2022. 12. 21 ISAAC'22 @ Seoul, Korea

On the Parameterized Intractability of Determinant Maximization

Naoto Ohsaka

CyberAgent AI Lab

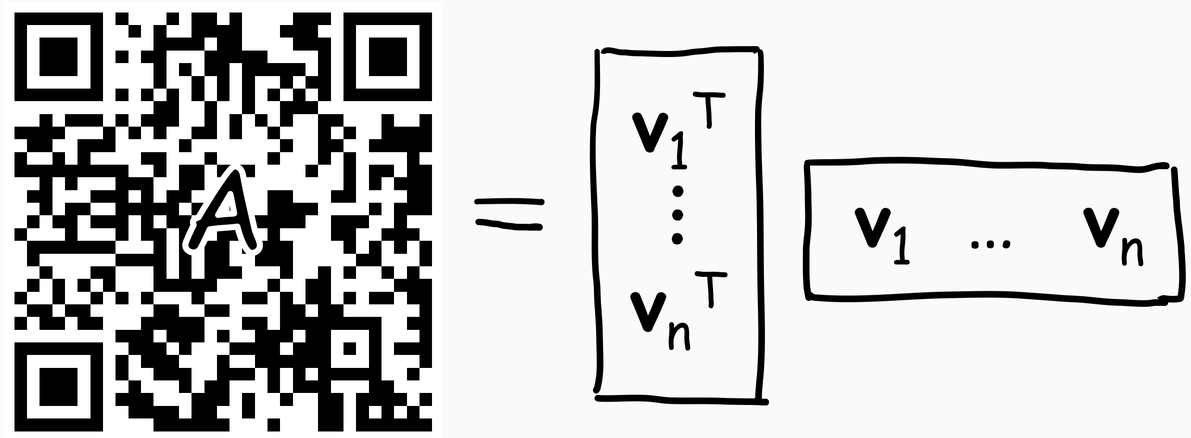
Slides available https://todo314.github.io/ →

What is Determinant Maximization?

- **Input:** $n \times n$ positive semi-definite $\mathbf{A}$ in $\mathbb{Q}^{n \times n}$ & $k \in [n]$
- **Output:** $S \in \binom{[n]}{k}$
- **Goal:** maximize principal minor $\det(\mathbf{A}_S)$

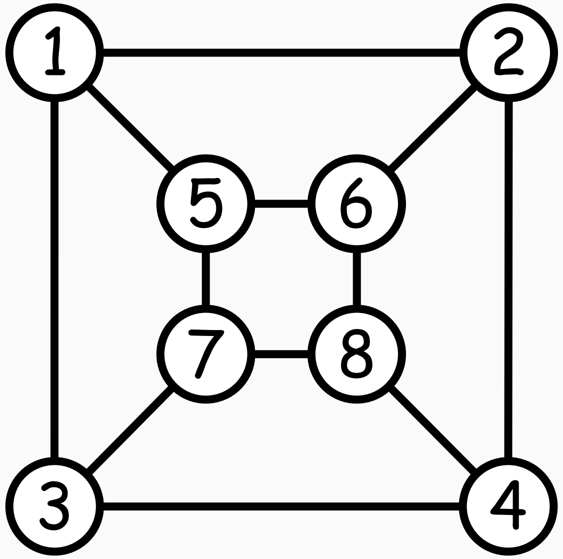
$\mathbf{A}$ is typically given as Gram matrix for n vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in $\mathbb{Q}^d$

$$\mathbf{A} \stackrel{\text{def}}{=} [\mathbf{v}_1, \ldots, \mathbf{v}_n]^\top [\mathbf{v}_1, \ldots, \mathbf{v}_n], \text{ or } A_{i,j} \stackrel{\text{def}}{=} \langle \mathbf{v}_i, \mathbf{v}_j \rangle$$



Example 1: Independent set

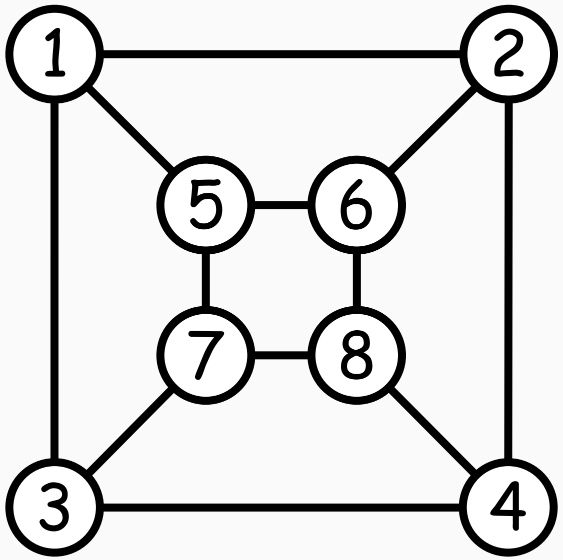
- $Q_3 = (V = [8], E)$: Hypercube graph
- $\mathbf{v}_i \in \{0,1\}^E$: $v_i(e) \stackrel{\text{def}}{=} [\![\text{ i is incident to e }]\!]$



	1	2	3	4	5	6	7	8
1								
2								
3								
4								
5								
6								
7								
8								

Example 1: Independent set

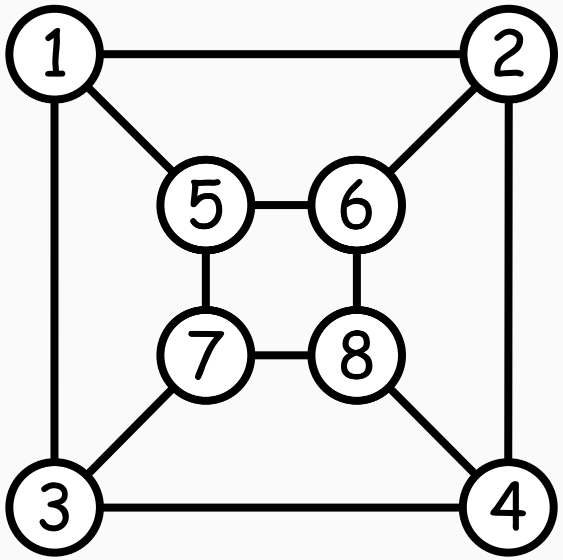
- $Q_3 = (V = [8], E)$: Hypercube graph
- $\mathbf{v}_i \in \{0,1\}^E$: $v_i(e) \stackrel{\text{def}}{=} [\![\ i \text{ is incident to } e\ ]\!]$



	1	2	3	4	5	6	7	8
1	3							
2		3						
3			3					
4				3				
5					3			
6						3		
7							3	
8								3

