Research on Quadratic 0-1 Programming

Duan Li

Department of Systems Engineering & Engineering
Management
The Chinese University of Hong Kong

At National Cheng Kung University
March 24, 2011

Outline

- Binary quadratic program and maximum cut problem

- Polynomially solvable subclasses of BQP

- Exploiting geometric properties of BQP

- SDP relaxation

- Goemanns and Williamson's bound and Nesterov's SDP bound

- Further improvement of the SDP bound

Binary quadratic optimization and max-cut problem

- A general 0-1 or binary quadratic problem can be expressed as:

$$(P) \qquad \min \ f(x) = \frac{1}{2}x^T Q x + c^T x$$
$$\text{s.t. } x \in X,$$

where Q is a symmetric $n \times n$ matrix, $c \in \mathbb{R}^n$ and X is either $\{-1, 1\}^n$ or $\{0, 1\}^n$. Note

$$x_i^2 = 1, \ x_i \in \{-1, 1\} \text{ or } x_i^2 = x_i, \ x_i \in \{0, 1\}.$$

Problem (P) is in general NP-hard even if the matrix Q is positive definite, since $x^T Q x = x^T(Q + diag(\lambda))x - e^T \lambda$ if $x_i \in \{-1, 1\}$ or $x^T Q x = x^T(Q + diag(\lambda))x - \lambda^T x$ if $x_i \in \{0, 1\}$.

Max-cut problem

- Given a graph $G = (V, E)$, find a partition of the nodes of V into two disjoint sets V_1 and V_2 ($V_1 \cap V_2 = \emptyset$, $V_1 \cup V_2 = V$) in such a way to maximize the weights of edges that have one endpoint in V_1 and the other in V_2.
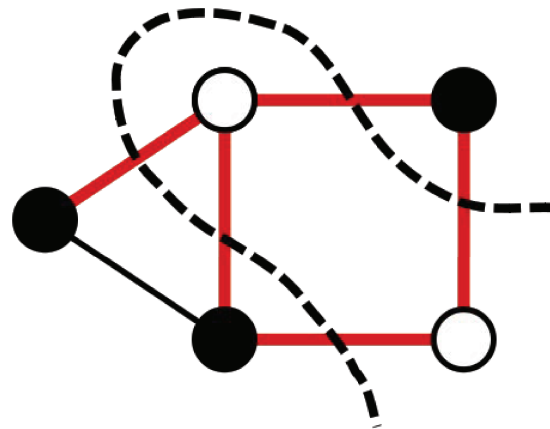
– Let w_{ij} be the weight corresponding to the (i, j) edge, and is zero if the nodes i and j are not connected. Define:

$$y_i = -1 \Longleftrightarrow i \in V_1, \quad y_i = 1 \Longleftrightarrow i \in V_2.$$

– The max-cut problem can be written as:

$$\max \frac{1}{4} \sum_{i,j} w_{ij}(1 - y_i y_j)$$

$$\text{s.t. } y_i \in \{-1, 1\}, \quad i = 1, \ldots, n,$$

– The max cut problem is also APX-hard, meaning that there is no polynomial-time (ϵ-) approximation scheme (PTAS) for it unless P = NP.

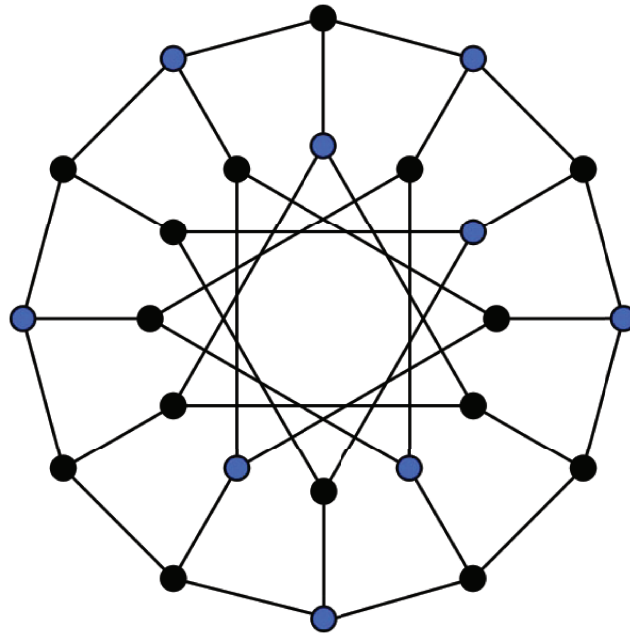## Maximal stable set

– Given an undirected graph $G = (V, E)$.

Figure 1: Maximal stable set of 9 blue nodes in a 24-node graph

– Independent subset of a graph: no two nodes from it are linked by an arc in E. Maximal stable set : the independent subset with maximum size of nodes $\alpha(G)$.

– Define

$$(SQP) \quad \min\ x^T(I+A)x$$
$$\text{s.t. } e^T x = 1,\ x \geq 0,$$

where A is the adjacency matrix of G. Then $\alpha(G) = \frac{1}{v(SQP)}$.

– Maximal stable set problem is NP-hard and is closely related to quadratic knapsack problem.

Polynomially solvable subclasses of BQP

- As the binary quadratic programming program is NP-hard in general, identifying polynomially solvable subclasses of binary quadratic programming problems not only offers theoretical insight into the complicated nature of the problem, but also provides platforms to design relaxation schemes for exact solution methods.

Problems with all off-diagonal elements of Q being non-positive

- Consider a subclass of problem $(0\text{-}1QP)$ where all off-diagonal elements of Q are non-positive and $q_{ii} = 0$, $i = 1, \ldots, n$.

- As $x_i x_j = \min(x_i, x_j)$ when $x_i, \ x_j \in \{0, 1\}$. and $q_{ij} \leq 0$ for $1 \leq i < j \leq n \Rightarrow$ A linear programming problem with a totally unimodular constraint matrix and an integral right-hand side

$$\min \ \sum_{i=1}^{n} c_i x_i + \sum_{1 \leq i < j \leq n} q_{ij} z_{ij}$$

$$\text{s.t.} \ \ z_{ij} \leq x_i, \ 1 \leq i < j \leq n,$$

$$z_{ij} \leq x_j, \ 1 \leq i < j \leq n,$$

$$x_i, \ x_j, \ z_{ij} \in \{0, 1\}, \ 1 \leq i < j \leq n.$$

- LP generates integer solution!

