1 2018-05-01

inner product $\text{Tr}(X^TS) = \text{Tr}(S^TX)$, Frobenius norm $\|X\|_F = \langle X, X\rangle^{\frac{1}{2}}$, trace
The trace is invariant under the similarity transform.
Defn of eigenvalues

Let $\mathbb{S}^n$ denote the subspace of n by n symmetric matrices (in $\mathbb{R}^{n\times n}$).

$$\mathbb{S}^n \simeq R^{\frac{n(n+1)}{2}}$$

We sort the real eigenvalues

$$\lambda_1(X) \geq \lambda_2(X) \geq \cdots \geq \lambda_n(X)$$

$\text{diag}(\mathbb{S}^n) \to \mathbb{R}^n$ is a linear transformation

Theorem (Spectral / Schur Decomposition Theorem). For every $X \in \mathbb{S}^n$, $\exists Q \in \mathbb{R}^{n\times n}$, orthogonal ($Q^TQ = I$), such that $X = Q\,\text{diag}(\lambda(X))Q^T$.
In the above spectral decomposition of X, the columns of Q are the eigenvectors of X.
(Note: vectors will be column vectors.)

e_j denotes the j-th unit vector.
Let

$$Q := [q^{(1)}q^{(2)}\cdots q^{(n)}]$$

$$\begin{aligned} Xq^{(j)} &= Q\,\text{diag}(\lambda(X)) \underbrace{Q^Tq(j)}_{=e_j,\text{ since } Q^TQ=I} \\ &= Q\underbrace{\text{diag}(\lambda(X))e_j}_{\lambda_j(X)e_j} \\ &= \lambda_j(X)\underbrace{Qe_j}_{=q^{(j)}} \end{aligned}$$

So

$$\|X\|_F = \left(\sum_{j=1}^n \lambda_j^2(X)\right)^{\frac{1}{2}} = \|\lambda(X)\|_2$$

We can extend p-norms to $\mathbb{S}^n$: for $X \in \mathbb{S}^n$,

$$\|X\|_p := \sup\{\|Xh\|_p : \|h\|_p = 1, h \in \mathbb{R}^n\}.$$

(Side remark: can also define p, q-norms.)
Note $\|X\|_2 = \max_{j\in\{1,2,\ldots,n\}}\{|\lambda_j(X)|\}$.

In the course, we'll mostly deal with symmetric positive semi-definite matrices and won't explicitly say they're symmetric.

Defn of square root: Let $X \in \mathbb{S}^n$ be positive definite. Every diagonal entry of D is positive.

$$X^{\frac{1}{2}} := QD^{\frac{1}{2}}Q^T \text{ (unique)}$$

Extend to positive semidefinite matrices.

Given $X \in \mathbb{S}^n$, if X is not PSD, then $\exists h \in \mathbb{R}^n$ such that $h^T X h < 0$.

Claim: If $X \in \mathbb{S}^n$ and p.s.d., then $x_{ii} = 0 \Rightarrow x_{ij} = 0 \forall j \in \{1, 2, \ldots, n\}$.

Proof. Let $X \in \mathbb{S}^n$, PSD, $x_{ii} = 0$. For the sake of reaching a contradiction, suppose $x_{ij} = \alpha \neq 0$.

$$X = \begin{bmatrix} 0 & \cdots & \alpha \\ \vdots & & \vdots \\ \alpha & \cdots & x_{jj} \end{bmatrix}$$

Consider $h := \beta e_i + e_j$, $\beta \in \mathbb{R}$ to be chosen later.

$$\begin{aligned} h^T X h &= (\beta e_i + e_j)^T (\beta X e_i + X e_j) \\ &= \beta^2 \cdot 0 + 2\alpha\beta + x_{jj} \longrightarrow \text{ can choose } \beta \text{ to make this negative} \end{aligned}$$

□

Theorem (Choleski Decomposition). Let $X \in \mathbb{S}^n$. Then X is PSD iff $\exists B \in \mathbb{R}^{n \times n}$, lower triangular ($B_{ij} = 0$, $\forall j > i$) such that $X = BB^T$.

Proof. Let $X \in \mathbb{S}^n$. We will prove the theorem by induction on n.
$n = 1$: X is PSD $\iff$ $X \in \mathbb{R}_+$. If X is PSD, $B = \sqrt{x_{11}}$ works. If X is not PSD, then $x_{11} < 0, h = 1$ works (i.e. $h^T X h < 0$).
Induction hypothesis: The claim holds for all $n \leq k - 1$.
$n = k$:
If $x_{11} < 0$, then $h = e_1$, $h^T X h = x_{11} < 0$.
If $x_{11} = 0$, if X is PSD, by the claim before the theorem, $x_{ij} = 0$, $\forall j$ and we are done by induction hypothesis. (For certificate of non-PSD, concatenate a 0).
So, we may assume $x_{11} > 0$.

$$\begin{aligned} X &:= \begin{bmatrix} x_{11} & x^T \\ x & \overline{X} \end{bmatrix} \\ b &:= \frac{1}{\sqrt{x_{11}} x} \\ \widetilde{X} &:= \overline{X} - \frac{1}{x_{11}} x x^T. \end{aligned}$$

If $\widetilde{X} \in S^{k-1}$ is PSD, then by induction hypothesis $\exists \widetilde{B} \in \mathbb{R}^{(k-1)\times(k-1)}$ lower triangular s.t. $\widetilde{X} = \widetilde{B}\widetilde{B}^T$. Then,

$$X = \begin{bmatrix} \sqrt{x_{11}} & 0 \\ b & \widetilde{B} \end{bmatrix} \begin{bmatrix} \sqrt{x_{11}} & b^T \\ 0 & \widetilde{B}^T \end{bmatrix}.$$

If $\widetilde{X}$ is not PSD then $\exists \widetilde{h} \in \mathbb{R}^{k-1}$ s.t. $\widetilde{h}^T \widetilde{X} \widetilde{h} < 0$.

$$h := \begin{bmatrix} -\frac{x^T \widetilde{h}}{x_{11}} \\ \widetilde{h} \end{bmatrix} \in \mathbb{R}^k$$

Then

$$\begin{aligned} h^T X h &= \frac{(x^T \widetilde{h})^2}{x_{11}} - 2\frac{x^T \widetilde{h}}{x_{11}} + \widetilde{h}^T \widetilde{X} \widetilde{h} + \frac{1}{x_{11}}(x^T \widetilde{h})^2 \\ &= \widetilde{h}^T \widetilde{X} \widetilde{h} < 0. \end{aligned}$$

□

2 2018-05-03

$$\begin{aligned} X &= \begin{bmatrix} \sqrt{x_{11}} & 0 \\ b & \widetilde{B} \end{bmatrix} \begin{bmatrix} \sqrt{x_{11}} & b^T \\ 0 & \widetilde{B}^T \end{bmatrix} \\ &= \begin{bmatrix} \sqrt{x_{11}} \\ b \end{bmatrix} \begin{bmatrix} \sqrt{x_{11}} & b^T \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \underbrace{\widetilde{X}}_{\widetilde{B}\widetilde{B}^T} \end{bmatrix} \end{aligned}$$

So, we also proved, for every $X \in \mathbb{S}^n_+$, $\exists h^{(1)}, h^{(2)}, \dots, h^{(n)} \in \mathbb{R}^n$ s.t.

$$X = h^{(1)} h^{(1)^T} + h^{(2)} h^{(2)^T} + \cdots + h^{(n)} h^{(n)^T}.$$

(Further, note the first $j-1$ entries of $h^{(j)}$ are zero.)

Proposition. Let $X \in \mathbb{S}^n$. Then TFAE:

(a) X is p.s.d.

(b) $\lambda(X) \geq 0$

(c) $\exists \mu \in \mathbb{R}^n_+$ and $h^{(1)}, h^{(2)}, \dots, h^{(n)} \in \mathbb{R}^n$ s.t. $X = \sum_{i=1}^n \mu_i h^{(i)} h^{(i)^T}$

(d) $\exists B \in \mathbb{R}^{n \times n}$ lower triangular s.t. $X = BB^T$

(e) $\forall J \subseteq \{1, 2, \dots, n\}$, $\det(X_J) \geq 0$ (where $X_J := [X_{ij} : i, j \in J]$)

(f) $\forall S \in \mathbb{S}^n_+$, $\langle X, S \rangle \geq 0$

Defn: $\mathbb{S}^n_{++} :=$ the set of positive definite matrices in $\mathbb{S}^n$

Proposition.

(1) $\mathbb{S}^n_{++} = \text{int}(\mathbb{S}^n_+)$

(2) Let $X \in \mathbb{S}^n$. Then TFAE:

  (a) X is positive definite
  (b) $\lambda(X) > 0$
  (c) $\exists \mu \in \mathbb{R}^n_{++}$ and $h^{(1)}, h^{(2)}, \ldots, h^{(n)} \in \mathbb{R}^n$ linearly independent s.t. $X = \sum_{i=1}^n \mu_i h^{(i)} {h^{(i)}}^T$
  (d) $\exists B \in \mathbb{R}^{n \times n}$ nonsingular, lower triangular s.t. $X = BB^T$
  (e) $\forall k \in \{1, 2, \ldots, n\}$, $\det(X_{J_k}) > 0$ (where $J_k := \{1, 2, \ldots, k\}$)
  (f) $\forall S \in \mathbb{S}^n_+ \setminus \{0\}$, $\langle X, S \rangle > 0$
  (g) $X \in \mathbb{S}^n_+$ and $\text{rank}(X) = n$

$X \in \mathbb{S}^n$ is diagonally dominant if $X_{ii} \geq \sum_{j=1, j \neq i}^n |X_{ij}|, \forall i \in \{1, 2, \ldots, n\}$
$X \in \mathbb{S}^n$ is strictly diagonally dominant if $X_{ii} > \sum_{j=1, j \neq i}^n |X_{ij}|, \forall i \in \{1, 2, \ldots, n\}$

If X is diagonally dominant then $X \in \mathbb{S}^n_+$ (converse is false).
If X is strictly diagonally dominant then $X \in \mathbb{S}^n_{++}$ (converse is false).
$\forall X \in \mathbb{S}^n, \exists \overline{\mu} \in \mathbb{R}$ s.t. $(X + \mu I) \in \mathbb{S}^n_+, \forall \mu \geq \overline{\mu}$
$\forall X \in \mathbb{S}^n, \exists \overline{\mu} \in \mathbb{R}$ s.t. $(X + \mu I) \in \mathbb{S}^n_{++}, \forall \mu > \overline{\mu}$
Note that $\forall X \in \mathbb{S}^n_+$, $\forall \varepsilon > 0$, $(X + \varepsilon I) \in \mathbb{S}^n_{++}$.

$K \subseteq \mathbb{R}^n$ is a convex cone if

(i) it is a cone $(\forall x \in K, \forall \alpha \in \mathbb{R}_+, \alpha x \in K)$, and

(ii) it is convex $(\forall u, v \in K, \forall \lambda \in [0, 1], \lambda u + (1 - \lambda) v \in K)$ [in the presence of (i), this is equivalent to $\forall u, v \in K, (u + v) \in K$]

A convex set is pointed if it does not contain any lines.

A pointed closed convex cone $K \subseteq \mathbb{R}^n$ with nonempty interior is

- self-dual if $\exists$ an inner-product on $\mathbb{R}^n$ such that

$$\underbrace{K^* := \{s \in \mathbb{R}^n : \langle x, s \rangle \geq 0, \forall x \in K\}}_{\text{dual cone of } K} = K$$

A pointed closed convex cone $K \subseteq \mathbb{R}^n$ with nonempty interior is homogeneous if $\forall u, v \in \text{int}(K), \exists L \in \text{Aut}(K)$ such that $Lu = v$, where

$$\text{Aut}(K) := \{L : \mathbb{R}^n \to \mathbb{R}^n, \text{ linear, nonsingular} : L(K) = K\}.$$

$\mathsf{Aut}(K)$: Automorphism group of K

A cone is called symmetric if it is homogeneous & self-dual.

Given a convex set $K \subseteq \mathbb{R}^n$, a ray of K is $R := \{\alpha\bar{x} : \alpha \geq 0\} \subseteq K$ for some $\bar{x} \in K \setminus \{0\}$.
A ray of K, R is an extreme ray of K if $\forall$ pairs of R_1, R_2 of K,

$$R_1 + R_2 \supseteq R \Rightarrow \text{ either } R_1 = R \text{ or } R_2 = R, \text{ or possibly both}$$

$R_1 + R_2 := \{r_1 + r_2 : r_1 \in R_1, r_2 \in R_2\}$ (Minkowski sum)
For $K_1 \in \mathbb{R}^{n_1}, K_2 \in \mathbb{R}^{n_2}$,

$$K_1 \oplus K_2 := \left\{ \begin{pmatrix} u \\ v \end{pmatrix} \in \mathbb{R}^{n_1} \oplus \mathbb{R}^{n_2} : u \in K_1, v \in K_2 \right\}$$

$\mathsf{ext}(K)$ denotes the set of normalized extreme rays of cone K
$\mathrm{Ext}(K)$ denotes the set of extreme rays of K

Theorem (1.16). $\mathbb{S}^n_+$ is a pointed, closed convex cone with nonempty interior. Moreover, $\mathbb{S}^n_+$ is homogeneous and self-dual (hence symmetric). The set of normalized extreme rays of $\mathbb{S}^n_+$ is given by $\mathsf{ext}(\mathbb{S}^n_+) = \{hh^T : h \in \mathbb{R}^n, \|h\|_2 = 1\}$.

$$\mathrm{Ext}(\mathbb{S}^n_+) = \{\{\alpha hh^T\} : \alpha \geq 0, hh^T \in \mathsf{ext}(\mathbb{S}^n_+)\}$$

Proof. Claim 1: $(\mathbb{S}^n_+)^* = \mathbb{S}^n_+$. (Recall $(\mathbb{S}^n_+)^* = \{S \in \mathbb{S}^n : \langle X, S\rangle \geq 0, \forall X \in \mathbb{S}^n_+\}$).
Proof: Let $\bar{S} \in \mathbb{S}^n$. Then $\bar{S}^{\frac{1}{2}}$ exists (and is unique),

$$\forall X \in \mathbb{S}^n_+, \langle X, \bar{S}\rangle = \mathrm{Tr}(X\bar{S}) = \mathrm{Tr}(\underbrace{\bar{S}^{\frac{1}{2}} X \bar{S}^{\frac{1}{2}}}_{\in \mathbb{S}^n_+}) \geq 0$$

Therefore $\bar{S} \in (\mathbb{S}^n_+)^*$. Hence, $(\mathbb{S}^n_+)^* \supseteq \mathbb{S}^n_+$.
Now, let $\widehat{S} \in (\mathbb{S}^n_+)^*$, let $h^{(1)}, h^{(2)}, \ldots, h^{(n)} \in \mathbb{R}^n$ be eigenvectors of $\widehat{S}$, then using Theorem 1.8, $\forall i \in \{1, 2, \ldots, n\}$,

$$\begin{aligned} \lambda_i(\widehat{S}) &= (h^{(i)})^T \widehat{S} h^{(i)} \\ &= \mathrm{Tr}((h^{(i)})^T \widehat{S} h^{(i)}) \\ &= \mathrm{Tr}(\widehat{S} \underbrace{h^{(i)} (h^{(i)})^T}_{\in \mathbb{S}^n_+}) \\ &\geq 0 \end{aligned}$$

because $\widehat{S} \in (\mathbb{S}^n_+)^*$ and $h^{(i)}(h^{(i)})^T \in \mathbb{S}^n_+$.
By Prop 1.10, $\widehat{S} \in \mathbb{S}^n_+$ (since $\lambda(\widehat{S}) \geq 0$). Thus, $(\mathbb{S}^n_+)^* \subseteq \mathbb{S}^n_+$. Therefore $(\mathbb{S}^n_+)^* =$

$\mathbb{S}^n_+$. ♢

Claim 2: $\mathbb{S}^n_+$ is a homogeneous cone.

Proof: Note that $\forall \overline{X} \in \mathbb{S}^n_{++}$, $T_{\overline{X}} : \mathbb{S}^n \to \mathbb{S}^n$, $T_{\overline{X}}(\cdot) := \overline{X}^{-\frac{1}{2}} \cdot \overline{X}^{-\frac{1}{2}}$, i.e. $\forall Z \in \mathbb{S}^n$, $T_{\overline{X}}(Z) = \overline{X}^{-\frac{1}{2}} Z \overline{X}^{-\frac{1}{2}}$.

Claim: $T_{\overline{X}} \in \text{Aut}(\mathbb{S}^n_+), \forall \overline{X} \in \mathbb{S}^n_{++}$. (Check!)

Note that $I \in \mathbb{S}^n_{++}$ and $\forall U \in \mathbb{S}^n_{++}$, $T_{\overline{U}}(U) = U^{-\frac{1}{2}} U U^{-\frac{1}{2}} = I$.

So, $\forall U, V \in \mathbb{S}^n_{++}$,

$$T_{\overline{V}^{-1}}(T_{\overline{U}}(\cdot)) \in \text{Aut}(\mathbb{S}^n_+)$$

and it maps U to V. The composition of automorphisms is again an automorphism.

$$[T_{\overline{V}^{-1}}(T_{\overline{U}}(Z)) = \overline{V}^{\frac{1}{2}} \overline{U}^{-\frac{1}{2}} Z \overline{U}^{-\frac{1}{2}} \overline{V}^{\frac{1}{2}}]$$

Therefore, $\mathbb{S}^n_+$ is homogeneous. ♢

Therefore, $\mathbb{S}^n_+$ is a symmetric cone. The rest of the claims are left as exercises. □

For a pair of matrices $U, V \in \mathbb{S}^n$, we write $U \succeq V$ to mean $(U - V) \in \mathbb{S}^n_+$ (Löwner (partial) order), and $U \succ V$ to mean $(U - V) \in \mathbb{S}^n_{++}$.

Note that any linear function $f : \mathbb{S}^n \to \mathbb{R}$ can be written as $f(X) = \langle A, X \rangle$ for some $A \in \mathbb{S}^n$. $A \in \mathbb{S}^n$ (otherwise we can take $(A + A^T)/2$).

So linear equations and inequalities on $\mathbb{S}^n$ are

$$\langle A_i, X \rangle = b_i, \langle A_i, X \rangle \leq b_i, \text{ etc. for } A_i \in \mathbb{S}^n, b_i \in \mathbb{R}.$$

Recall, a linear programming problem is a problem of optimizing (minimizing or maximizing) a linear function of finitely many real variables subject to finitely many linear equations and inequalities. Every LP can be put into the form

$$\begin{array}{rl} \min\limits_{x} & c^T x \\ \text{s.t.} & Ax \leq b \end{array}$$

$A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, c \in \mathbb{R}^n$ all given. A Semidefinite Programming Problem (SDP) is a problem of optimizing a linear function of finitely many matrix variables (real-valued entries) subject to finitely many linear equations and inequalities on these matrix variables and p.s.d.ness constraints on some of these matrix variables.

Every SDP can be put into the form

$$\begin{array}{rrcll} (P) \inf\limits_{X} & \langle C, X \rangle & & & \\ \text{s.t.} & \langle A_i, X \rangle & = & b_i & \forall i \in \{1, 2, \ldots, m\} \\ & X & \succeq & 0 & \end{array}$$

$C, A_1, A_2, \ldots, A_m \in \mathbb{S}^n, b \in \mathbb{R}^m$ are all given.
We define the dual SDP as

$$\begin{array}{lll} \sup_y & b^T y & \\ \text{s.t.} & \sum_{i=1}^m y_i A_i & \preceq & C \end{array}$$

or equivalently

$$\begin{array}{llll} (D) & \sup_y & b^T y & \\ & \text{s.t.} & \sum_{i=1}^m y_i A_i + S & = & C \\ & & S & \succeq & 0. \end{array}$$

Theorem (Weak Duality Relation). For every feasible solution $\overline{X}$ of (P) and for every feasible solution $(\overline{y}, \overline{S})$ of (D), we have

$$\langle C, \overline{X} \rangle - b^T \overline{y} = \langle \overline{X}, \overline{S} \rangle \geq 0.$$

Proof. Suppose $\overline{X}, (\overline{y}, \overline{S})$ are feasible in (P) and (D) respectively.
Define $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$ linear,

$$[\mathcal{A}(X)]_i := \langle A_i, X \rangle, \forall i \in \{1, 2, \ldots, m\}.$$

For every such linear map, its adjoint (another linear transformation) $\mathcal{A}^* : \mathbb{S}^n \to \mathbb{R}^m$ is defined by

$$\langle \mathcal{A}^*(y), X \rangle := [\mathcal{A}(X)]^T y, \ \forall X \in \mathbb{S}^n, \forall y \in \mathbb{R}^m.$$

For our choice of $\mathcal{A}$ above, $\mathcal{A}^*(y) = \sum_{i=1}^m y_i A_i$.

