On the Extension Complexity of the TSP Polytope

Speaker: Rakesh Venkat

Online CS Theory Seminar Series, IIT Hyderabad

May 2020

Attribution

This talk is primarily based on the following paper:

Linear vs. Semidefinite Extended Formulations: Exponential Separation and Strong Lower Bounds

By: Fiorini, Massar, Pokutta, Tiwary, de Wolf

Appeared in STOC, 2012.

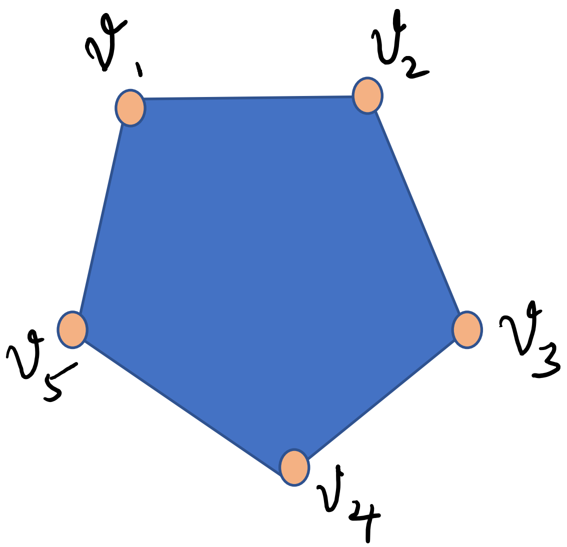
Outline

1. Polytopes, Extension Complexity, Linear Programming
2. Yannakakis' Theorem
3. The Correlation polytope and its extension complexity
4. From Corr(n) to the TSP Polytope
5. Related work and Open problems

Polytopes, Extension Complexity and Combinatorial optimization

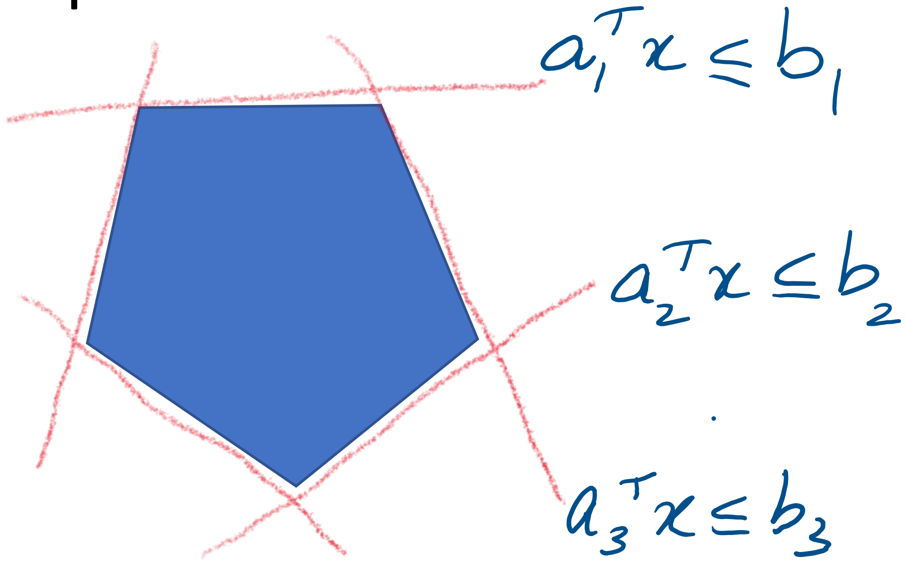
Polytopes: Two views

- Convex Hull of *vertices* in $\mathbb{R}^d$



- $P = \text{conv}(\{v_1, v_2, \ldots, v_N\})$

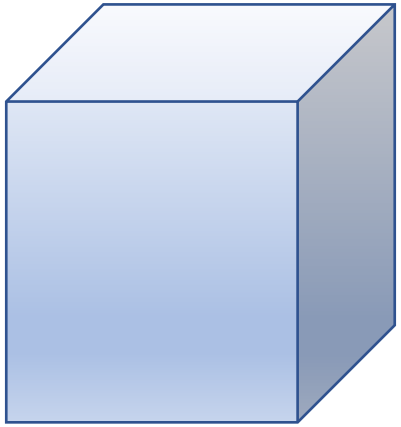
- Intersection of Finite Number of halfspaces in $\mathbb{R}^d$



- $P = \{x | Ax \leq b\}$ where $A \in \mathbb{R}^{m \times d}$, $b \in \mathbb{R}^m$

Vertices and Halfspaces

- Number of vertices may be *exponential* in the number of halfspaces
  - E.g. The hypercube in $\mathbb{R}^d$



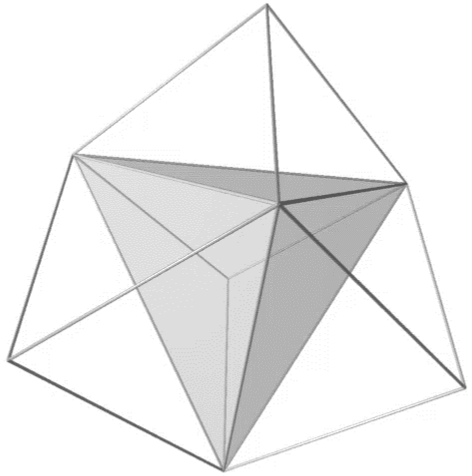
$$\begin{aligned} 0 \le x_1 \le 1 \\ 0 \le x_2 \le 1 \\ \vdots \\ 0 \le x_d \le 1 \end{aligned}$$

$$H_d = \text{conv}(\{0,1\}^d)$$

- Broad question for this talk: Can some polytopes that have an exponential number of vertices be expressed with a polynomial number of inequalities?

A non-trivial example

- The Parity Polytope $PP = \text{conv}(x \in \{0,1\}^d : \sum_i x_i = \text{odd})$



An exponential-sized halfspace description:

$$\sum_{i \in S} x_i - \sum_{i \notin S} x_i \leq |S| - 1 \qquad \forall S \subseteq$$
$$[d]: \text{even}-\text{sized sets}$$
$$0 \leq x_i \leq 1$$

Check: Above is violated for even-weighted x.

- Above is still an exponential-sized description; can we get a polysized description?

