Semi-definite programming approaches

Definition

A semi-definite program is an optimization program of the form

Minimize	$\langle C, X\rangle$
Over	$X \in \mathcal{S}_n^+$
Such that	$\langle A_i, X\rangle = b_i,\ i = 1, \dots, m,$

where $\mathcal{S}_n^+$: cone of semi-definite positive symmetric $n \times n$ matrices,
$\langle A, B\rangle = \mathrm{Tr}(AB^\top)$: Frobenius scalar product between matrices,
$C, A_1, \dots, A_m$: symmetric $n \times n$ matrices, and $b_1, \dots, b_m \in \mathbb{R}$.

Key properties:

- it is a convex minimization problem, since $\mathcal{S}_n^+$ is a convex cone.
- It can be solved in polynomial time (e.g. with the ellipsoid method initially developed by Kamarkar to solve linear problems): a solution within additive error of ϵ can be found in time polynomial in $n,\ m,\ \log(1/\epsilon)$.

Approach:

NP-hard combinatorial optimization problem $\mathcal{C} \rightarrow$ relaxation into (convex) SDP $\mathcal{RC} \rightarrow$ solution of $\mathcal{RC}$ + post-processing $\rightarrow$ approximate solution of $\mathcal{C}$ with bounded sub-optimality

Example: Max-cut

Given graph $G = (V, E)$, max-cut problem: find partition of V into V_-, V_+ maximizing number $|E(V_+, V_-)|$ of edges across partition.

Max-cut is NP-complete. In contrast, Min-cut is in P (thanks to the max-flow min-cut theorem).

Let A: adjacency matrix of G. Size $MC(G)$ of Max-cut: solution of

Maximize	$\frac{1}{4}\sum_{u,v\in[n]} A_{uv}[1 - Y_{uv}]$
Over	$Y \in \mathcal{S}_n^+$
Such that	$Y_{uv} = \sigma_u \sigma_v,\ u, v \in [n]$
where	$\sigma_u \in \{+1, -1\},\ u \in [n]$.

Equivalent formulation:

Maximize $\frac{1}{4} \sum_{u,v \in [n]} A_{uv}[1 - Y_{uv}]$
Over $Y \in \mathcal{S}_n^+$
Such that $Y_{uv} \in \{-1, +1\}, \; \mathrm{rank}(Y) = 1.$

Goemans-Williamson algorithm: consider SDP relaxation

Maximize $\frac{1}{4} \sum_{u,v \in [n]} A_{uv}[1 - Y_{uv}]$
Over $Y \in \mathcal{S}_n^+$
Such that $Y_{uu} = 1, \; u \in [n].$

Corresponding value: $GW(G)$. By construction, $GW(G) \geq MC(G)$.

Let Y^*: optimal solution of SDP, and $Z = (z_1 \cdots z_n)$: a symmetric square root of Y^*. Hence $Y^*_{uv} = \langle z_u, z_v \rangle$.

Constraint $Y_{uu} = 1$ guarantees that $\|z_u\| = 1$, i.e. $z_u \in \mathcal{S}^{n-1}$, unit sphere of $\mathbb{R}^n$.

Goemans-Williamson algorithm (continued)

Randomized algorithm:

1. Pick vector z uniformly at random on $\mathcal{S}^{n-1}$. Construct the sign vector $\sigma := \{\text{sign}(\langle z, z_u \rangle)\}_{u \in [n]} \in \{-1, 1\}^n$. Let $C(\sigma)$ denote the corresponding cut-size.
2. Repeat N times.
3. Output the largest cut obtained over the N iterations.

Theorem

The expected value $\mathbb{E}(C(\sigma))$ of the cut resulting from one random sign vector σ as above is larger than $0.878MC(G)$.

Proof.