3 2018-05-10

Proof (cont). Let $\overline{X}, (\overline{y}, \overline{S})$ be feasible in (P) & (D) respectively. Then,

$$\begin{aligned} \langle C, \overline{X} \rangle - b^T \overline{y} &= \langle \mathcal{A}^*(\overline{y}) + \overline{S}, \overline{X} \rangle - b^T \overline{y} \\ &= \langle \overline{S}, \overline{X} \rangle + \langle \mathcal{A}^*(\overline{y}), \overline{X} \rangle - b^T \overline{y} \\ &= \langle \overline{S}, \overline{X} \rangle + \overline{y}^T \mathcal{A}(\overline{X}) - b^T \overline{y} \\ &= \langle \overline{S}, \overline{X} \rangle + \overline{y}^T b - b^T \overline{y} \\ &= \langle \overline{S}, \overline{X} \rangle + b^T \overline{y} - b^T \overline{y} \\ &= \langle \overline{S}, \overline{X} \rangle \\ &\geq 0 \text{ since } \overline{X}, \overline{S} \succeq 0. \end{aligned}$$

□

A corollary is: if for a pair of feasible $\overline{X}, (\overline{y}, \overline{S})$, $\langle C, \overline{X} \rangle = b^T \overline{y}$, then $\overline{X}, (\overline{y}, \overline{S})$ are optimal in (P) & (D).

(P) unbounded $\Rightarrow$ (D) is infeasible.
(D) unbounded $\Rightarrow$ (P) is infeasible.

Dual of (D) is equivalent to (P). Usually, we will assume $\mathcal{A}$ is surjective (equivalently, $A_1, A_2, \ldots, A_m$ are linearly independent). In this situation, every S satisfying linear equations of (D) determines a unique y. So, sometimes, when we talk about dual feasible solutions, we may refer to only y, or only S.

It is better to think about the constraint $X \in \mathbb{S}^n_+$ as

$$X \in \mathbb{S}^{n_1}_+ \oplus \mathbb{S}^{n_2}_+ \oplus \cdots \oplus \mathbb{S}^{n_r}_+.$$

I.e.

$$X = \begin{bmatrix} n_1 \times n_1 & 0 & 0 & \cdots & 0 \\ 0 & n_2 \times n_2 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & & \vdots \\ \vdots & \vdots & & & 0 \\ 0 & 0 & \cdots & 0 & n_r \times n_r \end{bmatrix}.$$

There are at least two ways to embed LPs as SDPs:

(1) Write linear constraints $X_{ij} = 0, \forall i \neq j$ together with $X \in \mathbb{S}^n_+$

(2) Write $X \in \mathbb{S}^n_+$ as $X \in \underbrace{\mathbb{S}^1_+ \oplus \cdots \oplus \mathbb{S}^1_+}_{n \text{ times}}$

Proposition (1.19, complementary slackness). Let $X, S \in \mathbb{S}^n_+$. Then,

$$\langle X, S \rangle = 0 \iff XS = 0.$$

Proof. ($\Leftarrow$) $XS = 0 \Rightarrow \underbrace{\mathrm{Tr}(XS)}_{=\langle X,S \rangle} = \mathrm{Tr}(0) = 0.$

($\Rightarrow$) Suppose $X, S \in \mathbb{S}^n_+$, $\langle X, S \rangle = 0$.
$0 = \mathrm{Tr}(XS) = \mathrm{Tr}(\underbrace{X^{\frac{1}{2}} S X^{\frac{1}{2}}}_{\succeq 0 \text{ since } S \succeq 0,\ X^{\frac{1}{2}} \in \mathbb{S}^n}) \geq 0$. By Prop 1.10, $\lambda(X^{\frac{1}{2}} S X^{\frac{1}{2}}) = 0$. By Thm 1.8 (spectral decomposition theorem), $0 = X^{\frac{1}{2}} S X^{\frac{1}{2}} = (X^{\frac{1}{2}} S^{\frac{1}{2}})(X^{\frac{1}{2}} S^{\frac{1}{2}})^T$. Therefore $X^{\frac{1}{2}} S^{\frac{1}{2}} = 0$, and thus $XS = X^{\frac{1}{2}}(X^{\frac{1}{2}} S^{\frac{1}{2}}) S^{\frac{1}{2}} = X^{\frac{1}{2}} 0 S^{\frac{1}{2}} = 0$.
Note that $XS = 0$ implies $XS \in \mathbb{S}^n$ and that $\exists Q \in \mathbb{R}^{n \times n}$ orthogonal s.t.

$$X = Q \, \mathrm{Diag}(\lambda(X)) Q^T, \ S = Q \, \mathrm{Diag}(\lambda(S)) Q^T.$$

□

Lemma (1.22 (Schur Complement)). Let $T \in \mathbb{S}^m_{++}, U \in \mathbb{R}^{n\times m}, X \in \mathbb{S}^n$.

$$M := \begin{bmatrix} T & U^T \\ U & X \end{bmatrix} \in S^{m+n}.$$

Then $M \succeq 0$ iff $X - UT^{-1}U^T \succeq 0$ and
$M \succ 0$ iff $X - UT^{-1}U^T \succ 0$.

Proof. Let T, U, X, M be as above. Note

$$\begin{aligned}
&\underbrace{\begin{bmatrix} I & 0 \\ UT^{-1} & I \end{bmatrix}}_{=:L} \begin{bmatrix} T & 0 \\ 0 & X - UT^{-1}U^T \end{bmatrix} \underbrace{\begin{bmatrix} I & T^{-1}U^T \\ 0 & I \end{bmatrix}}_{=L^T} \\
&= \begin{bmatrix} T & 0 \\ U & X - UT^{-1}U^T \end{bmatrix} \begin{bmatrix} I & T^{-1}U^T \\ 0 & I \end{bmatrix} \\
&= \begin{bmatrix} T & U^T \\ U & X \end{bmatrix} \\
&= M
\end{aligned}$$

$\det(L) = 1 \Rightarrow L$ is a linear isomorphism,

$$h^T L \begin{bmatrix} * & 0 \\ 0 & * \end{bmatrix} L^T h \geq 0, \forall h \in \mathbb{R}^{m\times n} \iff h^T \begin{bmatrix} * & 0 \\ 0 & * \end{bmatrix} h \geq 0, \forall h \in \mathbb{R}^{m\times n}.$$

Therefore $M \succeq 0$ iff $T \succeq 0$ and $X - UT^{-1}U^T \succeq 0$.
The argument for the second part is similar. □

This lemma shows how some nonlinear and nonconvex "looking" constraints may be included in SDPs exactly.
Suppose we have an optimization problem with vector variables $u^{(1)}, u^{(2)}, \ldots, u^{(n)} \in \mathbb{R}^n$. Further assume that the objective function and the constraints only involve linear or affine functions of $\langle u^{(i)}, u^{(j)} \rangle,\ i, j \in \{1, 2, \ldots, n\}$. E.g.

$$\begin{array}{lll}
\inf & \langle u^{(1)}, u^{(2)} \rangle - 5\langle u^{(2)}, u^{(2)} \rangle + 7\langle u^{(3)}, u^{(10)} \rangle + \cdots & \\
\text{s.t.} & \langle u^{(5)}, u^{(6)} \rangle + 2\langle u^{(1)}, u^{(8)} \rangle - 12\langle u^{(6)}, u^{(6)} \rangle & \leq \quad 10 \\
& \vdots &
\end{array}$$

Such problems are SDPs.

$$\begin{aligned}
U &:= \begin{bmatrix} u^{(1)} & u^{(2)} & \cdots & u^{(n)} \end{bmatrix} \in \mathbb{R}^{n\times n} \\
X &:= U^T U
\end{aligned}$$

Note $X_{ij} = \langle u^{(i)}, u^{(j)} \rangle, \forall i, j \in \{1, 2, \ldots, n\}$. We form the SDP

$$\begin{array}{lrcl}
\inf & X_{1,2} - 5X_{2,2} + 7X_{3,10} + \cdots & & \\
\text{s.t.} & X_{5,6} + 2X_{1,8} - 12X_{6,6} & \leq & 10 \\
& \vdots & & \\
& X & \succeq & 0
\end{array}$$

4 2018-05-15

4.1 Duality Theory

For any set $K \subset \mathbb{R}^d$, we can define the dual cone of K:

$$K^* := \{s \in \mathbb{R}^d : \langle x, s \rangle \geq 0 \forall x \in K\}.$$

Note that by definition, K^* is always a closed convex cone; $\forall K \subseteq \mathbb{R}^d$, K^{**} is the smallest closed convex cone in $\mathbb{R}^d$, containing K.

polar of K:

$$K^\circ := \{s \in \mathbb{R}^d : \langle x, s \rangle \leq 1 \forall x \in K\}$$

Note: K° is always a closed convex set.

For cones K, $K^\circ = \{s \in \mathbb{R}^d : \langle x, s \rangle \leq 0 \forall x \in K\} = -K^*$.
(If $\langle x, s \rangle \geq c > 0$ for $x \in K$, then $\langle \alpha x, s \rangle \geq \alpha \cdot c$ for all $\alpha > 0$; $\alpha x \in K$ for K a cone).

For any function $f : \mathbb{R}^d \to (-\infty, +\infty]$,
Legendre-Fenchel conjugate of f:

$$f_*(s) := \sup_{x \in \mathbb{R}^d} \{-\langle x, s \rangle - f(x)\}$$

epigraph of f:

$$\text{epi}(f) := \left\{ \begin{pmatrix} u \\ x \end{pmatrix} \in \mathbb{R} \oplus \mathbb{R}^d : f(x) \leq u \right\}.$$

$f(x)$ is a convex function $\iff$ $\text{epi}(f)$ is a convex set.

Why do we care about automorphisms?
– inequalities: multiplying by a positive factor to both sides preserves the inequality
– Löwner inequalities, operator inequalities: applying an automorphism to both sides preserves the inequality

Theorem (2.8, Hyperplane Separation Theorem). Let $G \subseteq \mathbb{R}^d$ be a nonempty closed convex set and $O \in \mathbb{R}^d \setminus G$. Then, $\exists a \in \mathbb{R}^d \setminus \{O\}$ and $\alpha \in \mathbb{R}_{++}$ such that

$$G \subset \{x \in \mathbb{R}^d : \langle a, x \rangle \geq \alpha\}.$$

Proof. Suppose G is nonempty, closed convex, $0 \notin G$. Since $G \neq \varnothing$, $\exists \overline{x} \in G$,

$$\begin{aligned} G_{\overline{x}} &:= \{x \in G : \|x\|_2 \leq \|\overline{x}\|_2\} \\ &= G \cap B(0, \|\overline{x}\|_2) \end{aligned}$$

Claim: $G_{\overline{x}}$ is nonempty and compact. $\inf\{\|x\|_2^2 : x \in G_{\overline{x}}\}$ is uniquely attained.
Proof of claim:

- $G_{\overline{x}}$ is nonempty, since $\overline{x} \in G_{\overline{x}}$
- $G_{\overline{x}}$ is closed, since $G_{\overline{x}} = G \cap B(0, \|\overline{x}\|_2)$
- $G_{\overline{x}}$ is bounded, since $G_{\overline{x}} \subseteq B(0, \|\overline{x}\|_2)$. Since $f(x) := \|x\|_2^2$ is continuous on $\mathbb{R}^d$, the infimum is attained. Since f is strictly convex, the minimizer is unique. $\diamondsuit$

Let $a \in \mathbb{R}^d$ be the unique minimizer. Since $a \in G,\ 0 \notin G,\ \alpha := \|a\|_2^2 > 0$.
$\forall x \in G,\ \forall \lambda \in (0,1], [\lambda x + (1-\lambda)a] \in G$ (since G is convex).
Since a is the minimum norm element of G, $\forall x \in G,\ \forall \lambda \in (0,1]$, $\|\lambda x + (1-\lambda)a\|_2^2 \geq \|a\|_2^2$.
$\forall x \in G,\ \forall \lambda \in (0,1]$,

$$\begin{aligned}
0 &\leq \|\lambda(a-x) - a\|_2^2 - \|a\|_2^n \\
&= \lambda^2\|a-x\|_2^2 + \|a\|_2^2 - 2\lambda\langle a, a-x\rangle - \|a\|_2^2 \\
&= \lambda^2\|a-x\|_2^2 - 2\lambda(\|a\|_2^2 - \langle a, x\rangle)
\end{aligned}$$

$$\iff \forall x \in G,\ \forall \lambda \in (0,1],$$
$$\langle a, x\rangle - \|a\|_2^2 \geq -\frac{\lambda}{2}\|a-x\|_2^2.$$

Taking limit of both sides as $\lambda \to 0^+$, we obtain $\langle a, x\rangle \geq \|a\|_2^2 = \alpha \forall x \in G$. $\square$

Corollary (2.9). Let $G_1, G_2 \subseteq \mathbb{R}^d$ be nonempty, disjoint, closed convex sets such that at least one of G_1, G_2 is bounded. Then, $\exists a \in \mathbb{R}^d \setminus \{0\}$ such that

$$\inf\{\langle a, x\rangle : x \in G_1\} > \sup\{\langle a, x\rangle : x \in G_2\}.$$

Proof sketch: Define $G := G_1 - G_2 = \{g_1 - g_2 : g_1 \in G_1,\ g_2 \in G_2\}$. G is nonempty, convex, $0 \notin G$ (since G_1, G_2 are disjoint), and prove that G is closed if at least one of G_1, G_2 is bounded. Then apply Thm 2.8 and translate back to the language of G_1, G_2. $\square$

Note that if both G_1, G_2 are unbounded, trouble may ensue.

Corollary (2.12). Let G_1, G_2 be nonempty convex sets that are disjoint. Then, $\exists a \in \mathbb{R}^d \setminus \{0\}$ such that

$$\inf\{\langle a, x\rangle : x \in G_1\} \geq \sup\{\langle a, x\rangle : x \in G_2\}.$$

2.14 A Strong Duality Theorem for SDP

$$\begin{array}{rlrcl} (P) & \inf & \langle C, X\rangle & & \\ & \text{s.t.} & \langle \mathcal{A}, X\rangle & = & b \\ & & X & \succeq & 0 \end{array}$$

$$\begin{array}{rlrcl} (D) & \sup & b^T y & & \\ & \text{s.t.} & \mathcal{A}^*(y) + S & = & C \\ & & S & \succeq & 0. \end{array}$$

A Slater point of $\{\mathcal{A}(X) = b,\ X \succeq 0\}$ is $\overline{X} \in \mathbb{S}^n$ such that $\mathcal{A}(\overline{X}) = b,\ \overline{X} \succ 0$.
A Slater point of $\{\mathcal{A}^*(y) + S = C,\ S \succeq 0\}$ is $(\overline{y}, \overline{S}) \in \mathbb{R}^n \oplus \mathbb{S}^n$ such that $\mathcal{A}^*(\overline{y}) + \overline{S} = C,\ \overline{S} \succ 0$.

5 2018-05-17

Suppose (D) has a Slater point and that the optimal objective value of (D) is bounded above. Then (P) has an optimal solution and the optimal objective values are the same.

Proof. Suppose $\exists(\overline{y}, \overline{S}) \in \mathbb{R}^m \oplus \mathbb{S}^n_{++}$ such that $\mathcal{A}^*(\overline{y}) + \overline{S} = C,\ \overline{S} \succ 0$.
Claim 1: We may assume $b \neq 0$.
Proof: Suppose $b = 0$. Then $\overline{X} := 0$ is feasible in (P) with objective value 0, $(\overline{y}, \overline{S})$ is feasible in (D) with objective value $b^T\overline{y} = 0^T y = 0$, thus by Corollary 1.18, $\overline{X}, (\overline{y}, \overline{S})$ are optimal in their respective problems. $\Diamond$
From now on, $b \neq 0$.
Suppose the objective function value of (D) is bounded from above on the feasible region of (D).
$\Longrightarrow$ (D) has a finite optimal value. Call it z^*.

$$G_1 := \{S \in \mathbb{S}^n : S = C - \mathcal{A}^*(y), \text{ for some } y \in \mathbb{R}^m \text{ s.t. } b^T y \geq z^*\}$$
$$G_2 := \mathbb{S}^n_{++}$$

Claim 2: $G_1 \neq \varnothing, G_2 \neq \varnothing;\ G_1, G_2$ are convex; $G_1 \cap G_2 = \varnothing$.
Proof: Consider

$$\begin{array}{rlrcl} (LP_1) & \max & b^T y & & \\ & \text{s.t.} & \mathcal{A}^*(y) + S & = & C \\ & & b^T y & \leq & z^*. \end{array}$$

This LP has feasible solutions (e.g. $(\overline{y}, \overline{S})$) so it is not unbounded. Therefore, by the Fundamental Theorem of LPs, it has an optimal solution. $\therefore G_1 \neq \varnothing$.
To prove $G_1 \cap G_2 = \varnothing$, suppose not (we are seeking a contradiction). Suppose $\exists \widetilde{S} \in \mathbb{S}^n_{++}$ s.t. $\widetilde{S} = C - \mathcal{A}^*(\widetilde{y})$, $b^T\widetilde{y} \geq z^*$ for some $\widetilde{y} \in \mathbb{R}^m$. For $\varepsilon > 0$, and small

enough, consider $\widehat{y}_\varepsilon := \widetilde{y} + \varepsilon b$. Note that for all $\varepsilon > 0$ and small enough, $\widehat{y}_\varepsilon$ is feasible in (D) and its objective value is:

$$\begin{aligned} b^T\widehat{y}_\varepsilon &= \underbrace{b^T\widetilde{y}}_{\geq z^*} + \underbrace{\varepsilon\|b\|_2^2}_{>0,\text{ by Claim 1}} \\ &> z^*, \end{aligned}$$

a contradiction to the definition of z^*.
The rest of the claim is left as an exercise. $\Diamond$
Using Claim 2, we apply Corollary 2.12 to the sets G_1, G_2. $\exists \widetilde{X} \in \mathbb{S}^n \setminus \{0\}$ such that

$$\inf\{\langle \widetilde{X}, S\rangle : S \in \mathbb{S}^n_{++}\} \geq \sup\{\langle \widetilde{X}, S\rangle : S \in G_1\}.$$

Since $G_2 \neq \varnothing$, the infimum is bounded below. Since $\mathbb{S}^n_{++}$ is a cone, the infimum is bounded below by zero. By taking a sequence $\{S^{(k)}\} \subset \mathbb{S}^n_{++}$ s.t. $S^{(k)} \to 0$, we see that the infimum is zero.
Since the infimum is zero,

$$\begin{aligned} &\langle \widetilde{X}, S\rangle \geq 0,\ \forall S \in \mathbb{S}^n_{++} \\ \implies &\langle \widetilde{X}, S\rangle \geq 0,\ \forall S \in \text{cl}(\mathbb{S}^n_{++}) = \mathbb{S}^n_{+} \\ \implies &\widetilde{X} \in \mathbb{S}^n_{+} \text{ (Prop. 1.10)}. \end{aligned}$$

Also,

$$\begin{aligned} &\langle \widetilde{X}, C - \mathcal{A}^*(y)\rangle \leq 0,\ \forall y \in \mathbb{R}^m \text{ s.t. } b^Ty \geq z^* \\ \iff &\langle C, \widetilde{X}\rangle \leq [\mathcal{A}(\widetilde{X})]^T y. \end{aligned}$$

<u>Claim 3:</u> $\exists \alpha \in \mathbb{R}_+$ such that $\mathcal{A}(\widetilde{X}) = \alpha b$.
<u>Proof:</u> Consider

$$(LP_2)\ \begin{array}{lll} \min & [\mathcal{A}(\widetilde{X})]^T y & \\ \text{s.t.} & b^T y \ \geq & z^*. \end{array} \qquad (LD_2)\ \begin{array}{llll} \max & \alpha z^* & & \\ \text{s.t.} & \alpha b & = & \mathcal{A}(\widetilde{X}) \\ & \alpha & \geq & 0. \end{array}$$

Since (LP_2) has feasible solutions and is not unbounded, it has an optimal solution. By the LP strong duality theorem, its dual (LD_2) has an optimal solution. $\Diamond$
<u>Case 1:</u> $\alpha = 0$. Then $\mathcal{A}(\widetilde{X}) = 0$, and $\forall y \in \mathbb{R}^m$ s.t. $b^Ty \geq z^*$, we have

$$\begin{aligned} 0 &= [\mathcal{A}(\widetilde{X})]^T y \\ &\geq \langle C, \widetilde{X}\rangle \\ &= \langle \overline{S} - \mathcal{A}^*(\overline{y}), \widetilde{X}\rangle \\ &= \underbrace{[\mathcal{A}(\widetilde{X})]^T\overline{y}}_{=0} + \underbrace{\langle \overline{S}, \widetilde{X}\rangle}_{>0,\text{ by Prop 1.11}} \\ &> 0, \end{aligned}$$

a contradiction.
Therefore, Case 1 never happens.
Case 2: $\alpha > 0$. Then $\widehat{X} := \frac{1}{\alpha}\widetilde{X} \in \mathbb{S}^n_+$ and $\mathcal{A}(\widehat{X}) = b$ (by Claim 3).
We have

$$\begin{aligned} &\langle C, \widehat{X}\rangle \le \underbrace{[\mathcal{A}(\widehat{X})]^T y}_{=b^T y}, \ \forall y \in \mathbb{R}^m \text{ s.t. } b^T y \ge z^* \\ \Longrightarrow\ &\langle C, \widehat{X}\rangle \le z^*. \end{aligned}$$

By the Weak Duality Relation, we conclude that $\widehat{X}$ is an optimal solution of (P), and the optimal objective values of (P) and (D) are the same. ☐

Ex:

$$n := 2, \quad m := 1, \quad C := \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad A := \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad b := 2$$

$$\begin{array}{lllll} (P) & \inf & \langle C, X\rangle = x_{11} & & \\ & \text{s.t.} & 2x_{21} & = & x_{21} + x_{12} = \langle A_1, X\rangle = 2 \\ & & X & \succeq & 0. \end{array} \equiv \begin{array}{ll} \inf & x_{11} \\ \text{s.t.} & \begin{bmatrix} x_{11} & 1 \\ 1 & x_{22} \end{bmatrix} \succeq 0. \end{array}$$

$$\begin{array}{lll} (D) \inf & 2y & \\ \text{s.t.} & \begin{bmatrix} 0 & y \\ y & 0 \end{bmatrix} \preceq & \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}. \end{array}$$

(D) is equivalent to

$$\begin{array}{ll} (D) \inf & 2y \\ \text{s.t.} & \begin{bmatrix} 1 & -y \\ -y & 0 \end{bmatrix} \succeq 0 \iff y = 0. \end{array}$$

Thus, $y = 0$ is the only feasible solution in (D); it is optimal with objective value zero.

$$X_\varepsilon := \begin{bmatrix} \varepsilon & 1 \\ 1 & \frac{1}{\varepsilon} \end{bmatrix}$$

is feasible in (P), $\forall \varepsilon > 0$.
Even though the optimal objective values are the same, (P) does not attain its optimal value.