Example 1: Independent set

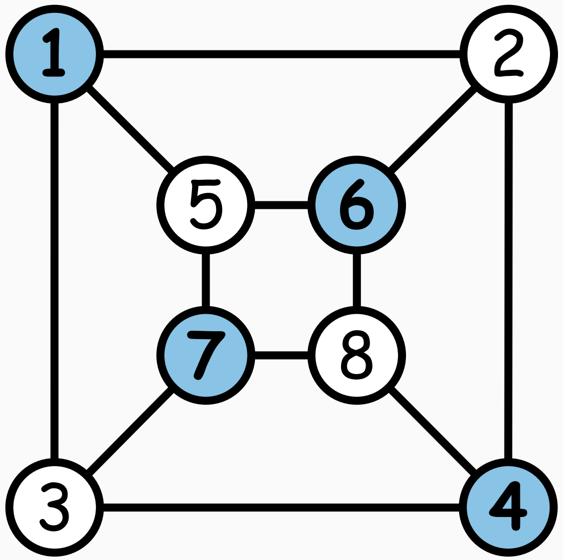
- Q_3 = (V = [8], E): Hypercube graph
- $\mathbf{v}_i \in \{0,1\}^E$: $v_i(e) \stackrel{\text{def}}{=} [\![$ i is incident to e $]\!]$



	1	2	3	4	5	6	7	8
1	3	1	1	0	1	0	0	0
2	1	3	0	1	0	1	0	0
3	1	0	3	1	0	0	1	0
4	0	1	1	3	0	0	0	1
5	1	0	0	0	3	1	1	0
6	0	1	0	0	1	3	0	1
7	0	0	1	0	1	0	3	1
8	0	0	0	1	0	1	1	3

Example 1: Independent set

- $Q_3 = (V = [8], E)$: Hypercube graph
- $\mathbf{v}_i \in \{0,1\}^E$: $v_i(e) \stackrel{\text{def}}{=} [\![\text{ i is incident to e }]\!]$



	1	2	3	**4**	5	**6**	**7**	8
1	**3**	1	1	**0**	1	**0**	**0**	0
2	1	3	0	1	0	1	0	0
3	1	0	3	1	0	0	1	0
4	**0**	1	1	**3**	0	**0**	**0**	1
5	1	0	0	0	3	1	1	0
6	**0**	1	0	**0**	1	**3**	**0**	1
7	**0**	0	1	**0**	1	**0**	**3**	1
8	0	0	0	1	0	1	1	3

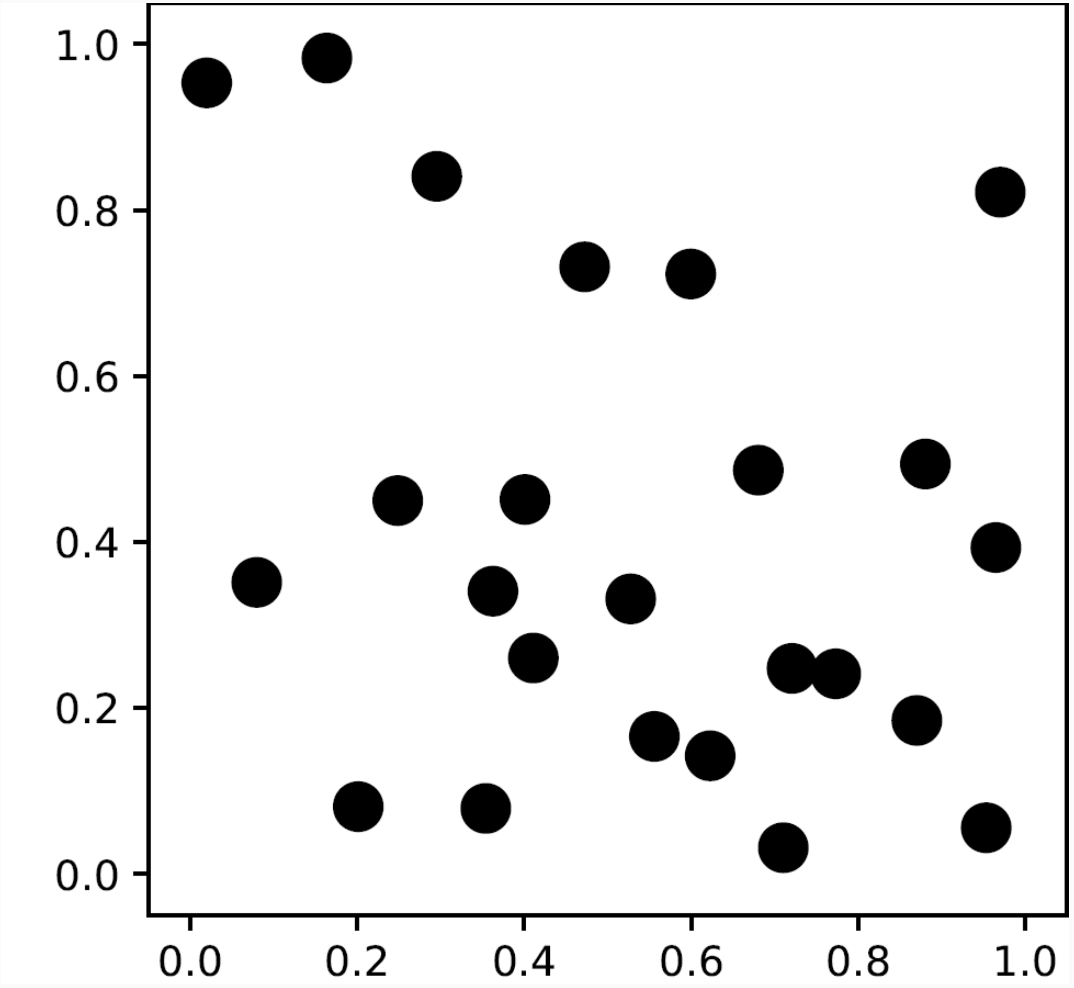
😋 $\det(\boldsymbol{A}_S) = 3^{|S|}$ ➜ S is independent!

e.g., $S = \{1,4,6,7\}$

Example 2: Selecting dispersed points

- $\mathbf{p}_1, \ldots, \mathbf{p}_n$: (random) points on $\mathbb{R}^2$
- Let $A_{i,j} \stackrel{\text{def}}{=} \exp(|\mathbf{p}_i - \mathbf{p}_j|^2)$
  - Known as Gaussian/RBF kernel
  - $\boldsymbol{A}$ is positive semi-definite