For $u \neq v \in [n]$, $\mathbb{P}(\sigma_u = +, \sigma_v = -) = \frac{\arccos\langle z_u, z_v\rangle}{2\pi}$. Thus:

$$\mathbb{E}C(\sigma) = \frac{1}{2}\sum_{u,v\in[n]} A_{uv} 2\frac{\arccos\langle z_u, z_v\rangle}{2\pi} = \frac{1}{2\pi}\sum_{u,v\in[n]} A_{uv} \arccos\langle z_u, z_v\rangle.$$

It can be verified by calculus that

$$\forall x \in [-1, 1],\ \frac{1}{\pi}\arccos(x) \geq \beta\frac{1}{2}(1 - x), \text{ where } \beta = 0.87856.$$

This implies that $\mathbb{E}C(\sigma) \geq \beta GW(G)$. The result then follows from $GW(G) \geq MC(G)$. □

Corollary

For $\epsilon > 0$, the above randomized algorithm outputs a cut that is less than $(1-\epsilon)\beta MC(G)$ with probability at most $[1+\epsilon\bar{C}/(n^2/4-\bar{C})]^{-N}$, where $\bar{C} := \mathbb{E}C(\sigma)$. Thus for large enough N, it produces w.h.p. a cut achieving a fraction at least $\beta(1-\epsilon)$ of the optimum $MC(G)$.

Proof: let $p = \mathbb{P}(C(\sigma) \leq (1-\epsilon)\bar{C})$. Since $C(\sigma) \in [0, n^2/4]$, one has $p(1-\epsilon)\bar{C} + (1-p)n^2/4 \geq \bar{C}$, so that $p \leq [n^2/4 - \bar{C}]/[n^2/4 - \bar{C} + \epsilon\bar{C}]$, i.e. $\frac{1}{p} \geq 1 + \frac{\epsilon\bar{C}}{n^2/4-\bar{C}}$.

Remark

See [Mohar 97, 'Some applications of Laplace eigenvalues of graphs'] for more details, e.g. a derandomization scheme due to Goemans and Williamson allowing to find a cut of size at least $\beta MC(G)$ with probability 1 in polynomial time.

Grothendieck inequality

For rectangular matrix $M \in \mathbb{R}^{n\times m}$, define its norm $\|M\|_{\infty\to 1}$ as $\|M\|_{\infty\to 1} := \sup_{x_i,y_j\in\{-1,1\}} \sum_{i\in[n],j\in[m]} M_{ij}x_iy_j$.

Semi-definite relaxation of combinatorial optimization defining $\|M\|_{\infty\to 1}$:

Maximize	$\sum_{i\in[n],j\in[m]} M_{ij}\langle u_i, v_j\rangle$
Over	$u_i, v_j \in \mathbb{R}^{n+m}$
Such that	$\|u_i\| = 1,\ \|v_j\| = 1.$

It is indeed an SDP, in view of equivalent formulation:

Maximize	$\langle \hat{M}, Y\rangle$
Over	$Y \in \mathcal{S}^+_{n+m}$
Such that	$Y_{uu} = 1,\ u \in [n+m]$,

where $\hat{M} = \frac{1}{2}\begin{pmatrix} 0 & M^\top \\ M & 0 \end{pmatrix}$:

Considering a square root of matrix Y, its first m columns as vectors v_j, its last n columns as vectors u_i, then $Y_{m+i,j} = \langle u_i, v_j\rangle$, hence equivalence of the two optimization problems.

Grothendieck inequality

Denote by $f(M)$ value of previous SDP. Clearly, $f(M) \geq \|M\|_{\infty\to 1}$. One then has:

Theorem (Grothendieck inequality)

For any $n \times m$ real matrix M, the optimal value $f(M)$ satisfies
$f(M) \leq K_G \|M\|_{\infty\to 1}$
for some universal constant K_G, with $K_G \leq \frac{\pi}{2\ln(1+\sqrt{2})} = 1.783\cdots$.

Denote $\mathcal{M} = \{(\langle u_i, v_j \rangle)_{i\in[n], j\in[m]} \in \mathbb{R}^{n\times m}, \|u_i\| \equiv \|v_j\| \equiv 1\}$, so that $f(M) = \sup_{Y\in\mathcal{M}} \langle M, Y\rangle$.