6 2018-05-24

Ex: $n := 3, m := 2$,

$$C := \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A_1 := \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A_2 := \begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \quad b := \begin{bmatrix} 0 \\ 2\gamma \end{bmatrix},$$

$\gamma \in \mathbb{R}_+$, a parameter.

$$
\begin{array}{lll}
(P_\gamma) & \inf & 2x_{21} \\
& \text{s.t.} & \begin{bmatrix} x_{11} & 0 & x_{31} \\ 0 & 0 & 0 \\ x_{31} & 0 & \gamma \end{bmatrix} \succeq 0.
\end{array}
$$

Optimal objective value of (P_γ) is zero.
$\forall \gamma \in \mathbb{R}_+$,

$$
X_\gamma^* := \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \gamma \end{bmatrix}
$$

$$
\begin{array}{lll}
(D) & \sup & 2\gamma y_2 \\
& \text{s.t.} & \begin{bmatrix} 0 & 1+y_2 & 0 \\ 1+y_2 & -y_1 & 0 \\ 0 & 0 & -2y_2 \end{bmatrix} \succeq 0.
\end{array}
$$

$\forall$ feasible solutions of (D), $1+y_2 = 0 \iff y_2 = -1$.
The set of feasible solutions $= \left\{ \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} : y_1 \leq 0,\ y_2 = -1 \right\}$.
Optimal objective value of (D) is -2γ. There is a duality gap of 2γ.

6.1 Infeasibility/Unboundedness Certificates:

Recall from LP theory a "Theorem of the Alternative":
Let $A \in \mathbb{R}^{m\times n},\ c \in \mathbb{R}^n$. Then exactly one of the following systems has a solution:

(I) $A^\top y \leq c,\ y \in \mathbb{R}^m$,

(II) $Ad = 0,\ d \geq 0,\ c^\top d < 0,\ d \in \mathbb{R}^n$.

An exact generalization of this would have been:
Let $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m,\ C \in \mathbb{S}^n$. Then exactly one of the following systems has a solution:

(I) $\mathcal{A}^*(y) \preceq C$,

(II) $\mathcal{A}(D) = 0,\ D \succeq 0,\ \langle C, D\rangle < 0$.

- False, in general!

Ex: $n := 2,\ m := 1,\ C := \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},\ A_1 := \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},\ b := 1$.

$$
\begin{array}{lll}
(P) & \inf & 2x_{21} \\
& \text{s.t.} & \begin{bmatrix} 1 & x_{21} \\ x_{12} & x_{22} \end{bmatrix} \succeq 0.
\end{array}
\qquad
\begin{array}{lll}
(D) & \sup & y \\
& \text{s.t.} & \begin{bmatrix} -y & 1 \\ 1 & 0 \end{bmatrix} \succeq 0.
\end{array}
$$

(D) is infeasible.
However, it is almost feasible.
$X(t) := \begin{bmatrix} 1 & -t \\ -t & t^2 \end{bmatrix}$ is feasible for all $t \in \mathbb{R}$. The objective value of $X(t)$, $\langle C, X(t) \rangle = -2t \to -\infty$ as $t \to +\infty$. $\implies (P)$ is unbounded.

Feasible region of (P) can be represented as $x_{22} \geq x_{21}^2$.
However, $\nexists D \in \mathbb{S}^2_+$ s.t. $\mathcal{A}(D)$ [[TODO!]]

Defn: Given $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$, $C \in \mathbb{S}^n$, the system $\mathcal{A}^*(y) \preceq C$ is almost feasible if $\forall \varepsilon > 0$, $\exists C_\varepsilon \in \mathbb{S}^n$ such that $\|C - C_\varepsilon\|_F < \varepsilon$ and the system $\mathcal{A}^*(y) \preceq C_\varepsilon$ is feasible.

Theorem (2.22). Let $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$ linear, $C \in \mathbb{S}^n$. Then exactly one of the following holds:

(I) $\mathcal{A}(D) = 0,\ D \succeq 0,\ \langle C, D \rangle < 0.$

(II) $\mathcal{A}^*(y) \preceq C$ is almost feasible.

Proof. Suppose (I) holds. Wlog, we may assume $\exists D \in \mathbb{S}^n_+$ s.t. $\mathcal{A}(D) = 0, \langle C, D \rangle = -1$. For the sake of reaching a contradiction, suppose (II) also holds.
Then $\forall \varepsilon > 0$, $\exists C_\varepsilon \in \mathbb{S}^n$ s.t. $\mathcal{A}^*(y_\varepsilon) \preceq C_\varepsilon$, for some $y_\varepsilon \in \mathbb{R}^m \implies \mathcal{A}^*(y_\varepsilon) \preceq C + (C - \varepsilon - C)$.
Take inner product of both sides with D.

$$\begin{aligned} \implies 0 &= \underbrace{[\mathcal{A}(D)]}_{=0}^{\top} y_\varepsilon \\ &= \langle D, \mathcal{A}^*(y_\varepsilon) \rangle \\ &\leq \underbrace{\langle C, D \rangle}_{=-1} + \underbrace{\langle C_\varepsilon - C, D \rangle}_{\leq \|C - C_\varepsilon\|_F \|D\|_F < \varepsilon} \\ &< -\frac{1}{2} \quad \forall \varepsilon < \tfrac{1}{2\|D\|_F} \end{aligned}$$

We reached [[TODO!]]
Therefore, (I) holds $\implies$ (II) does not hold.
Now, suppose (I) does not hold. I.e., $\nexists D \succeq 0$ s.t. $\mathcal{A}(D) = 0,\ \langle C, D \rangle < 0$.
Consider

$$\begin{array}{lllll} (D) & \sup & \eta & & \\ & \text{s.t.} & \mathcal{A}^*(y) + \eta I & \succeq & C \\ & & \eta & \leq & 0 \end{array} .$$

Its dual is

$$\begin{array}{lllll} (P) & \inf & \langle C, X \rangle & & \\ & \text{s.t.} & \mathcal{A}(X) & = & 0 \\ & & \langle I, X \rangle & \leq & 1 \\ & & X & \succeq & 0. \end{array}$$

Since (D) has a Slater point ($\bar{y} := 0$, $\bar{\eta} := -(\|C\|_2 + 1)$) and its objective value is bounded above (by zero, recall the constraint "$\eta \leq 0$"), our Strong Duality Theorem applies. Note that (P) has a feasible solution $\overline{X} := 0$ with objective value zero. Since $\nexists D \succeq 0$ s.t. $\mathcal{A}(D) = 0$, $\langle C, D \rangle < 0$, zero is the optimal value of (P).
By the Strong Duality Theorem, the optimal objective value of (D) is also zero. If (D) attains this optimal value then $\mathcal{A}(y) \preceq C$ has a feasible solution.
Otherwise, $\exists$ a sequence $(y^{(k)}, \eta_k)$ of feasible solutions of (D) such that $\eta_k \to 0^-$ as $k \to +\infty$.

$$\mathcal{A}^*(y^{(k)}) \preceq C - \eta_k I$$

For every $\varepsilon > 0$, choosing k large enough we extract $C_\varepsilon := C - \eta_k I$ and verify that $\mathcal{A}^*(y) \preceq C$ is almost feasible. $\square$

6.2 Some Geometry for the Cone $\mathbb{S}^n_+$

Let $K \subseteq \mathbb{R}^n$ be a closed convex cone. A convex cone $F \subseteq K$ is a face of K if $\forall u, v \in K$ s.t. $(u + v) \in F$, we have $u, v \in F$.

A face F of K is exposed if $\exists a \in \mathbb{R}^n$ such that

$$K \subseteq \{x \in \mathbb{R}^n : \langle a, x \rangle \geq 0\} \text{ and } F = \{x \in K : \langle a, x \rangle = 0\}.$$

A face F of K is called proper if $F \neq \varnothing$, and $F \neq K$.

K is called facially exposed if every proper face of K is exposed.

$\mathbb{S}^n_+$ is facially exposed, but in general feasible regions of SDPs are not facially exposed.

7 2018-05-29

Faces of convex cones

Exposed faces

Facially exposed cones

If F is a face of K then we write $F \trianglelefteq K$. This relation is transitive: $F_1 \trianglelefteq F_2$ and $F_2 \trianglelefteq K \implies F_1 \trianglelefteq K$.

We will use the notion of relative interior of a set.
For $G \subseteq \mathbb{R}^n$, affine hull of G is the smallest affine space which contains G.

$$\text{affine. hull}(G) := \{\sum_{i=1}^{n} \lambda_i v^i : v^1, v^2, \ldots, v^n \in G, \sum_{i=1}^{n} \lambda_i = 1\}$$

relative interior of G is the interior of G with respect to affine hull of G. We denote $\text{relint}(G)$.

Theorem (2.25).

(a) Every nonempty face G of $\mathbb{S}^n_+$ is uniquely characterized by a linear subspace $L \subseteq \mathbb{R}^n$ such that

$$\begin{aligned} G &= \{X \in \mathbb{S}^n_+ : \text{Null}(X) \supseteq L\} \\ \text{relint}(G) &= \{X \in \mathbb{S}^n_+ : \text{Null}(X) = L\} \end{aligned}$$

($\text{Null}(X)$ = null space / kernel of X).

(b) $\mathbb{S}^n_+$ is facially exposed

(c) Every proper face of $\mathbb{S}^n_+$ is projectionally exposed, in particular, $G = (I - Q)\mathbb{S}^n_+(I - Q)$, where Q is the orthogonal projection onto the linear subspace L defining G via part (a).

As a consequence of this theorem, we see that every proper face of $\mathbb{S}^n_+$ is isomorphic to $\mathbb{S}^k_+$ for some $k < n$:

$$G = \left\{ Q \begin{bmatrix} X & 0 \\ 0 & 0 \end{bmatrix} Q^\top : X \in \mathbb{S}^k_+ \right\}$$

for some $Q \in \mathbb{R}^{n \times n}$ orthogonal.

7.1 Back to Duality Theory

If Slater condition holds, then our Strong Duality Theorem applies. What do we do if it fails but (P) is feasible?

7.2 Borwein and Wolkowicz Approach

Restrict the problem to the minimal face of $\mathbb{S}^n_+$ which contains the feasible region. When restricted to the minimal face, we have Slater condition.

A key lemma in this approach is

Lemma (2.27). Suppose (P) is feasible. Then exactly one of the following holds:

(I) $\mathcal{A}(X) = b,\ X \in \mathbb{S}^n_{++}$

(II) $\exists y \in \mathbb{R}^m$ s.t. $\mathcal{A}^*(y) \in \mathbb{S}^n_+ \setminus \{0\},\ b^\top y = 0$

Note that if (II) has a solution $\overline{y} \in \mathbb{R}^m$, then for every $\overline{X} \in \{X \in \mathbb{S}^n_+ : \mathcal{A}(X) = b\}$, we have $\underbrace{\overline{y}^\top \mathcal{A}(\overline{X})}_{=\langle \mathcal{A}^*(y\overline{y}), \overline{X}\rangle} = \overline{y}^\top b = 0.$

So, every solution $\overline{y}$ of system (II) gives us a linear equation $\langle \mathcal{A}^*(y\overline{y}), X\rangle = 0$ that can be added to the constraints in (P).

In LP, adding redundant constraints to (P) does not lead to same kind of consequences as in SDP.

Ex: $n := 3$, $m := 2$, $C := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $A_1 := \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $A_2 := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$, $b := \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

$$(P)\ \inf\ x_{11} \quad \text{s.t.} \quad \begin{bmatrix} 1 & 0 & x_{21} \\ 0 & 0 & 0 \\ x_{21} & 0 & x_{33} \end{bmatrix} \succeq 0$$

$$(D)\ \inf\ y_2 \quad \text{s.t.} \quad \begin{bmatrix} 1-y_2 & 0 & 0 \\ 0 & -y_1 & -y_2 \\ 0 & -y_2 & 0 \end{bmatrix} \succeq 0.$$

For every feasible solution $y_2 = 0$, optimal value $= 0$.

Adding the redundant linear equation $\langle A_3, X\rangle = 0$ for $A_3 := \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ to (P) does not change the feasible region of (P) or the set of optimal solutions, but (D) becomes

$$(D)\ \inf\ y_2 \quad \text{s.t.} \quad \begin{bmatrix} 1-y_2 & 0 & 0 \\ 0 & -y_1 & -y_2-y_3 \\ 0 & -y_2-y_3 & 0 \end{bmatrix} \succeq 0.$$

This dual has no duality gap $(y_1^* = 0,\ y_2^* = 1,\ y_3^* = -1)$.

7.3 Ramana's Extended Lagrange Slater Dual (ELSD)

Our main problem of interest is

$$(D)\ \sup\ b^\top y \quad \text{s.t.} \quad \mathcal{A}^*(y) \preceq 0.$$

$$\begin{aligned} (ELSD)\ \inf\ & \langle C, U+W\rangle \\ \text{s.t.}\ & \mathcal{A}(U+W) = b \\ & W \in \underbrace{\mathcal{W}_n}_{\text{linear subspace}}, \\ & U \succeq 0. \end{aligned}$$

$$\begin{aligned}\mathcal{C}_k := \{(U_1, W_1, U_2, W_2, \ldots, U_k, W_k) : W_0 :&= 0,\\ \mathcal{A}(U_i + W_{i-1} + W_{i-1}^\top) &= 0,\\ \langle C, U_i + W_{i-1} + W_{i-1}^\top \rangle &= 0,\\ U_i \succeq W_i W_i^\top, &\forall i \in \{1, 2, \ldots, k\}\}\end{aligned}$$

Note that $U_i \succeq W_i W_i^\top \iff \begin{bmatrix} I & W_i^\top \\ W_i & U \end{bmatrix} \succeq 0.$

$$\mathcal{W}_k := \{W_k + W_k^\top : (U_1, W_1, \ldots, U_k, W_k) \in \mathcal{C}_k, \text{ for some } (U_1, W_1, \ldots, W_{k-1}, U_k)\}$$

Theorem (2.28). If (D) has a finite optimal value, then $(ELSD)$ has an optimal solution, and the optimal values of (D) & $(ELSD)$ coincide.

Theorem (2.29). Given $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$ linear, $C \in \mathbb{S}^n$, exactly one of the following has a solution:

(I) $\mathcal{A}^*(y) \preceq C$,

(II) $\mathcal{A}(U + W) = 0,\ U \succeq 0,\ W \in \mathcal{W}_n,\ \langle C, U + W \rangle = -1.$

8 2018-05-31

Assigned reading: finish reading Chapter 2
We proved

- a Strong Duality Theorem

have seen how to remove the Slater point assumption (at least in theory)

- by Borwein-Wolkowicz approach (a.k.a Facial Reduction)
- or by Ramana's Extended Lagrange-Slater Dual

(the two are closely related)

In the majority of the applications, SDP problems arise as a relaxation of a typically nonconvex optimization problem.

8.1 When does the Slater Condition hold in SDP relaxations?

Let $F \subset \mathbb{R}^n$ denote (nonconvex) set of feasible solutions. Our application problem is

$$\begin{array}{ll} \inf & c^\top x \\ & x \in F \end{array} \quad \text{or} \quad \begin{array}{ll} \inf & c^\top x + x^\top C x \\ & x \in F, \end{array}$$

where $c \in \mathbb{R}^n,\ C \in \mathbb{S}^n$ are given.

8.2 Homogeneous Equality Form

If $\exists$ a linear transformation $\mathcal{A} : \mathbb{S}^{n+1} \to \mathbb{R}^m$ such that

$$F = \{x \in \mathbb{R}^n : \mathcal{A} \underbrace{\begin{pmatrix} 1 & x^\top \\ x & xx^\top \end{pmatrix}}_{\begin{pmatrix} 1 \\ x \end{pmatrix} \begin{pmatrix} x & x^\top \end{pmatrix}} = 0\}$$

we say that $\mathcal{A}$ is a homogeneous equality form representation of F.

Any finite system of multivariate quadratic equations (their solution set) can be expressed in this form.
For $i \in \{1, 2, \ldots, m\}$, let $Q^{(i)} \in \mathbb{S}^n,\ q^{(i)} \in \mathbb{R}^n,\ \gamma_i \in \mathbb{R}$ be given such that our system is:

$$\left\langle \begin{bmatrix} \gamma & {q^{(i)}}^\top \\ q^{(i)} & Q^{(i)} \end{bmatrix} \begin{bmatrix} 1 & x^\top \\ x & xx^\top \end{bmatrix} \right\rangle = x^\top Q^{(i)} x + 2{q^{(i)}}^\top x + \gamma_i = 0,\ \forall i \in \{1, 2, \ldots, m\}.$$

We can also handle quadratic inequalities:

$$x^\top Q^{(i)} x + 2{q^{(i)}}^\top x + \gamma_i \leq 0 \iff \underbrace{x^\top Q^{(i)} x + 2{q^{(i)}}^\top x + \gamma_i + s_i^2}_{= \left\langle \begin{bmatrix} \gamma_i & {q^{(i)}}^\top & 0 \\ {q^{(i)}}^\top & q^{(i)} & 0 \\ 0 & 0 & 1 \end{bmatrix} \underbrace{\begin{bmatrix} 1 & x^\top & s_i \\ x & xx^\top & s_i x \\ s_i & s_i x^\top & s_i^2 \end{bmatrix}}_{= \begin{bmatrix} 1 \\ x \\ s_i \end{bmatrix} \begin{bmatrix} 1 & x^\top & s_i \end{bmatrix}} \right\rangle} = 0$$

Consider a multivariate polynomial inequality:

$$x_1^6 x_2^4 + x_2^3 + x_1^2 x_3^2 + x_1 - 7 \leq 0$$

$$\begin{aligned} x_4 &= x_1^2 \\ x_5 &= x_4^2 \\ x_6 &= x_5^2 \\ x_7 &= x_2^2 \\ x_8 &= x_7^2 \\ x_9 &= x_3^2 \\ x_6 x_8 + x_2 x_7 + x_4 x_9 + x_1 - 7 &= 0 \end{aligned}$$

Note that the solution set of the quadratic system projected onto the first three coordinates is the solution set of the original inequality.

Proposition (2.32). Any finite system of multivariate polynomial equations and inequalities can be put into Homogeneous Equality Form.

$$\widehat{\mathcal{P}} := \left\{ \begin{pmatrix} 1 & x^\top \\ x & X \end{pmatrix} \in \mathbb{S}^{n+1}_+ : \mathcal{A}\begin{pmatrix} 1 & x \\ x & X \end{pmatrix} = 0 \right\}$$
$$\mathcal{F} := \operatorname{conv}\left\{ \begin{pmatrix} 1 & x^\top \\ x & xx^\top \end{pmatrix} : x \in F \right\} \subseteq \widehat{\mathcal{P}}$$

So, $\widehat{\mathcal{P}}$ is an SDP relaxation of F.