Homogenous problems with fixed rank Q

- We consider now a subclass of homogenous problems where Q is negative semidefinite and $rank(Q) = d$.

- Let $G = -Q \Rightarrow$ There a row full rank $d \times n$ matrix, V, such that $G = V^T V$.

- $\Rightarrow$ Equivalent problem:

$$(BQP_{fr}) \qquad \max_{x \in \{0,1\}^n} x^T G x = x^T V^T V x = \sum_{i=1}^{d} (v_i x)^2, \quad (1)$$

where v_i is the i-th row of matrix V.

- If d is equal to 1, i.e., the matrix G is of rank one with $G = v_1^T v_1$, the solution to (BQP_{fr}) can be easily found by inspection. More specifically, we only need to select x such that the absolute value of $v_1 x$ is maximized on $\{0,1\}^n$.

- In general cases with $rank(G) = d > 1$, we consider a linear map $\Phi \colon x \in \mathbb{R}^n \to z = Vx \in \mathbb{R}^d$, in which Φ maps the hypercube $[0,1]^n$ into a convex polytope $Z(V) = \Phi([0,1]^n) = \{z \in \mathbb{R}^d \mid z = Vx, \ x \in [0,1]^n\}$, known as a *zonotope*.

- Note that

$$\max_{x \in \{0,1\}^n} x^T G x = \max_{x \in \{0,1\}^n} \sum_{i=1}^{d} (v_i x)^2 = \max_{z \in Z(V)} \sum_{i=1}^{d} z_i^2 = \max_{z \in Z(V)} \|z\|^2.$$

Thus, (BQP_{fr}) reduces to a problem of finding the maximum norm in a zonotope.

- Let $N_{ep}(Z)$ denote the number of extreme points of the zonotope $Z(V)$. Then $N_{ep}(Z) = O(n^{d-1})$.

- All the extreme points of the zonotope $Z(V)$ can be found by cell enumeration algorithms in discrete geometry with complexity $O(n^d)$.

- $\Rightarrow$ Homogenous problems with fixed rank Q is polynomially solvable.

Problems with Q Being a tridiagonal matrix are polynomially solvable

(BQP) defined by a series-parallel graph are polynomially solvable

- (BQP) can be always reduced to a max-cut problem. Consider an instance of (BQP) with

$$
Q = \begin{pmatrix} 0 & 1 & 0 & -1.5 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.5 \\ -1.5 & 0 & -0.5 & 0 \end{pmatrix}, \quad c = \begin{pmatrix} 2.5 \\ -2 \\ 3 \\ 1.5 \end{pmatrix}
$$
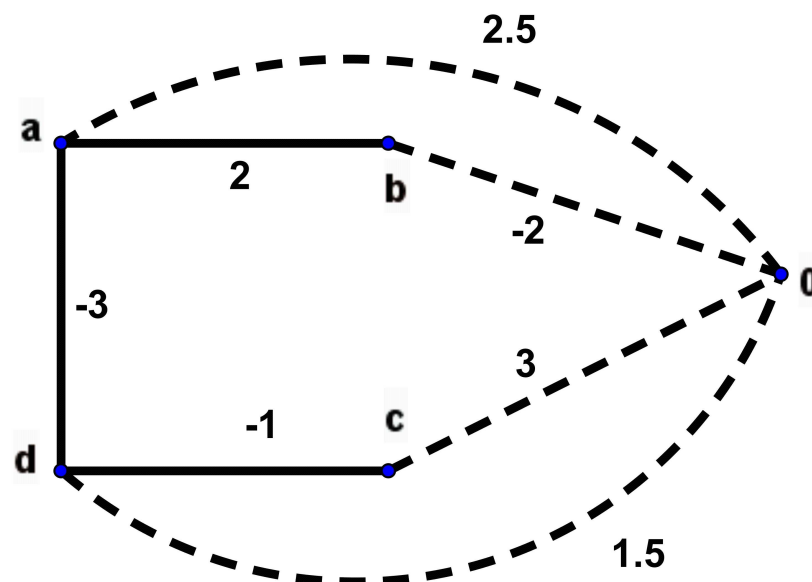
- A graph is called *series-parallel* if it is not contractible to K_4.

- If graph $G(Q)$ is series-parallel, then graph $G(Q, c)$ is not contractible to K_5.

Exploiting Geometric Properties of QBP

- Equivalent perturbed problem:

$$(P_\mu) \quad \min_{x \in \{0,1\}^n} f_\mu(x) = \frac{1}{2} x^T (Q + \mu I) x + (c - \frac{1}{2}\mu e)^T x,$$

where $\mu \geq 0$, I is the $n \times n$ identity matrix and e is the vector with all elements equal to 1.

- The shape of contour changes when μ changes.

- The properties of (P_μ), e.g., the conditional number, change when μ changes.