Denote $\mathcal{M}^+ := \{Y \in \mathcal{S}_n^+ : \mathrm{Diag}(Y) \preceq I_n\}$.
Then for $n = m$, $\mathcal{M}^+ \subset \mathcal{M}$ and $-\mathcal{M}^+ \subset \mathcal{M}$, yielding

Corollary

For $M \in \mathbb{R}^{n\times n}$, $\sup_{Y\in\mathcal{M}^+} |\langle M, Y\rangle| \leq K_G \|M\|_{\infty\to 1}$.

Proof of Grothendieck's inequality

Lemma

Let $u, v \in \mathcal{S}^{n-1}$, normed vectors of $\mathbb{R}^n$, and z: uniformly distributed over $\mathcal{S}^{n-1}$. Then $\frac{\pi}{2}\mathbb{E}\left[Sign(\langle u, z\rangle)Sign(\langle v, z\rangle)\right] = arcsin(\langle u, v\rangle)$.

Lemma

Let $u_i,\ i \in [n]$, $v_j,\ j \in [m]$ be normed vectors of some Hilbert space H (wlog, $H = \mathbb{R}^{n+m}$). Let $c := sinh^{-1}(1) = \ln(1+\sqrt{2})$.
Then there exist normed vectors $u'_i,\ i \in [n]$, $v'_j,\ j \in [n]$ of some Hilbert space H' such that for z uniformly distributed on the unit sphere of H', $\forall i \in [n],\ j \in [m],\ \frac{\pi}{2}\mathbb{E}\left[Sign(\langle u'_i, z\rangle)Sign(\langle v'_j, z\rangle)\right] = c\langle u_i, v_j\rangle$.

Proof of second lemma

Write $\sin(c\langle u,v\rangle) = \sum_{k\geq 0}(-1)^k \frac{c^{2k+1}}{(2k+1)!}\langle u,v\rangle^{2k+1}$.
For $w \in H$, $j \in \mathbb{N}$, write $w^{\otimes j} = w \otimes \cdots \otimes w$.

Use fact $\langle u^{\otimes j}, v^{\otimes j}\rangle = \langle u,v\rangle^j$ to obtain
$\sin(c\langle u,v\rangle) = \sum_{k=0}^{\infty}(-1)^k \frac{c^{2k+1}}{(2k+1)!}\langle u^{\otimes 2k+1}, v^{\otimes 2k+1}\rangle \quad (*)$.

Let $H' := \oplus_{k=0}^{\infty} H^{\otimes 2k+1}$.

Define functions $S, T : H \to H'$ by

$$\begin{cases} T(u) = \{T(u)_k\}_{k\geq 0} \text{ where } T(u)_k & = (-1)^k \sqrt{\frac{c^{2k+1}}{(2k+1)!}}\, u^{\otimes 2k+1}, \\ S(v) = \{S(v)_k\}_{k\geq 0} \text{ where } S(v)_k & = \sqrt{\frac{c^{2k+1}}{(2k+1)!}}\, v^{\otimes 2k+1}. \end{cases}$$

Hence by $(*)$: $\langle T(u), S(v)\rangle = \sin(c\langle u,v\rangle)$.

Note $\|T(u)\|^2 = \sum_{k\geq 0} \frac{c^{2k+1}}{(2k+1)!} \|u^{\otimes 2k+1}\|^2 = \sinh(c\|u\|^2)$.
Similarly, $\|S(v)\|^2 = \sinh(c\|v\|^2)$.
Hence for normed $u_i,\ v_j$ and $c = \sinh^{-1}(1)$, vectors $u'_i := T(u_i)$, $v'_j := S(v_j)$ are normed, and such that
$\forall i \in [n],\ j \in [m],\ \arcsin(\langle u'_i, v'_j\rangle) = c\langle u_i, v_j\rangle$.
Together with first lemma, this concludes proof of second lemma.