Q. What happens if $\det(\boldsymbol{A}_S)$ is max?



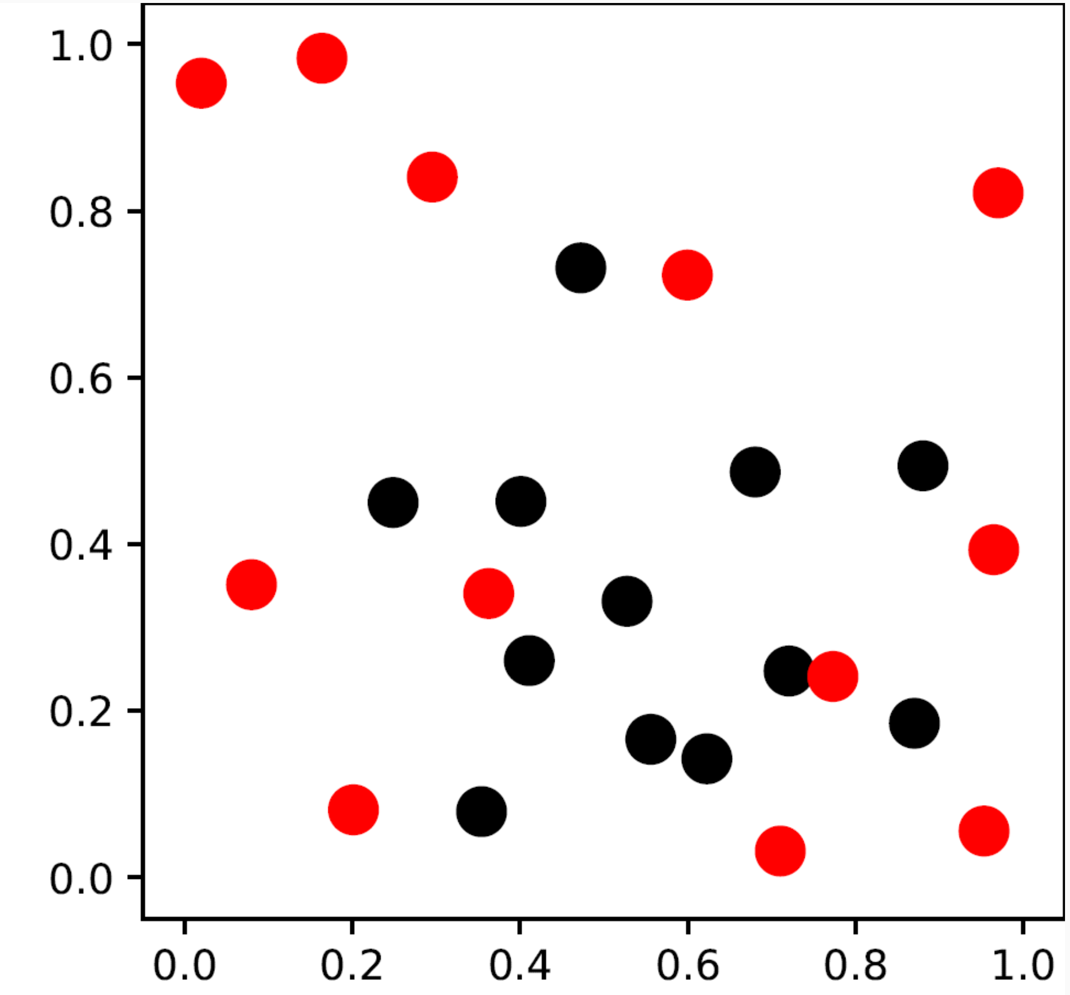
Example of n=24 & k=12

Example 2: Selecting dispersed points

- $\mathbf{p}_1, \ldots, \mathbf{p}_n$: (random) points on $\mathbb{R}^2$
- Let $A_{i,j} \stackrel{\text{def}}{=} \exp(|\mathbf{p}_i - \mathbf{p}_j|^2)$
  - Known as Gaussian/RBF kernel
  - $\boldsymbol{A}$ is positive semi-definite

Q. What happens if $\det(\boldsymbol{A}_S)$ is max?

A. Select "dispersed" points



Example of n=24 & k=12

Why study DETERMINANT MAXIMIZATION?

Various interpretations and applications

- **Parallelepiped volume**
- Diversity promotion in Machine Learning ... many applications!
  [Kulesza-Taskar. *Found. Trends Mach. Learn.*'12]
- Simplex volume [Nikolov. *STOC*'15]
- Maximum-entropy sampling
  [Ko-Lee-Queyranne. *Oper. Res.*'95]

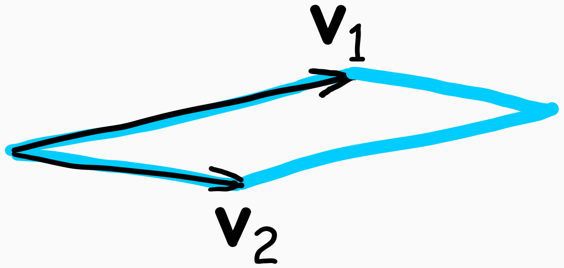
One interpretation: Parallelepiped volume

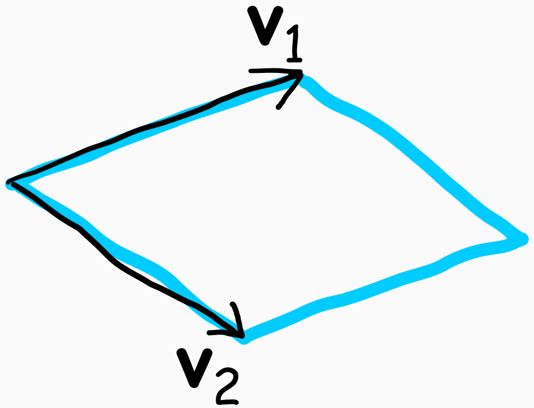
Gram matrix $\boldsymbol{A} \stackrel{\text{def}}{=} [\mathbf{v}_1, \ldots, \mathbf{v}_n]^\top [\mathbf{v}_1, \ldots, \mathbf{v}_n]$

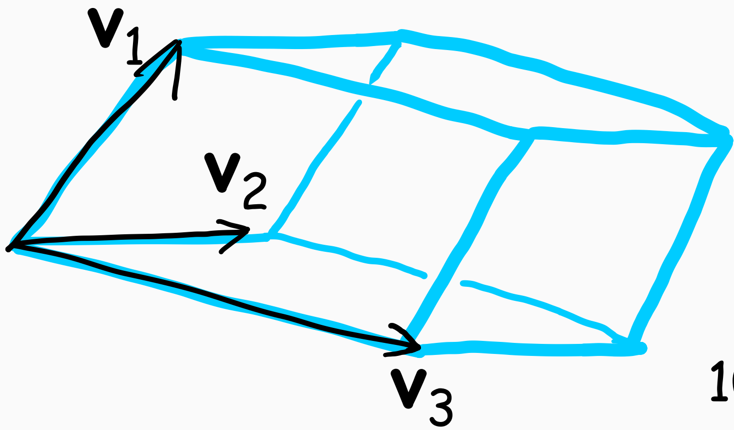
$$\boldsymbol{A} = \begin{bmatrix} \mathbf{v}_1^\top \\ \vdots \\ \mathbf{v}_n^\top \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}$$

$$\det(\boldsymbol{A}_S) = \mathrm{vol}^2(\{\mathbf{v}_i : i \in S\})$$

Determinant Maximization = Volume Maximization







Known results in **polynomial-time** regime

- 😖 NP-hard [Ko-Lee-Queyranne. *Oper. Res.*'95]
- 😋 Greedy is $k!$-approx. [Çivril & Magdon-Ismail. *Theor. Comput. Sci.*'09]
- 😖 NP-hard to $2^{O(k)}$-approx. [Koutis. *Inf. Process. Lett.*'06] [Çivril & Magdon-Ismail. *Algorithmica*'13] [Di Summa-Eisenbrand-Faenza-Moldenhauer. *SODA*'14]

