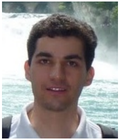
Sparse sum-of-squares certificates on finite abelian groups

Pablo A. Parrilo

Joint work w/ Hamza Fawzi and James Saunderson



Laboratory for Information and Decision Systems
Dept. of Electrical Engineering and Computer Science
Massachusetts Institute of Technology

IMA, May 2015
arXiv:1503.01207

Graphical models and SDP

Semidefinite programming (SDP) is a nice class of convex optimization problems.

Many interesting applications in graphical models, e.g.:

- Approximate inference, marginal polytopes (e.g., Wainwright)
- Clustering/Community detection (e.g., Abbe/Bandeira/Hall, Hajek/Wu/Xu, Deshpande/Montanari, etc)
- Latent-variable Gaussian graphical model selection (w/Chandrasekaran, Willsky)

Graphical models and SDP

Semidefinite programming (SDP) is a nice class of convex optimization problems.

Many interesting applications in graphical models, e.g.:

- Approximate inference, marginal polytopes (e.g., Wainwright)
- Clustering/Community detection (e.g., Abbe/Bandeira/Hall, Hajek/Wu/Xu, Deshpande/Montanari, etc)
- Latent-variable Gaussian graphical model selection (w/Chandrasekaran, Willsky)

However, we don't yet *really* understand what SDP is...

Question: representability of convex sets

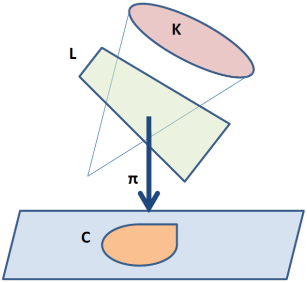
Existence and efficiency:

- When is a convex set representable by conic optimization?
- How to quantify the number of additional variables that are needed?

Given a convex set C, is it possible to represent it as

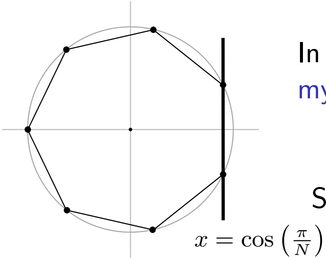
$$C = \pi(K \cap L)$$

where K is a cone, L is an affine subspace, and π is a linear map?



Constructive methods?

Motivation: regular polygons and sparse SOS

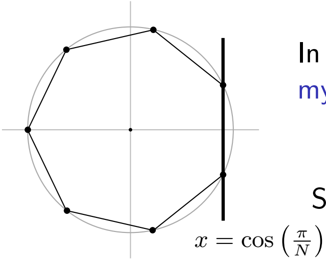


In earlier work (arXiv:1409.4379) mysterious recursion gave

non-negative with freq. $\{0, 1\}$ =
SOS of functions with freq. $\{0, 1, 2, 4, 8, \ldots\}$

Implication: exponentially better relaxations than "standard" sum of squares (SOS) hierarchy!

Motivation: regular polygons and sparse SOS



In earlier work (arXiv:1409.4379) mysterious recursion gave

$$\text{non-negative with freq. } \{0,1\} =$$
$$\text{SOS of functions with freq. } \{0,1,2,4,8,\ldots\}$$

Implication: exponentially better relaxations than "standard" sum of squares (SOS) hierarchy!

(However, only true for certain problems. For cut polytope, strong exponential lower bounds, e.g., Lee-Raghavendra-Steurer).

Motivation

How to do this more systematically?

Motivation

How to do this more systematically?

Abstraction: 'frequencies' $\rightarrow$ *characters of* $\mathbb{Z}_N$

$$\chi_k(x) = \exp\left(\frac{2\pi ikx}{N}\right)$$

Key properties:

- Characters take values in unit complex numbers
- Characters form a group under multiplication

Main question

Setup:

G finite abelian group.

A nonnegative function $f : G \to \mathbb{R}$.

Want a certificate of nonnegativity of f.

Main question

Setup:

G finite abelian group.

A nonnegative function $f : G \to \mathbb{R}$.

Want a certificate of nonnegativity of f.