Theorem (2.33). Suppose $F \subset \mathbb{R}^n$ is such that $\dim(\operatorname{conv}(F)) = n$. Then $\widehat{\mathcal{P}}$ has Slater points.

Proof. Suppose $\operatorname{conv}(F)$ is full dimensional. Then, $\exists v^{(1)}, v^{(2)}, \ldots, v^{(n+1)} \in F$ such that $v^{(1)}, v^{(2)}, \ldots, v^{(n+1)}$ is affinely independent (equivalently, $\begin{pmatrix} 1 \\ v^{(1)} \end{pmatrix}, \begin{pmatrix} 1 \\ v^{(2)} \end{pmatrix}, \ldots \begin{pmatrix} 1 \\ v^{(n+1)} \end{pmatrix} \in \mathbb{R}^{n+1}$ are linearly independent.)
Then for every $\lambda \in \mathbb{R}^{n+1}_{++}$ such that $\bar{e}^\top \lambda = 1$, we have

$$V_\lambda := \sum_{i=1}^{n+1} \lambda_i \begin{pmatrix} 1 \\ v^{(i)} \end{pmatrix} \begin{pmatrix} 1 & {v^{(i)}}^\top \end{pmatrix} \in \mathcal{F} \subseteq \widehat{\mathcal{P}};$$

moreover, by Prop 1.11, $V_\lambda \succ 0$.
Therefore, $\widehat{\mathcal{P}}$ has Slater points. □

If the $\dim(\operatorname{conv}(F)) =: d < n$, but we know d, we can construct an SDP relaxation that has Slater points.
Suppose we know the d-dimensional affine subspace that contains F. That is, we are given $\ell \in \mathbb{R}^n$, $L \in \mathbb{R}^{d \times n}$ such that $x \in F \implies x = \ell + L^\top y$ for some $y \in \mathbb{R}^d$.
Define $\mathcal{L} : \mathbb{S}^{n+1} \to \mathbb{S}^{d+1}$

$$\mathcal{L}(Z) := \begin{pmatrix} 1 & \ell^\top \\ 0 & L \end{pmatrix} Z \begin{pmatrix} 1 & 0 \\ \ell & L^\top \end{pmatrix}.$$

Its adjoint is $\mathcal{L}^* : \mathbb{S}^{d+1} \to \mathbb{S}^{n+1}$

$$\mathcal{L}^*(W) = \begin{pmatrix} 1 & 0 \\ \ell & L^\top \end{pmatrix} W \begin{pmatrix} 1 & \ell^\top \\ 0 & L \end{pmatrix}.$$

$\overline{\mathcal{A}} : \mathbb{S}^{d+1} \to \mathbb{R}^m$,

$$\overline{\mathcal{A}}(W) := \mathcal{A}(\mathcal{L}^*(W))$$

$$F_{\mathcal{L}} := \left\{ y \in \mathbb{R}^d : \overline{\mathcal{A}}\begin{pmatrix} 1 & y^\top \\ y & yy^\top \end{pmatrix} = 0 \right\}, \widehat{\mathcal{P}}_{\mathcal{L}} := \left\{ \begin{pmatrix} 1 & y^\top \\ y & Y \end{pmatrix} \in \mathbb{S}^{d+1} : \overline{\mathcal{A}}\begin{pmatrix} 1 & y^\top \\ y & Y \end{pmatrix} = 0 \right\}.$$

Theorem (2.34). Slater condition holds for $\widehat{\mathcal{P}}_{\mathcal{L}}$.
SDP relaxation for a set of given $c \in \mathbb{R}^n,\ C \in \mathbb{S}^n$ is

$$\begin{array}{rll} \inf & \left\langle \mathcal{L}\begin{pmatrix} 0 & c^\top \\ c & C \end{pmatrix}, \begin{pmatrix} 1 & y^\top \\ y & Y \end{pmatrix} \right\rangle & \\ \text{s.t.} & \overline{\mathcal{A}}\begin{pmatrix} 1 & y^\top \\ y & Y \end{pmatrix} & = 0, \\ & \begin{pmatrix} 1 & y^\top \\ y & Y \end{pmatrix} & \in \mathbb{S}^{d+1}_+ \end{array}$$

8.3 Nonhomogeneous Equality Form

Suppose $\underbrace{\mathcal{A}}_{\text{linear}} : \mathbb{S}^n \to \mathbb{R}^m,\ b \in \mathbb{R}^m$ are given such that

$$F = \{x \in \mathbb{R}^n : \mathcal{A}(xx^\top) = b\}$$
$$\widehat{\mathcal{P}} = \{X \in \mathbb{S}^n_+ : \mathcal{A}(X) = b\}$$

Theorem (2.35). Suppose there exists a linearly independent set of vectors

$$\{v^{(1)}, v^{(2)}, \ldots, v^{(n)}\} \subseteq F.$$

Then $\widehat{\mathcal{P}}$ has Slater points.

9 2018-06-05

9.1 Ellipsoid Method

Given $c \in \mathbb{R}^d$ (defining the center) and $A \in \mathbb{S}^d_{++}$ (determining the shape and size) the set $E := \{x \in \mathbb{R}^d : (x-c)^\top A^{-1}(x-c) \leq 1\}$ defines an ellipsoid, and every full-dimensional ellipsoid can be expressed this way.

Note that every ellipsoid is an affine image of a Euclidean ball

$$E = c + A^{\frac{1}{2}} B_d(0,1)$$
$$\implies \operatorname{vol}(E) = \sqrt{\det(A)}\operatorname{vol}(B_d(0,1))$$

Theorem (3.1). For every compact convex set $G \subset \mathbb{R}^d$ with nonempty interior, there exists a unique minimum volume ellipsoid E which contains G. Moreover, shrinking E around its centre by a factor of d results in an ellipsoid contained in G.

The ellipsoid E in the theorem is called Löwner-John ellipsoid. Suppose $c \in \mathbb{R}^d$ is the centre of the Löwner-John ellipsoid. Translate both sets (E, G) such that c is the origin.

$$\frac{1}{d}(E-c) \subseteq (G-c) \subseteq (E-c)$$

The factor d is tight, take simplex in $\mathbb{R}^d$ as G.

Let's discuss some ingredients for a proof of this theorem.

$$(x-c)^\top A(x-c) \le 1 \quad \forall x \in G$$

$$A \succ 0,\ c \in \mathbb{R}^d$$

Objection function: minimize the volume of E.

$$\text{vol}(E) = [\det(A)]^{-\frac{1}{2}} \text{vol}(B_d(0,1))$$

$\ln(\cdot)$ is monotone on $\mathbb{R}_{++}$

$$\begin{array}{lllll} (P_{\bar{A}}) & \inf & -\ln\det(A) & & \\ & \text{s.t.} & (x-c)^\top A(x-c) & \le & 1 \quad \forall x \in G \\ & & A & \succ & 0,\ c \in \mathbb{R}^d \end{array}$$

$$\begin{array}{lllll} (P_A) & \inf & -\ln\det(A) \longleftarrow \text{ strictly convex on } \mathbb{S}^d_{++} & & \\ & \text{s.t.} & \underbrace{\begin{bmatrix} 1 & x^\top \end{bmatrix} \begin{bmatrix} \alpha & a^\top \\ a & A \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix}}_{=\text{Tr}\left(\left(\begin{bmatrix} 1 \\ x \end{bmatrix}\begin{bmatrix} 1 & x^\top \end{bmatrix}\right)\begin{bmatrix} \alpha & a^\top \\ a & A \end{bmatrix}\right)} & \le & 1 \quad \forall x \in G \\ & & \begin{bmatrix} \alpha & a^\top \\ a & A \end{bmatrix} & \succeq & 0,\ c \in \mathbb{R}^d \end{array}$$

Given an optimal solution $(\bar{A}, c)$ of $(P_{\bar{A}})$, we can construct a feasible solution (α, a, A) of (P_A) with the same objective value.

$$\begin{aligned} A &:= \bar{A}, \\ a &:= -\bar{A}c, \\ \alpha &:= c^\top \bar{A} c. \end{aligned}$$

Consider a problem of computing a minimizer of a convex function $f : \mathbb{R} \to \mathbb{R}$. We are given that a minimizer of f lies in an interval $[a, b] \subset \mathbb{R}$. We have access to an oracle for f which takes as input $\bar{x} \in [a, b]$ and outputs one of the following:

- $\bar{x}$ is a minimizer
- minimizer lies in $\{x : x > \bar{x}\}$
- minimizer lies in $\{x : x < \bar{x}\}$

We will use an Information Complexity approach to prove that bisection is an optimal algorithm for this problem. An algorithm that deviates from bisection can be fooled by a convex function which is constructed after the interaction with the algorithm takes place.

If an algorithm uses $\bar{x}$ to the left of the midpoint of the current interval, the oracle responds "the minimizer is in $\{x \in \mathbb{R} : x > \bar{x}\}$". If it uses $\bar{x}$ to the right of the midpoint of the current interval, the oracle responds "the minimizer is in $\{x : x < \bar{x}\}$". At the end, $\exists f : \mathbb{R} \to \mathbb{R}$ strictly convex which is consistent with these answers.

One can view the Ellipsoid Method as a generalization of bisection to $\mathbb{R}^d$. First, we will start with the problem of "Given a separation oracle for G and an ellipsoid $E \supseteq G$, find a point in G". $G \subset \mathbb{R}^d$ is a compact convex set.

$$\tilde{E} = \{x \in E : \langle a, x\rangle \leq \langle a, c\rangle\}$$

10 2018-06-07

10.1 Ellipsoid Algorithm I (Convex feasibility)

"Input" access to a weak separation oracle for a closed convex set $G \subseteq \mathbb{R}^d$, $E_0 := E(A_0, c^{(0)})$ such that $E_0 \cap G \neq \varnothing$, $\varepsilon > 0$

$k := 0$,

Step 1. Ask the oracle "is $c^{(k)} \in G$?" If "YES", stop $\overline{x} := c^{(k)} \in G$. If "NO" and $\mathrm{vol}(E_k) < \varepsilon$, STOP and report $\mathrm{vol}(E_k)$

Step 2. Oracle returns $a \in \mathbb{R}^d \setminus \{0\}$ s.t. $\widetilde{E} := \{x \in E_k : \langle a, x\rangle \leq \langle a, c^{(k)}\rangle\} \supseteq G \cap E_k$

Step 3.

$$c^{(k)} := c^{(k)} - \frac{1}{(d+1)\sqrt{a^\top A_k a}} A_k a$$
$$A_{k+1} := \frac{d^2}{d^2-1}\left[A_k - \frac{2}{(d+1)a^\top A_k a} A_k a a^\top A_k\right]$$

$E_{k+1} := E(A_{k+1}, c^{(k+1)})$
$k := k+1$
Go to Step 1.

Theorem (3.4). For every $k \in \mathbb{Z}_{++}$, we have $\widetilde{E} \subseteq E_{k+1}$ and $\ln\left(\frac{\mathrm{vol}(E_{k+1})}{\mathrm{vol}(E_k)}\right) \leq -\frac{1}{2d}$.

After k iterations,

$$\ln\left(\frac{\text{vol}(E_{k+1})}{\text{vol}(E_k)}\right) = \ln\left(\frac{\text{vol}(E_k)}{\text{vol}(E_0)}\frac{\text{vol}(E_k)}{\text{vol}(E_{k-1})}\frac{\text{vol}(E_{k-1})}{\text{vol}(E_{k-2})}\cdots\frac{\text{vol}(E_1)}{\text{vol}(E_0)}\right)$$
$$= \sum_{j=0}^{k-1}\ln\left(\frac{\text{vol}(E_1)}{\text{vol}(E_0)}\right)$$
$$\leq -\frac{k}{2d} \text{ by Thm 3.4}$$

Suppose $\text{vol}(E_0) =: R$. If $k \geq 4d\ln(\frac{R}{\varepsilon})$ then we know that $\text{vol}(E_k) < \varepsilon$. We have

$$\ln\left(\frac{\text{vol}(E_1)}{\text{vol}(E_0)}\right) \leq -2\ln(\frac{R}{\varepsilon})$$
$$\Longrightarrow \leq -2\ln R + 2\ln\varepsilon + \ln R < \ln\varepsilon$$

We are assuming $R > 1,\ \varepsilon \in (0,1)$

Theorem (3.5). Let $G \subseteq \mathbb{R}^d$ be a closed convex set. Ellipsoid $E_0 := E(A_0, c^{(0)})$ of volume $R > 1$ be given such that $G \cap E_0 \neq \varnothing$, and suppose we have access to a separation oracle for G. Then for every $\varepsilon \in (0,1)$ in $O(d\ln(\frac{R}{\varepsilon}))$ iterations, either the algorithm returns $\bar{x} \in G \cap E_0$ or proves that $\text{vol}(G \cap E_0) < \varepsilon$.

If we are interested in solving

$$\begin{array}{ll} \inf & f(x) \\ \text{s.t.} & x \in G, \end{array}$$

where $f : \mathbb{R}^d \to \mathbb{R}$ a convex function.
Introduce a new parameter $t \in \mathbb{R}$ and consider

$$\begin{array}{lrcl} \inf & 0 & & \\ \text{s.t.} & x & \in & G, \\ & f(x) & \leq & t \end{array}$$

We have a convex feasibility problem on $\widetilde{G}_t := \{x \in G : f(x) \leq t\}$ and we can do bisection on t.
Another way to deal with this problem is to have access to a subgradient oracle for f.

For a convex function $f : \mathbb{R}^d \to \mathbb{R}$ a subgradient of f at $\bar{x}$ is $h \in \mathbb{R}^d$ such that $f(x) \geq f(\bar{x}) + \langle h, x - \bar{x}\rangle\ \forall x \in \mathbb{R}^d$.

Given $\bar{x} \in \mathbb{R}^d$, the subgradient oracle returns $f(\bar{x})$ and $h \in \partial f(\bar{x})$, where

$$\partial f(\bar{x}) := \{h \in \mathbb{R}^d : h \text{ is a subgradient of } f \text{ at } \bar{x}\} \text{ (subdifferential of } f \text{ at } \bar{x}).$$

We can modify our Algorithm I (Convex feasibility) to Ellipsoid Algorithm 2 (Convex Optimization) in the following way:

In each iteration, we have $E_k := E(A_k, c^{(k)})$ we ask the separation oracle for G "is $c^{(k)} \in G$?" If "NO" proceed as before, if "YES" call the subgradient oracle with $\bar{x} := c^{(k)}$.

If $h = 0$ then STOP $c^{(k)}$ is optimal; otherwise $\widetilde{E} := \{x \in E_k : h^\top x \leq h^\top \bar{x}\}$ and continue as before.

Theorem (3.7). Let $G \subset \mathbb{R}^d$ be a closed convex set such that for some $0 < r < 1 < R$ we have

$$B(\tilde{x}, r) \subseteq G \subseteq B_d(0, R), \text{ where } \tilde{x} \in \mathbb{R}^d \text{ \underline{exists} but is not given.}$$

Suppose we have access to a separation oracle for G and a subgradient oracle for f, and $\varepsilon \in \mathbb{Q}_+, \varepsilon \in (0,1)$. Then in $O(d^2(\ln(R/r) + \ln(\mu_0/\varepsilon)))$ iterations, Ellipsoid Algorithm 2 returns $\bar{x} \in G$ such that $f(\bar{x}) \leq \min_{x \in G} f(x) + \varepsilon$, where $\mu_0 = \varepsilon + \sup_{x \in B_d(0,R)}\{f(x)\} - \inf_{x \in B_d(0,R)}\{f(x)\}$.

11 2018-06-12

11.1 Interior Point Method for SDP

We will study an algorithm which generates $(X^k, y^k, S^k) \in \Sigma^2_{++} \oplus \mathbb{R}^m \oplus \Sigma^n_{++}$, $\mathcal{A}(X^k) = b,\ \mathcal{A}^*(y^k) + S^k = C$, such that at every iteration, the number of calculations is in the order of solving a linear system of size $O(n)$.

$$f(x) : -\ln\det(X) : \Sigma^n_{++} \to \mathbb{R}$$

$f(x)$ is a <u>self-concordant</u> function introduced by Nesterov-Nemirovski.

Proposition (4.1). For every $X \in \Sigma^n_{++},\ H \in \Sigma^n$,

$$\begin{aligned} f'(X)[H] &= \frac{\partial}{\partial\alpha} f(X + \alpha H)\big|_{\alpha=0} \\ &= -\langle X^{-1}, H\rangle \\ f''(X)[H,H] &= \frac{\partial^2}{\partial\alpha^2} f(X + \alpha H)\big|_{\alpha=0} \\ &= \mathrm{Tr}(X^{-1}HX^{-1}H) \text{ which, since } X, H \in \Sigma^n_{++}, \text{ shows that } f \text{ is strictly convex.} \end{aligned}$$

Assume both (P) and (D) have Slater points, A is surjective. For every $\mu > 0$, define

$$\begin{aligned} (P_\mu)\ \inf \quad & \tfrac{1}{\mu}\langle C, X\rangle - \ln\det X \\ \text{s.t.} \quad & \mathcal{A}(X) \ = \ b. \end{aligned}$$

Ex: (P_μ) has a unique solution $X(\mu)$ for every $\mu > 0$ (using the existence of Slater points $(P), (D)$).
If we write the optimality conditions (KKT),

$$\frac{1}{\mu}C - X^{-1} - \mathcal{A}^*(y) = 0$$
$$\mathcal{A}(X) = b,\ X \succ 0.$$

$y := \mu y,\ S = \mu X^{-1}$.

$$\begin{aligned} \mathcal{A}(X) = b,\ X \succ 0, \\ \mathcal{A}^*(y) + S = C, \\ XS = \mu I \end{aligned}$$

(call this system $(*)$) $\implies$ For every $\mu > 0$, this system has a unique solution $(X(\mu), y(\mu), S(\mu))$ that defines the primal-dual central path.

Exercise: $\langle X(\mu), S(\mu)\rangle = n\mu \to 0$.
Newton direction:

$$F(x) = 0$$
$$\nabla F(x^0)d = -F(x^0)$$

– is a first-order approximation
The Newton direction (D_X, dy, D_S) at (X, y, S) for $(*)$ satisfies

$$\begin{aligned} \mathcal{A}(D_X) = 0 \\ \mathcal{A}^*(dy) + D_S = 0 \\ XD_S + SD_X = \mu I - XS. \end{aligned}$$

Ex: This system has a unique solution for (X, S).

We can exploit the symmetric structure of PSD cone to design a primal-dual symmetric and scale-invariant algorithm.
For some $W \in \Sigma^n$, non-singular, let us define $W \bullet W \in \mathbf{Aut}(\Sigma^n_{++})$,

$$\begin{aligned} V &= WSW \\ &= W^{-1}XW^{-1} \end{aligned}$$

Ex: $W^2 = S^{-\frac{1}{2}}(S^{\frac{1}{2}}XS^{\frac{1}{2}})^{\frac{1}{2}}S^{-\frac{1}{2}}$

Define

$$\begin{aligned} \overline{\mathcal{A}}(\bullet) &= \mathcal{A}(W \bullet W) \\ \overline{D}_X &:= W^{-1}D_XW^{-1}, \\ \overline{D}_S &= WD_SW. \end{aligned}$$

The system for $(*)$ becomes

$$\begin{aligned}\overline{\mathcal{A}}(\overline{D}_X) &= 0\\ \overline{\mathcal{A}}^*(dy) + \overline{D}_S &= 0\\ X\overline{D}_S + S\overline{D}_X &= \gamma V^{-1} - V.\end{aligned}$$

We need a proximity measure that quantifies the distance to the central path.

Theorem (4.2). For every $X, S \in \Sigma^n_{++}$,

$$\Psi(X,S) := n \ln\left(\frac{\langle X, S\rangle}{n}\right) - \ln\det(X) - \ln\det(S) \geq 0.$$

Moreover, equality holds $\iff XS = \mu I$.

Primal-dual potential function:

$$\Phi_{\sqrt{n}}(X,S) := \sqrt{n}\ln(\langle X,S\rangle) + \Psi(X,S), \quad \forall X, S \in \Sigma^n_{++}.$$

Idea: drive the value of $\Phi_{\sqrt{n}}(X,S) \to -\infty$.

Algorithm 1: Primal-dual Potential Reduction

Input: $X^0, S^0 \in \Sigma^n_{++}$ feasible in $(P)\&(D)$, and $\epsilon \in (0,1)$ s.t. $\Psi(X^0, S^0) \leq n \ln \frac{1}{\epsilon}$.