  ⇅ nearly tight
- 😋 Can find e^k-approx. [Nikolov. *STOC*'15]

$k=|S|$ is the output size

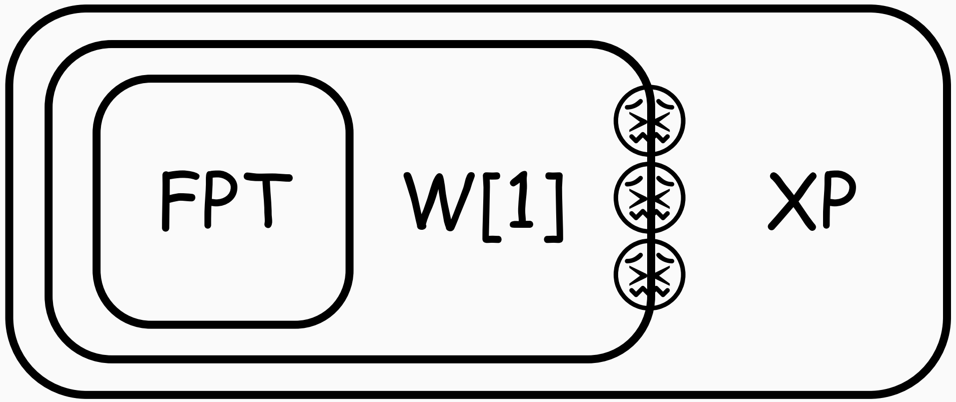
Known results in **parameterized** regime

Measure complexity w.r.t. input size n & parameter k

- Fixed-parameter tractable (FPT): Solvable in $f(k)n^{O(1)}$ time

- $n^{O(k)}$-time brute-force alg. ➜ said to be XP w.r.t. k (very natural param.)
- 😫 But W[1]-hard w.r.t k [Ko-Lee-Queyranne. *Oper. Res.* '95] [Koutis. *Inf. Process. Lett.* '06]

➜ No FPT alg. unless Exponential Time Hypothesis is false (unlikely!)



Q. How can we make DETERMINANT MAXIMIZATION tractable?

Three possible scenarios (we expect)

1. Structural restriction

- (Underlying graph of) **A** is very sparse
- e.g., PERMANENT is #P-hard in general, but FPT w.r.t. treewidth
  [Courcelle-Makowsky-Rotics. *Discrete Appl. Math.*'01] [Cifuentes-Parrilo. *Linear Algebra Appl.*'16]

2. Strong parameter

- rank(**A**) ≥ output size k (always!)
- Room for consideration of f(rank)$n^{O(1)}$-time FPT alg.

3. FPT approximation [Feldmann-Karthik-Lee-Manurangsi. *Algorithms*'20]

- Some W[1]-hard problems are approximable in FPT time
- e.g., PARTIAL VERTEX COVER & MINIMUM k-MEDIAN
  [Har-Peled & Soham Mazumdar. *STOC*'04]

Three possible scenarios (we expect)

1. Structural restriction
- (Underlying graph of) A is very sparse
- e.g., PERMANENT is #P-hard in general, but FPT … t. treewidth
  [Courcelle-Makowsky-… Discrete Appl… …ilo. Lin… …gebra Appl.'16]

😱 All hopes are dashed! 😱

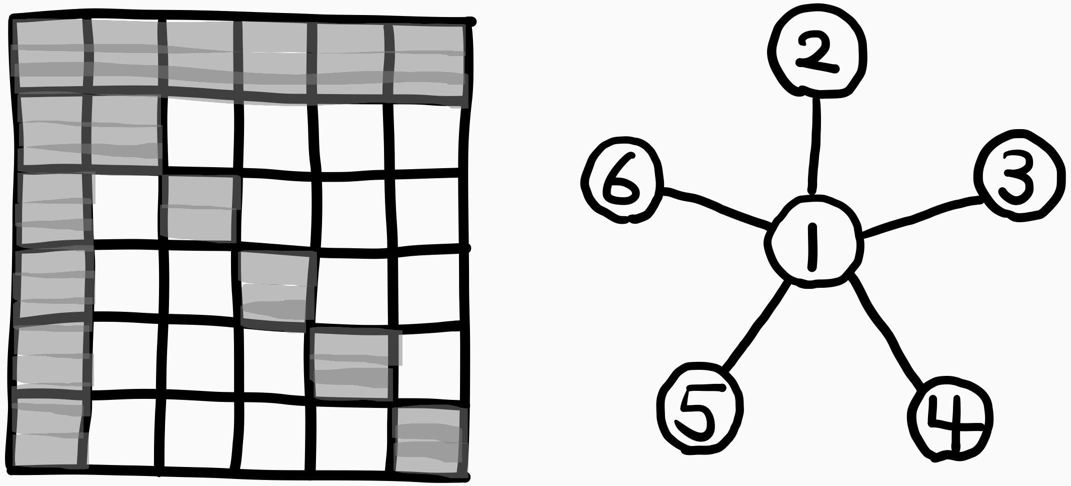
2. Strong …
- rank(A) …
- Room for consi… …of f(rank)n… time FPT alg.

3. FPT approximation [Feldmann-Karthik-Lee-Manurangsi. *Algorithms'20*]
- Some W[1]-hard problems are approximable in FPT time
- e.g., PARTIAL VERTEX COVER & MINIMUM k-MEDIAN
  [Har-Peled & Soham Mazumdar. *STOC'04*]

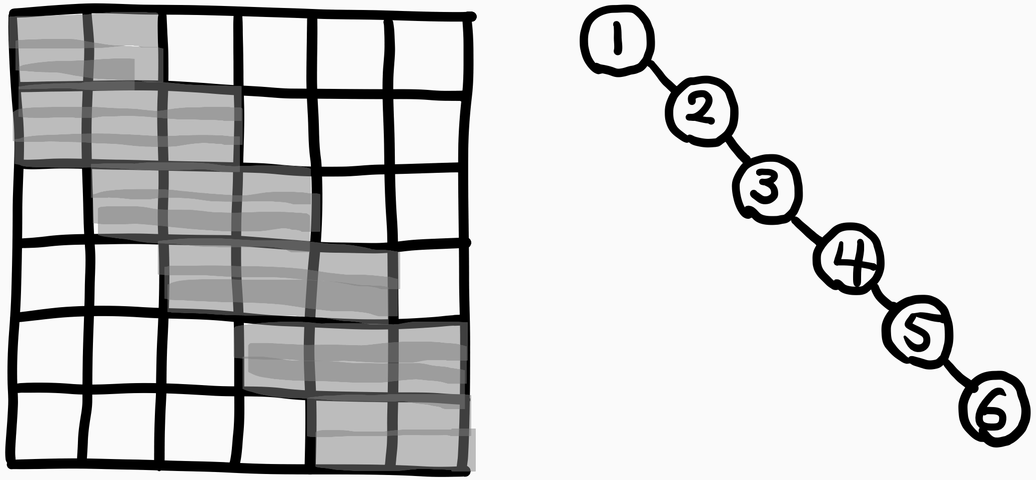
Our first result:

Hardness on arrowhead matrices ⇱ ⇱ ⇱

Arrowhead = Star graph



Tridiagonal = Path graph



W[1]-hard & NP-hard

Treewidth & pathwidth = 1

vertex cover number = 1

Polytime solvable

[Al-Thani & Lee. *LAGOS'21*]

Structural sparsity is NOT very helpful

Our second & third results

😖 W[1]-hard when parameterized by rank of A

→ W[1]-hard w.r.t. output size k even if rank only depends on k

😖 W[1]-hard to $2^{O(\sqrt{k})}$-approx. w.r.t. k under Parameterized Inapproximability Hypothesis

[Lokshtanov-Ramanujan-Saurab-Zehavi. *SODA'20*]

Binary Constraint Satisfaction Problem is W[1]-hard to approx. w.r.t. # variables

Proof overview

(1) Proof overview on arrowhead matrices

(Thm) Determinant Maximization on **arrowhead matrices** is W[1]-hard

- k-Sum: Parameterized version of Subset Sum [Abboud-Lewi-Williams. *ESA'14*]

⬇

⚠ Sophisticated construction of arrowhead matrix

⬇

- Determinant Maximization on arrowhead matrices

(1) Proof overview on W[1]-hardness on arrowhead matrices

k-Sum [Abboud-Lewi-Williams. *ESA'14*] & reduction strategy

- **Input:** n integers $x_1, \ldots, x_n$, $t \in [0, n^{2k}]$, $k \in [n]$
- **Find:** $S \in \binom{[n]}{k}$ s.t. $\sum_{i \in S} x_i = t$

- W[1]-complete w.r.t. k [Downey-Fellows. *Theor. Comput. Sci.'95*] [Abboud-Lewi-Williams. *ESA'14*]

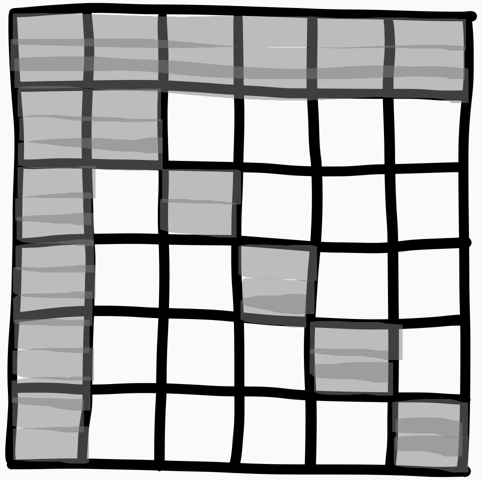
Construct n+1 vectors $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_n$ s.t.

- Gram matrix in $\mathbb{R}^{[0..n]\times[0..n]}$ is arrowhead
- $\det(\mathbf{A}_S)$ s.t. $S \in \binom{[n]}{k+1}$ is maximum when $\sum_{i \in S-\{0\}} x_i = t$ (if exists)

  i.e., $\mathbf{v}_i$ corresponds to x_i

(1) Proof overview on W[1]-hardness on arrowhead matrices

Key finding on arrowhead matrices



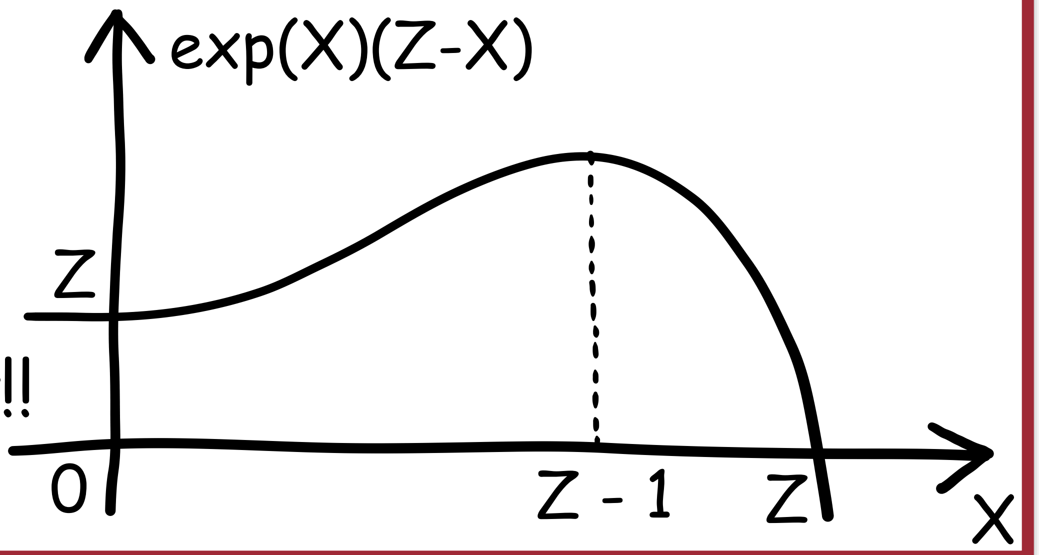
- If $\boldsymbol{A}$ in $\mathbb{R}^{[0..n]\times[0..n]}$ is arrowhead and $0 \in S$:

$$\det(\boldsymbol{A}_S) = \prod_{i \in S-\{0\}} A_{i,i} \cdot \left(A_{0,0} - \sum_{i \in S-\{0\}} \frac{A_{0,i} \cdot A_{0,i}}{A_{i,i}} \right)$$

(Lem) Carefully choose $\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_n \in \mathbb{R}_+^{2n}$ s.t. for $0 \in S \in \binom{[n]}{k}$

$$\det(\boldsymbol{A}_S) \propto \exp\left(\sum_{i \in S-\{0\}} x_i \right) \cdot \left(Z - \sum_{i \in S-\{0\}} x_i \right)$$

Maximized at $\sum_{i \in S-\{0\}} x_i = Z - 1$ ☜ set t!!