The function $f : G \to \mathbb{R}$ has Fourier decomposition:

$$f(x) = \sum_{\chi \in \hat{G}} \hat{f}(\chi)\, \chi(x) \quad \text{where } \chi : G \to U(1) \text{ are characters}$$

[Characters are group homomorphisms $\chi(st) = \chi(s)\chi(t)$)

Examples

- Discrete trigonometric polynomials:

$$G = \mathbb{Z}_N, \qquad f(x) = \sum_{k \in \mathbb{Z}_N} \hat{f}(k) e^{2i\pi kx/N}$$

Characters are the N complex exponentials $e^{2i\pi kx/N}$ for $k = 0, \dots, N-1$.

- Boolean hypercube:

$$G = \{-1, 1\}^n \cong \mathbb{Z}^n, \qquad f(x) = \sum_{S \subseteq [n]} \hat{f}(S) \prod_{i \in S} x_i.$$

Characters are the 2^n monomials $\chi_S(x) = \prod_{i \in S} x^i$.

Main question

Setup:
G finite abelian group, $f : G \to \mathbb{R}$ has Fourier decomp.

$$f(x) = \sum_{\chi \in \hat{G}} \hat{f}(\chi)\, \chi(x) \quad \text{where } \chi : G \to U(1) \text{ are characters}$$

[Characters are group homomorphisms $\chi(st) = \chi(s)\chi(t)$)

Assume:

- f is non-negative and
- f is sparse w.r.t. $\mathcal{S} \subseteq \hat{G}$

Problem: Find a 'small' set $\mathcal{T} \subseteq \hat{G}$ such that

- $f(x) = \sum_i |f_i(x)|^2$
- each $f_i : G \to \mathbb{C}$ is sparse w.r.t. $\mathcal{T}$

Flavor of result

Main result: describes combinatorial way to take
G and $\mathcal{S}$ and construct (many different) $\mathcal{T}$ s.t.

$$\begin{aligned} & f : G \to \mathbb{R} \text{ non-negative and sparse w.r.t. } \mathcal{S} \\ \implies \quad & f \text{ is a sum of squares of functions sparse w.r.t. } \mathcal{T} \end{aligned}$$

(Lofty) aim: minimize $|\mathcal{T}|$ w.r.t. choices in construction

Consequence:

$$\{f : f \geq 0 \text{ and sparse w.r.t. } \mathcal{S}\} \quad \text{has a psd lift of size } \mathcal{T}$$

Main result

Assume: $f : G \to \mathbb{R}$ non-negative, sparse w.r.t. $\mathcal{S}$

Let: $\mathrm{Cay}(\hat{G}, \mathcal{S})$ be the Cayley graph of $\hat{G}$ w.r.t. $\mathcal{S}$

Choose:

- a chordal cover Γ of $\mathrm{Cay}(\hat{G}, \mathcal{S})$
- and for each maximal clique C of Γ choose:
  - a character $\chi_C \in \hat{G}$

Then: f is a sum of squares of functions supported on

$$\mathcal{T} = \bigcup_C (\chi_C \cdot C)$$

Aim: make choice of Γ and the $(\chi_C)_C$ so $\mathcal{T}$ as small as possible!

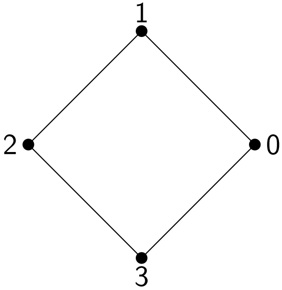
Example:

Assume: $f : \mathbb{Z}_4 \to \mathbb{R}$ is non-negative and

$$f(x) = a_0\chi_0(x) + a_1\chi_1(x) + \overline{a_1}\chi_{-1}(x)$$

i.e. sparse w.r.t. $\mathcal{S} = \{-1, 0, 1\}$.

- $\mathrm{Cay}(\hat{\mathbb{Z}}_4, \mathcal{S})$ is the 4-cycle
- Choose: Γ a chordal cover maximal cliques $C_1 = \{0, 1, 2\}, C_2 = \{0, 2, 3\}$
- Choose: character for each maximal clique



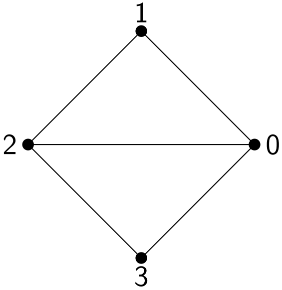
Example:

Assume: $f : \mathbb{Z}_4 \to \mathbb{R}$ is non-negative and

$$f(x) = a_0\chi_0(x) + a_1\chi_1(x) + \overline{a_1}\chi_{-1}(x)$$

i.e. sparse w.r.t. $\mathcal{S} = \{-1, 0, 1\}$.