$v = f_\mu(x)$

μ=5

μ=3

μ=1

μ=2
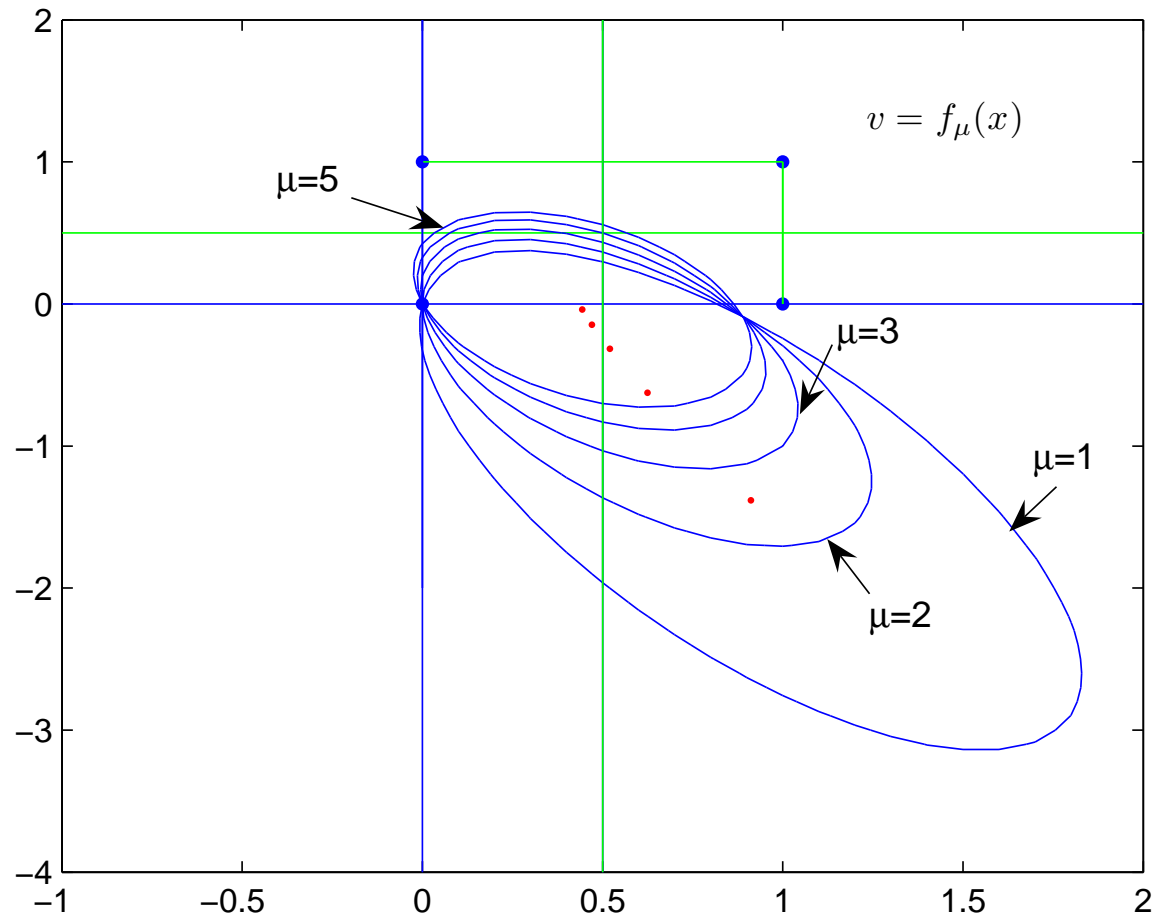
Properties of Perturbation Problem

- The ellipse contour of $f_\mu(x)$ with level $\tilde{v}$:

$$E_\mu(\tilde{v}) = \{x \in \mathbb{R}^n \mid f_\mu(x) = \tilde{v}\}.$$

- The center of $E_\mu(\tilde{v})$:

$$x^0(\mu) \;=\; -(Q + \mu I)^{-1}(c - \frac{1}{2}\mu e). \qquad (2)$$

- Denote $x^0 = x^0(0) = -Q^{-1}c$ and $t_0(\mu) = f_\mu(x^0(\mu))$. The perturbed objective function can be rewritten as

$$f_\mu(x) = \frac{1}{2}(x - x^0(\mu))^T(Q + \mu I)(x - x^0(\mu)) + t_0(\mu). \quad (3)$$

Properties of Perturbation Problem

Theorem 1 *Assume that $x^0 \neq -\frac{1}{2}e$. (i) $t_0(\mu)$ is a strictly concave function on $[0, \infty)$.*

(ii) Let $d_0(\mu) = \|x^0(\mu) - \frac{1}{2}e\|$. Then $d_0(\mu)$ is a strictly decreasing function on $[0, \infty)$.

(iii) Let $S = \{x \in \mathbb{R}^n \mid \|x - \frac{1}{2}e\| \leq \frac{\sqrt{n}}{2}\}$. Let $\bar{\mu}$ be the critical point on the sphere such that $\|x^0(\mu_0) - \frac{1}{2}e\| = \frac{\sqrt{n}}{2}$. If $x^0 \notin S$, then $t_0(\mu)$ is strictly increasing on $[0, \bar{\mu}]$ and strictly decreasing on $[\bar{\mu}, \infty)$. Otherwise, if $x^0 \in S$, then $t_0(\mu)$ is strictly decreasing on $[0, \infty)$.

Lower Bound Derived from the Minimum Distance Sphere
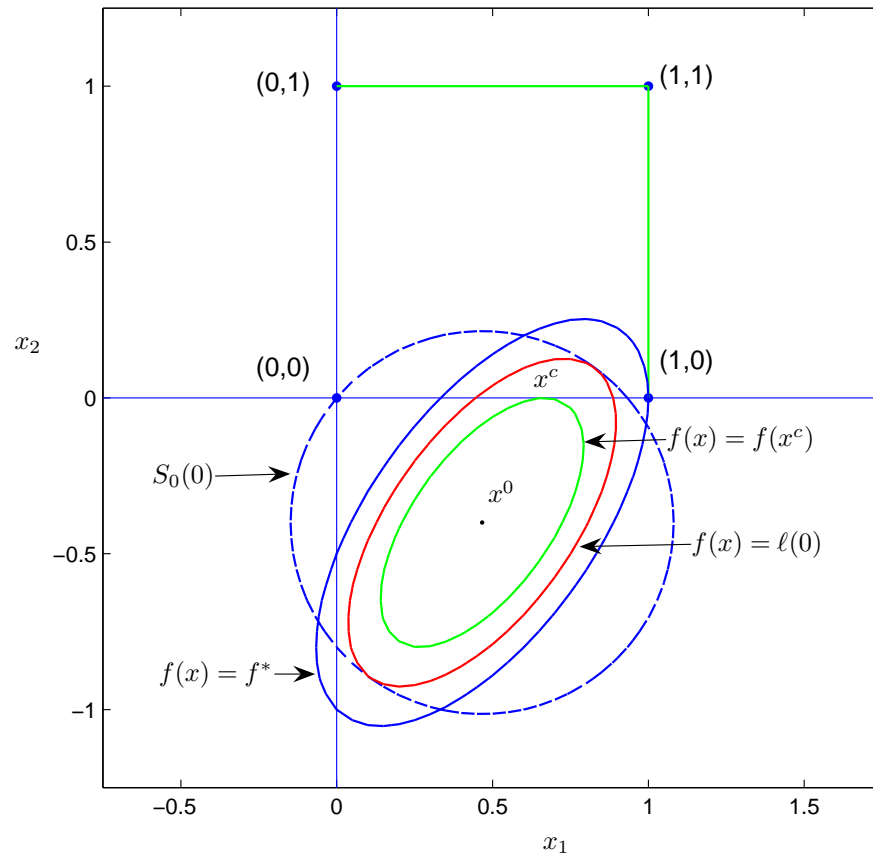
Lower Bound Derived from the Minimum Distance Sphere