Animation from: Anup Rao, UWash. Link

A polynomial-sized description of PP_d

- The Parity Polytope $PP = \text{conv}(x \in \{0,1\}^d : \sum_i x_i = \text{odd})$

$$\sum_{k:odd} \alpha_k = 1$$

$$\sum_{k:odd} z_{ik} = x_i \qquad \forall i \in [d]$$

$$\frac{1}{k} \sum_i z_{ik} = \alpha_k \qquad \forall k \in [d], odd$$

Intuition:

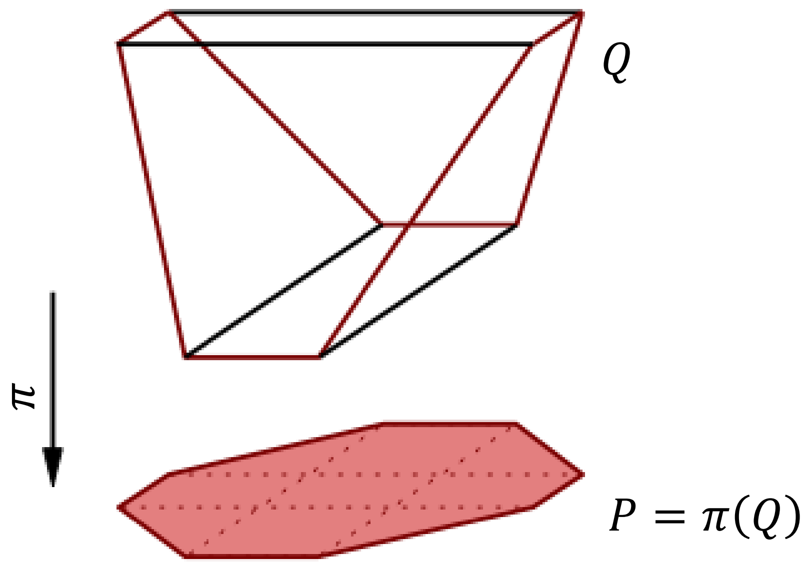
- α_k selects which k we are looking at
- z_{ik} is 1, if x_i is 1 and $\sum_i x_i = k$

Key point: Introduction of auxiliary variables z, α.

- The above polytope lies in higher dimension ($\mathbb{R}^{O(d^2)}$), the *projection* onto the x variables gives us our desired polytope.

Extension Complexity and Projections

We will mainly deal with coordinate projections, but the theory holds for general linear maps.



- The polytope P is a projection of a polytope Q along a subset of the dimensions.
- The description of Q in terms of halfspaces is called an *extended formulation* of the polytope P
- As we have seen, the extended formulation may have fewer *facets/faces.*

Question: Is it always possible to find such compact extended formulations?

Figure courtesy: Fiorini-Rothvoss-Tiwary, Extended Formulations for Polygons, In Discrete and Computational Geometry, 2012.

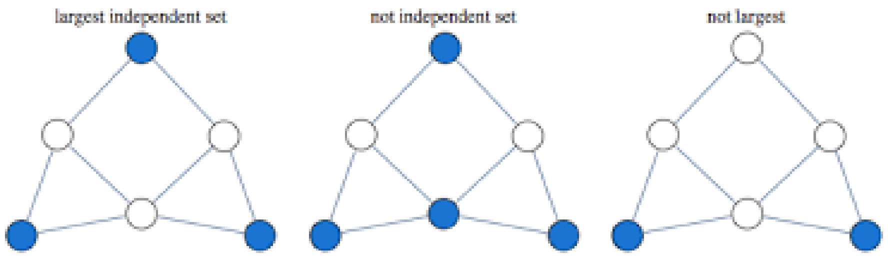
Extension Complexity: Definition and a result

Definition of Extension Complexity: $\mathrm{xc}(P)$ is the *minimum* number of inequalities required to describe P, even when allowed to use auxiliary variables/extended formulations.

- Rothvoss[2011]: There *exist* polytopes in $\mathbb{R}^d$ that require $2^{\Omega(\frac{d}{2})}$ inequalities to describe.
  - Shown using a probabilistic counting argument

- But why should we (as computer scientists) care?

Linear Programming (LP) Relaxations



- Linear optimization can be solved in polytime!

- Express a combinatorial optimization problem as a *linear optimization problem*
  - E.g. Consider the maximum independent-set problem on a graph $G = (V, E)$:
    - $\max \sum_i x_i$

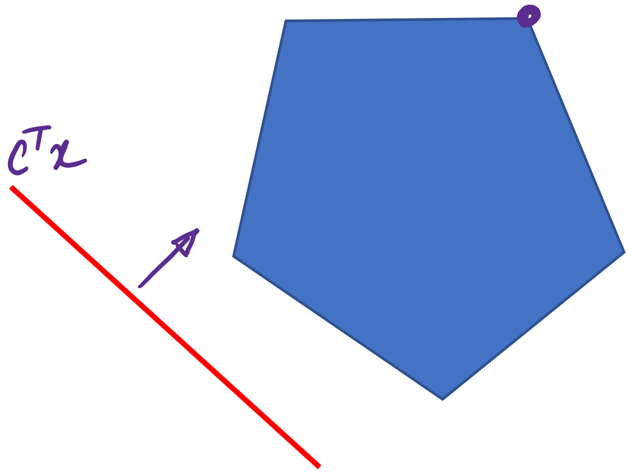
$$\text{s.t.} \quad x_i + x_j \leq 1 \qquad \forall i, j \in E(G)$$
$$0 \leq x_i \leq 1 \qquad \forall i \in V$$

- If the variables were restricted to $x_i \in \{0,1\}$, then clearly, any feasible solution to the above is an independent set.
- The *relaxation* of each x_i to the real interval introduces spurious solutions.
  - E.g. setting all $x_i = \frac{1}{2}$ is a feasible solution! Doesn't mean anything.

- **Question:** When would a relaxation be useful/ideal?

An Ideal Relaxation

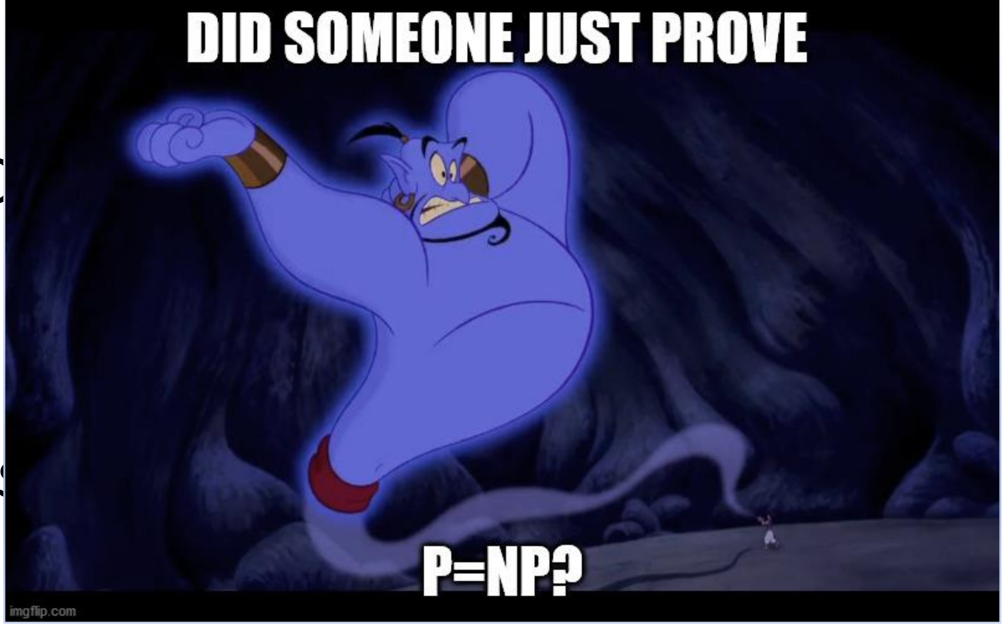
An ideal relaxation for the independent set problem: the feasible polytope is exactly the convex hull of the independent sets in G.



