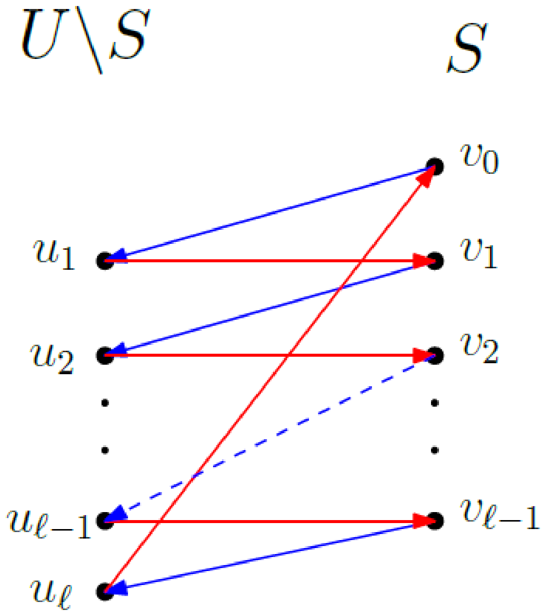


- **Claim**: Let C be a cycle in D(S). Let $S' = S\Delta C$. Then $Vol(S') = Vol(S) \cdot \det(B)$ where B is $|C| \times |C|$ matrix indexed by $C \setminus S$ and $S \setminus C$ and entries a_{uv}.

- Observe that weights on the edges of the cycle

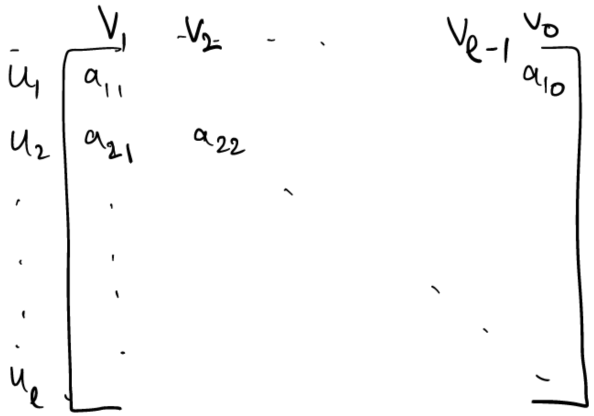
  are exactly the diagonal entries of this cycle.

VOLUME CHANGE IN CYCLE EXCHANGE



Lemma [Cycle to Determinant]: If C is the shortest hop length cycle with $\sum_{e \in C} w_e \leq -5|C| \log |C|$ then $\frac{\det(S \Delta C)}{\det(S)} \geq 2$. Moreover, $S \Delta C$ is a basis of constraint matroid.

- Claim: If C is the shortest length cycle with $\sum_{e \in C} w_e \leq -5|C| \log |C|$ then det(B)> 2.



Why? Use the fact that this is the shortest cycle. This allows to bound non-diagonal entries of the matrix B. A technical argument then shows that the determinant is close to the product of diagonal entries.

FINAL GUARANTEE

- Theorem [Brown, Laddha, Pittu, S', Tetali '22, 23] : There exists a polynomial time algorithm for the determinant maximization problem under a matroid constraint that returns a solution S such that

$$\det\left(V_S V_S^T\right) \geq \frac{1}{d^{O(d)}} \det(V_T V_T^T)$$ where T is the optimal solution.

CONCLUSION AND OPEN QUESTIONS

- Introduced the determinant maximization problem.
- Both combinatorial and convex programming methods are applicable.
- Combinatorial methods are used in practice and have provable guarantees.
- Better guarantees?
- Here we focused on determinant objective. Other spectral objectives?