Theorem 2 *Let f^* denote the optimal value of (P) and $\bar{x}(\mu)$ be the nearest 0-1 point to $x_0(\mu)$.*
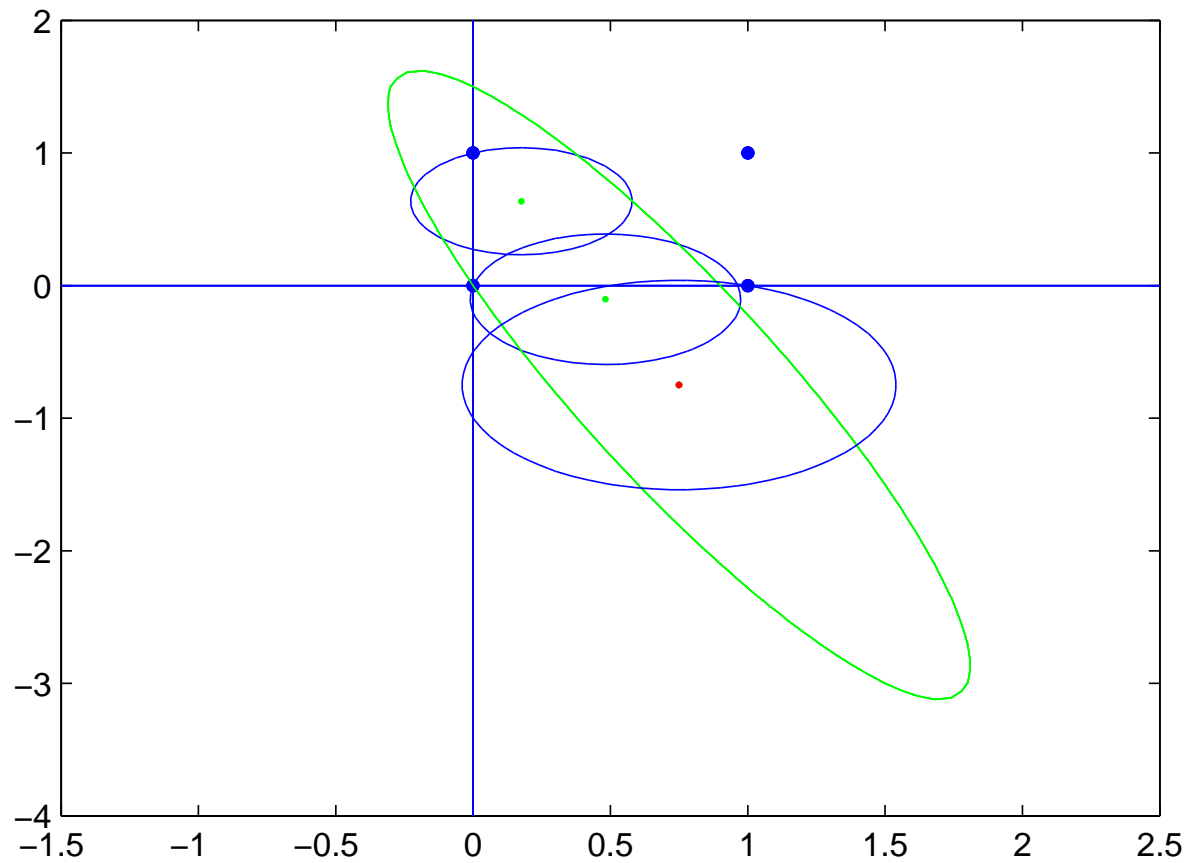
(i) *For any $\mu \geq 0$, a lower bound of f^* is given by*