- We may still get convex-combinations of the vertices as fractional solutions, but that is fine
  - Optimal value will be attained at some vertex always.

LPs have found widespread use in the design of algorithms. Can we hope to capture *all* problems using LPs?

NP-Hard problems and Extension Co

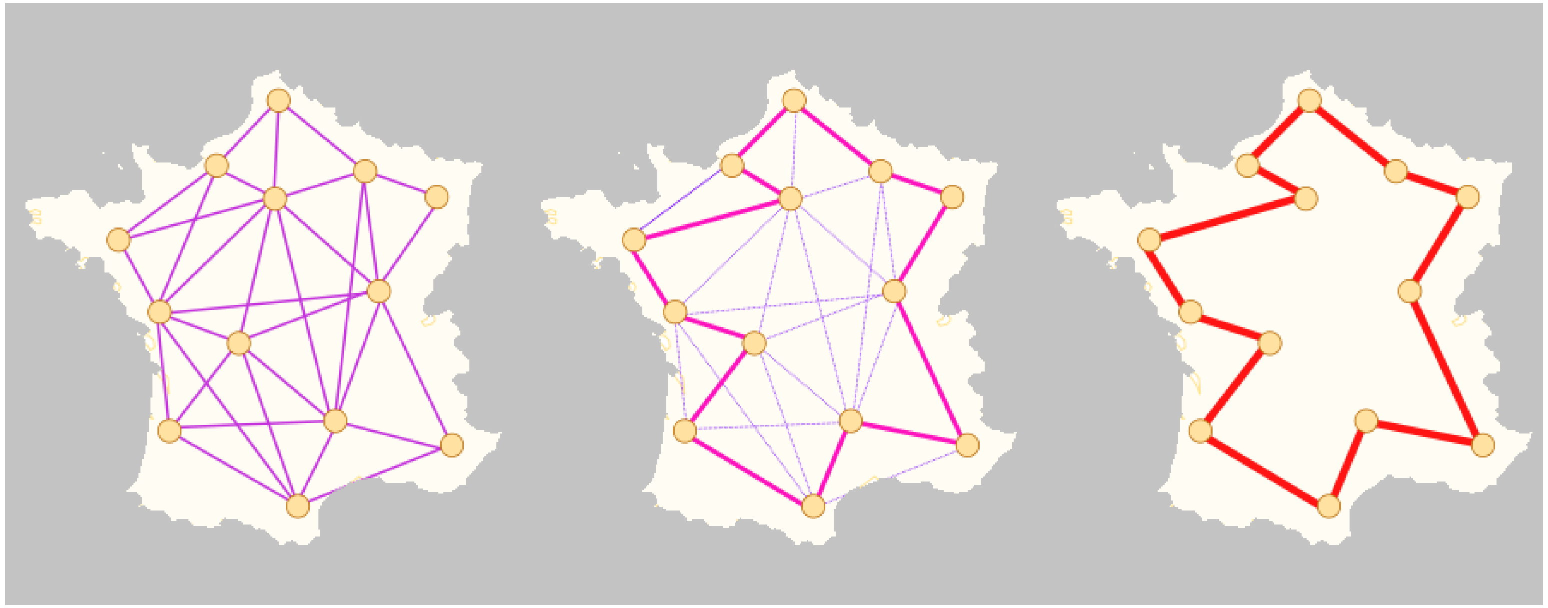


- Objective: Prove P=NP / Conquer the world
- Method: Pick a favourite NP-Hard problem, s
  Problem (TSP).
- Come up with an extremely clever Polynomial-sized LP formulation, using auxiliary variables.
- Show that the vertices of the feasible region are exactly Hamiltonian cycles in the input instance
- Congratulations, you've have just released the NP-genie; infinite riches await!

- ***If the world is sane (P≠NP), the above should not work.***

The Travelling Salesperson Problem (TSP)

- Given n cities and distances w_{ij} between them, find the shortest tour that visits all cities and returns back to the starting city.



- Input can be viewed as a weighted complete graph K_n. Select edges that form a Hamiltonian cycle of minimum weight.

Figure courtesy: Wikipedia, TSP page.

The TSP Polytope, and a relaxation formulation

- Use variables x_{ij} for each distinct pair $i, j \in V$. Thus, $x \in \mathbb{R}^{\binom{n}{2}}$.

$$\min \sum_{ij} w_{ij} x_{ij} \quad s.t. \quad x \text{ is a HamCycle in } K\_n$$

$$\begin{aligned} \min \quad & \sum_{ij} w_{ij} x_{ij} \\ s.t. \quad & x(\delta(S)) \geq 2 \quad \forall S \neq \emptyset, V \\ & x\left(\delta(i)\right) = 2 \quad \forall i \in V \\ & x \geq 0 \end{aligned}$$

- Above, $\delta(S)$ denotes the edges crossing S.
- The above LP has a conjectured gap of 4/3 to the optimum integer solution.

The TSP story

- In 1986, Swart came up with a claimed polynomial-sized EF for the TSP polytope (in a draft titled "P=NP")
- With some effort, researchers found bugs in the LP
- Swart claimed to fix these in a new draft; but more bugs were were again fixed… and this went on.

- Yannakakis came up with an ingenious method that showed that *any* symmetric extended formulation that captures the TSP Polytope is necessarily exponential-sized, disproving Swart's approach.
- However, symmetry can be powerful:
  - [Kaibel-Pashkovich-Theis 2011]: There are explicit polytopes with exponential symmetric extension complexity, but polynomial asymmetric extension complexity.

Settling the TSP story

- 20 years after Yannakakis' result, Fiorini-Massar-Pokutta-Tiwary-deWolf (2012) showed that *any* extended formulation for the TSP polytope has size $2^{\Omega(\sqrt{n})}$, answering the open question.
- [Rothvoss-2013] showed that this can be improved to $2^{\Omega(n)}$.
- Yannakakis' result has lead to a number of other breakthroughs, tying in diverse fields
  - Communication complexity, quantum computation, Fourier analysis, etc.
- Many interesting questions remain open (we will see a few at the end)

Today's Result

Theorem: [Fiorini et al., 2012]

$$\text{xc}(TSP_n) \geq 2^{\Omega(\ n)}$$

- Step 1: Yannakakis' Factorization Theorem
- Step 2: The Correlation Polytope and its Extension Complexity
- Step 3: From the Correlation Polytope to TSP

Why prove lower bounds for NP-hard problems?