Proof of Grothendieck's inequality: For matrix $M \in \mathbb{R}^{n+m}$, normed vectors u_i, v_j such that $f(M) = \sum_{i,j} M_{ij}\langle u_i, v_j\rangle$, construct vectors u'_i, v'_j as per second lemma. Then for random vector z uniform on unit sphere of H',
$f(M) = \sum_{i,j} M_{ij}\langle u_i, v_j\rangle = \frac{\pi}{2c}\mathbb{E}\sum_{i,j} M_{ij}\mathrm{Sign}(\langle u'_i, z\rangle)\mathrm{Sign}(\langle v'_j, z\rangle)$.
Thus there must exist some signs $s_i, t_j \in \{-1, 1\}$ such that
$\frac{\pi}{2c}\sum_{ij} M_{ij}s_it_j \geq f(M)$.

Remark

Proof provides a probabilistic algorithm for obtaining an approximate value of $\|M\|_{\infty\to 1} = \sup_{s_i,t_j=\pm} M_{ij}s_it_j$*: solve SDP for vectors* u_i, v_j*, construct (in* $\mathbb{R}^{n+m}$*) vectors* u'_i, v'_j *such that* $\arcsin(\langle u'_i, v'_j\rangle) = c\langle u_i, v_j\rangle$*, then sample* z *uniformly on* $\mathcal{S}^{n+m-1}$ *and let* $s_i = Sign(\langle u'_i, z\rangle)$*,* $t_j = Sign(\langle v'_j, z\rangle)$*.*

SDP for SBM reconstruction

Consider symmetric two-block, sparse SBM with parameters a, b, for which reconstruction feasible iff $\alpha < \beta^2$, where $\alpha = \frac{a+b}{2}$, $\beta = \frac{a-b}{2}$.
For simplicity assume $\beta > 0$, and n even.

Minimum bisection approach to recover clustering $\sigma^* \in \{-,+\}^n$: letting $J = ee^\top$, $e = (1, \cdots, 1)^\top$:

$$\begin{array}{ll} \text{Maximize} & \langle A, Y\rangle \\ \text{Over} & Y \in \mathcal{S}_n^+ : Y = \sigma\sigma^\top,\ \sigma \in \{-,+\}^n \\ \text{such that} & \forall u \in [n],\ Y_{uu} = 1 \text{ and } \langle J, Y\rangle = 0 \text{ (i.e. } \sum_i \sigma_i = 0) \end{array}$$

Since Min-bisection is NP-hard, consider SDP relaxation

$$\begin{array}{ll} \text{Maximize} & \langle A, Y\rangle \\ \text{Over} & Y \in \mathcal{S}_n^+ \\ \text{such that} & \forall u \in [n],\ Y_{uu} = 1 \text{ and } \langle J, Y\rangle = 0. \end{array}$$

Theorem (Guédon-Vershynin 2014)

Let $\hat{Y}$*: solution of SDP, and* y*: eigenvector of* $\hat{Y}$ *associated with eigenvalue* $\rho(\hat{Y})$*.*
Assume that $\alpha > 8\ln(2)$, $\frac{\beta^2}{\alpha} > 2^{17}\ln(2)$*.*
Then for some $\delta > 0$, $\exists$ *sign* $\epsilon = \pm$ *such that w.h.p.,*
$\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\sigma_i^* = \epsilon Sign(y_i)} \geq \frac{1}{2} + \delta$*, i.e. partition based on* $Sign(y_i)$ *achieves non-vanishing overlap.*

Proof: Let $Y^* = \sigma^* \sigma^{*\top}$. W.h.p., $\left|\sum_{i\in[n]} \sigma_i^*\right| \leq \sqrt{n\ln(n)}$, hence $\exists\tilde{\sigma}$: $\sum_i |\sigma_i^* - \tilde{\sigma}_i| = O(\sqrt{n\ln(n)})$ and $\tilde{Y} := \tilde{\sigma}\tilde{\sigma}^\top$ such that $\left\langle \tilde{Y}, J \right\rangle = 0$.
Since $\sum_i |\sigma_i^* - \tilde{\sigma}_i| = O(\sqrt{n\ln(n)})$, $\left\langle A, \tilde{Y} - Y^* \right\rangle$: at most $O(n\ln(n))$ pairs u, v with $(\tilde{Y} - Y^*)_{uv} \neq 0$, and corresponding entry $A_{uv} \leq \mathrm{Ber}(a/n)$.
$\Rightarrow$ w.h.p., $\left\langle A, \tilde{Y} - Y^* \right\rangle = O(\ln(n))$.