$k = 0$

while $\langle X^k, S^k\rangle > \epsilon\langle X^0, S^0\rangle$ do

- $W^2 = (S^k)^{-\frac{1}{2}}((S^k)^{\frac{1}{2}}X^k(S^k)^{\frac{1}{2}})^{\frac{1}{2}}(S^k)^{-\frac{1}{2}}$
- $\overline{\mathcal{A}} = \mathcal{A}(W \bullet W)$,
- $V = WS^kW = W^{-1}X^kW^{-1}$
- $\tilde{U} := -\frac{n+\sqrt{n}}{\langle X^k, S^k\rangle}V + V^{-1}$
- $U := \frac{\tilde{U}}{\|\tilde{U}\|_F}$
- Solve the system

  $$\begin{aligned}\overline{\mathcal{A}}(\overline{D}_X) &= 0\\ \overline{\mathcal{A}}^*(dy) + \overline{D}_S &= 0\\ X\overline{D}_S + S\overline{D}_X &= U.\end{aligned}$$

  Compute

  $$\begin{aligned}\bar{\alpha} &:= \operatorname{argmin}\{\Phi_{\sqrt{n}}(V + \alpha\overline{D}_X, V + \alpha\overline{D}_S) : \alpha > 0\}\\ X^{k+1} &= X^k + \bar{\alpha}W\overline{D}_XW\\ S^{k+1} &= S^k + \bar{\alpha}W^{-1}\overline{D}_SW^{-1}\end{aligned}$$

  // Can use line search to approximately solve for $\bar{\alpha}$
- $k = k + 1$

Theorem (4.13). The above algorithm terminates in at most $24\sqrt{n}\ln\frac{1}{\epsilon}$ iterations with X^k, S^k feasible in (P) and (D) such that $\langle X^k, S^k\rangle < \epsilon\langle X^0, S^0\rangle$.

In the above algorithm,

$$\Phi_{\sqrt{n}}(X^{k+1}, S^{k+1}) - \Phi_{\sqrt{n}}(X^k, S^k) \le -\frac{1}{12} = -\delta.$$

$$\Phi_{\sqrt{n}}(X^k, S^k) - \Phi_{\sqrt{n}}(X^0, S^0) = \sqrt{n}\ln\frac{\langle X^k, S^k\rangle}{\langle X^0, S^0\rangle} + \underbrace{\Psi(X^k, S^k)}_{\ge 0} - \underbrace{\Psi(X^0, S^0)}_{\le\sqrt{n}\ln\frac{1}{\epsilon}}.$$

By the above assumption,

$$-\frac{k}{12} \ge \sqrt{n}\ln\frac{\langle X^k, S^k\rangle}{\langle X^0, S^0\rangle} - \sqrt{n}\ln\frac{1}{\epsilon}.$$

Therefore, for every $k \ge \frac{2\sqrt{n}}{\delta}\ln\frac{1}{\epsilon}$, we get $\langle X^k, S^k\rangle \le \epsilon\langle X^0, S^0\rangle$.

$$\Phi_{\sqrt{n}}(X(\alpha), S(\alpha)) - \Phi_{\sqrt{n}}(X, S) = (n + \sqrt{n})\ln\frac{\langle X(\alpha), S(\alpha)\rangle}{\langle X, S\rangle} + f(X(\alpha)) - f(X) + f(S(\alpha)) - f(S)$$

where $f = -\ln\det X$

Lemma (4.6). Let $X \in \Sigma^n_{++}$. Suppose $D \in \Sigma^n$ satisfies

$$\|D\|_X := \langle D, X^{-1}DX^{-1}\rangle^{\frac{1}{2}} \le 1$$
$$f(X+D) \le f(X) + \langle f'(X), D\rangle + \frac{\|D_X\|_X^2}{2(1-\|D_X\|_X)^2}$$

12 2018-06-14

Finish reading Chapter 4 and start reading Chapter 5.

Central Path is defined by solutions (X_μ, y_μ, S_μ) of

$$\mathcal{A}(X) = b,\ X \succ 0$$
$$\mathcal{A}^*(y) + S = C$$
$$S = \mu X^{-1}$$

Proximity measure: $\Psi(X, S) := n\ln\left(\frac{\langle X, S\rangle}{n}\right) - \ln\det(X) - \ln\det(S)$.
$\Psi : \mathbb{S}^n_{++} \oplus \mathbb{S}^n_{++} \to \mathbb{R}$

Theorem (4.2). For every pair $X, S \in \mathbb{S}^n_{++}$, $\Psi(X,S) \geq 0$. Equality holds above iff $S = \mu X^{-1}$ (with $\mu := \frac{\langle X,S\rangle}{n}$).

Theorems 4.5 and 4.13 assume

- $X^{(0)} \succ 0,\ S^{(0)} \succ 0$ feasible in (P) & (D) respectively are given
- $\Psi(X^{(0)}, S^{(0)}) \leq \sqrt{n}\ln\left(\frac{1}{\epsilon}\right)$

Let's consider an auxiliary problem (pick a large constant $M > 0$, add a new variable ζ)

$$\begin{array}{llrcl}(P_{\text{aux}}) & \inf & \zeta & & \\ & \text{s.t.} & \mathcal{A}(X) + [b - \mathcal{A}(I)]\zeta & = & b \\ & & \langle I, X\rangle & \leq & M \\ & & X & \succeq & 0 \\ & & \zeta & \geq & 0\end{array}$$

$$\begin{array}{llrcl}(D_{\text{aux}}) & \sup & b^\top y + M\eta & & \\ & \text{s.t.} & \mathcal{A}^*(y) + \eta I & \preceq & 0 \\ & & [b - \mathcal{A}(I)]^\top y & \leq & 1 \\ & & \eta & \leq & 0\end{array}$$

$X^{(0)} := I,\ \zeta_0 = 1$ is a Slater point.
$y^{(0)} := 0,\ \eta_0 := -1$ is feasible in (D_{aux}). $S^{(0)} = I \succ 0$
$\Psi(X^{(0)}, \zeta_0, S^{(0)}, \eta_0) = 0$
If we prove that the optimal objective value of (P_{aux}) is positive, then we would have proven "(P) has no feasible solutions in $\{X \in \mathbb{S}^n_{++} : \text{Tr}(X) \leq M\}$".

To make our discussion more detailed, let's consider LP as a special case.

$$\begin{array}{llrcl}(LP) & \min & c^\top x & & \\ & \text{s.t.} & Ax & = & b \\ & & x & \geq & 0\end{array}$$

$A \in \mathbb{Q}^{m\times n}, b \in \mathbb{Q}^m, c \in \mathbb{Q}^n$ are given.
Given $\beta \in \mathbb{Z}$, $\text{size}(\beta) := \lceil \log_2(|\beta| + 1)\rceil + 1$
We can write every $\gamma \in \mathbb{Q}$ as $\gamma = \frac{p}{q}$, $p, q \in \mathbb{Z}$ relatively prime, $\text{size}(\gamma) := \text{size}(p) + \text{size}(q)$
$\text{size}(A) := \sum_{i=1}^m \sum_{j=1}^n \text{size}(A_{ij}), \text{size}(LP) := \text{size}(A) + \text{size}(b) + \text{size}(c) =: L.$
We may assume $A \in \mathbb{Z}^{m\times n}, b \in \mathbb{Z}^m, c \in \mathbb{Z}^n$
$F := \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$ (feasible region of the LP)

Proposition (4.14). (a) For every extreme point $\bar{x}$ of F, $\text{size}(\bar{x}) = O(L)$

(b) For every pair of extreme points $\bar{x}, \hat{x}$ of F, either $c^\top\bar{x} = c^\top\hat{x}$ or $\left|c^\top\bar{x} - c^\top\hat{x}\right| > 2^{-2L}$.

Some general ideas for the proof: We may assume $\text{rank}(A) = m$
If $\bar{x} \in F$ is an extreme point of F, then $\exists$ a basis B of A s.t. $N := \{1, 2, \ldots, n\} \setminus B$

$\bar{x}_B = A_B^{-1}b,\ \bar{x}_N = 0$

$$
\begin{aligned}
(\bar{x}_B)_j &= \frac{\text{subdet}[A'_{B_i} b]}{\det(A_B)} \\
&\leq \frac{(m!)[\max_{i,j}[A_B : b]_{ij}]^m}{1}
\end{aligned}
$$

[EN: not sure what the expression above was supposed to be.]
Take $\log_2(\cdot)$ of both sides.

Corollary (4.5). If we have x feasible in LP such that $\left|c^\top x - \nu(LP)\right| \leq 2^{-2L}$ then in $O(n^3)$ arithmetic operations, we can compute an exact optimal solution of LP. ($\nu(LP)$ denotes the optimal value of LP.)

If x is an extreme point of F then it is optimal by Prop. 4.14. Otherwise, we will round (purify) x to an extreme point of F with at least as good objective value.
If $x \in F$ is not an extreme point of F, then $B := \{j : x_j > 0\}$. Then $\exists \bar{d}$ nonzero such that $A_B\bar{d} = 0$ (B is not a basis of A). This gives $d \in \mathbb{Q}^n \setminus \{0\}$ such that $Ad = 0$. Choose a sign for d such that $c^\top d \leq 0$.
$\bar{\alpha} := \max\{\alpha \in \mathbb{R}_+ : x + \alpha d \geq 0\}$ (if $c^\top d = 0$ make sure $d \not\geq 0$ – if $d \geq 0$ replace d by $-d$).
$x \leftarrow x + \bar{\alpha} d$
New x has at least one fewer nonzero entry. So, this algorithm stops in at most n iterations.
The last algorithm (rounding to an extreme point with at least as good objective value) generalizes to SDP except that $\bar{\alpha}$ may be irrational.
Prop 4.14 does not nicely generalize to SDPs.
Consider the SDP:

$$
y_1 \geq 2, \begin{pmatrix} 1 & y_1 \\ y_1 & y_2 \end{pmatrix} \succeq 0, \begin{pmatrix} 1 & y_2 \\ y_2 & y_3 \end{pmatrix} \succeq 0, \ldots, \begin{pmatrix} 1 & y_{n-1} \\ y_{n-1} & y_n \end{pmatrix} \succeq 0
$$

For every feasible solution, $y_n \geq 2^{2^{n-1}}$.

13 2018-06-19

In the case of LP problems having a feasible solution $\bar{X}$ with objective value $\langle C, \bar{X} \rangle \leq \nu(LP) + \varepsilon$ for small enough $\varepsilon > 0$ ($\log\left(\frac{1}{\varepsilon}\right) = O(\text{poly}(L))$) allowed us to compute an exact optimal solution in polynomial time in the Turing machine model (moreover, it was a strongly polynomial time subroutine).

Let's generalize the purification algorithm to the SDPs

$$\begin{array}{rrcl} (P) \ \inf & \langle C, X\rangle & & \\ \text{s.t.} & \mathcal{A}(X) & = & b \\ & X & \succeq & 0 \end{array}$$

We are given $\bar{X}$ feasible in (P).

Let G be the minimal face of $\mathbb{S}^n_+$ which contains $\bar{X}$. By lemma 2.25 $\exists$ a unique linear subspace $L \subseteq \mathbb{R}^n$ s.t. $\text{relint}(G) = \{X \in \mathbb{S}^n_+ : \text{Null}(\bar{X}) = L\}$.

Find $D \in \mathbb{S}^n$ such that $\text{Null}(D) \supseteq L$ and $\mathcal{A}(D) = 0,\ D \neq 0$. If no such D, then STOP, $\bar{X}$ is an extreme point of $\{X \succeq 0 : \mathcal{A}(X) = b\}$.

Choose the sign of D such that $\langle C, D\rangle < 0$ (or $\langle C, D\rangle = 0$ and D has a negative eigenvalue).
$\bar{\alpha} := \sup\{\alpha : \bar{X} + \alpha D \succeq 0\}$ ($\bar{\alpha}$ may be irrational even if $\bar{X}, D$ are rational).
If $\bar{\alpha} = +\infty$ then (P) is unbounded. STOP. (proof: $\bar{X}, D$).
$\bar{X} \longleftarrow \bar{X} + \bar{\alpha} D$.
Note that rank of $\bar{X}$ decreases by at least one.
Repeat the iteration.

In SDP problems it is possible that

- every feasible solution has norm $\geq 2^{2^{\Omega(L)}}$
- the feasible region contains a ball but the largest radius ball has radius $\leq 2^{-2^{\Omega(L)}}$
- it has a unique optimal solution that is irrational.

<u>Ex:</u>

$$\begin{array}{rrcl} \inf & \left\langle \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, X \right\rangle & & \\ \text{s.t.} & \left\langle \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, X \right\rangle & = & 1 \\ & \left\langle \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, X \right\rangle & = & 2 \\ & X & \in & \mathbb{S}^2_+ \end{array}$$

$\bar{X} := \begin{bmatrix} 1 & -\sqrt{2} \\ -\sqrt{2} & 2 \end{bmatrix}$ is the unique optimal solution.

<u>SDP-Feasibility:</u> Given $A_1, A_2, \ldots, A_m \in \mathbb{S}^n \cap \mathbb{Z}^{n\times n}, b \in \mathbb{Z}^m$, does there exist $\bar{X} \in \mathbb{S}^n_+$ s.t. $\mathcal{A}(\bar{X}) = b$?
<u>Open Problem:</u> Is SDP-Feasibility in $\mathcal{P}$?
<u>Also open:</u> Is SDP-Feasibility in $\mathcal{NP}$?

13.1 Approximation Algorithms Based on SDP

13.1.1 Approximation Algorithms for MaxCut

Given a simple graph $G = (V, E)$.
Every $U \subseteq V$ identifies a <u>cut</u> $(U, V \setminus U)$ in G. ($U, V \setminus U$ are called the <u>shores</u> of the cut.)
The set of cut edges is

$$\delta(U) := \{ij \in E : i \in U,\ j \in V \setminus U\}.$$

Given a simple graph $G = (V, E)$, and nonnegative weights $w_{ij} \geq 0$ on the edges, we want to find a <u>cut</u> in G of maximum total weight.

$$\text{weight of the cut} := \sum_{ij \in \delta(U)} w_{ij}$$

MAXCUT is NP hard.

13.1.2 An approximation algorithm for MaxCut

$U_i := \varnothing,\ U_2 := \varnothing$
For each $v \in V$,
if $\sum_{u \in U_1} w_{uv} > \sum_{u \in U_2} w_{uv}$ then $U_2 := U_2 \cup \{v\}$
otherwise $U_1 := U_1 \cup \{v\}$
Repeat until $U_1 \cup U_2 = V$.

<u>Fact:</u> this algorithm runs in strongly polynomial time and always delivers a cut U such that weight of $\delta(U) \geq \frac{1}{2}$ MaxCut.

$$\begin{aligned} \text{weight of } \delta(U) &\geq \frac{1}{2} \sum_{ij \in E} w_{ij} \\ &\geq \frac{1}{2} \text{ MaxCut.} \end{aligned}$$

This algorithm can be viewed as derandomization of a beautiful and very simple randomized algorithm (for each vertex independently flip a fair coin).

Compute the expected value of the total weight of a cut delivered by the randomized algorithm.

Let's write a formulation for MaxCut.

$$u_i := \begin{cases} 1 & \text{if } i \in U, \\ -1 & \text{if } i \in V \setminus U \end{cases}.$$

$$\begin{array}{rl} \max & \frac{1}{4} \sum_{i,j} w_{ij}(1 - u_i u_j) \\ \text{s.t.} & u \in \{-1, 1\}^n \end{array}$$

$w_{ij} := 0 \ \forall i, j$ s.t. $ij \notin E$.
$W \in \mathbb{S}^n$, $W_{ij} := w_{ij} \ \forall i, j \in V$
Last problem is equivalent to

$$\begin{array}{lrcl} \max & -\frac{1}{4}\langle W, uu^\top \rangle & (+\frac{1}{4}\langle W, \bar{e}\bar{e}^\top \rangle) & \\ \text{s.t.} & u_i^2 & = & 1 \ \forall i \in V \end{array}$$

$(\bar{e} = \mathbb{1})$
Last problem is equivalent to

$$\begin{array}{lrcl} \max & -\frac{1}{4}\langle W, X \rangle \ (+\frac{1}{4}\langle W, \bar{e}\bar{e}^\top \rangle) & & \\ \text{s.t.} & \operatorname{diag}(X) & = & \bar{e} \\ & X & \in & \mathbb{S}^n_+ \\ & \operatorname{rank}(X) & = & 1 \quad \longleftarrow \text{nonconvex constraint} \end{array}$$

SDP relaxation:

$$\begin{array}{lrcl} \max & -\frac{1}{4}\langle W, X \rangle \ (+\frac{1}{4}\langle W, \bar{e}\bar{e}^\top \rangle) & & \\ \text{s.t.} & \operatorname{diag}(X) & = & \bar{e} \\ & X & \in & \mathbb{S}^n_+ \end{array}$$

14 2018-06-21

$G = (V, E), w \in \mathbb{R}^E_+$ given. $n := |V|, W_{ij} := \begin{cases} w_{ij} & \forall \{i, j\} \in E \\ 0 & \text{otherwise} \end{cases}, W \in \mathbb{S}^n$.

$u_i := \begin{cases} +1, & i \in U \\ -1, & i \in V \setminus U \end{cases}$

$$\begin{array}{llrcl} (P) & \max & -\frac{1}{4}\langle W, X \rangle \ (+\frac{1}{4}\langle W, \bar{e}\bar{e}^\top \rangle) & & \\ & \text{s.t.} & \operatorname{diag}(X) & = & \bar{e} \\ & & X & \succeq & 0 \end{array}$$

$$\begin{array}{llrcl} (D) & \min & \bar{e}^\top y \ (+\frac{1}{4}\langle W, \bar{e}\bar{e}^\top \rangle) & & \\ & \text{s.t.} & \operatorname{Diag}(y) - S & = & -\frac{1}{4}W \\ & & S & \succeq & 0 \end{array}$$

Both (P) & (D) have Slater points:

$$\bar{X} := I$$
$$\bar{\eta} := \frac{1}{4}\langle W, \bar{e}\bar{e}^\top \rangle + 1$$
$$\bar{y} := \bar{\eta}\bar{e}$$
$$\underbrace{\operatorname{Diag}(\bar{y}) + \frac{1}{4}W}_{=:\bar{S}} \succ 0 \text{ by strict diag. dominance}$$

Therefore, by Corollary 2.17, both (P) & (D) attain their optimal values <u>and</u> their optimal values are the same. Suppose we solved (P) and have an optimal (or near optimal) solution $\bar{X}$.

$$\begin{aligned}
\bar{X} &=: BB^\top,\ B \in \mathbb{R}^{n\times n} \text{ (B exists by Prop. 1.10 since $\bar{X} \succeq 0$)}\\
B^\top &=: [v^{(1)}, v^{(2)}, \dots, v^{(n)}],\ v^{(i)} \in \mathbb{R}^n\\
\bar{X}_{ij} &= \langle v^{(i)}, v^{(j)} \rangle,\ \forall i, j\\
&\implies \left\| v^{(i)} \right\|_2 = 1,\ \forall i \in V \quad (\mathrm{diag}(\bar{X}) = \bar{e})
\end{aligned}$$

14.1 Randomized Hyperplane Technique (RHT)

Pick $r \in \mathbb{R}^n, \|r\|_2 = 1$ uniformly randomly.

$$U := \{v \in V : \langle r, v^{(i)} \rangle \geq 0\}.$$

For $a \in \mathbb{R}^n$, $\mathrm{sign}(a) \in \{-1, 1\}^n$ is defined by $[\mathrm{sign}(a)]_j := \begin{cases} +1, & \text{if } a_j \geq 0; \\ -1, & \text{if } a_j < 0 \end{cases}$.

Lemma (5.1). With the above definitions,

Lemma (5.2). $\forall u \in [-1, 1]$, we have

(i) $\frac{\arccos(u)}{\pi} \geq \frac{\rho}{2}(1-u)$

(ii) $1 - \frac{\arccos(u)}{\pi} \geq \frac{\rho}{2}(1+u)$

where $\rho \approx 0.87856$.

Theorem (5.4). For every graph $G = (V, E)$ and $w \in \mathbb{Q}^E_+$ we can obtain in polynomial time a cut of total weight at least ρ(MaxCut value in G).

Proof. RLT has been derandomized. □

Feasible region of the MaxCut SDP:

$$\{X \in \mathbb{S}^n_+ : \mathrm{diag}(X) = \bar{e}\} \text{ is called \underline{elliptope}.}$$

15 2018-06-26

15.1 Satisfiability, Max k-SAT, derandomization

<u>variables:</u> $x_1, x_2, \dots, x_n$ $\quad x_j \in \{0, 1\}$ (0 = false, 1 = true)
<u>literals:</u> $x_j, \bar{x}_j$ (represents the complement of x_j)
<u>clauses:</u> $C_1, C_2, \dots, C_m$ $\quad C_i$:= some disjunction of literals e.g. $C_1 := (x_1 \vee x_2 \vee \bar{x}_5)$
Formula in Conjunctive Normal Form (CNF):

$$C_1 \wedge C_2 \wedge \cdots \wedge C_m$$

Satisfiability: Given a formula, does there exist a truth assignment $x \in \{0,1\}^n$ such that the formula evaluates to 1 ("True")?

Optimization version: Together with the formula, we are given $w \in \mathbb{R}_+^m$ and we want to find $x \in \{0,1\}^n$ such that total weight of satisfied clauses is maximized (Max SAT).