- $\mathrm{Cay}(\hat{\mathbb{Z}}_4, \mathcal{S})$ is the 4-cycle
- Choose: Γ a chordal cover maximal cliques $C_1 = \{0, 1, 2\}, C_2 = \{0, 2, 3\}$
- Choose: character for each maximal clique



$$\mathcal{T} = (\chi_0 \cdot \{0, 1, 2\}) \cup (\chi_0 \cdot \{0, 2, 3\}) = \{0, 1, 2, 3\} \quad \to \quad \text{size } 4$$

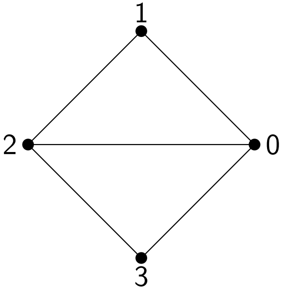
Example:

Assume: $f : \mathbb{Z}_4 \to \mathbb{R}$ is non-negative and

$$f(x) = a_0\chi_0(x) + a_1\chi_1(x) + \overline{a_1}\chi_{-1}(x)$$

i.e. sparse w.r.t. $\mathcal{S} = \{-1, 0, 1\}$.

- $\mathrm{Cay}(\hat{\mathbb{Z}}_4, \mathcal{S})$ is the 4-cycle
- Choose: Γ a chordal cover maximal cliques $C_1 = \{0,1,2\}, C_2 = \{0,2,3\}$
- Choose: character for each maximal clique



$$\mathcal{T} = (\chi_0 \cdot \{0,1,2\}) \cup (\chi_0 \cdot \{0,2,3\}) = \{0,1,2,3\} \quad \to \quad \text{size } 4$$

$$\mathcal{T} = (\chi_{-1} \cdot \{0,1,2\}) \cup (\chi_1 \cdot \{0,2,3\}) = \{-1,0,1\} \quad \to \quad \text{size } 3$$

Next: Proof sketch by example

Aim: explain origin in main result of

- Cayley graph $\mathrm{Cay}(\hat{G}, \mathcal{S})$
- Chordal cover Γ
- choice of χ_C

Work with $f : \mathbb{Z}_4 \to \mathbb{R}$,

$$f(x) = 2\chi_0(x) + (1+i)\chi_1(x) + (1-i)\chi_{-1}(x)$$

- non-negative
- support $\mathcal{S} = \{-1, 0, 1\}$

Trivial SOS certificate → sparse Gram matrix

$$f(x) = 2\chi_0(x) + (1+i)\chi_1(x) + (1-i)\chi_{-1}(x)$$
$$= \begin{bmatrix}\delta_0(x)\\ \delta_1(x)\\ \delta_2(x)\\ \delta_3(x)\end{bmatrix}^* \begin{bmatrix} f(0) & & & \\ & f(1) & & \\ & & f(2) & \\ & & & f(3)\end{bmatrix} \begin{bmatrix}\delta_0(x)\\ \delta_1(x)\\ \delta_2(x)\\ \delta_3(x)\end{bmatrix}$$

Observations:

- If $f : G \to \mathbb{R}$ is non-negative then it is a sum of squares.

Trivial SOS certificate → sparse Gram matrix

$$
\begin{aligned}
f(x) &= 2\chi_0(x) + (1+i)\chi_1(x) + (1-i)\chi_{-1}(x) \\
&= \begin{bmatrix}\delta_0(x)\\ \delta_1(x)\\ \delta_2(x)\\ \delta_3(x)\end{bmatrix}^* \begin{bmatrix} f(0) & & & \\ & f(1) & & \\ & & f(2) & \\ & & & f(3)\end{bmatrix} \begin{bmatrix}\delta_0(x)\\ \delta_1(x)\\ \delta_2(x)\\ \delta_3(x)\end{bmatrix} \\
&= \frac{1}{4} \begin{bmatrix}\chi_0(x)\\ \chi_1(x)\\ \chi_2(x)\\ \chi_3(x)\end{bmatrix}^* \begin{bmatrix} 2 & 1+i & 0 & 1-i \\ 1-i & 2 & 1+i & 0 \\ 0 & 1-i & 2 & 1+i \\ 1+i & 0 & 1-i & 2\end{bmatrix} \begin{bmatrix}\chi_0(x)\\ \chi_1(x)\\ \chi_2(x)\\ \chi_3(x)\end{bmatrix}
\end{aligned}
$$