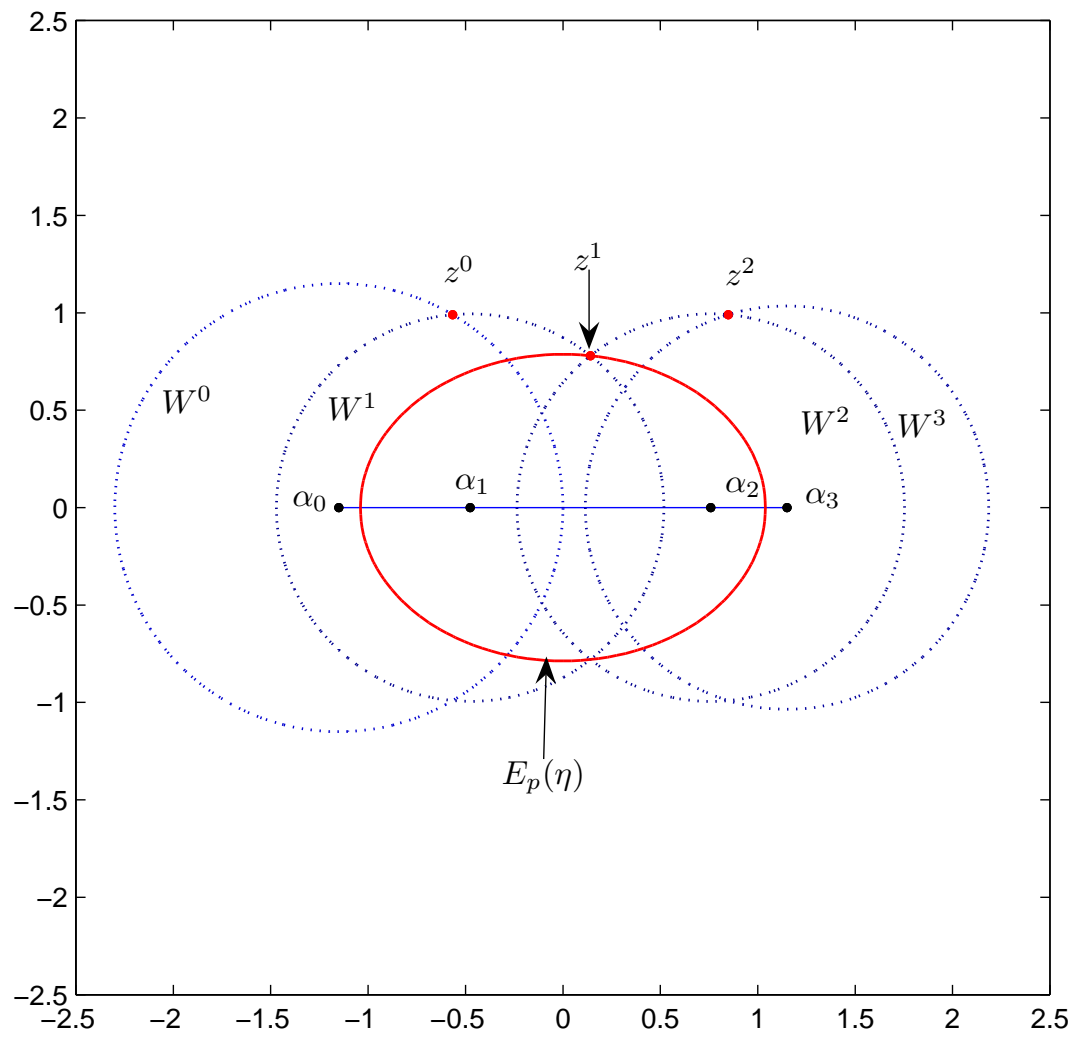$$\ell(\mu) = \frac{\lambda_1 + \mu}{2} \|\bar{x}(\mu) - x^0(\mu)\|^2 + t_0(\mu). \qquad (4)$$

(ii) *The limit of $\ell(\mu)$ exists:*

$$\ell^\infty \equiv \lim_{\mu \to \infty} \ell(\mu) = f(\frac{1}{2}e) + \frac{n\lambda_1}{8} - (\bar{x}^\infty - \frac{1}{2}e)^T Q(x^0 - \frac{1}{2}e).$$

Lower Bound derived from Multiple Minimum Distance Spheres centered at the Longest Axis of the Contour

Lower Bound

- Let z^i be the intersection point of Ω^i and Ω^{i+1} and define $\eta = \min\{\min_{i=0,\dots,m} h(z^i), h(u^1), h(u^2)\}$.

- Let Ψ be the set of all points of $\{0,1\}^n$ in $\cup_{i=0}^{m+1} S^i$.

Theorem 3 *Let $P(\bar{v}) = \{x \in \mathbb{R}^n \mid f(x) \le \bar{v}\}$, where $\bar{v} = \frac{1}{2}\eta + f(x^0)$*

(i) If $P(\bar{v}) \cap \Psi \ne \emptyset$, Then

$$\tilde{x} \in \arg\min_{x \in P(\bar{v}) \cap \Psi} f(x)$$

is an optimal solution to (P).

(ii) If $P(\bar{v}) \cap \Psi = \emptyset$, then $\bar{v}$ is a lower bound of f^*.

SDP Problem

- Example of Linear program and semidefinite program (SDP):

$$(LP) \qquad \min 2x_1 + x_2 + x_3$$
$$\text{s.t. } x_1 + x_2 + x_3 = 1$$
$$(x_1, x_2, x_3) \geq 0.$$

$$(SDP) \qquad \min 2x_1 + x_2 + x_3$$
$$\text{s.t. } x_1 + x_2 + x_3 = 1$$
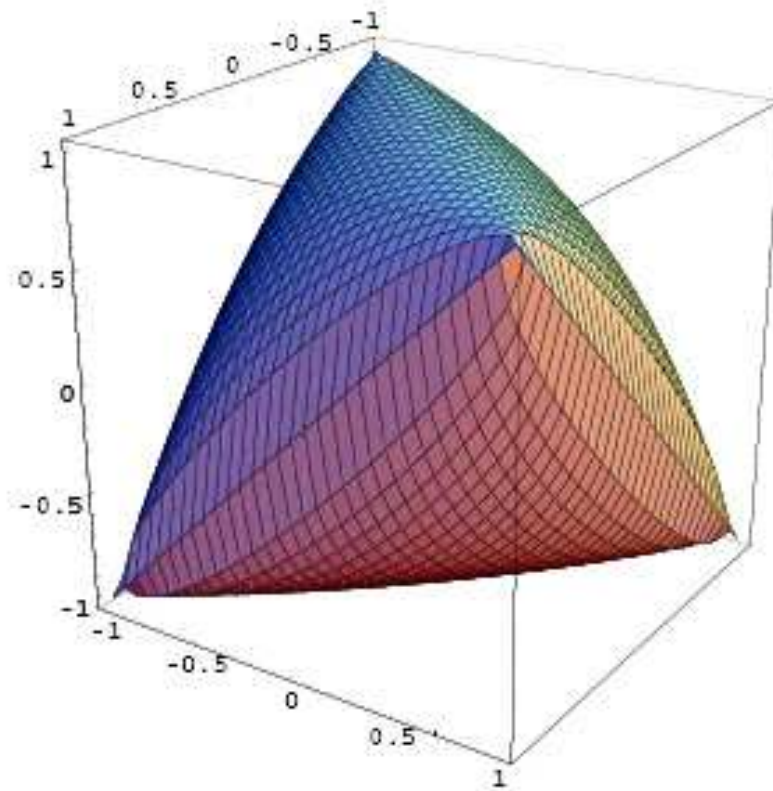$$\begin{pmatrix} x_1 & x_2 \\ x_2 & x_3 \end{pmatrix} \succeq 0.$$