Max k-SAT: Max SAT where each clause has exactly k-literals.
2-SAT is easy; 3-SAT and Max 2-SAT are $\mathcal{NP}$-hard.
Let's give an Integer Programming formulation for Max SAT.

$$z_i := \begin{cases} 1 & \text{if clause } C_i \text{ is satisfied} \\ 0 & \text{otherwise.} \end{cases}$$

$$\begin{array}{lrcl} \max & \sum_{i=1}^m w_i z_i & & \\ \text{s.t.} & z_i & \leq & \sum_{x_j \in C_i} x_j + \sum_{\bar{x}_j \in C_i} (1 - x_j), \ \forall i \in \{1, 2, \ldots, m\} \ \longleftarrow \text{the \# of literals in } C_i \text{ set to "True"} \\ & x & \in & \{0,1\}^n \\ & z & \in & \{0,1\}^m \end{array}$$

15.2 A Simple Randomized Approximation Algorithm for Max k-SAT

David Johnson: author of Computers and Intractability, bin-packing
Johnson [1974]: $p_1, p_2, \ldots, p_n$ are independently chosen probabilities.
Assign $x_j := 1$ with probability p_j.

$$\begin{aligned} u_i : = & \text{ probability that clause } C_i \text{ is not satisfied} \\ = & \prod_{\bar{x}_j \in C_i} p_j \prod_{x_j \in C_i} (1 - p_j) \end{aligned}$$

The expected total weight of satisfied clauses is

$$\begin{aligned} E[w, p] &= E[\sum_{i=1}^m w_i \Pr(C_i \text{ is satisfied})] \\ &= \sum_{i=1}^m w_i (1 - u_i) \end{aligned}$$

For $p_j = \frac{1}{2} \forall j$, we get $u_i = \frac{1}{2^k} \Longrightarrow$

$$\begin{aligned} E[w, \tfrac{1}{2}\bar{e}] &= (1 - 2^{-k}) \sum_{i=1}^m w_i \\ &\geq (1 - 2^{-k}) \operatorname{opt}(\text{Max } k\text{-SAT}). \end{aligned}$$

(Assume clauses involve distinct variables; i.e. do preprocessing first.)

We can derandomize this algorithm. We will make choices for the values of x_j's so that with each choice, the corresponding conditional expectation is at least the overall expectation.

$$
\begin{aligned}
&E[w;p] = p_j E[w;p|x_j = 1] + (1-p_j)E[w;p|x_j = 0] \\
&\implies \max\{E[w;p|x_j = 1], E[w;p|x_j = 0]\} \geq E[w;p]
\end{aligned}
$$

To derandomize the algorithm, we compute $E[w;p|x_j = 1], E[w;p|x_j = 0]$ and assign x_j 0 or 1 depending on which conditional expectation is larger.

15.3 Generalizations to Quadratic Optimization over vertices of hypercubes

Given $W \in \mathbb{S}^n$

$$
\begin{aligned}
\bar{f}(W) := \max \quad & x^\top W x \\
\text{s.t.} \quad & x \in \{-1,1\}^n
\end{aligned}
$$

Computing $\bar{f}(W)$ is $\mathcal{NP}$-hard.

$$
\begin{aligned}
\max \quad & \langle W, xx^\top \rangle \\
\text{s.t.} \quad & \operatorname{diag}(xx^\top) = \bar{e} \\
& x \in \mathbb{R}^n
\end{aligned}
$$

SDP relaxation:

$$
\begin{aligned}
\max \quad & \langle W, X \rangle \\
\text{s.t.} \quad & \operatorname{diag}(X) = \bar{e} \\
& X \succeq 0
\end{aligned}
\quad \underbrace{=}_{\text{Both } (P) \text{ and } (D) \text{ have Slater points}^\dagger} \quad
\begin{aligned}
\min \quad & \bar{e}^\top y \\
\text{s.t.} \quad & \operatorname{Diag}(y) \succeq W
\end{aligned}
$$

($^\dagger \bar{X} := I, \bar{y} = \bar{\eta}\bar{e}, \bar{\eta} = \|W\|_2 H$)
Similarly, we define

$$
\begin{aligned}
\underline{f}(W) := \min \quad & x^\top W x \\
\text{s.t.} \quad & x \in \{-1,1\}^n,
\end{aligned}
\qquad
\begin{aligned}
\underline{F}(W) := \min \quad & \langle W, X \rangle \\
\text{s.t.} \quad & \operatorname{diag}(X) = \bar{e} \\
& X \succeq 0
\end{aligned}
\quad = \quad
\begin{aligned}
\max \quad & \bar{e}^\top y \\
\text{s.t.} \quad & \operatorname{Diag}(y) \succeq W
\end{aligned}
$$

For approximation ratios, we will consider the relative approximation ratio (for a given $\bar{x} \in \{-1,1\}^n$),

$$
\frac{\bar{f}(W) - \bar{x}^\top W \bar{x}}{\bar{f}(W) - \underline{f}(W)}.
$$

Proposition (5.10). For every $W \in \mathbb{S}^n$, we have

- $\bar{f}(W) = -\underline{f}(-W), \bar{F}(W) = -\underline{F}(-W)$
- $\underline{F}(W) \leq \underline{f}(W) \leq \bar{f}(W) \leq \bar{F}(W)$

Special Case $W \in \mathbb{S}^n_+$

$$\bar{f}(W) = \max \quad x^\top W x \quad \text{s.t.} \quad x \in \{-1,1\}^n \qquad = \max \quad x^\top W x \quad \text{s.t.} \quad x \in [-1,1]^n$$

We are maximizing a convex function over a non empty closed convex set (whenever $\exists$ an optimal solution, then $\exists$ one that is an extreme point).

Suppose $\exists$ a maximizer $\bar{x}$ that is not an extreme point of the feasible region $\implies \exists x^{(1)}, x^{(2)} \neq \bar{x}$ feasible, s.t. $\bar{x} = \frac{1}{2}x^{(1)} + \frac{1}{2}x^{(2)}$. Let $g(x) := x^\top W x$.

$$\begin{aligned} g \text{ is convex } &\implies g(\bar{x}) \leq \tfrac{1}{2}g(x^{(1)}) + \tfrac{1}{2}g(x^{(2)}) \\ &\implies \max\{g(x^{(1)}), g(x^{(2)})\} \geq g(\bar{x}). \end{aligned}$$

Choose extreme points $x^{(1)}, x^{(2)}, \ldots, x^{(k)}$ such that $\bar{x} = \sum_{i=1}^k \lambda_i x^{(i)}, \sum_{i=1}^k \lambda_i = 1, \lambda \geq 0$.

16 2018-06-28

Finish reading Chapter 5 and read first two sections of Chapter 6.

For every $W \in \mathbb{S}^n_+$,

$$\bar{f}(W) = \max \quad x^\top W x \quad \text{s.t.} \quad x \in \{-1,1\}^n$$

Special case: $W \in \mathbb{S}^n_+$.
The maximum value of a convex function over a nonempty closed, bounded convex set is attained at an extreme point of the feasible region.

Lemma (5.11). For every $W \in \mathbb{S}^n$, we have

$$\begin{aligned} \bar{f}(W) = \max \quad & \zeta^\top W \zeta \\ \text{s.t.} \quad & \zeta = \operatorname{sign}(Br) \\ & \|B^\top e_i\|_2 = 1 \quad \forall i \in \{1,2,\ldots,n\} \\ & \|r\|_2 = 1 \\ & B \in \mathbb{R}^{n\times n} \\ & r \in \mathbb{R}^n \end{aligned}$$

Proof. For every feasible solution of the RHS, the optimization problem has $\zeta \in \{-1,1\}^n$ by definition of $\operatorname{sign}(\cdot)$. Therefore, $f(\bar{W}) \geq$ RHS.
Let $\bar{x} \in \{-1,1\}^n$ such that $\bar{f}(W) = \bar{x}^\top W \bar{x}$.
Pick any $\bar{r} \in \mathbb{R}^n$ s.t. $\|\bar{r}\|_2 = 1$.

$$\bar{B}^\top e_i := \begin{cases} \bar{r} & \text{if } \bar{x}_i = 1 \\ -\bar{r} & \text{if } \bar{x}_i = -1. \end{cases}$$

Let $\bar{\zeta} = \underbrace{\text{sign}(\bar{B}\bar{r})}_{=\bar{x}}$.
Thus, the objective value of $(\bar{\zeta}, \bar{B}, \bar{r})$ in the RHS is $\bar{x}^\top W \bar{x} = \bar{f}(W)$. □

Lemma (5.12). For every $W \in \mathbb{S}^n$, we have

$$\begin{array}{rrcl} \bar{f}(W) = \max & E_r[\zeta^\top W \zeta] & & \\ \text{s.t.} & \zeta & = & \text{sign}(Br) \\ & \left\|B^\top e_i\right\|_2 & = & 1 \quad \forall i \in \{1, 2, \dots, n\} \\ & \|r\|_2 & = & 1 \\ & B & \in & \mathbb{R}^{n \times n} \\ & r & \in & \mathbb{R}^n. \end{array}$$

Proof. As in the proof of Lemma 5.11, in every feasible solution of the RHS, $\zeta \in \{-1, 1\}^n$,

$$E_r[\zeta^\top W \zeta] \leq \max_{\zeta \in \{-1,1\}} \zeta^\top W \zeta = \bar{f}(W).$$

Thus, $\bar{f}(W) \geq$ RHS.
Let $\bar{x} \in \{-1, 1\}^n$ such that $\bar{f}(W) = \bar{x}^\top W \bar{x}$.
$\bar{B} := \frac{1}{\sqrt{n}} \bar{x}\bar{x}^\top$ (then, $\bar{B}\bar{B}^\top = \bar{x}\bar{x}^\top$).

$$\bar{B}^\top e_i = \begin{cases} \frac{1}{\sqrt{n}} \bar{x} & \text{if } \bar{x}_i = 1 \\ -\frac{1}{\sqrt{n}} \bar{x} & \text{if } \bar{x}_i = -1 \end{cases}$$

$\left\|\bar{B}^\top e_i\right\|_2 = 1 \ \forall i \in \{1, 2, \dots, n\}$. Thus, $\bar{B}$ is a feasible solution of the RHS.

$$\begin{aligned} E_r[\text{sign}(\bar{B}r)^\top W \, \text{sign}(\bar{B}r)] &= E_r[\sum_{i=1}^n \sum_{j=1}^n \text{sign}(r^\top \bar{B} e_i) \, \text{sign}(r^\top \bar{B} e_j) W_{ij}] \\ &= \sum_{i=1}^n \sum_{j=1}^n W_{ij} E_r[\text{sign}(\underbrace{r^\top \bar{B} e_i}_{=\bar{x}_i \left(\frac{r^\top \bar{x}}{\sqrt{n}}\right)}) \, \text{sign}(\underbrace{r^\top \bar{B} e_j}_{=\bar{x}_j \left(\frac{r^\top \bar{x}}{\sqrt{n}}\right)})] \\ &= \bar{x}_i \bar{x}_j. \end{aligned}$$

If $r^\top \bar{x} \neq 0$ this is clear, noting that $\dim\{r \in \mathbb{R}^n : \|r\|_2 = 1, r^\top \bar{x} = 0\} = n - 2 < n - 1$, we conclude the equality. □

For every matrix $X \in \mathbb{R}^{n \times n}$ $\left|X_{ij}\right| \leq 1 \ \forall i, j$, define $\arcsin(X) \in \mathbb{R}^{n \times n}$ componentwise:

$$[\arcsin(X)]_{ij} = \arcsin(X_{ij}).$$

Theorem (5.13). For every $W \in \mathbb{S}^n$, we have

$$\begin{array}{rrcl} \bar{f}(W) = \frac{2}{\pi} \max & \langle W, \arcsin(X) \rangle & & \\ \text{s.t.} & \text{diag}(X) & = & \bar{e} \\ & X & \succeq & 0. \end{array}$$

Proof. Note that for every feasible solution $\bar{X}$ of the RHS problem, $|\bar{X}ij| \leq 1 \ \forall i, j \in \{1, 2, \ldots, n\}$ (every 2-by-2 symmetric minor is positive semidefinite) $\Longrightarrow$ RHS is well-defined.
Since the feasible region is nonempty and compact, the objective function is continuous, max value in the RHS is attained.
Let $\bar{B} \in \mathbb{R}^{n \times n}$ be an optimal solution of the problem from Lemma 5.12. Thus,

$$\bar{f}(W) = E_r[\text{sign}(\bar{B}r)^\top W \,\text{sign}(\bar{B}r)].$$

$\bar{B}^\top =: [v^{(1)} v^{(2)} \cdots v^{(n)}]$. Then,

$$\begin{aligned}
&E_r[\text{sign}(\underbrace{r^\top \bar{B} e_i}_{=r^\top v^{(i)}}) \,\text{sign}(\underbrace{r^\top \bar{B} e_j}_{=r^\top v^{(j)}})] \\
&= -\Pr[\text{sign}(r^\top v^{(i)}) \neq \text{sign}(r^\top v^{(j)})] + \Pr[\text{sign}(r^\top v^{(i)}) = \text{sign}(r^\top v^{(j)})] \\
&= 1 - 2\Pr[\text{sign}(r^\top v^{(i)}) \neq \text{sign}(r^\top v^{(j)})] \\
&= 1 - \tfrac{2}{\pi} \arccos\langle v^{(i)}, v^j \rangle \text{ by Lemma 5.1.}
\end{aligned}$$

Since $\arcsin(u) + \arccos(u) = \frac{\pi}{2} \ \forall u \in [-1, 1]$, thus,

$$E_r[\text{sign}(r\bar{B}e_i)^\top \,\text{sign}(r\bar{B}e_j)] = \frac{2}{\pi} \arcsin \underbrace{\langle v^{(i)}, v^{(j)} \rangle}_{=(\bar{B}\bar{B}^\top)_{ij}}, \quad \forall i, j \in \{1, 2, \ldots, n\}.$$

So, $\bar{f}(W) = \frac{2}{\pi} \langle W, \arcsin(\bar{B}\bar{B}^\top) \rangle$.
$\bar{X} := \bar{B}\bar{B}^\top$ is a feasible solution of the RHS problem in the statement of the theorem. Therefore, $\bar{f}(W) \leq$ RHS.
For the remaining inequality, let $\hat{X}$ be an optimal solution of the RHS problem, define $\hat{B} \in \mathbb{R}^{n \times n}$ by $\hat{X} =: \hat{B}\hat{B}^\top$. Then, consider $\hat{B}$ as a feasible solution of the stochastic optimization problem from Lemma 5.12. Its objective value is $\frac{2}{\pi} \langle W, \arcsin(\hat{B}\hat{B}^\top) \rangle$ by the above expectation computation. Therefore, RHS $\leq \bar{f}(W)$ by Lemma 5.12. □

Lemma (5.14). For every $X \in \mathbb{S}^n_+$ such that $|X_{ij}| \leq 1 \ \forall i, j \in \{1, 2, \ldots, n\}$, we have $\arcsin(X) \succeq X$.

Proof. Use a Taylor expansion of $\arcsin(u)$:

$$\begin{aligned}
\arcsin(u) &= \sum_{k=0}^{\infty} \frac{(2k-1)!!}{(2k)!!(2k+1)} u^{2k+1} \\
\Longrightarrow \arcsin(X) &= \sum_{k=0}^{\infty} \frac{(2k-1)!!}{(2k)!!(2k+1)} X^{\odot 2k+1} \\
&= X + \tfrac{1}{6} X \odot X \odot X + \tfrac{3}{40} X \odot X \odot X \odot X \odot X + \cdots \quad \text{– all positive semidefinite.}
\end{aligned}$$

□

17 2018-07-10

Theorem (5.15). For every $W \in \mathbb{S}^n_+$, $\frac{2}{\pi}\bar{F}(W) \leq \bar{f}(W) \leq \bar{F}(W)$.

Proof. We already noted the RHS inequality. For the LHS inequality, take an optimal solution $\bar{X}$ defining $\bar{F}(W)$. Then $\bar{X}$ is feasible in the nonlinear SDP of Theorem 5.13.

$$\begin{aligned}\frac{2}{\pi}\bar{F}(W) &= \frac{2}{\pi}\langle W, \bar{X}\rangle \\ &\leq \frac{2}{\pi}\langle W, \arcsin(\bar{X})\rangle \text{ since } \bar{X} \succeq 0, \text{ Lemma 5.13 and } W \succeq 0 \\ &\leq \bar{f}(W) \text{ since } \bar{X} \text{ is feasible in nonlinear SDP.}\end{aligned}$$

□

Note that the MaxCut problem arises as a special case of $W \in \mathbb{S}^n_+$. Given a graph $G = (V, E)$, and $w \in \mathbb{R}^E$, the weighted Laplacian of G with respect to w as $\mathcal{L}_G : \mathbb{R}^E \to \mathbb{S}^V$.

$$[\mathcal{L}_G(w)]_{ij} := \begin{cases} \sum_{k\,:\,ik \in E} w_{ik} & \text{if } i = k \\ -w_{ij} & \text{if } ij \in E \\ 0 & \text{otherwise.}\end{cases}$$

If $w \in \mathbb{R}^E_+$ then $\mathcal{L}_G(w) \succeq 0$ (diagonally dominant).
$W := \frac{1}{4}\mathcal{L}_G(w)$ covers the MaxCut case.

Next, let's consider $W \in \mathbb{S}^n$ (not necessarily PSD). Note that the dual SDPs related to $\bar{F}(W)$ and $\underline{F}(W)$ have constraints that look like: $[\mathrm{Diag}(y) - W] \succeq 0, [W - \mathrm{Diag}(y)] \succeq 0$. Consider

$$x^\top (W + \mathrm{Diag}(y))x = x^\top W x + \sum_{i=1}^n y_i x_i^2 \underbrace{=}_{x \in \{-1,1\}^\top} x^\top W x + \underbrace{\bar{e}^\top y}_{\substack{\text{does not} \\ \text{depend on } x}}$$

$$\langle (W + \mathrm{Diag}(y)), X\rangle = \langle W, X\rangle + y^\top \mathrm{diag}(X) \underbrace{=}_{\mathrm{diag}(X) = \bar{e}} \langle W, X\rangle + \bar{e}^\top y.$$

We conclude $\forall y \in \mathbb{R}^n$,

$$\begin{aligned}\underline{f}(W + \mathrm{Diag}(y)) &= \underline{f}(W) + \bar{e}^\top y \\ \bar{f}(W + \mathrm{Diag}(y)) &= \bar{f}(W) + \bar{e}^\top y \\ \underline{F}(W + \mathrm{Diag}(y)) &= \underline{F}(W) + \bar{e}^\top y \\ \bar{F}(W + \mathrm{Diag}(y)) &= \bar{F}(W) + \bar{e}^\top y.\end{aligned}$$

Theorem (5.16). For every $W \in \mathbb{S}^n$, we have

$$\underline{F}(W) \leq \underline{f}(W) \leq \frac{2}{\pi}\underline{F}(W) + (1 - \frac{2}{\pi})\bar{F}(W) \leq (1 - \frac{2}{\pi})\underline{F}(W) + \frac{2}{\pi}\bar{F}(W) \leq \bar{f}(W) \leq \bar{F}(W).$$

Proof. We have observed three of the inequalities. Let $W \in \mathbb{S}^n$, let $\bar{y} \in \mathbb{R}^n$ be an optimal solution of the SDP (dual to the one defining $\bar{F}(W)$):

$$\begin{array}{rl} \min & \bar{e}^\top y \\ \text{s.t.} & \operatorname{Diag}(y) - W \succeq 0. \end{array}$$

We compute

$$\begin{aligned} \bar{F}(W) - \underline{f}(W) &= \bar{e}^\top \bar{y} - \underline{f}(W) \text{ by definition of } y \\ &= \bar{e}^\top \bar{y} + \underline{f}(-W) \\ &= \bar{f}(\underbrace{\operatorname{Diag}(\bar{y}) - W}_{\succeq 0}) \\ &\geq \tfrac{2}{\pi}\bar{F}(\operatorname{Diag}(y) - W) \text{ by Theorem 5.15} \\ &= \tfrac{2}{\pi}\bar{F}(-W) + \tfrac{2}{\pi}\bar{e}^\top \bar{y} \\ &= -\tfrac{2}{\pi}\underline{F}(W) + \tfrac{2}{\pi}\bar{F}(W). \end{aligned}$$

Thus,

$$\underline{f}(W) \leq \tfrac{2}{\pi}\underline{F}(W) + (1 - \tfrac{2}{\pi})\bar{F}(W).$$

The remaining inequality can be proved similarly. □

Corollary. For every $W \in \mathbb{S}^n$, with $c' := (1 - \frac{2}{\pi})\underline{F}(W) + \frac{2}{\pi}\bar{F}(W)$, we have

$$\bar{f}(W) - c' \leq \bar{f}(W) - \underline{f}(W).$$

What if there is a linear term in the objective function? $W \in \mathbb{S}^n, q \in \mathbb{R}^n$ given.

$$\begin{array}{rl} \max\limits_x & x^\top W x + q^\top x \\ \text{s.t.} & x \in \{-1,1\}^n \end{array} = \begin{array}{rl} \max\limits_x & \tilde{x}^\top \widetilde{W} \tilde{x} \\ \text{s.t.} & \tilde{x} \in \{-1,1\}^n \end{array} \quad \widetilde{W} := \begin{bmatrix} 0 & \frac{1}{2}q^\top \\ \frac{1}{2}q & W \end{bmatrix} \quad \tilde{x} := \begin{bmatrix} x_0 \\ x \end{bmatrix} \in \mathbb{R}^{n+1}.$$

$\tilde{x}^\top \widetilde{W} \tilde{x} = x^\top W x + x_0(q^\top x)$.
(If $x_0 = 1 \rightarrow$ okay. If $x_0 = -1, x \leftarrow -x$.)
A related generalization leads to sufficient conditions for a "Matrix Cube" to be contained in $\mathbb{S}^n_+$: Given $A_0, A_1, \ldots, A_k \in \mathbb{S}^n$, fine the largest $r \in \mathbb{R}_+$ such that $\{A_0 + \sum_{i=1}^k y_i A_i : \|y\|_2 \leq r\} \subseteq \mathbb{S}^n_+$.