- Bounds are independent of complexity-theoretic assumptions (P$\neq$NP).
- Can view LPs as a computational model/proof system. This gives lower bounds in this particular proof system.
- Yields insights into the computational difficulty of the problem at hand
  - Bounds in the LP (or SDP) world have been translated to lower bounds using hardness assumptions [Raghavendra 08].

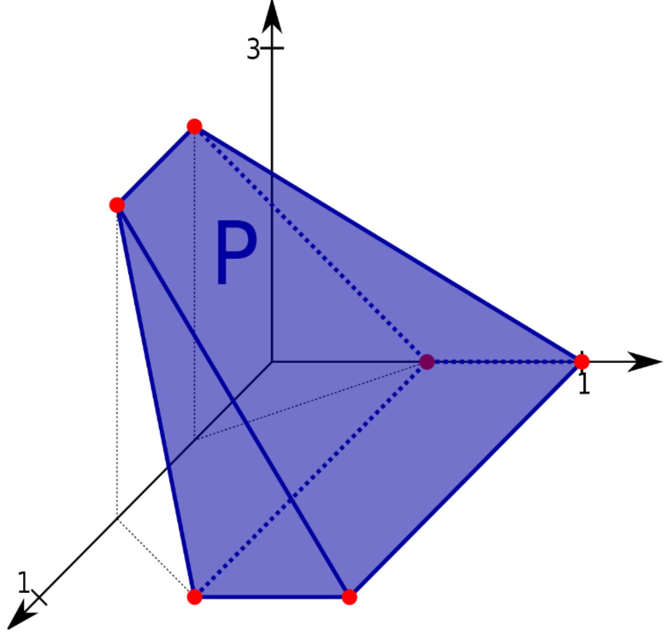
Step 1: Yannakakis' Factorization Theorem

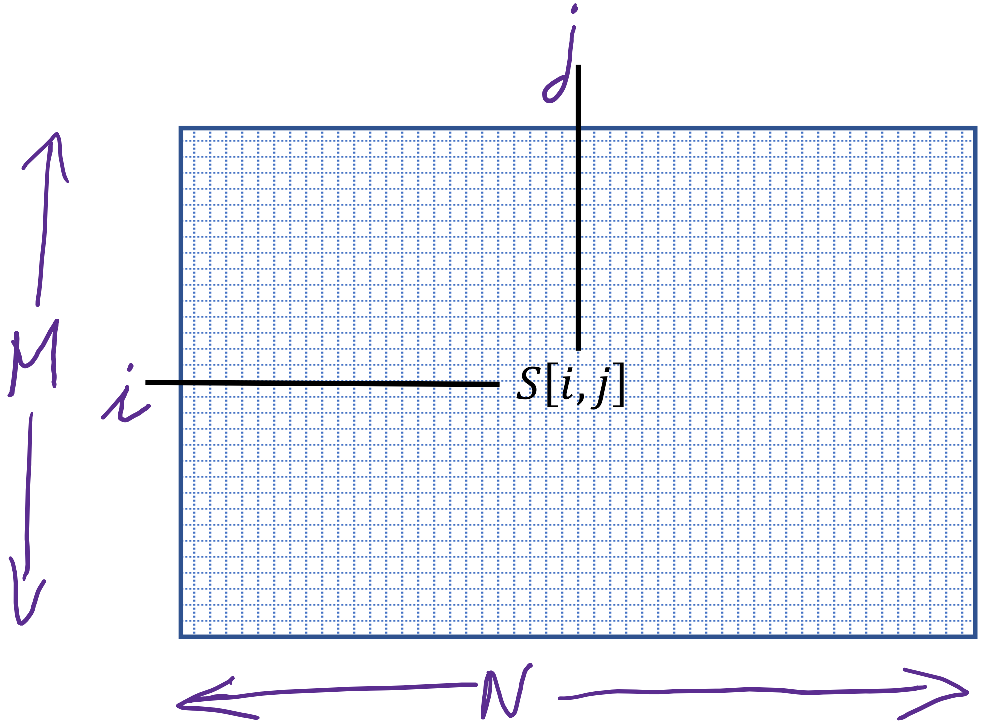
The Slack Matrix of a Polytope

- Consider a polytope given by M inequalities: $Ax \leq b$
- Suppose its vertices are $u_1, \ldots, u_N$
- The Slack matrix S is defined as:

$$S[i,j] = b_i - \langle A^i, u_j \rangle$$

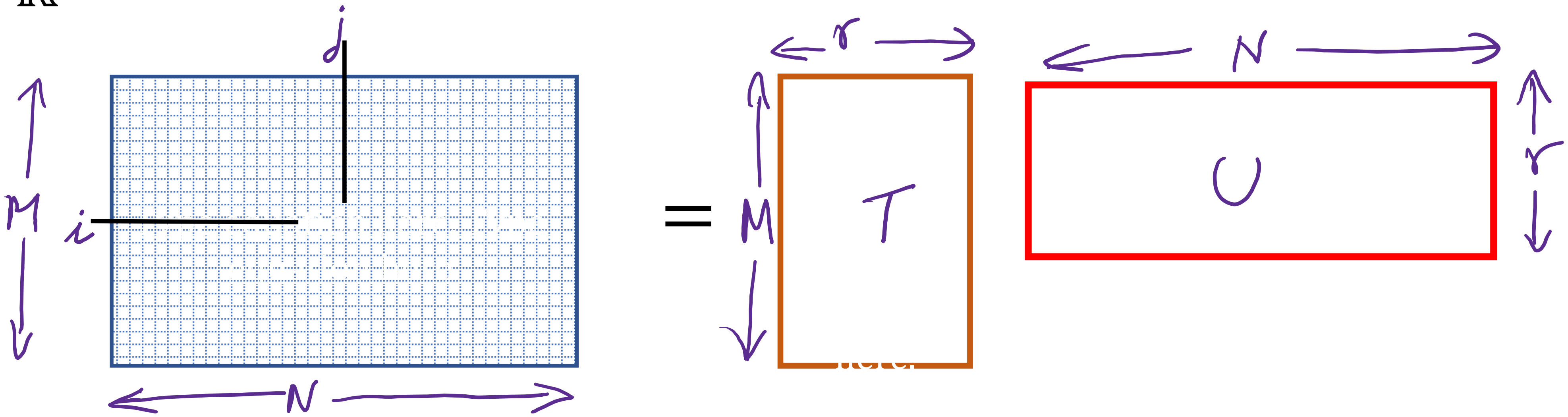
- $S \in \mathbb{R}^{M \times N}$ (A^i is i^{th} row of A)
- Every entry is non-negative
- Slack of the j^{th} vertex on i^{th} constraint