Figure 2: Set of 3×3 positive semidefinite matrices with unit diagonal

- General form of SDP problem:

$$(SDP) \qquad \min C \bullet X$$
$$\text{s.t. } A_i \bullet X = b_i, \ \ i = 1, \ldots, m,$$
$$X \succeq 0,$$

where C, A_i are given $n \times n$ symmetric and b_is are given scalars, and

$$A \bullet X = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_{ij} = Tr(A^T X).$$

- The dual of (SDP) is

$$(SDD) \qquad \max b^T y$$
$$\text{s.t. } \sum_{i=1}^{m} y_i A_i \preceq C,$$

where $y \in \mathbb{R}^m$. Or equivalently

$$(SDD) \qquad \max b^T y$$
$$\text{s.t. } \sum_{i=1}^{m} y_i A_i + S = C,$$
$$S \succeq 0.$$

- Strong duality: $v(SDP) = v(SDD)$ if (SDP) or (SDD) is strictly feasible.

- The SDP interior-point algorithm finds an ϵ-approximate solution where solution time is linear in $\log(1/\epsilon)$ and polynomial in m and n.

- SDP solvers and software based on MatLab: SeDuMi, CVX $(n \leq 1000)$.

SDP relaxation for 0-1 QP

Consider the unconstrained case of (P):

$$(BQP) \qquad \min \ f(x) = \frac{1}{2} x^T Q x + c^T x$$
$$x \in \{-1, 1\}^n,$$

Two ways of constructing SDP relaxations:

- Lifting method. Note that

$$f(x) = \frac{1}{2}(1, x^T) \begin{pmatrix} 0 & c \\ c & Q \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix} = \frac{1}{2} \tilde{Q} \bullet [\begin{pmatrix} 1 \\ x \end{pmatrix} (1, x^T)].$$

Lifting x to a symmetric matrix $X = \begin{pmatrix} 1 \\ x \end{pmatrix} (1, x^T)$, we get

$$\frac{1}{2}x^T Q x + c^T x = \frac{1}{2}\tilde{Q} \bullet X,$$

$$x \in \{-1, 1\}^n \Longleftrightarrow X_{ii} = 1, \ i = 2, \ldots, n+1.$$

- Note also

$$X = \begin{pmatrix} 1 \\ x \end{pmatrix} (1, x^T) \Longleftrightarrow X \succeq 0, \ X_{11} = 1, \ \mathrm{rank}(\mathrm{X}) = 1.$$

Dropping $\mathrm{rank}(\mathrm{X}) = 1$, we obtain the following SDP relaxation for (BQP):

$$(SDP_0) \qquad \min \ \frac{1}{2}\tilde{Q} \bullet X$$

$$\text{s.t. } X \succeq 0, \quad X_{ii} = 1, \ i = 1, \ldots, n+1.$$

- **Shor's relaxation scheme.** Note that (BQP) is equivalent to

$$(BQP_c) \qquad \min \ f(x) = \frac{1}{2} x^T Q x + c^T x$$
$$x_i^2 = 1, \ i = 1, \dots, n.$$

The dual problem of (BQP) is

$$(D) \qquad \max_{\lambda \in \mathbb{R}^n} \min_{x \in \mathbb{R}^n} \frac{1}{2} x^T [Q + 2 diag(\lambda)] x + c^T x - e^T \lambda$$

which is equivalent to an SDP problem:

$$(D) \qquad \max \ \tau$$
$$\text{s.t.} \ \begin{pmatrix} Q + 2 diag(\lambda) & c \\ c & -2 e^T \lambda - 2\tau \end{pmatrix} \succeq 0.$$

- $(SDP_0) \Longleftrightarrow (D)$ (conic dual).

- The nature of the SDP relaxation is a nonlinear lifting to a higher dimensional space. Indeed, rather than solving the original quadratic problem in the n-dimensional vector space, we are instead solving a linear problem in the $n \times n$ matrix space.

- If we find an optimal solution X of the SDP that has rank one, $\Rightarrow$ we solve the original problem.

- In general, it is not the case that the optimal solution of the SDP relaxation will be rank one. However, it is possible to use rounding schemes to obtain nearby rank one solutions. Furthermore, in some cases, it is possible to do so while obtaining some approximation guarantees on the quality of the rounded solutions.

Goemanns and Williamson's bound

- Basic questions:

  - Approximation guarantees: Is it possible to prove general properties on the quality of the bounds obtained by SDP?

  - Feasible solutions: Can we (somehow) use the SDP relaxations to provide not just bounds, but actual feasible points with good (or optimal) values of the objective?

- In their celebrated MAXCUT paper (JACM, 1995), Goemans and Williamson developed the following randomized method for finding a good feasible cut from the solution of the SDP:

– Factorize X as $X = V^T V$, where $V = [v_1, \ldots, v_n] \in \mathbb{R}^{r \times n}$, where r is the rank of X.

– Then $X_{ij} = v_i^T v_j$, and, since $X_{ii} = 1$, this factorization gives n vectors v_i on the unit sphere in $\mathbb{R}^r$.

– Now, choose a random hyperplane (passing through the origin) in $\mathbb{R}^r$, and assign to each variable x_i either $+1$ or -1, depending on which side of the hyperplane the point v_i lies.

• It turns out that this procedure gives a solution that, on average, is quite close to the value of the SDP bound. The random hyperplane can be characterized by its normal vector p, which is chosen to be uniformly distributed on the unit sphere.