The feasibility of the following SDP guarantees that r given below, works above:

$$\begin{aligned} (X^{(i)} \pm rA_i) \in \mathbb{S}^n_+; \forall i \in \{1, 2, \ldots, k\} \\ \sum_{i=1}^k X^{(i)} \preceq A_0. \end{aligned}$$

There are related problems in algebraic geometry that go back to Grothendieck (his work from 1950s). Consider, given $W \in \mathbb{R}^{m \times n}$,

$$\begin{array}{rl} \max & u^\top W v \\ \text{s.t.} & u \in \{-1,1\}^m \\ & v \in \{-1,1\}^n. \end{array}$$

An SDP relaxation is

$$\begin{array}{rrcl} \max & \langle \bar{W}, X \rangle & & \\ \text{s.t.} & \operatorname{diag}(X) & = & \bar{e} \\ & X & \in & \mathbb{S}_{+}^{m+n} \end{array} \qquad \bar{W} := \begin{bmatrix} 0 & W \\ W^{\top} & 0 \end{bmatrix} \in \mathbb{S}^{m+n}.$$

18 2018-06-12

18.1 Geometric Representations of Graphs

Given a graph $G = (V, E)$, a geometric representation of G is $v : V \to \mathbb{R}$. A unit distance representation of $G = (V, E)$ is a geometric representation v of G such that $\|v(i) - v(j)\|_2 = 1 \forall i, j \in E$.
Ex: $G := K_3 :=$ clique on three vertices, $d := 2$.
$t_b(G) :=$ the square of the smallest radius Euclidean Ball which contains a unit distance representation of G.
Given a geometric representation v of G, define $n := |V|$,

$$\begin{aligned} B^\top &:= \begin{bmatrix} v(1) & v(2) & \dots & v(n) \end{bmatrix} \in \mathbb{R}^{d \times n} \\ X &:= BB^\top \in \mathbb{S}_+^V \end{aligned}$$

Suppose $ij \in E$, then $\|v(i) - v(j)\|_2 = 1 \iff X_{ii} + X_{jj} - 2X_{ij} = 1$.
$\forall i \in V, \|v(i)\|_2^2 \leq t \iff \text{diag}(X) \leq t\bar{e}$.

Theorem (6.2). For every graph $G = (V, E)$, $t_b(G) =$

$$\begin{array}{lrcl} \min & t & & \\ \text{s.t.} & \text{diag}(X) - t\bar{e} & \leq & 0 \\ & X_{ii} + X_{jj} - 2X_{ij} & = & 1 \ \forall ij \in E \\ & X & \in & \mathbb{S}_+^V \end{array}$$

When the graph G has many symmetries the underlying SDPs can be greatly simplified. For example, let G be the Petersen Graph.

- For every pair of vertices, there is an automorphism of G which maps one to the other.
- For every pair of edges, there is an automorphism of G which maps one to the other.
- For every pair of non-edges, there is an automorphism of G which maps one to the other.

Using these symmetries, the SDP for $t_b(G_{\text{Petersen}})$ reduces to an LP problem with three variables.

Theorem (6.3). Suppose $C, A_1, A_2, \dots, A_m \in \mathbb{S}_+^n$ are such that they pairwise commute. Then for every $b \in \mathbb{R}^m$, the underlying SDP (P) and its dual (D) are equivalent to a pair of primal-dual LP problems.

Proof. Suppose $C, A_1, A_2, \dots, A_m \in \mathbb{S}_+^n$ are such that they pairwise commute. Then, $\exists Q \in \mathbb{R}^{n \times n}$ orthogonal s.t. $QC, Q^\top, QA, Q^\top, QA_m Q^\top$ are all diagonal matrices. For every $b \in \mathbb{R}^m$,

$$
\begin{array}{llcll}
(D) \ \sup & b^\top y & & & \\
\text{s.t.} & \sum_{i=1}^m y_i A_i & \preceq & C & \iff \sum_{i=1}^m y_i (QA_iQ^\top) \preceq QCQ^\top \\
& & & & \text{since } Q, Q^\top \in A_m + (\mathbb{S}^n_+) \\
& & & & \iff \sum_{i=1}^m y_i \operatorname{diag}(QA_iQ^\top) \\
& & & & \leq \operatorname{diag}(QCQ^\top)
\end{array}
$$

Taking the dual of this resulting LP gives an LP problem equivalent to (P). Or, we can do it directly:

$$
\begin{array}{llcl}
\inf & \langle C, X \rangle & & \\
\text{s.t.} & \langle A_i, X \rangle & = & b_i, \ \forall i \in \{1, 2, \ldots, m\} \\
& X & \succeq & 0
\end{array}
$$

$$
\begin{array}{llcl}
\inf & \langle QCQ^\top, QXQ^\top \rangle & & \\
\text{s.t.} & \langle QA_iQ^\top, QXQ^\top \rangle & = & b_i, \ \forall i \in \{1, 2, \ldots, m\} \\
& X & \succeq & 0
\end{array}
$$

$\tilde{x} := \operatorname{diag}(QXQ^\top) \in \mathbb{R}^n$. Then SDP (P) is equivalent to the LP

$$
\begin{array}{llcl}
\min & \operatorname{diag}(QCQ^\top)^\top \tilde{x} & & \\
\text{s.t.} & \operatorname{diag}(QA_iQ^\top)^\top \tilde{x} & = & b_i, \ \forall i \\
& \tilde{x} & \geq & 0.
\end{array}
$$

□

A nicer version of unit distance representation:
$t_h(G)$:= square of the minimum radius hypersphere which contains a unit distance representation of G.

Theorem (6.4). For every graph $G = (V, E)$,

$$
\begin{array}{rrcl}
t_h(G) = \min & t & & \\
\text{s.t.} & \operatorname{diag}(X) & = & t\bar{e} \\
& X_{ii} + X_{jj} - 2X_{ij} & = & 1 \ \forall ij \in E \\
& X & \in & \mathbb{S}^V_+.
\end{array}
$$

In fact, $t_b(G) = t_h(G), \forall$ graphs G.
$\forall$ graphs G, $t_h(G) < \frac{1}{2}$.
Let $G = (V, E), n := |V|$.
$\bar{X} := \frac{1}{2} I - \frac{1}{2n} \bar{e}\bar{e}^\top \in \mathbb{S}^V, \operatorname{diag}(\bar{X}) = \underbrace{\frac{1}{2}(1 - \frac{1}{n})}_{< \frac{1}{2}} \bar{e}$
$\bar{X}_{ii} + \bar{X}_{jj} - 2\bar{X}_{ij} = 1 \forall i \neq j$.
$\bar{X} \succeq 0 \iff \forall h \in \mathbb{R}^n, \|h\|_2 = 1, \ 0 \leq h^\top \bar{X} h$,
$h^\top \bar{X} h = \underbrace{\frac{1}{2} - \frac{1}{2n} (\underbrace{\bar{e}^\top h}_{\leq n})^2}_{\geq -\frac{1}{2}} \geq 0.$

$[|e^\top h| \le \frac{n}{\sqrt{n}}, \forall h \in \mathbb{R}^n : \|h\|_2 = 1]$
$\implies \bar{X} \in \mathbb{S}^n_+, t_h(G) \le \frac{1}{2}(1 - \frac{1}{n}) < \frac{1}{2}$.

19 2018-07-14 (Make-up Lecture)

19.1 Hypersphere representation of G:

$v : V \to \mathbb{R}^d$,

$$\|v(i)\|_2^2 = t \ \forall i \in V,$$
$$\|v(i) - v(j)\|_2 = 1 \ \forall ij \in E.$$

19.2 Orthonormal representation of $G = (V, E)$

$u : V \to \mathbb{R}^d$ such that

$$\|u(i)\|_2 = 1 \ \forall i \in V$$
$$\langle u(i), u(j) \rangle = 0 \ \forall ij \in \bar{E}$$
$$\overline{E} := \{ij : i, j \in V,\ i \neq j,\ ij \notin E\}$$

The complement of G is $\overline{G} := (V, \overline{E})$.

Given $G = (V, E)$ let $v : V \to \mathbb{R}^d$ be a hypersphere representation of G with $t < \frac{1}{2}$.
Claim: Let $u : V \to \mathbb{R}^{d+1}, u(i) := \sqrt{2} \begin{bmatrix} \sqrt{\frac{1}{2} - t} \\ v(i) \end{bmatrix} \forall i \in V$. Then $u : V \to \mathbb{R}^{d+1}$ is an orthonormal representation of $\bar{G}$.
Proof of claim:

$$\begin{aligned} \forall i \in V, \qquad \|u(i)\|_2^2 \qquad &= 2((\tfrac{1}{2} - t) + \underbrace{\|v(i)\|_2^2}_{=t}) = 1 \\ \forall ij \in E, \quad \langle u(i), u(j) \rangle \quad &= 2(\tfrac{1}{2} - t + \underbrace{\langle v(i), v(j) \rangle}_{=t - \frac{1}{2}}) = 0. \end{aligned}$$

$(\forall ij \in E,\ \|v(i) - v(j)\|_2^2 = 2t - 2\langle v(i), v(j) \rangle = 1.)$ $\Diamond$
Also, every orthonormal representation $u : V \to \mathbb{R}^d$ of G yields a hypersphere representation via

$$v(i) := \frac{1}{\sqrt{2}} u(i) \ \forall i \in V, \ \text{ of } \overline{G}.$$

19.3 Orthonormal Representations and Stable Set Problem

Given a graph $G = (V, E), \mathcal{S} \subseteq V$ is a stable set (independent set) in G if $\forall ij \in E$, at most one of $i, j \in \mathcal{S}$.

Note that $\mathcal{S} \subseteq V$ is a stable set in $G \iff \mathcal{S}$ is a clique in $\overline{G}$.
Incidence vector:

$$(x^{\mathcal{S}})_i := \left.\begin{cases} 1 & \text{if } i \in \mathcal{S} \\ 0 & \text{otherwise} \end{cases}\right\} \in \{0,1\}^V$$

$\text{STAB}(G) := \text{conv}\{x^{\mathcal{S}} : \mathcal{S} \text{ is a stable set in } G\} \leftarrow$ Stable set polytope of G

$\alpha(G) := \max\{|\mathcal{S}| : \mathcal{S} \text{ is a stable set in } G\} \leftarrow \mathcal{NP}$-hard to approximate let alone compute

$\alpha(G)$: stability number of G

$$\alpha(G) = \max\{\bar{e}^\top x : x \in \text{STAB}(G)\}$$

Elementary IP formulation based relaxation:

$\text{FRAC}(G) := \{x \in \mathbb{R}^V : 0 \leq x \leq \bar{e},\ x_i + x_j \leq 1\ \forall ij \in E\} \leftarrow$ Fractional Stable Set polytope

$\text{CLQ}(G) := \{x \in \mathbb{R}^V : 0 \leq x,\ \sum_{i \in \mathcal{C}} x_i \leq 1 \text{ for all cliques } \mathcal{C} \text{ in } V\}$

$\text{CLQ}(G)$: clique polytope of G

$$\text{TH}(G) := \left\{ x \in \mathbb{R}^V_+ : \sum_{i \in V} [c^\top u(i)]^2 x_i \leq 1 \quad \begin{array}{l} \forall \text{ ortho. repr. } u : V \to \mathbb{R}^V \text{ of } G \\ \text{and } \forall c \in \mathbb{R}^V \text{ s.t. } \|c\|_2 = 1 \end{array} \right\}$$

$\text{TH}(G)$: Theta Body of G
$\text{TH}(G)$ is the intersection of $\mathbb{R}^V_+$ with a collection (possibly uncountable) of closed half spaces. Therefore, $\text{TH}(G)$ is a closed convex set.

Theorem (6.6). For every graph G,

$$\text{STAB}(G) \subseteq \text{TH}(G) \subseteq \text{CLQ}(G) \subseteq \text{FRAC}(G).$$

Proof. We already observed $\text{CLQ}(G) \subseteq \text{FRAC}(G)$.
$\text{TH}(G) \subseteq \text{CLQ}(G)$: It suffices to show that for every clique $\mathcal{C}$ in G, the inequality $\sum_{i \in \mathcal{C}} x_i \leq 1$ arises as an orthonormal representation constraint for some $u : V \to \mathbb{R}^V$ and some unit vector c.
Let $\mathcal{C} \subseteq V$ be an arbitrary clique in G. Pick any $c \in \mathbb{R}^V$ s.t. $\|c\|_2 = 1$.
$u(i) := c\ \forall i \in \mathcal{C}$, for the vertices $i \in V \setminus \mathcal{C}$, pick an orthonormal system in $\mathbb{R}^V_+ \cap \{x \in \mathbb{R}^V : c^\top x = 0\}$. Then, $u : V \to \mathbb{R}^V$ is an orthonormal representation of G. The orthonormal representation constraint for u and c is

$$\begin{aligned} 1 &\geq \sum_{i \in V} [c^\top u(i)]^2 x_i \\ &= \sum_{i \in \mathcal{C}} (\underbrace{c^\top c}_{=1})^2 x_i + 0 \\ &= \sum_{i \in \mathcal{C}} x_i. \end{aligned}$$

STAB$(G) \subseteq$ TH(G): We will pick an arbitrary stable set $\mathcal{S}$ in G and show that $x^{\mathcal{S}} \in$ TH(G). (Then since TH(G) is a convex set, by definition of the convex hull and STAB(G), TH$(G) \supseteq$ STAB(G).)
$x^{\mathcal{S}} \in \mathbb{R}^V_+$, pick an arbitrary orthonormal representation $u : V \to \mathbb{R}^V$ of G and an arbitrary $c \in \mathbb{R}^V$ such that $\|c\|_2 = 1$. Then

$$\begin{aligned}\sum_{i \in V} [c^\top u(i)]^2 (x^{\mathcal{S}})_i &= \sum_{i \in \mathcal{S}} [c^\top u(i)]^2 \\ &= \|\mathcal{U}_{\mathcal{S}} c\|_2^2 \\ &\leq \underbrace{\|\mathcal{U} c\|_2^2}_{=\|c\|_2^2 = 1}\end{aligned}$$

Aside:

$$\begin{aligned}\mathcal{U}_{\mathcal{S}}^\top &:= [u(i) : i \in \mathcal{S}] \in \mathbb{R}^{V \times \mathcal{S}} \\ \mathcal{U}^\top &:= [u(i) : i \in \mathcal{S}, \underbrace{*}_{\text{complete to an orthonormal basis for } \mathbb{R}^V}] \in \mathbb{R}^{V \times V}\end{aligned}$$

□

Given $G = (V, E), w \in \mathbb{R}^V_+$,

$$\theta(G, w) := \max\{w^\top x : x \in \text{TH}(G)\}.$$

Note: $\max\{w^\top x : x \in \text{STAB}(G)\} \leq \theta(G, w)$ since STAB$(G) \subseteq$ TH(G).

Theorem (6.7). For every graph $G = (V, E)$ and for every $w \in \mathbb{R}^V_+$, the following are all equal:

(i)

$$\theta(G, w)$$

(ii)

$$\min_{\substack{\forall u : V \to \mathbb{R}^V \\ \text{ortho. repr.} \\ u \text{ of } G, \\ \forall c \in \mathbb{R}^V : \|c\|_2 = 1}} \max_{i \in V} \left\{ \frac{w_i}{[c^\top u(i)]^2} \right\}.$$

(iii)

$$\min\{\lambda_1(S + W) : \text{diag}(S) = 0,\ S_{ij} = 0\ \forall ij \in \overline{E},\ S \in \mathbb{S}^V\}$$

where $W \in \mathbb{S}^V$, $W_{ij} = \sqrt{w_i w_j}\ \forall i, j \in V$.

(iv)

$$\max\{\langle W, X \rangle : X_{ij} = 0, \forall \{i, j\} \in E;\ \text{Tr}(X) = 1;\ X \succeq 0\}$$

19.4 Stable Set Problem and Shannon Capacity of a Channel

Suppose two people are communicating over a noisy channel. We have an alphabet where some pairs of letters may be confused with each other. Construct a graph $G = (V, E)$ with one vertex for each letter and put an edge between vertex i and vertex j if the corresponding letters may be confused with each other. Then $\alpha(G)$ is the maximum number of letters we may use without confusion.
Two k-letter words may not be confused with each other if there is a position ℓ in which these two words differ and the corresponding ℓ-th letters may not be confused with each other.

Strong Products of Graphs: for $G = (V, E), H = (W, F)$,

$$(G \otimes H) := (V \times W, E(G \otimes H))$$
$$E(G \otimes H) = \left\{ \{(i,u),(j,v)\} \ : \begin{array}{l} ij \in E \text{ and } uv \in F \text{ or} \\ ij \in E \text{ and } u = v \text{ or} \\ i = j \text{ and } uv \in F \end{array} \right\}$$

$$G^k := \underbrace{G \otimes G \otimes \cdots \otimes G}_{k \text{ times}}$$

The maximum number of k-letter words that can be communicated without confusion is $\alpha(G^k)$.
Shannon Capacity of $G := \Theta(G) := \lim_{k \to +\infty} [\alpha(G^k)]^{\frac{1}{k}}$.
Note that if $\mathcal{S}_1 \subseteq V$ is a stable set in G and $\mathcal{S}_1 \subseteq W$ is a stable set in H, then $(\mathcal{S}_1 \times \mathcal{S}_2)$ is a stable set in $G \otimes H$.
$\Longrightarrow \alpha(G^k) \geq [\alpha(G)]^k$.
This last observation implies $\Theta(G) \geq \alpha(G)$.
Ex: $G = C_5$ (the 5-cycle), $\alpha(C_5) = 2,\ \alpha(C_5^2) = 5$.
Lovász [1979] proved $\Theta(C_5) = \sqrt{5}$ via computing $\theta(C_5, \bar{e})$.

Lemma (stronger version of 6.11). For all graphs G, H,

$$\theta(G \otimes H) = \theta(G)\theta(H),$$

where $\theta(G) := \theta(G, \bar{e})$.

Theorem (6.12). $\forall$ graphs $G = (V, E)$,

$$\theta(G) = \begin{array}{rrcl} \max & \langle \bar{e}\bar{e}^\top, X \rangle & & \\ \text{s.t.} & X_{ij} & = & 0 \ \forall ij \in E, \\ & \text{Tr}(X) & = & 1, \\ & X & \in & \mathbb{S}_+^V \end{array} = \begin{array}{rrcl} \min & t & & \\ \text{s.t.} & \text{diag}(Z) & = & (t-1)\bar{e}, \\ & Z_{ij} & = & -1 \ \forall ij \in \overline{E}, \\ & Z & \succeq & 0. \end{array}$$

Moreover, $\alpha(G) \leq \Theta(G) \leq \theta(G) \leq \chi(\overline{G})$. Equality holds if G is perfect.

20 2018-07-17

k-colouring of G is $\sigma : V \to \{1, 2, \ldots, k\}$ such that $\forall ij \in E, \sigma(i) \neq \sigma(j)$.
A graph G is perfect if for every vertex-induced subgraph H of G,

$$\alpha(\overline{H}) = \chi(H).$$

Theorem (6.13). For every graph $G = (V, E)$,

$$\theta(G) = \begin{array}{rrcl} \max & \langle \bar{e}\bar{e}^\top, X \rangle & & \\ \text{s.t.} & X_{ij} & = & 0 \ \forall ij \in E, \\ & \mathrm{Tr}(X) & = & 1, \\ & X & \in & \mathbb{S}_+^V \end{array} = \begin{array}{rrcl} \min & t & & \\ \text{s.t.} & \mathrm{diag}(Z) & = & (t-1)\bar{e}, \\ & Z_{ij} & = & -1 \ \forall ij \in \overline{E}, \\ & Z & \succeq & 0. \end{array}$$

Moreover, $\alpha(G) \leq \Theta(G) \leq \theta(G) \leq \chi(\overline{G})$. Equality holds if G is perfect.