Slack Matrix: Non-Negative Rank

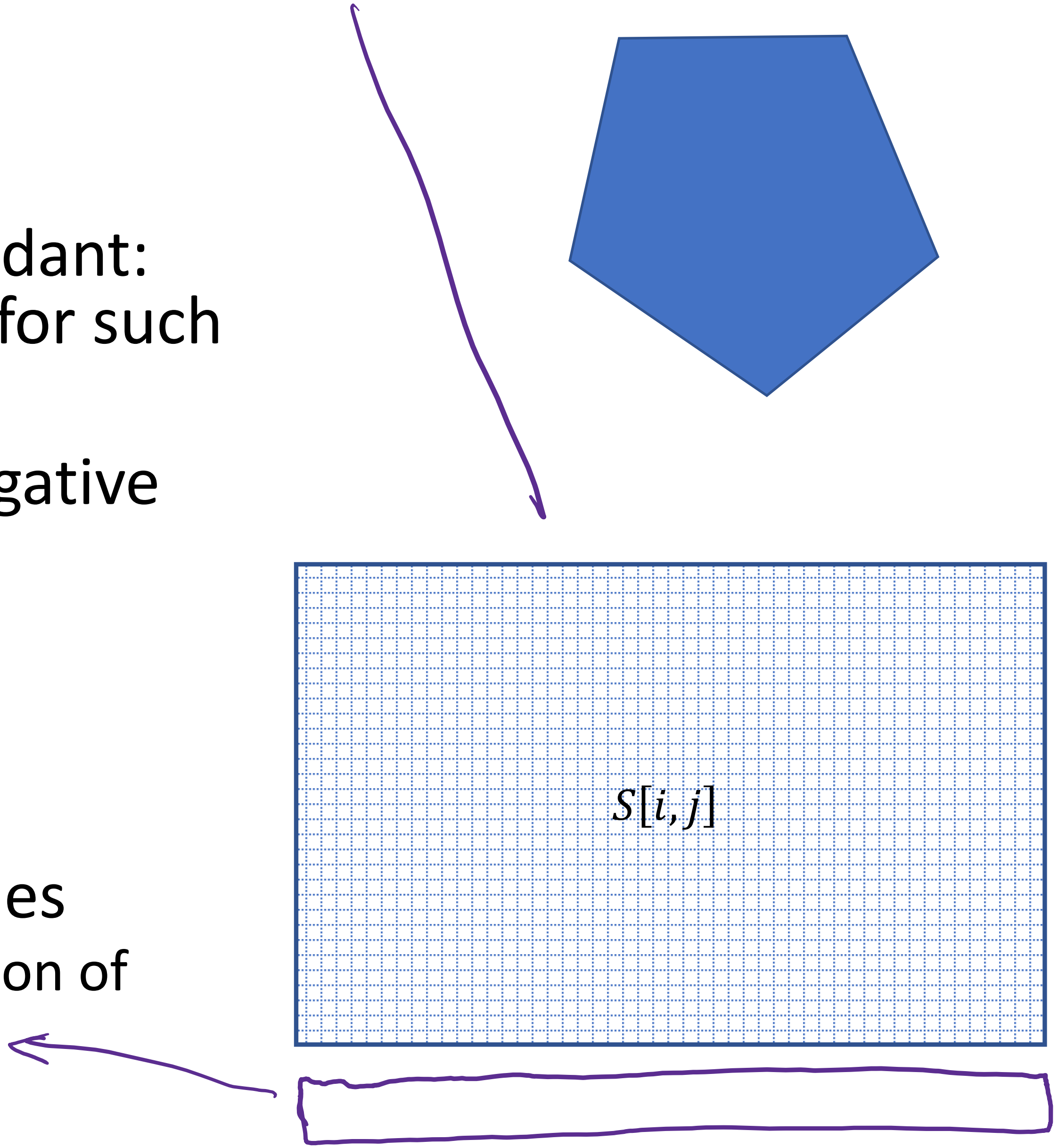
Definition: Non-negative rank $rk_+(S)$ of S is the smallest r for which S is the product of two (entry-wise) non-negative matrices $S = TU$, where $T \in \mathbb{R}^{M\times r}$, and $U \in \mathbb{R}^{r\times N}$



Remark: Without the non-negativity condition on T, U this would just be the usual rank of S

Slack Matrix: Observation

- Some facets / inequalities may be redundant: the slack matrix may also includes rows for such inequalities.
- **Claim:** This does not change the non-negative rank!
- Proof: Farkas' Lemma!
- Idea: The redundant inequality is a non-negative combination of other inequalities
  - The added row is a non-negative combination of other rows



Farkas' Lemma

- Let $P \in \mathbb{R}^n$ be a polyhedron defined by a set of M inequalities $\mathrm{A}x \leq b$, bounded along at least one direction.
- Then any inequality $c^T x \leq \delta$ that is valid for all points of P can be derived by a non-negative linear combination of the given inequalities: i.e. there exist non-negative $\lambda \in \mathbb{R}^m$:

$$\lambda^T A = c^T \qquad \textbf{and} \qquad \lambda^T b = \delta$$

- Conversely, if $c^T x \leq \delta$ is invalid, then it can be refuted by deriving a contradiction using non-negative linear combinations (derive 0=-1)

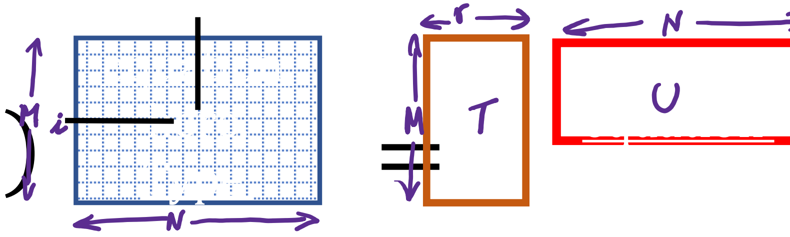
Yannakakis' Factorization Theorem

Theorem: Let $P = \{Ax \leq b | \; x \in \mathbb{R}^n\}$ be a polytope with $\dim(P) \geq 1$. Let V be the set of vertices of P. Let S be the slack matrix of P with respect to the given inequalities. Then the following are equivalent:

1. S has non-negative rank at most r
2. $xc(P) \leq r + O(n)$

In words, the extension complexity is **characterized** by the non-negative rank of the slack matrix.