- The rounded solution is given by $x_i = sign(p^T v_i)$. $\Rightarrow$ The expected value of this solution in $x^T W x$:

$$
\begin{aligned}
E_p[x^T W x] &= \sum_{i,j} w_{ij} E_p[x_i x_j] \\
&= \sum_{i,j} w_{ij} E_p[sign(p^T v_i) \cdot sign(p^T v_j)] \\
&= \frac{2}{\pi} \sum_{i,j} w_{ij} \arcsin X_{ij}.
\end{aligned}
$$

- The objective of maximum cut is $\frac{1}{4} \sum_{i,j} w_{ij}(1 - y_i y_j)$. $\Rightarrow$

$$
c_{\text{sdp}-\text{expected}} = \frac{1}{4} \cdot \frac{2}{\pi} \sum_{i,j} w_{ij} \arccos X_{ij}.
$$

- On the other hand, the solution of the SDP gives an upper bound on the cut capacity equal to:

$$c_{\text{sdp-upperbound}} = \frac{1}{4} \sum_{i,j} w_{ij}(1 - X_{ij}).$$
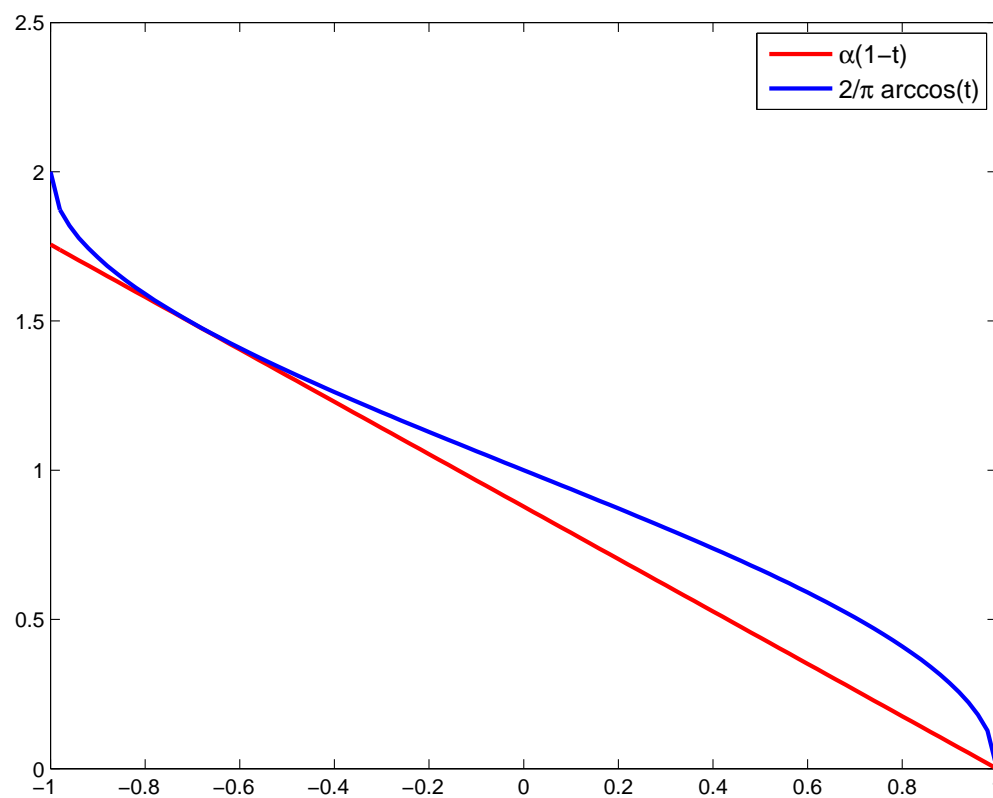
- Consider the problem of finding a constant α such that

$$\alpha(1 - t) \le \frac{2}{\pi} \arccos(t), \quad \forall t \in [-1, 1].$$

This is

$$\alpha = \min_{t \in [-1,1]} \frac{2}{\pi} \frac{\arccos(t)}{1 - t} = \min_{\theta \in [0,\pi]} \frac{2}{\pi} \frac{\theta}{1 - \cos\theta}.$$

- It can be shown that $0.87856 < \alpha < 0.87857$.

- Thus

$$c_{\text{sdp-upperbound}} \le \frac{1}{4} \cdot \frac{1}{\alpha} \sum_{i,j} w_{ij} \arccos(X_{ij}) = \frac{1}{\alpha} c_{\text{sdp-expected}}.$$

- So far we have the following inequalities:

$$c_{\text{sdp-upperbound}} \leq \frac{1}{\alpha} c_{\text{sdp-expected}}$$

$$c_{\text{sdp-expected}} \leq c_{\text{max}}$$

$$c_{\text{max}} \leq c_{\text{sdp-upperbound}}$$

- Therefore

$$\boxed{0.878 \cdot c_{\text{sdp-upperbound}} \leq c_{\text{max}}}$$

(approximation ratio for SDP bound)

$$\boxed{0.878 \cdot c_{\text{max}} \leq c_{\text{sdp-expected}}}$$

(approximation ratio for feasible solution)

Nesterov's SDP bound

- In the MAXCUT problem, we are in fact maximizing the homogeneous quadratic form (omitting $\frac{1}{4}$):

$$x^T A x = \sum_{i,j=1}^{n} w_{ij}(1 - x_i x_j) = \sum_{i=1}^{n}(\sum_{j=1}^{n} w_{ij}) x_i^2 - \sum_{i,j=1}^{n} w_{ij} x_i x_j$$

  over $\{-1, 1\}^n$.

- Special properties of A:

  - $A \succeq 0$;
  - $A_{ij} \leq 0$ for all $i \neq j$;
  - $\sum_{j=1}^{n} A_{ij} = 0$.

- What happens if A is a general positive semidefinite matrix?

- Let $A \succeq 0$, consider the problem:

$$(P) \qquad \max_{x \in \{-1,1\}^n} x^T A x.$$

- The SDP relaxation of (P) is

$$(SDP) \qquad \max A \bullet X$$
$$\text{s.t. } diag(X) = e,$$
$$X \succeq 0.$$

$\frac{v(SDP)}{v(P)} \geq 1.$

- Nesterov (1998). Let $A \succeq 0$. Then
$$\boxed{v(P) \leq v(SDP) \leq \frac{\pi}{2} v(P)}$$

Improvement of the SDP bound

– Saddle point characterization of optimality conditions:
Let $x^* \in \{-1, 1\}^n$ and $\lambda^* \in \mathbb{R}^n$. Then x^* solves (P),
λ^* solves (D) and $v(P) = v(D) \Leftrightarrow$

(1) $Q + 2\mathrm{diag}(\lambda^*) \succeq 0$;

(2) $[Q + 2\mathrm{diag}(\lambda^*)]x^* = -c$.