Note that SDPs in the statement of the theorem are dual to each other. Let $\mathcal{S} \subseteq V$ be a stable set in G.

$$\overline{X} := \begin{bmatrix} \frac{1}{\mathcal{S}}\bar{e}\bar{e}^\top & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{S}_+^V \text{ is feasible in the primal SDP.}$$

The objective value of $\overline{X}$ is: $\frac{(\bar{e}^\top \bar{e})^2}{|\mathcal{S}|} = |\mathcal{S}|$
$\implies \theta(G) \geq \alpha(G)$.
Take a k-colouring of $\overline{G}$.

$$Z_{ij} := \begin{cases} (k-1) & \text{if colour}(i) = \text{colour}(j) \\ -1 & \text{if colour}(i) \neq \text{colour}(j) \end{cases}$$

$$Z = \left[\begin{array}{ccc:ccc:c:ccc} (k-1) & \cdots & (k-1) & -1 & \cdots & -1 & & & & -1 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & & & & \\ (k-1) & \cdots & (k-1) & -1 & \cdots & -1 & & & & \\ \hdashline -1 & & & (k-1) & \cdots & (k-1) & & & & \\ & & & \vdots & \ddots & \vdots & & & & \\ & & & (k-1) & \cdots & (k-1) & & & & \\ \hdashline \vdots & & & & & & \ddots & & & \\ \hdashline & & & & & & & (k-1) & \cdots & (k-1) \\ & & & & & & & \vdots & \ddots & \vdots \\ -1 & & & & & & -1 & (k-1) & \cdots & (k-1) \end{array}\right]$$

Note that nonsingular symmetric minors of $\overline{Z}$ may have at most one row-column from each colour class. Therefore, every nonsingular symmetric minor of $\overline{Z}$ is:

$$\begin{bmatrix} (k-1) & -1 & -1 & \cdots & -1 \\ -1 & (k-1) & & & \\ \vdots & & \ddots & & \\ -1 & & \cdots & -1 & (k-1) \end{bmatrix}$$

and all such minors are psd since they are diagonally dominant.

$\implies \overline{Z} \succeq 0$ and $(\overline{Z}, \bar{t} := k)$ is feasible in the dual.

This proves $\theta(G) \leq \chi(\overline{G})$.
Recall for $G \subseteq \mathbb{R}^d$ the polar of G is $G^\circ := \{s \in \mathbb{R}^d : x^\top s \leq 1 \ \forall x \in G\}$.

Theorem (6.9). For every graph $G = (V, E)$,

$$[\mathrm{TH}(G)]^\circ \cap \mathbb{R}^V_+ = \mathrm{TH}(\overline{G}).$$

TH(G) can be represented as a projection of the feasible region of an SDP. For every graph $G = (V, E)$,

$$\widehat{\mathrm{TH}}(G) := \{Y \in \mathbb{S}_+^{\{0\}\cup V} : Y_{00} = 1,\ \mathrm{diag}(Y) = Ye_0,\ Y_{ij} = 0\ \forall ij \in E\}.$$

$Y \in \widehat{\mathrm{TH}}(G)$ then

$$Y = \left[\begin{array}{c|cccc} 1 & x_1 & x_2 & \cdots & x_n \\ \hline x_1 & x_1 & 0 & \cdots & 0 \\ x_2 & 0 & x_2 & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ x_n & 0 & \cdots & 0 & x_n \end{array}\right].$$

Theorem (6.10). For every graph $G = (V, E)$,

$$\mathrm{TH}(G) = \{x \in \mathbb{R}^V : Ye_0 = \begin{pmatrix} 1 \\ x \end{pmatrix},\ Y \in \widehat{\mathrm{TH}}(G).\}$$

An odd-hole is an odd cycle of length at least 5 with no chords.
An odd-antihole is the complement of an odd hole.

Theorem (6.8). Let $G = (V, E)$ be a graph. Then TFAE:

(i) G is perfect

(ii) $\overline{G}$ is perfect

(iii) G does not contain an odd-hole, or odd-antihole

(iv) $\mathrm{CLQ}(G) = \mathrm{STAB}(G)$

(v) defining linear inequality system for $\mathrm{CLQ}(G)$ is TDI (Totally Dual Integral)

(vi) $\mathrm{TH}(G) = \mathrm{STAB}(G)$

(vii) $\mathrm{TH}(G) = \mathrm{CLQ}(G)$

(viii) $\mathrm{TH}(G)$ is a polytope

(ix) defining system of $\widehat{\mathrm{TH}}(G)$ is TDI

(x) $\{x_i^2 - x_i \ \forall i \in V;\ x_i x_j \ \forall ij \in E\}$ is $(1,1)$-SOS (Algebraic Geometry)

(xi) $\forall$ probability distributions p on V,

$$H(p) = H(G,p) + H(\overline{G},p)$$

where $H(G,p)$ is the graph entropy. $H(p) := -\sum_{i\in V} p_i \ln p_i$. (Information theory)

21 2018-07-19

Finish reading Chapters 7 & 9. Start reading chapters 10, 8, 12.

21.1 Lift-and-Project Methods for Combinatorial Optimization

Recall, for every graph $G = (V,E)$ we have

$$\widehat{\mathrm{TH}}(G) := \{Y \in \mathbb{S}_+^{\{0\}\cup V} : Y_{00} = 1,\ \mathrm{diag}(Y) = Ye_0,\ Y_{ij} = 0\ \forall ij \in E\}.$$

Consider $Y \in \widehat{TH}(G)$ with $\mathrm{rank}(Y) = 1$. Then

$$Y = \begin{bmatrix} 1 & x^\top \\ x & xx^\top \end{bmatrix} \text{ for some } x \underbrace{\in \{0,1\}^V}_{\text{(we used } Ye_0 = \mathrm{diag}(Y))} \cap \mathrm{STAB}(G)$$

$$xx^\top = \left[\ x_1 x \ \vdots\ x_2 x \ \vdots\ \cdots \ \vdots\ x_n x\ \right] = \left[x_i x_j \ :\ i,j \in V.\right]$$

We can add more constraints on Y to tighten our relaxation of $\mathrm{STAB}(G)$ given by $\mathrm{TH}(G)$. We can require that the columns of Y satisfy the constraints of $\mathrm{FRAC}(G)$. We can enforce $Ye_i, Y(e_0 - e_i) \in \mathrm{cone}(\{1\} \oplus \mathrm{FRAC}(G))$ (the smallest convex cone containing the argument.)

$$\left\{ \begin{pmatrix} x_0 \\ x \end{pmatrix} \in \mathbb{R} \oplus \mathbb{R}^V \ :\ x_0 = 1 \right\}$$

Suppose $P := \{x \in \mathbb{R}^n \ :\ Ax \leq b,\ 0 \leq x \leq \bar{e}\}$

$$\mathrm{cone}(\{1\} \oplus P) = \left\{ \begin{pmatrix} x_0 \\ x \end{pmatrix} \in \mathbb{R} \oplus \mathbb{R}^V \ :\ Ax \leq x_0 b,\ 0 \leq x \leq x_0 \bar{e},\ x_0 \geq 0 \right\}$$

If $P \neq \varnothing$,

$$\text{cone}(\{1\} \oplus P) = \text{cl}\left\{\begin{pmatrix} x_0 \\ x \end{pmatrix} \in \mathbb{R} \oplus \mathbb{R}^V \ : \ Ax \leq x_0 b,\ 0 \leq x \leq x_0 \bar{e},\ x_0 > 0\right\}$$

$$\underbrace{\text{LS}_+(G)}_{\text{(Lovász \& Schrijver)}} = \left\{x \in \mathbb{R}^V \ : \ \begin{pmatrix} 1 \\ x \end{pmatrix} = Ye_0,\ \text{diag}(Y) = Ye_0,\right.$$

$$\left. Ye_i, Y(e_0 - e_i) \in \text{cone}(\{1\} \oplus \text{FRAC}(G)) \forall i \in V,\ Y \in \mathbb{S}_+^{\{0\} \cup V}\right\}$$

We can apply this construction to any combinatorial optimization problem. Consider the 0,1 IP problem

$$\begin{array}{lrcl} \max & c^\top x & & \\ \text{s.t.} & Ax & \leq & b, \\ & 0 \leq x & \leq & \bar{e}, \\ & x & \in & \{0,1\}^n. \end{array}$$

$$P := \{x \in \mathbb{R}^n \ : \ Ax \leq b,\ 0 \leq x \leq \bar{e}\}.$$

We want $\max\{c^\top x \ : \ x \in \text{conv}(P \cap \{0,1\}^n)\}$.

$$\underbrace{\text{LS}_+(G)}_{\text{(Lovász \& Schrijver)}} = \left\{x \in \mathbb{R}^n \ : \ \begin{pmatrix} 1 \\ x \end{pmatrix} = Ye_0,\ \text{diag}(Y) = Ye_0,\right.$$

$$\left. Ye_i, Y(e_0 - e_i) \in \text{cone}(\{1\} \oplus P) \forall i \in \{1, 2, \ldots, n\},\ Y \in \mathbb{S}_+^{1+n}\right\}$$

Note that LS_+ : subsets of $[0,1]^n \to$ convex subsets of $[0,1]^n$. We can apply LS_+ iteratively:

$$\text{LS}_+^{k+1}(P) := \text{LS}_+^k(\text{LS}_+(P)),\ k \in \mathbb{Z}_+,\ \text{LS}_+^0(P) := P.$$

Take $x \in \text{LS}_+(P)$. Then $\exists Y$ s.t. $\begin{pmatrix} 1 \\ x \end{pmatrix} = \underbrace{Ye_i}_{\in \text{cone}(\{1\} \oplus P)} + \underbrace{Y(e_0 - e_i)}_{\in \text{cone}(\{1\} \oplus P)}$ for all $i \in \{1, \ldots, n\}$.

$$\begin{pmatrix} 1 \\ x \end{pmatrix} = \underbrace{Ye_i}_{\in \text{cone}(\{1\} \oplus P) \cap \{\left(\begin{smallmatrix} x_0 \\ x \end{smallmatrix}\right) : x_i = x_0\}} + \underbrace{Y(e_0 - e_i)}_{\in \text{cone}(\{1\} \oplus P) \cap \{\left(\begin{smallmatrix} x_0 \\ x \end{smallmatrix}\right) : x_i = 0\}} \quad \text{for all } i \in \{1, \ldots, n\}$$

$$Y(e_0 - e_i) = \begin{bmatrix} 1 - x_i \\ \vdots \\ x_i - x_i = 0 \\ \vdots \end{bmatrix} \qquad Ye_i = \begin{bmatrix} x_i \\ \vdots \\ x_i \\ \vdots \end{bmatrix}$$

In the space of P, this means:

$\mathrm{LS}_+(P) \subseteq \mathrm{conv}[(P \cap \{x \in \mathbb{R}^n : x_i = 1\}) \cup (P \cap \{x \in \mathbb{R}^n : x_i = 0\})] \forall i \in \{1, 2, \ldots, n\}$.

We can show $\mathrm{LS}_+^2(P) \subseteq \mathrm{conv}[(P \cap \{x \in \mathbb{R}^n : x_i = 1, x_j = 1\}) \cup (P \cap \{x \in \mathbb{R}^n : x_i = 1, x_j = 0\}) \cup (P \cap \{x \in \mathbb{R}^n : x_i = 0, x_j = 1\}) \cup (P \cap \{x \in \mathbb{R}^n : x_i = 0, x_j = 0\})]$.
This leads to

Lemma (7.8). For every polytope $P \subseteq [0,1]^n$,

$$\mathrm{LS}_+^n(P) = \mathrm{conv}(P \cap \{0,1\}^n).$$

22 2018-07-24

Finish reading Chapters 10, 8. Read Chapter 12.

Theorem (7.10). Let $P \subseteq [0,1]^d$ be a polytope. Then

$$\mathrm{LS}_+^d = \mathrm{conv}(P \cap \{0,1\}^d).$$

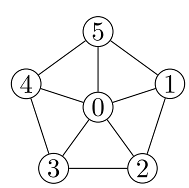
$G = (V, E)$ a given graph.

$$\underbrace{\mathrm{OC}(G)}_{\text{odd-cycle polytope}} := \{x \in [0,1]^V : \sum_{i \in H} x_i \leq \lfloor \tfrac{|H|-1}{2} \rfloor, \text{ for every odd-cycle } H \text{ in } G\}$$

$$\text{ANTI-HOLE}(G) := \{x \in [0,1]^V : \sum_{i \in H} x_i \leq 2, \text{ for every odd anti-hole } H \text{ in } G\}$$

An <u>odd-wheel</u> in G is a vertex induced subgraph $H =: \{v_0, v_1, \ldots, v_{2k+1}\}$ such that

$$\text{WHEEL}(G) := \{x \in [0,1]^V : kx_{v_0} + \sum_{i=1}^{2k+1} x_{v_i} \leq k, \text{ for every odd-wheel } \{v_0, v_1, \ldots, v_{2k+1}\} \text{ in } G\}$$



Theorem (8.21). For every graph G,

$$\mathrm{LS}_+(G) \subseteq \mathrm{TH}(G) \cap \underbrace{\mathrm{OC}(G) \cap \mathrm{WHEEL}(G) \cap \text{ANTI-HOLE}(G)}_{\text{Each of these require exponentially many linear inequalities to describe in the worst case}}.$$

LS_+ has been generalized to solve $\min f(x)$ (f continuous) over $x \in F$ (F compact) $\in \mathbb{R}^d$.
A nice special case is PoP (Polynomial Optimization Problems). Let $p_0, p_1, \ldots, p_m : \mathbb{R}^d \to \mathbb{R}$ be polynomials.

$$\begin{array}{llll} (PoP) & \inf & p_0(x) & \\ & \text{s.t.} & p_i(x) & \geq \ 0 \ \forall i \in \{1, 2, \ldots, m\} \end{array}$$

Every PoP can also be put into the form

$$\begin{array}{llll} (PoP) & \inf & p_0(x) & \\ & \text{s.t.} & p_i(x) & = \ 0 \ \forall i \in \{1, 2, \ldots, m\} \end{array}$$

This problem is equivalent to $\inf_{x \in \mathbb{R}^d} p_0 + \mu \sum_{i=1}^m [p_i(x)]^2$, where $\mu > 0$ is a parameter.
Deciding on the minimum value of a multivariate polynomial is equivalent to deciding on the optimal value of a PoP.
Given $\bar{z} \in \mathbb{R}$, is $[p(x) - \bar{z}] \geq 0 \ \forall x \in \mathbb{R}^d$?
Hilbert's 17th question was answered by Artin:

Theorem (10.1). Let $p : \mathbb{R}^d \to \mathbb{R}$ be a polynomial. Then $p(x) \geq 0 \ \forall x \in \mathbb{R}^d$ iff $\exists$ polynomials $h_0, h_1, \ldots, h_k : \mathbb{R}^d \to \mathbb{R}$ such that

$$p(x) = \sum_{i=1}^k \left(\frac{h_i(x)}{h_0(x)} \right)^2.$$

Obviously, if $\exists$ polynomials $h_1, \ldots, h_k$ such that $p(x) = \sum_{i=1}^k [h_i(x)]^2$ then $p(x) \geq 0 \ \forall x \in \mathbb{R}^d$. Given $p : \mathbb{R}^d \to \mathbb{R}$ polynomial of degree $2n$,

$$\underbrace{\begin{bmatrix} 1 & x_1 & x_2 & \cdots & x_d & x_1^2 & x_1 x_2 & \cdots & x_d^n \end{bmatrix}}_{=:[g(x)]^\top} \underbrace{X}_{\in \mathbb{S}^N} \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \cdots \\ x_d \\ x_1^2 \\ x_1 x_2 \\ \cdots \\ x_d^n \end{bmatrix}$$

$N := \binom{n+d}{d}$.
Using this equation, we get linear equations on the entries of X.

$$\mathcal{F}(p) := \{X \in \mathbb{S}_+^N : \underbrace{[g(x)]^\top X g(x) = p(x)}_{\iff \mathcal{A}(x) = b}\}$$

If $\mathcal{F}(p) \neq \varnothing$, then $\exists B \in \mathbb{R}^{N \times N}$ such that $X = BB^\top$ and $p(x) = \left\| B^\top g(x) \right\|_2^2 \geq 0$.

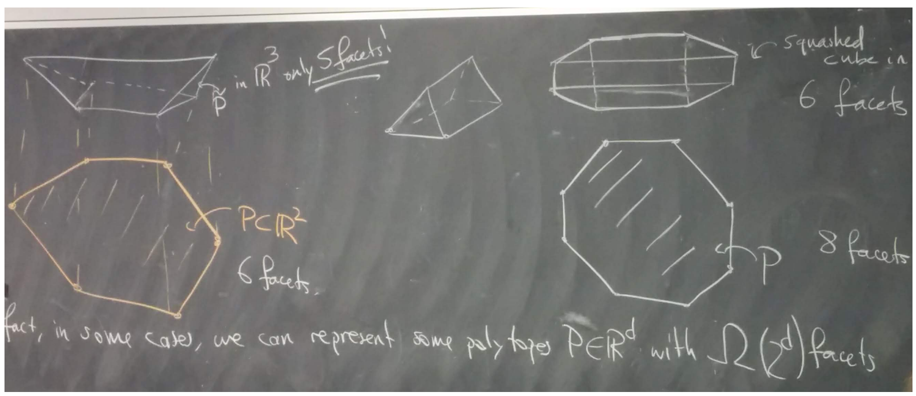
Theorem (10.2). Let $\bar{z} \in \mathbb{R}$ and $p : \mathbb{R}^d \to \mathbb{R}$ be a polynomial. Then, $p(x) - \bar{z}$ is SoS (a sum of squares of polynomials) iff $\{X \in \mathcal{F}(p) : X \succeq \bar{z} e_1 e_1^\top\} \neq \varnothing$.

Recall

Theorem (8.21). $\forall$ graphs G,

$$\mathrm{LS}_+(G) \subseteq \mathrm{TH}(G) \cap \mathrm{OC}(G) \cap \mathrm{WHEEL}(G) \cap \mathrm{ANTI\text{-}HOLE}(G).$$

Note that every d-dimensional polytope has a unique facetal description in $\mathbb{R}^d$. $p = \{x \in \mathbb{R}^d : Ax \leq b\}$.



In fact, in some cases, we can represent some polytopes $P \in \mathbb{R}^d$ with $\Omega(2^d)$ facets as a projection of $\widetilde{P} \subset \mathbb{R}^{O(d^2)}$ with $O(d^3)$ facets. E.g. $\forall$ graphs G, $\mathrm{LS}(G) = \mathrm{OC}(G)$.
Given a polytope $P \subset \mathbb{R}^d$, we can try to construct $\widetilde{P} \subset \mathbb{R}^N$ such that $P = \mathcal{L}(\widetilde{P} \cap U)$, where $U \subset \mathbb{R}^N$ is an affine subspace and $\mathcal{L} : \mathbb{R}^N \to \mathbb{R}^d$ is a linear map.
For example, $P = \{x \in \mathbb{R}^d : Ax + Fu = b,\ u \geq 0\}$. $\widetilde{P} := \{\binom{x}{u} : u \geq 0\}$. If we wanted to solve $\min_{x \in P} c^\top x$, we could equivalently solve

$$\begin{array}{lrcl} \min & [c^\top 0] \begin{bmatrix} x \\ u \end{bmatrix} & & \\ \text{s.t.} & Ax + Fu & = & b, \\ & u & \geq & 0. \end{array}$$

Let $P \subset \mathbb{R}^d$ be a given polytope such that $P = \{x \in \mathbb{R}^d : Ax \leq b\}$ (facetal description).
Let $m := |\mathcal{F}|$ (# of facets), $n := |\mathcal{V}|$ (extreme points of P).
$S \in \mathbb{R}_+^{m \times n}$, <u>slack matrix of P</u>, $S_{ij} := b_i - a_i^\top v^{(j)}$ $\forall i, j$ where $v^{(j)} \in \mathcal{V}$ and $a_i^\top x \leq b_i$ is a facet.
Given $S \in \mathbb{R}_+^{m \times n}$, a <u>nonnegative factorization</u> of S is $F \in \mathbb{R}_+^{m \times k}, V \in \mathbb{R}_+^{n \times k}$ for some positive integer k such that $S = FV^\top$. Smallest such k is callled the nonnegative rank of S (and P); $\mathrm{rank}_+(S), \mathrm{rank}_+(P)$.

Theorem (Yannakakis 1989). Let $P \subset \mathbb{R}^d$ be a polytope, and $k := \mathsf{rank}_+(P)$. Then every lifted representation of P uses at least k constraints. Moreover, P has a lifted representation using at most $(k+d)$ constraints on $(k+d)$ variables.