Note $\bar{A} = \mathbb{E}(A|\sigma^*) = \frac{\alpha}{n}J + \frac{\beta}{n}Y^* - \frac{\alpha+\beta}{n}I$.

Write $\left\langle \bar{A}, \hat{Y} - Y^* \right\rangle = \frac{\beta}{n}\left\langle Y^*, \hat{Y} - Y^* \right\rangle - \frac{\alpha}{n}\langle J, Y^* \rangle \leq \frac{\beta}{n}\left\langle Y^*, \hat{Y} - Y^* \right\rangle$.

By SDP optimality of $\hat{Y}$,

$0 \leq \left\langle A, \hat{Y} - \tilde{Y} \right\rangle = \left\langle A, \hat{Y} \right\rangle - \langle A, Y^* \rangle + O(\ln(n))$, so that

$0 \leq \left\langle \bar{A}, \hat{Y} - Y^* \right\rangle + \left\langle A - \bar{A}, \hat{Y} - Y^* \right\rangle + O(\ln(n))$, hence

$\frac{\beta}{n}\left\langle Y^*, Y^* - \hat{Y} \right\rangle \leq \left\langle \Delta, \hat{Y} - Y^* \right\rangle + O(\ln(n))$, where $\Delta = A - \bar{A}$.

$$\begin{aligned}
\left\| Y^* - \hat{Y} \right\|_F^2 &= \|Y^*\|_F^2 + \left\| \hat{Y} \right\|_F^2 - 2\left\langle Y^*, \hat{Y} \right\rangle \\
&\leq 2(n^2 - \left\langle Y^*, \hat{Y} \right\rangle) \\
&= 2\left\langle Y^*, Y^* - \hat{Y} \right\rangle
\end{aligned}$$

Hence $\frac{\beta}{2n}\left\| Y^* - \hat{Y} \right\|_F^2 \leq \left| \left\langle \Delta, \hat{Y} \right\rangle \right| + |\langle \Delta, Y^* \rangle| + O(\ln(n))$

Let $s^2 := \sum_{i<j} \mathrm{Var}(\Delta_{ij} | \sigma^*) = n\alpha(1/2 + o(1))$. By Bernstein's inequality, $\mathbb{P}(|\sum_{i<j} Y^*_{ij}\Delta_{ij}| \geq \sqrt{2s^2 t} + \frac{2}{3}t \mid \sigma^*) \leq 2e^{-t}$.
Thus for $t = \ln(n)$, w.h.p., $\left|\sum_{i<j} Y^*_{ij}\Delta_{ij}\right| = O(\sqrt{\alpha n \ln(n)})$.

Upper bounding $\left|\left\langle \Delta, \hat{Y} \right\rangle\right|$:

$$\begin{aligned}\left|\left\langle \Delta, \hat{Y} \right\rangle\right| &\leq \sup_{Y \in \mathcal{S}_n^+,\, \mathrm{Diag}(Y) \preceq I} |\langle \Delta, Y \rangle| \\ &\leq K_G \|\Delta\|_{\infty \to 1},\end{aligned}$$

by Corollary to Grothendieck's inequality. Write then:

$$\begin{aligned}\|\Delta\|_{\infty \to 1} &= \sup_{x,y \in \{+,-\}^n} x^\top \Delta y \\ &= \sup_{x,y \in \{+,-\}^n} \sum_{i<j} \Delta_{ij}(x_i y_j + x_j y_i).\end{aligned}$$

By Bernstein's inequality,
$\mathbb{P}(\sum_{i<j} \Delta_{ij}(x_i y_j + x_j y_i) \geq \sqrt{8s^2 t} + \frac{4}{3}t \mid \sigma^*) \leq e^{-t}$.