Observations:

- If $f : G \to \mathbb{R}$ is non-negative then it is a sum of squares.
- In character basis f has Gram matrix sparse w.r.t. $\mathrm{Cay}(\hat{G}, \mathcal{S})$

Aside: Sparse PSD matrices and chordality

A beautiful characterization [Grone-Johnson-Sa-Wolkowicz 84], in primal and dual forms. For a chordal sparsity pattern,

Aside: Sparse PSD matrices and chordality

A beautiful characterization [Grone-Johnson-Sa-Wolkowicz 84], in primal and dual forms. For a chordal sparsity pattern,

- The partially specified matrix

$$\begin{bmatrix} * & * & ? & ? \\ * & * & * & ? \\ ? & * & * & * \\ ? & ? & * & * \end{bmatrix}$$

can be completed to a psd matrix if and only if it is psd on every maximal clique.

Aside: Sparse PSD matrices and chordality

A beautiful characterization [Grone-Johnson-Sa-Wolkowicz 84], in primal and dual forms. For a chordal sparsity pattern,

- The partially specified matrix

$$\begin{bmatrix} * & * & ? & ? \\ * & * & * & ? \\ ? & * & * & * \\ ? & ? & * & * \end{bmatrix}$$

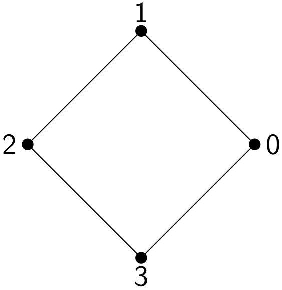
  can be completed to a psd matrix if and only if it is psd on every maximal clique.

- The matrix

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

  is psd if and only it is the sum of smaller psd matrices, each supported on a maximal clique.

Sparse Gram matrix $\to$ chordal cover



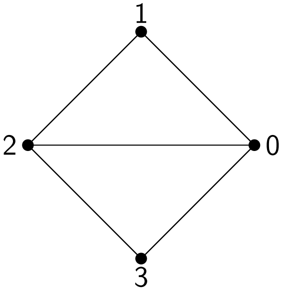
Gram matrix decomposes
as sum of psd matrices:

$$\begin{bmatrix} * & * & 0 & * \\ * & * & * & 0 \\ 0 & * & * & * \\ * & 0 & * & * \end{bmatrix} = \begin{bmatrix} * & * & * & 0 \\ * & * & * & 0 \\ * & * & * & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} * & 0 & * & * \\ 0 & 0 & 0 & 0 \\ * & 0 & * & * \\ * & 0 & * & * \end{bmatrix}$$

Chordal decomposition theorem:
If $X \succeq 0$ and sparse w.r.t. chordal graph Γ then X decomposes as sum of psd matrices supported on maximal cliques of Γ.

Sum of squares interpretation: $f(x) = f_{\{0,1,2\}}(x) + f_{\{0,2,3\}}(x)$
where f_C is a sum of squares of functions supported on C

Sparse Gram matrix $\to$ chordal cover



Gram matrix decomposes as sum of psd matrices:

$$\begin{bmatrix} * & * & 0 & * \\ * & * & * & 0 \\ 0 & * & * & * \\ * & 0 & * & * \end{bmatrix} = \begin{bmatrix} * & * & * & 0 \\ * & * & * & 0 \\ * & * & * & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} * & 0 & * & * \\ 0 & 0 & 0 & 0 \\ * & 0 & * & * \\ * & 0 & * & * \end{bmatrix}$$

Chordal decomposition theorem:
If $X \succeq 0$ and sparse w.r.t. chordal graph Γ then X decomposes as sum of psd matrices supported on maximal cliques of Γ.