(1) Proof overview on W[1]-hardness on arrowhead matrices

Sketch of construction

	1	…	i	…	n	n+1	…	n+i	…	n+n
$\mathbf{v}_0$	$\gamma\sqrt{x_1}$		$\gamma\sqrt{x_i}$		$\gamma\sqrt{x_n}$					
$\mathbf{v}_1$	$\sqrt{\alpha\, e^{x_1}}$					$\sqrt{\beta\, e^{x_1}}$				
$\mathbf{v}_i$			$\sqrt{\alpha\, e^{x_i}}$					$\sqrt{\beta\, e^{x_i}}$		
$\mathbf{v}_n$					$\sqrt{\alpha\, e^{x_n}}$					$\sqrt{\beta\, e^{x_n}}$

Parameterized by α, β, γ (to be determined appropriately)

Omitted details: We have to…

- efficiently approximate $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_n$ using rationals
- ensure that any optimal solution includes $\mathbf{v}_0$

(2) Proof overview on W[1]-hardness by rank

(Thm) Determinant Maximization is W[1]-hard w.r.t. **rank** of A

- Grid Tiling: W[1]-complete [Marx. *FOCS'07*]

⬇

⚠ Can use only f(k)-dimensional vectors / f(k)-rank matrices

e.g., vectors in $\mathbb{Q}^n$ are not allowed

⬇

- Determinant Maximization parameterized by rank of A

(2) Proof overview on W[1]-hardness by rank

GRID TILING [Marx. FOCS'07]

- **Input:** $\mathcal{S} = (S_{i,j} \subseteq [n]^2 : i,j \in [k])$
- **Find:** Select (x,y) in $S_{i,j}$ for all (i,j) s.t.
  - Vertical neighbors agree in 1st coordinate
  - Horizontal neighbors agree in 2nd coordinate

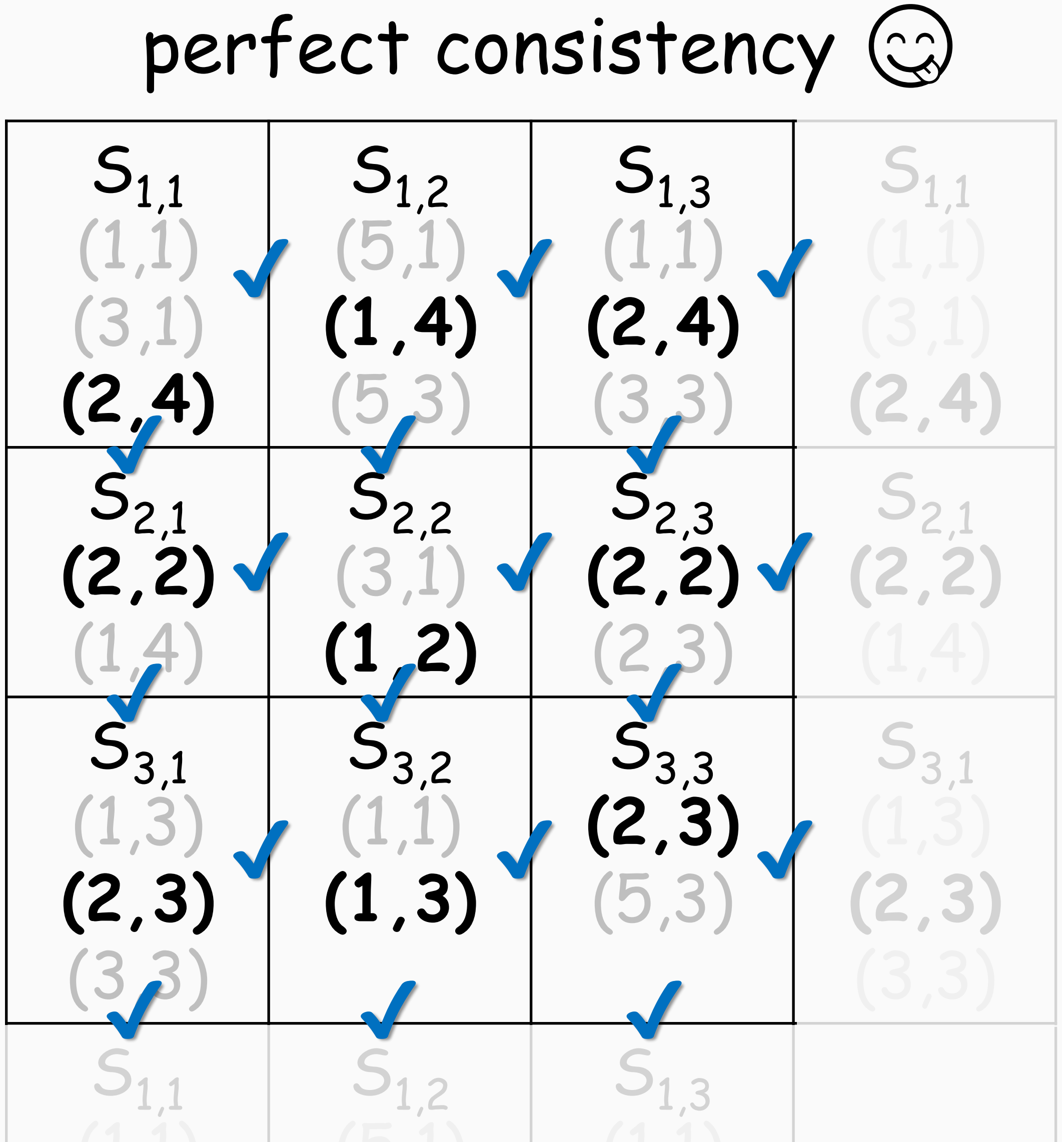
$S_{1,1}$ (1,1) (3,1) (2,4)	$S_{1,2}$ (5,1) (1,4) (5,3)	$S_{1,3}$ (1,1) (2,4) (3,3)
$S_{2,1}$ (2,2) (1,4)	$S_{2,2}$ (3,1) (1,2)	$S_{2,3}$ (2,2) (2,3)
$S_{3,1}$ (1,3) (2,3) (3,3)	$S_{3,2}$ (1,1) (1,3)	$S_{3,3}$ (2,3) (5,3)