Let $t = 2(1+\epsilon)n\ln(2)$, so that $e^{-t} = 2^{-2(1+\epsilon)n}$.

Then (union bound) w.h.p.,
$\|\Delta\|_{\infty\to 1} \leq \sqrt{8s^2 t} + \frac{4}{3}t \leq n\left[\frac{8}{3}(1+\epsilon)\ln(2) + \sqrt{8(1+\epsilon)\alpha(1+o(1))\ln(2)}\right]$.

By assumption $8\ln(2) < \alpha$, this gives $\|\Delta\|_{\infty\to 1} \leq \frac{4}{3}n\sqrt{8\alpha(1+\epsilon)\ln(2)}$,

hence w.h.p.: $\frac{1}{n^2}\left\|Y^* - \hat{Y}\right\|_F^2 \leq \frac{8}{3\beta}K_G\sqrt{8\alpha(1+\epsilon)\ln(2)} =: \theta$

Let y: eigenvector of $\hat{Y}$ associated with $\rho(\hat{Y})$ normalized so that $\|y\| = \sqrt{n}$. By Davis-Kahane, for some $\epsilon = \pm 1$,

$$\frac{1}{\sqrt{n}}\|y - \epsilon\sigma^*\| \leq 2\sqrt{2}\frac{\left\|Y^* - \hat{Y}\right\|_F}{\lambda_{max}(Y^*)}, \text{ i.e. } \frac{1}{n}\|y - \epsilon\sigma^*\|^2 \leq \frac{8\left\|Y^* - \hat{Y}\right\|_F^2}{n^2} \leq 8\theta.$$

Since $\frac{1}{n}\sum_i \mathbb{I}_{\sigma_i^* \neq \epsilon \mathrm{Sign}(y_i)} \leq \|y - \epsilon\sigma^*\|^2 \leq 8\theta$,

to conclude it suffices to ensure $8\theta < 1/2$, i.e. $16\frac{8}{3\beta}K_G\sqrt{8\alpha\ln(2)} < 1$.

Thus $\frac{\beta^2}{\alpha} > 8\ln(2)\left(\frac{2^7}{3}K_G\right)^2$ suffices.

Other application of Grothendieck's inequality

Definition

Cut norm of matrix $B \in \mathbb{R}^{n \times m}$: $\|B\|_{cut} := \max_{I \subseteq [n], J \subseteq [m]} \left| \sum_{i \in I, j \in J} B_{ij} \right|$.

Basic inequalities: $\|B\|_{cut} \leq \|B\|_{\infty \to 1} \leq 4\|B\|_{cut}$

Indeed for $s, t \in \{-, +\}^n$, letting $I_\pm = \{i : s_i = \pm\}$, $J_\pm = \{j : t_j = \pm\}$,

$$\sum_{i,j} s_i t_j B_{ij} = \sum_{i \in I_+, j \in J_+} B_{ij} + \sum_{i \in I_-, j \in J_-} B_{ij} - \sum_{i \in I_+, j \in J_-} B_{ij} - \sum_{i \in I_-, j \in J_+} B_{ij},$$

hence $\|B\|_{\infty \to 1} \leq 4\|B\|_{cut}$.

Also for $I \subseteq [n]$, $J \subseteq [m]$, let $x_i = 2\mathbb{I}_{i \in I} - 1$, $y_j = 2\mathbb{I}_{j \in J} - 1$. Then

$$\sum_{i \in I, j \in J} B_{ij} = \sum_{i,j} B_{ij} \frac{1 + x_i}{2} \frac{1 + x_j}{2} = \frac{1}{4} \sum_{i,j} [B_{ij} x_i y_j + B_{ij} x_i + B_{ij} y_j + B_{ij}],$$

hence $\|B\|_{cut} \leq \|B\|_{\infty \to 1}$. Approximation of $\|\cdot\|_{cut}$ can be done based on probabilistic algorithm for approximating $\|\cdot\|_{\infty \to 1}$ (see Alon-Naor 2004])