Proof Outline: $rk_+(S) \le r \Rightarrow$ EF of size $O(r)$



- Let P be given using inequalities. Suppose the slack matrix $S = TU$.
- Let T^i be the i^{th} row of T; U_j be the j^{th} col of U
- $S[i,j] = \langle T^i, U_j \rangle = b_i - \langle A^i, v_j \rangle$ by definition.
- Define polytope P' as:
  - $Ax + Ty = b; y \ge 0$ Here, $y \in \mathbb{R}^r$
- Among the above equalities, at most $n + r$ of them are relevant, as this is the number of variables. Thus we have at most $n + 2r$ constraints.
- Any extreme point v_j in P has an extreme point (v_j, U_j) in P'.

P: v_j extreme pt

P': $Av_j + TU_j$

$= Av_j + (b - Av_j) = b$.

Proof II: EF of size $r \Rightarrow rk_+(S) \leq O(r)$

- If P was specified using r inequalities, we can always bring it to the form:

$$Ex + Fy = t, y \geq 0$$

by adding at most $O(r)$ variables y.

  - e.g. every unconstrained $z = y^+ - y^-$, with $y^+, y^- \geq 0$),
  - inequality $c^T x \leq d$ gets $c^T x + y' = d$ with $y' \geq 0$

- For every vertex v_j, there is a y_j satisfying the above equalities.
- Since we can derive $A^i x \leq b_i$ from above, it implies there exist $\lambda_i \in \mathbb{R}^r_{\geq 0}$, with:

$\lambda_i^T E = A^i, \lambda_i^T t = b_i$, and $\lambda_i^T F \geq 0$. Further, $\lambda_i^T F y_j$ is the slack on vertex v_j of constraint i.

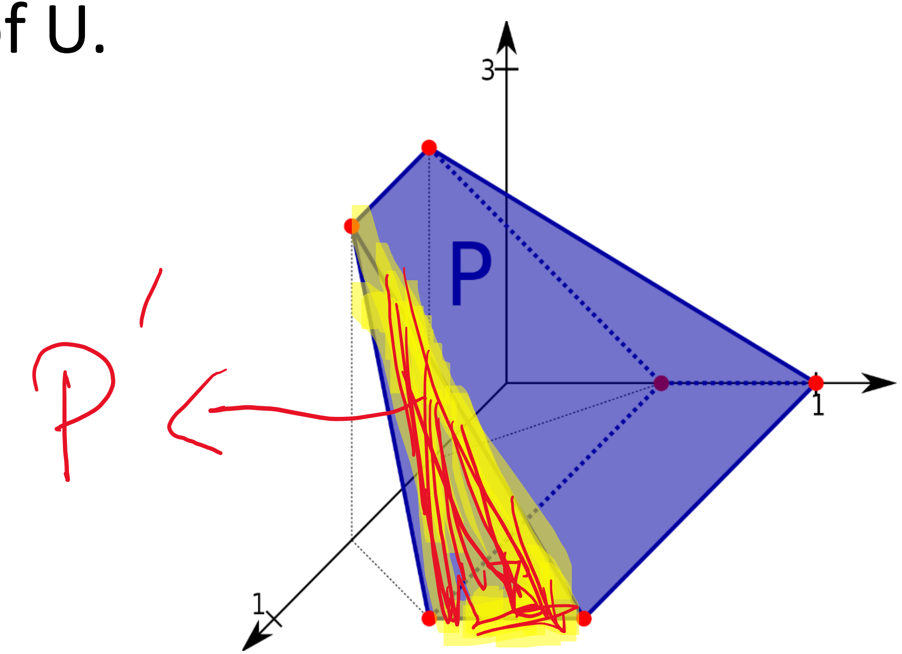
- Finally, define the matrices T, U: $T^i = \lambda_i F$ and $U_j = y_j$.

Details

A simple observation

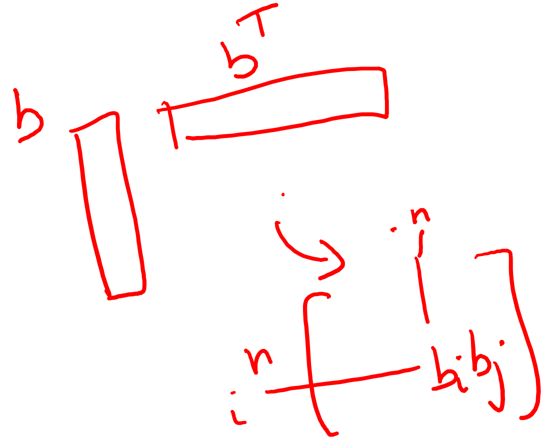
- If P' is a face of P, then xc(P)≥xc(P')

- Proof: The slack matrix of P' is a submatrix of the slack matrix of P. The non-negative rank factorization of P also holds for P' by keeping appropriate rows of T and columns of U.



Step 2: The Correlation Polytope and its Slack Matrix

The correlation polytope $CORR(n)$



$$CORR(n) = \operatorname{conv}\{bb^T \mid b \in \{0,1\}^n\}$$

- The polytope lies in $\mathbb{R}^{n^2}$. One way to think of a feasible point of the polytope is as a matrix $x \in \mathbb{R}^{n \times n}$.
- The extreme points are those for which $x = bb^T$, for some $b \in \{0,1\}^n$.

Main Result: $xc\big(CORR(n)\big) = 2^{\Omega(n)}$

Slack (sub) Matrix of the Correlation polytope

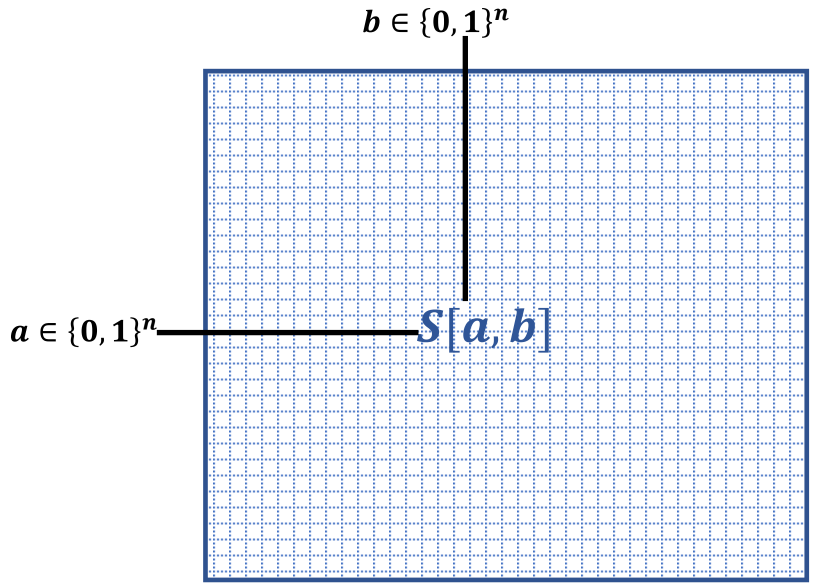
Notation: $X, Y \in \mathbb{R}^{n \times n}$
$\langle X, Y \rangle = \mathrm{Tr}(X^T Y)$