Sum of squares interpretation: $f(x) = f_{\{0,1,2\}}(x) + f_{\{0,2,3\}}(x)$
where f_C is a sum of squares of functions supported on C

Most important idea: support alignment

Given sum of squares with each term having 'small' support

$$2f = |\chi_0 + (1+i)\chi_1 + i\chi_2|^2 + |\chi_0 - i\chi_2 + (1-i)\chi_3|^2$$

Possible supports inside first square:

$$\{0, 1, 2\}$$

Possible supports inside second square:

$$\{0, 2, 3\}$$

Most important idea: support alignment

Given sum of squares with each term having 'small' support

$$
\begin{aligned}
2f &= |\chi_0 + (1+i)\chi_1 + i\chi_2|^2 + |\chi_0 - i\chi_2 + (1-i)\chi_3|^2 \\
&= |\chi_a(\chi_0 + (1+i)\chi_1 + i\chi_2)|^2 + |\chi_b(\chi_0 - i\chi_1 + (1-i)\chi_2)|^2
\end{aligned}
$$

Possible supports inside first square:

$$\{0,1,2\}+0, \{0,1,2\}+1, \{0,1,2\}+2, \{0,1,2\}+3$$

Possible supports inside second square:

$$\{0,2,3\}+0, \{0,2,3\}+1, \{0,2,3\}+2, \{0,2,3\}+3$$

Choose a and b to align supports as well as possible

Most important idea: support alignment

Given sum of squares with each term having 'small' support

$$
\begin{aligned}
2f &= |\chi_0 + (1+i)\chi_1 + i\chi_2|^2 + |\chi_0 - i\chi_2 + (1-i)\chi_3|^2 \\
&= |\chi_a(\chi_0 + (1+i)\chi_1 + i\chi_2)|^2 + |\chi_b(\chi_0 - i\chi_1 + (1-i)\chi_2)|^2 \\
&= |\chi_3 + (1+i)\chi_0 + i\chi_1|^2 + |-i\chi_3 + (1-i)\chi_0 + \chi_1|^2
\end{aligned}
$$

Possible supports inside first square:

$$\{0,1,2\}+0, \{0,1,2\}+1, \{0,1,2\}+2, \{0,1,2\}+3$$

Possible supports inside second square:

$$\{0,2,3\}+0, \{0,2,3\}+1, \{0,2,3\}+2, \{0,2,3\}+3$$

Choose a and b to align supports as well as possible

Main result (again)

Assume: $f : G \to \mathbb{R}$ non-negative, sparse w.r.t. $\mathcal{S}$

Let: $\text{Cay}(\hat{G}, \mathcal{S})$ be the Cayley graph of $\hat{G}$ w.r.t. $\mathcal{S}$

Choose:

- a chordal cover Γ of $\text{Cay}(\hat{G}, \mathcal{S})$
- and for each maximal clique C of Γ choose:
  - a character $\chi_C \in \hat{G}$

Then: f is a sum of squares of functions supported on

$$\mathcal{T} = \bigcup_C (\chi_C \cdot C)$$

Aim: make choice of Γ and the $(\chi_C)_C$ so $\mathcal{T}$ as small as possible!

Quadratic polynomials on $\{-1, 1\}^n$

Conjecture (Laurent 2003): If

$$f(x) = a_0 + \sum_{i<j} a_{ij} x_i x_j \quad \text{non-negative } \forall x \in \{-1, 1\}^n$$

then f is a sum of squares of polynomials of degree at most $\lceil n/2 \rceil$.

- Laurent (2003): degree at least $\lceil n/2 \rceil$ necessary
- Blekherman, Gouveia, Pfeiffer (2014): true if allow multipliers

Quadratic polynomials on $\{-1,1\}^n$

Conjecture (Laurent 2003): If

$$f(x) = a_0 + \sum_{i<j} a_{ij} x_i x_j \quad \text{non-negative } \forall x \in \{-1,1\}^n$$

then f is a sum of squares of polynomials of degree at most $\lceil n/2 \rceil$.