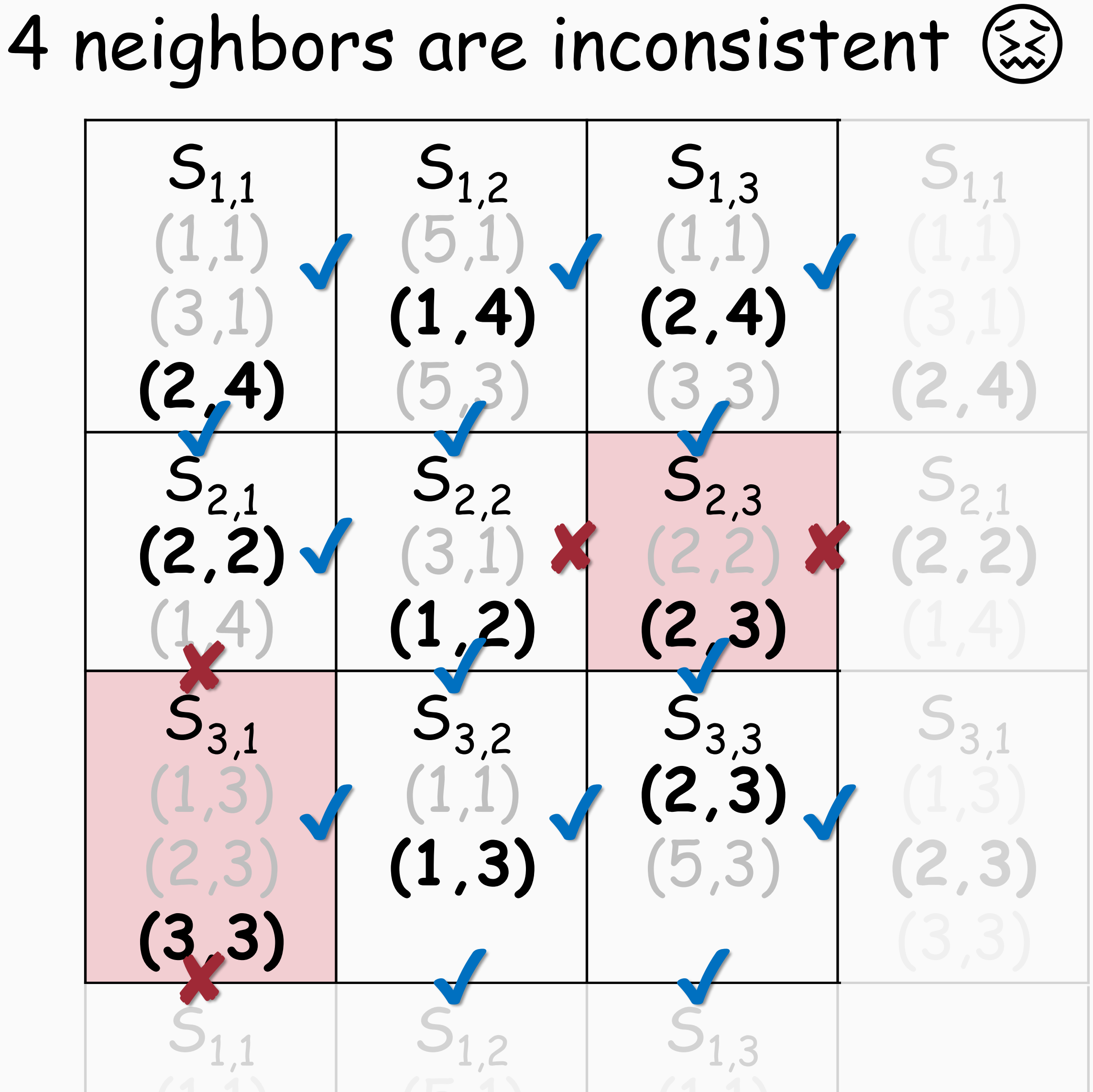
Example of k=3 & n=5
Taken from Fig. 14.2 of
[Cygan-Fomin-Kowalik-Lokshtanov-Marx-Pilipczuk-Pilipczuk-Saurabh.]

- Equality constraints are **SIMPLE**😋
- Cells (i,j) are adjacent to **FOUR** cells😋

(2) Proof overview on W[1]-hardness by rank

Grid Tiling [Marx. FOCS'07]





(2) Proof overview on W[1]-hardness by rank

Reduction from Grid Tiling

- **Input**: $\mathcal{S} = (S_{i,j} \subseteq [n]^2 : i,j \in [k])$
- **Find**: Select (x,y) in $S_{i,j}$ for all (i,j) s.t.
  - Vertical neighbors agree in 1st coordinate
  - Horizontal neighbors agree in 2nd coordinate

f(k)-dim. $\mathbf{v}^{(i,j)}_{x,y}$ for (x,y) in $S_{i,j}$ describing "consistency":

Conditions about "consistency"

(2) Proof overview on W[1]-hardness by rank

Reduction from Grid Tiling

- **Input**: $\mathcal{S} = (S_{i,j} \subseteq [n]^2 : i,j \in [k])$
- **Find**: Select (x,y) in $S_{i,j}$ for all (i,j) s.t.
  - Vertical neighbors agree in 1st coordinate
  - Horizontal neighbors agree in 2nd coordinate

🎯 f(k)-dim. $\mathbf{v}^{(i,j)}_{x,y}$ for (x,y) in $S_{i,j}$ describing "consistency":

- Vertical nbr. $\langle \mathbf{v}^{(i,j)}_{x,y}, \mathbf{v}^{(i+1,j)}_{x',y'} \rangle = 0$ iff $x=x'$
- Horizontal nbr. $\langle \mathbf{v}^{(i,j)}_{x,y}, \mathbf{v}^{(i,j+1)}_{x',y'} \rangle = 0$ iff $y=y'$
- Same cell $\langle \mathbf{v}^{(i,j)}_{x,y}, \mathbf{v}^{(i,j)}_{x',y'} \rangle \neq 0$

Focus in the next slide

😋 Gram matrix $A_{i,j,x,y,i',j',x',y'} \stackrel{\text{def}}{=} \langle \mathbf{v}^{(i,j)}_{x,y}, \mathbf{v}^{(i',j')}_{x',y'} \rangle$ satisfies...

- $\mathcal{S}$ is YES ➜ $\exists k^2 \times k^2$ diagonal submatrix ...select CORRECT $\mathbf{v}^{(i,j)}_{x,y}$ for each $(i,j) \in [k]^2$
- $\mathcal{S}$ is NO ➜ $\forall k^2 \times k^2$ submatrix is NOT diagonal

(2) Proof overview on W[1]-hardness by rank

Represent "consistency" at lower dimensions?

- Want $\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{w}_1, \ldots, \mathbf{w}_n$ in $\mathbb{Q}^{O(1)}$ s.t. $\langle \mathbf{v}_i, \mathbf{w}_j \rangle = 0$ iff i=j

  🙄 How to construct?