"Unroll the X and Y into vectors and take their inner product"

- What inequalities to consider?
- **Claim:** $\forall a \in \{0,1\}^n \ \forall x \in CORR(n)$:
  $\langle 2\,\mathrm{diag}(a) - aa^T, x \rangle \leq 1$
- **Proof:** Only show for vertices, rest follows by linearity

$$\left\langle 2\begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix} - \begin{bmatrix} a \end{bmatrix}\begin{bmatrix} a^T \end{bmatrix}, x \right\rangle \leq 1.$$

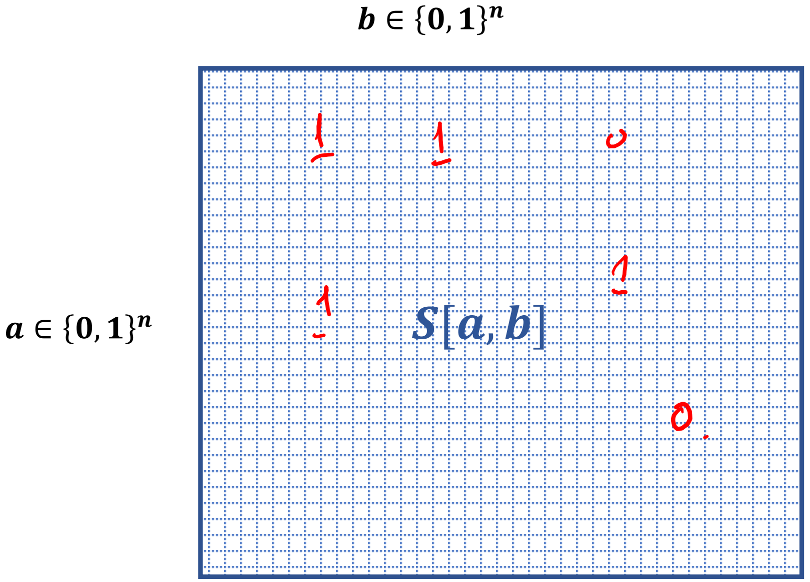
$$1 - \langle 2\,\mathrm{diag}(a) - aa^T, bb^T \rangle$$
$$= 1 - \langle 2\,\mathrm{diag}(a), bb^T \rangle + \langle aa^T, bb^T \rangle$$
$$= 1 - 2a^T b + (a^T b)^2 = (1 - a^T b)^2 \geq 0 \,!$$



$S[a,b] = 1 - \langle 2\,\mathrm{diag}(a) - aa^T, bb^T \rangle$

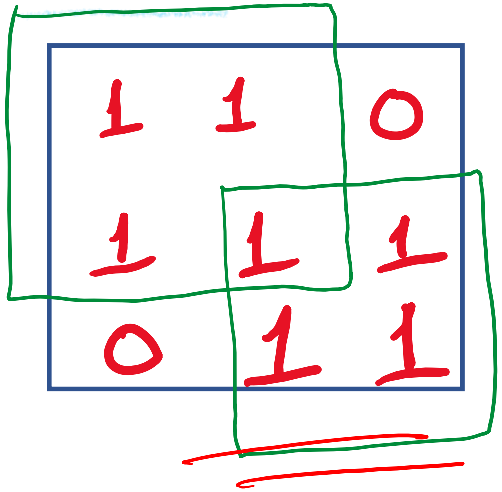
The support matrix

- $$suppmat(S) = \begin{cases} 1 & if \quad S[a,b] \neq 0 \\ 0 & if \quad S[a,b] = 0 \end{cases}$$



$S[a,b]$ =1-$\langle 2\ \mathrm{diag}(a) - aa^T, bb^T \rangle$

- **[Razborov]:** *Covering* only the 1's in SuppMat(S) using *rectangles* requires at least $2^{\Omega(n)}$ rectangles.
  - Suppmat(S) with an appropriate measure, is exactly the communication matrix of the Unique Disjointness function!



Example: Any cover of the 1's in this matrix using rectangles uses at least 2 rectangles

Extension Complexity and Covers

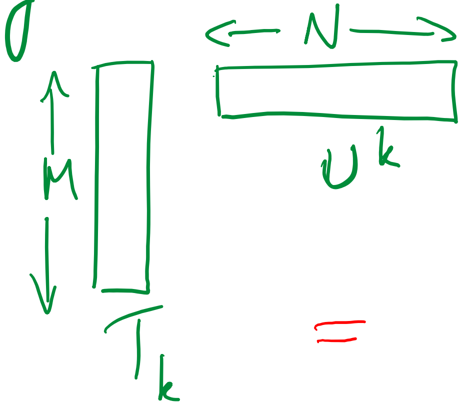
Theorem: $xc(S) \approx rk_+(S) \geq |\text{cover}(suppmat(S))|$ $\geq 2^{\Omega(n)}$.

Corr(n) seen

where, cover denotes the minimum-sized collection of rectangles needed to cover the 1's of suppmat(S).

Proof: Let S=TU, and $rk_+(S) = r$. Then we have: $S = \sum_{k\in[r]} T_k U^k$

every matrix here is



$$\Rightarrow Supp(S) = \bigcup_{k=1}^{r} Supp(T_k U^k)$$

$$= \bigcup_{k=1}^{r} \underbrace{Supp(T_k) \times Supp(U^k)}_{\text{a rectangle}} \Rightarrow r \geq |cover|$$

Step 3: Relating back $CORR(n)$ to TSP_n

How to relate $CORR(n)$ to TSP_n

- We have: $CORR(n) = \text{conv}\{bb^T \mid b \in \{0,1\}^n\}$, $xc(CORR) \geq 2^{\Omega(n)}$
- $TSP_n = \text{conv}\,\{x \in \mathbb{R}^{\binom{n}{2}}\colon x =$ Ham-cycle$(K_n)\}$

We will use: If P' is a face of P, then $xc(P) \geq xc(P')$. A face of TSP_n itself will have extension complexity as exponential.

$TSP_n \geqslant 2^{\Omega(\sqrt{n})}$.

Also, you might have guessed: we will show that $CORR(n)$ is a face of $TSP_{O(n^2)}$.

Use the standard NP-hardness reduction from 3-SAT to TSP.