- Laurent (2003): degree at least $\lceil n/2 \rceil$ necessary
- Blekherman, Gouveia, Pfeiffer (2014): true if allow multipliers

In our language:

- Group: $G = \{-1,1\}^n \cong \mathbb{Z}_2^n$
- Characters: $\chi_S(x) = \prod_{i \in S} x_i$ (square-free polynomials)
- non-negative functions with support $\mathcal{S} = \{S : |S| \in \{0,2\}\}$

Good choices in main result $\rightarrow$ prove Laurent's conjecture

Degree d polynomials on $\mathbb{Z}_N$

non-negative degree d trigonometric polynomials on $\mathbb{Z}_N$

$$\longleftrightarrow$$

valid linear inequalities for $TC_{N,2d}$

where $TC_{N,2d}$ is a cyclic polytope in $\mathbb{R}^{2d}$ with N vertices:

$$TC_{N,2d} = \text{conv}\left\{\left(\cos\left(\frac{2\pi x}{N}\right), \sin\left(\frac{2\pi x}{N}\right), \ldots, \cos\left(\frac{2\pi d x}{N}\right), \sin\left(\frac{2\pi d x}{N}\right)\right) : x \in \{0, 1, 2 \ldots, N-1\}\right\}$$

Degree d polynomials on $\mathbb{Z}_N$

non-negative degree d trigonometric polynomials on $\mathbb{Z}_N$

$$\longleftrightarrow$$

valid linear inequalities for $TC_{N,2d}$

where $TC_{N,2d}$ is a cyclic polytope in $\mathbb{R}^{2d}$ with N vertices:

$$TC_{N,2d} = \text{conv}\left\{\left(\cos\left(\tfrac{2\pi x}{N}\right), \sin\left(\tfrac{2\pi x}{N}\right), \ldots, \cos\left(\tfrac{2\pi d x}{N}\right), \sin\left(\tfrac{2\pi d x}{N}\right)\right) : x \in \{0, 1, 2 \ldots, N-1\}\right\}$$

Good choices in main result + duality:
If d divides N then $TC_{N,2d}$ has a PSD lift of size $\leq 3d\log_2(N/d)$.

Degree d polynomials on $\mathbb{Z}_N$

non-negative degree d trigonometric polynomials on $\mathbb{Z}_N$

$$\longleftrightarrow$$

valid linear inequalities for $TC_{N,2d}$

where $TC_{N,2d}$ is a cyclic polytope in $\mathbb{R}^{2d}$ with N vertices:

$$TC_{N,2d} = \text{conv}\left\{\left(\cos\left(\tfrac{2\pi x}{N}\right), \sin\left(\tfrac{2\pi x}{N}\right), \ldots, \cos\left(\tfrac{2\pi dx}{N}\right), \sin\left(\tfrac{2\pi dx}{N}\right)\right) : x \in \{0, 1, 2 \ldots, N-1\}\right\}$$

Good choices in main result + duality:
If d divides N then $TC_{N,2d}$ has a PSD lift of size $\leq 3d\log_2(N/d)$.

Related polytope: Bogomolov et al. $\rightarrow$ size $(\log N)^{\lfloor d/2 \rfloor}$ LP lift of:

$$\text{conv}\{(i, i^2, \ldots, i^d) : i = 1, 2, \ldots, N\}$$

Gap between LP and SDP xc for $TC_{d^2,d}$

Good choices in main result + duality:
If d divides N then $TC_{N,2d}$ has a PSD lift of size $\leq 3d \log_2(N/d)$.

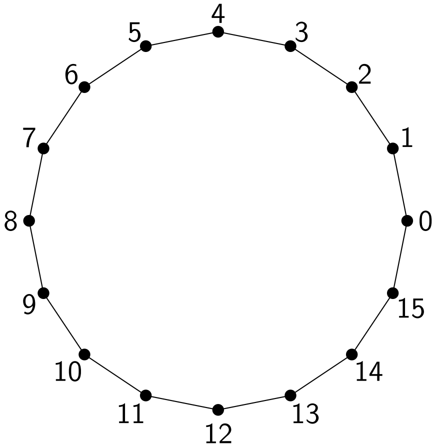
$TC_{d^2,d}$ polytope in $\mathbb{R}^{2d}$ with d^2 vertices:

- SDP extension complexity $O(d \log(d))$
- LP extension complexity $\Omega(d^2)$
  (lower bound due to Fiorini et al. for d-neighborly polytopes)