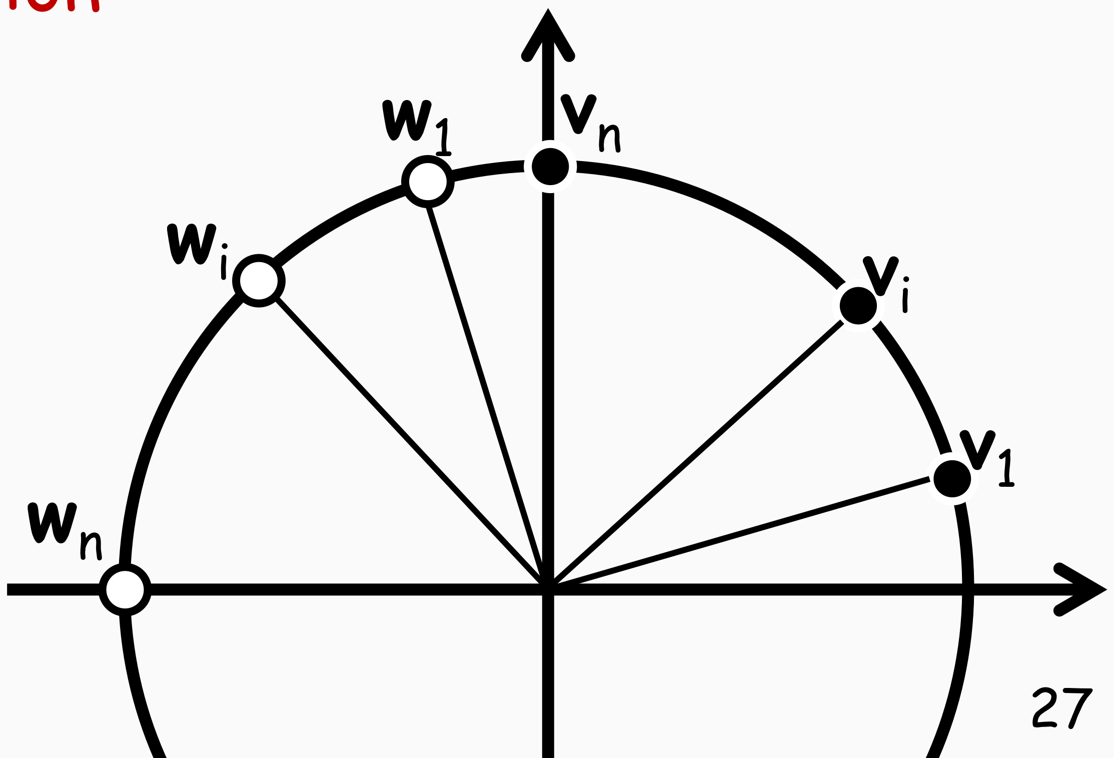
🚫 One-hot vectors require n-dimension

[0,…,0,1,0,…,0]

😋 Use points on the unit circle:

- $\mathbf{v}_i \stackrel{\text{def}}{=} \left(\cos\left(\frac{\pi i}{2n}\right), \sin\left(\frac{\pi i}{2n}\right)\right)$
- $\mathbf{w}_j \stackrel{\text{def}}{=} \left(\sin\left(\frac{\pi j}{2n}\right), -\cos\left(\frac{\pi j}{2n}\right)\right)$

Use Pythagorean triples to get rational vectors



(3) Proof overview on inapproximability

(Thm) Under PIH, $\exists\delta$, Determinant Maximization is W[1]-hard w.r.t. output size k to approx. within $0.999^{\delta\sqrt{k}}$-factor

- Parameterized Inapproximability Hypothesis (PIH)
  [Lokshtanov-Ramanujan-Saurab-Zehavi. *SODA'20*] I don't go into details in this talk

⬇

- Optimization version of Grid Tiling: W[1]-hard to approx. w.r.t. k

⬇ ⚠ Gap-preserving reduction (different from the last one)

- Determinant Maximization parameterized by k

(3) Proof overview on inapproximability

Optimization version of GRID TILING

- **Input:** $\mathcal{S} \stackrel{\text{def}}{=} (S_{i,j} \subseteq [n]^2 : 1 \leq i,j \leq k)$
- **Output:** Select (x,y) in $S_{i,j}$ for all (i,j)
- **Goal:** maximize (# vertical nbr. agreeing in 1st coordinate)
  + (# horizontal nbr. agreeing in 2nd coordinate)

  $\text{opt}(\mathcal{S}) \stackrel{\text{def}}{=}$ max. of ☝

(Lem) Under PIH, ∃δ, it is W[1]-hard to distinguish between

- Completeness: $\text{opt}(\mathcal{S}) = 2k^2$ ...$\mathcal{S}$ is YES
- Soundness: $\text{opt}(\mathcal{S}) \leq 2k^2 - \delta k$ ...$\mathcal{S}$ is much worse than YES

Sketch of reduction from GRID TILING

- Construct $\mathbf{v}^{(i,j)}_{x,y}$ in $\mathbb{Q}^{O(k^2n^2)}$ for each (x,y) of $S_{i,j}$ s.t. $|\mathbf{v}^{(i,j)}_{x,y}|^2 = 4$,

Undesirable cases impose **const.** penalty

Sketch of reduction from GRID TILING

Construct $\mathbf{v}^{(i,j)}_{x,y}$ in $\mathbb{Q}^{O(k^2n^2)}$ for each (x,y) of $S_{i,j}$ s.t. $|\mathbf{v}^{(i,j)}_{x,y}|^2 = 4$,

- Same cell $\langle \mathbf{v}^{(i,j)}_{x,y}, \mathbf{v}^{(i,j)}_{x',y'} \rangle$ is **≥2**
- Vertical nbr. $\langle \mathbf{v}^{(i,j)}_{x,y}, \mathbf{v}^{(i+1,j)}_{x',y'} \rangle$ is $\begin{cases} 0 & \text{if } x=x' \\ 1/2 & \text{otherwise} \end{cases}$
- Horizontal nbr. $\langle \mathbf{v}^{(i,j)}_{x,y}, \mathbf{v}^{(i,j+1)}_{x',y'} \rangle$ is $\begin{cases} 0 & \text{if } y=y' \\ 1/2 & \text{otherwise} \end{cases}$

Undesirable cases impose **const.** penalty

KEY: Gadget of [Çivril & Magdon-Ismail. *Algorithmica'13*]

(Lem) $\det(\mathbf{A}_S)$ exponentially decays in # duplicates & $2k^2$–opt($\mathcal{S}$); so,

- Completeness: $\text{opt}(\mathcal{S}) = 2k^2 \rightarrow \max_{|S|=k\times k} \det(\mathbf{A}_S) = 4^{k\times k}$
- Soundness: $\text{opt}(\mathcal{S}) \leq 2k^2-\delta k \rightarrow \max_{|S|=k\times k} \det(\mathbf{A}_S) \leq 4^{k\times k}\cdot 0.999^{\delta k}$

😋 Some tractable cases (see the paper)

1. Polytime solvable on **tridiagonal** matrices [Al-Thani & Lee. *LAGOS'21*]

- Dynamic programming

2. Orthogonal vectors in $\mathbb{Q}^d$ is FPT w.r.t. d for **nonnegative** vectors

- Reduce to SET PACKING

3. ε-additive approximation (bounded entries) is FPT w.r.t. **rank**

- Use standard rounding technique

Conclusion and future work

- Study parameterized hardness of DETERMINANT MAXIMIZATION

- 1. Boundary between P vs. NP (or FPT vs. W[1])

  Tridiagonal & spider of bounded legs ...Polytime

  [Al-Thani & Lee. *LAGOS'21*]

  Tree of bounded degree ...?

  Arrowhead ...NP-hard & W[1]-hard

- 2. Further strong parameters?

- 3. Strengthening inapprox. factor
  - W[1]-hardness of **$2^{O(k)}$-approx.**?

↖ Thank you! ↖