CORR(n) to 3-SAT

- The following formula ϕ_n on variables Z_{ij} for $i,j \in [n]$
  - $\phi_n = \bigwedge_{i,j\in[n]} \left((Z_{ii} \vee Z_{jj} \vee \bar{Z}_{ij}) \wedge (Z_{ii} \vee \bar{Z}_{jj} \vee \bar{Z}_{ij}) \wedge (\bar{Z}_{ii} \vee Z_{jj} \vee \bar{Z}_{ij}) \wedge (\bar{Z}_{ii} \vee \bar{Z}_{jj} \vee Z_{ij}) \right)$
  - Each set of 4 clauses encodes $Z_{ij} = b_i \wedge b_j$ for each i,j. for some $b_i, b_j \in \{0,1\}$.
- Satisfying assignments to ϕ_n are exactly $Z = bb^T$ for any $b \in \{0,1\}^n$.
  - Convex hull of satisfying assignments is CORR(n)

3-SAT (ϕ_n) to TSP_n

- Build a graph G_n on $\mathrm{O}(n^2)$ vertices. First start with a directed graph for simplicity; then add few vertices to make it undirected.

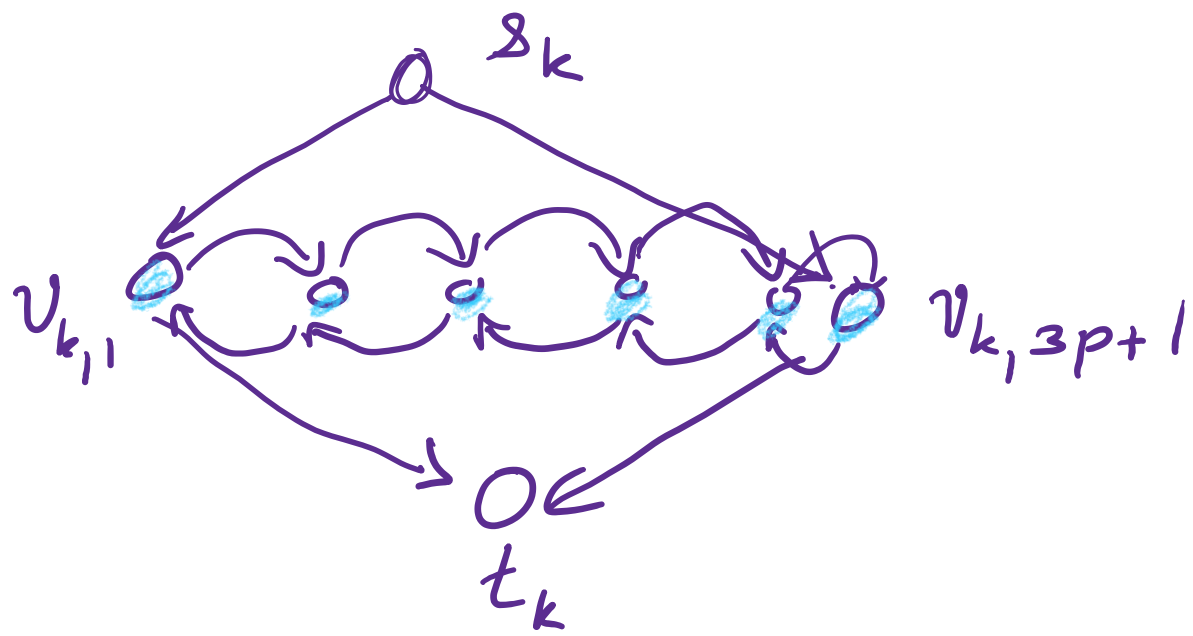
- Tours in G_n will be in one-one correspondence with satisfying assignments of ϕ_n.

- Each tour in G_n is also a tour in K_n. So, convex hull of the tours of G_n is a face of $TSP_{O(n^2)}$.

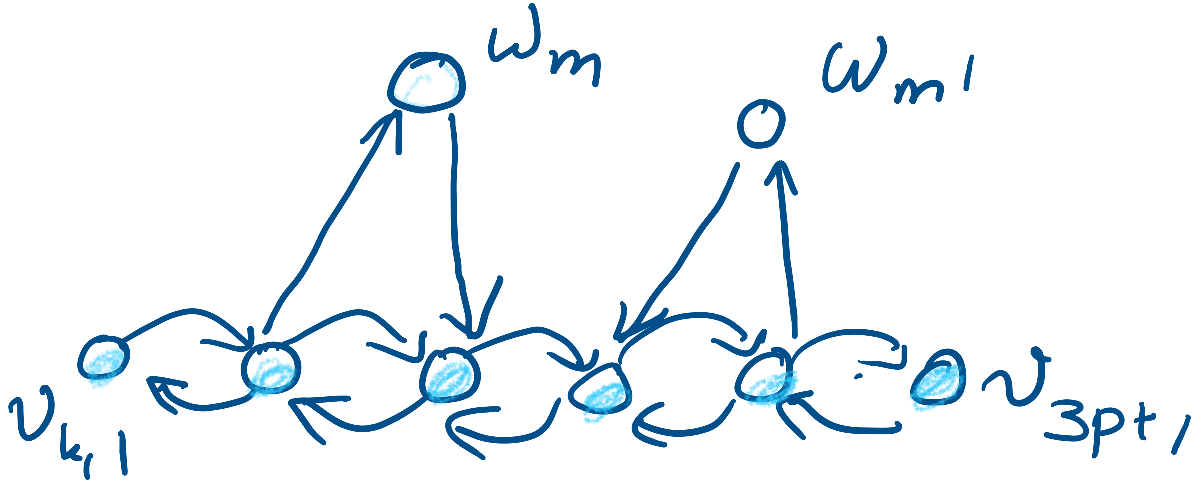
- Since this face is exactly CORR(n) $\Rightarrow xc(TSP_n) = 2^{\Omega(\sqrt{n})}$. □

The Gadget for reducing 3-SAT to TSP

- Variable Gadget:
  - For variable v_k in ϕ occuring in p clauses



- Clause Gadget
  - If clause m has variable k un-negated, and m' has variable k negated



Further developments

- Rothvoss[2013] showed that the perfect matching polytope has exponential extension complexity!
  - Note that perfect matching is solvable in polytime
  - This also improves the TSP lower bound to $2^{\Omega(n)}$.
- Semidefinite extension lower bounds:
  - There exist polytopes with exponential LP complexity, but polynomial SDP complexity.
  - CUT, TSP, Stable set polytopes also have exponential *semidefinite*-extension complexity
- *Approximately* capturing polytopes: indicates what approximation factor can be achieved using LPs/SDPs[Braun-Fiorini-Pokutta-Steurer12]
- Closely related to hierarchies of Linear and Semidefinite Programs (Sherali-Adams, Lasserre, etc.) [CLRS13, LRS15]

Open problems

- Most techniques work only when the base polytope is *independent* of the graph
- Extending known techniques to handle graph-dependent polytopes is a challenging open problem
- Techniques for approximate EFs do not work when there are hard constraints involved
    - For e.g. How well can we *approximate* TSP_n using Extended Formulations is still Open.

Thank You!