– Let

$$Q^* = Q + 2\mathrm{diag}(\lambda^*),$$
$$C = \{x \in \mathbb{R}^n \mid Q^* x = -c\}.$$

Then, $v(P) = v(D) \Leftrightarrow \in \{-1, 1\}^n \cap C \neq \emptyset.$

– A sufficient condition for the polynomial solvability of (P):

$Q^* \succ 0 \Rightarrow v(P) = v(D)$, i.e. there is no duality gap.

– What if $Q^* \nsucc 0$? The saddle point conditions motivate us to define the distance:

$$\delta = \mathrm{dist}(\{-1, 1\}^n, C).$$

– $\delta = 0 \Leftrightarrow v(P) = v(D)$.

– Can we improve the SDP bound $v(D)$ if $\delta \neq 0$?

— Let

$$Q^* = Q + 2\text{diag}(\lambda^*),$$
$$C = \{x \in \mathbb{R}^n \mid Q^* x = -c\}.$$

Then, $v(P) - v(D) = 0 \Leftrightarrow \exists\, x^* \in \{-1, 1\}^n \cap C.$

— This motivates us to define the distance:

$$\delta = \text{dist}(\{-1, 1\}^n, C).$$

— $\delta = 0 \Leftrightarrow v(P) = v(D).$

- **Main result**: $\delta > 0 \Rightarrow$ an improved lower bound:

$$\nu_s = v(D) + \frac{1}{4}\xi_{r+1}\delta^2,$$

  where $0 < \xi_{r+1} \leq \cdots \leq \xi_n$ are the positive eigenvalues of $Q^* = Q + 2\mathrm{diag}(\lambda^*)$.

- Now, the question is how to compute

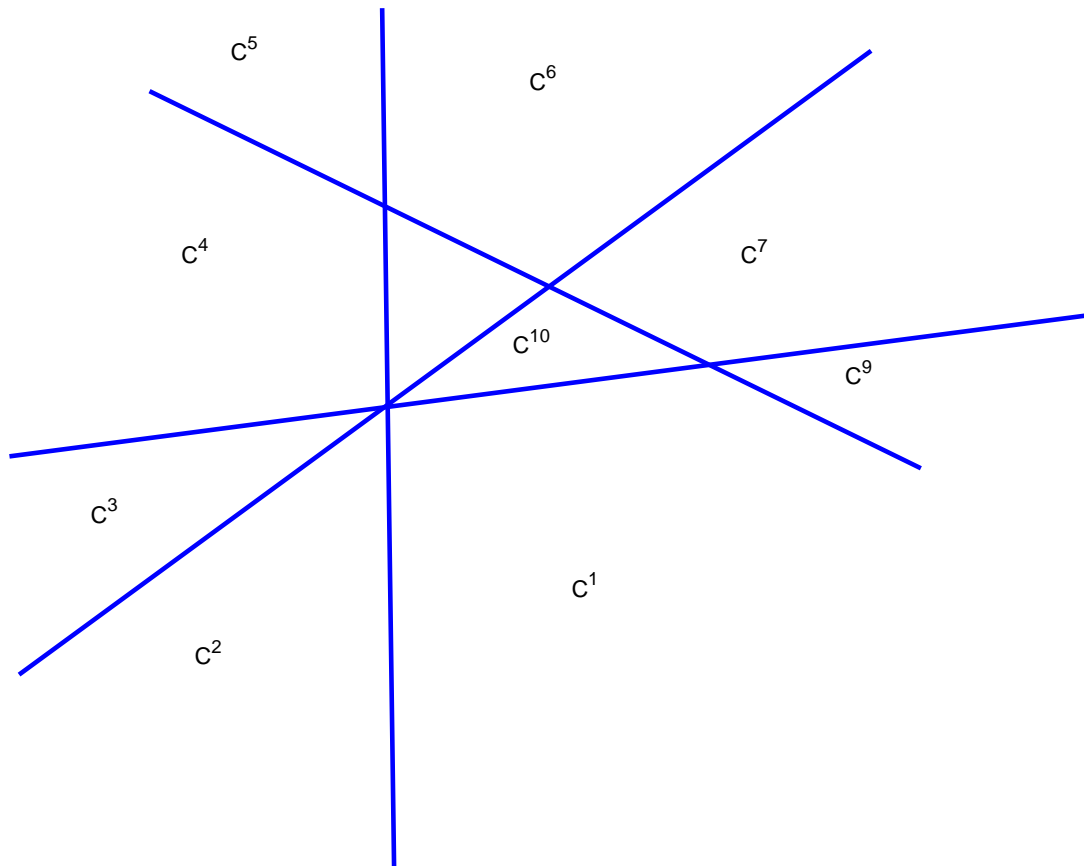$$\delta = \mathrm{dist}(\{-1, 1\}^n, C)?$$

Figure 3: Partition of C

Computation of distance δ

The idea is to partition the set C into certain subsets in each of which the signs of $x \in C$ are the same.

$\Downarrow$

The distance between $\{-1, 1\}^n$ and each subset must be achieved at $sign(\pi)$, where π is an interior point of the subset.

$\Downarrow$

Via enumerating all the subsets and finding an interior point in each of the subsets, we obtain δ.

- Let the solution in $\mathbb{R}^s$ of C be given as

$$x_j = x_j^0 + \sum_{i=1}^{s} G_{ij} z_i = 0, \ i = 1, \ldots, n. \qquad (5)$$

- Consider n hyperplanes in $\mathbb{R}^s$:

$$x_j^0 + \sum_{i=1}^{s} G_{ij} z_i = 0, \ i = 1, \ldots, n, \qquad (6)$$

which consist of an arrangement of hyperplanes (in a term of geometric computation). Finding all T_x's corresponds to finding all cells of the arrangement.

- It has been known that the number of cells of the hyperplane arrangement is $O(n^s)$.

- For fixed s, the distance δ can be computed in polynomial time. Another polynomially solvable case is $s = n - 1$.

- Solution method for cell enumeration: Reverse search algorithm (Avis and Fukuda (1996), Sleumer (1999))