Degree d polynomials on $\mathbb{Z}_N$

$d = 1$

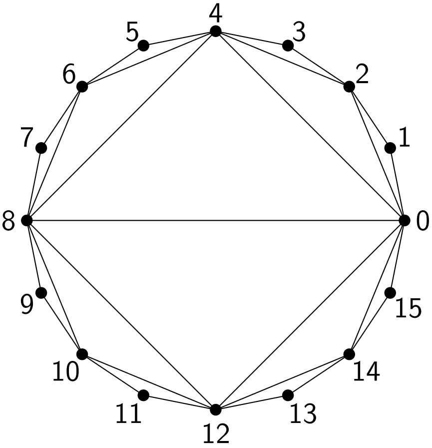
- $\mathcal{S} = \{-1, 0, 1\}$
- Cay$(\hat{\mathbb{Z}}_N, \mathcal{S})$ is N-cycle



Degree d polynomials on $\mathbb{Z}_N$

$d = 1$

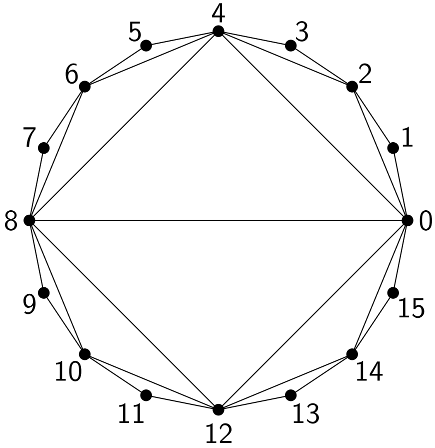
- $\mathcal{S} = \{-1, 0, 1\}$
- $\mathrm{Cay}(\hat{\mathbb{Z}}_N, \mathcal{S})$ is N-cycle

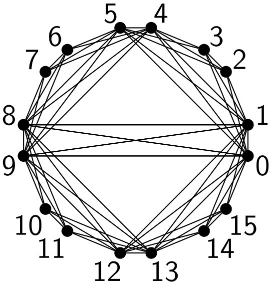


Degree d polynomials on $\mathbb{Z}_N$

$d = 1$

- $\mathcal{S} = \{-1, 0, 1\}$
- $\mathrm{Cay}(\hat{\mathbb{Z}}_N, \mathcal{S})$ is N-cycle

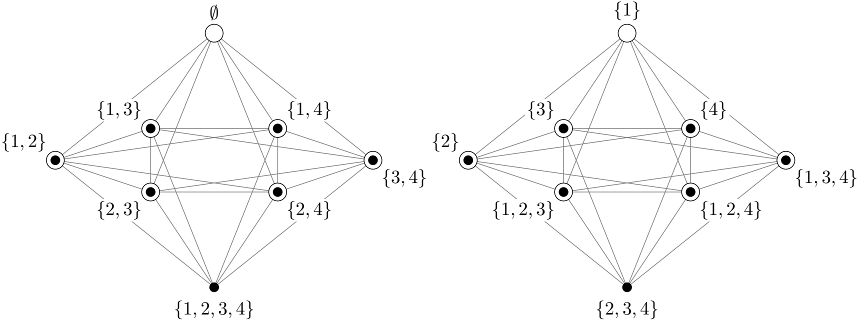




General d

- builds on $d = 1$ case
- more complicated

Quadratic functions on $\{-1,1\}^n$



- $G = \mathbb{Z}_2^n$, $\mathcal{S} = \{S : |S| \in \{0, 2\}\}$
- $\mathrm{Cay}(\hat{\mathbb{Z}}_2^n, \mathcal{S})$ is the half-cube graph
- the most obvious chordal cover *almost* works

Conclusion

Summary:

- If $f : G \to \mathbb{R}$ is non-negative and sparse w.r.t. $\mathcal{S}$ described way to construct $\mathcal{T}$ s.t. f is SOS of functions sparse w.r.t. $\mathcal{T}$.
- Applied to non-negative quadratics on $\{-1, 1\}^n$
  - all are SOS of functions of degree $\leq \lceil n/2 \rceil$
- Applied to non-negative degree d polynomials on $\mathbb{Z}_N$
  - explicit family of polytopes with $\mathrm{xc}_{SDP}/\mathrm{xc}_{LP} \in O(\log(d)/d)$.

Questions:

- Other interesting choices of group G and support $\mathcal{S}$?
- What else is 'symmetry-aware chordal completion' good for?
- Can improve by using *values* of coefficients?

For more information: **arXiv:1503.01207**