AMLDS Course Review: Optimization & Spectral Graph Theory

Apoorv Vikram Singh

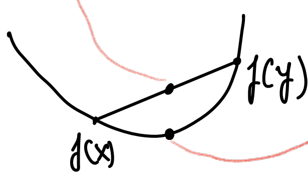
Dec 13, 2023

Optimization

1. Convex Function:

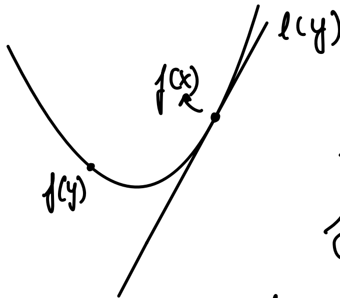
(a) **Def**: f is convex if for any x, y, $\lambda \in [0,1]$

$$(1-\lambda)\cdot f(x) + \lambda f(y) \geq f((1-\lambda)x + \lambda y)$$



(b) **Def**: For differentiable f:

$$f(x) - f(y) \leq \nabla f(x)^T (x-y) .$$



$$l(y) = f(x) + \nabla f(x)^T (y-x)$$

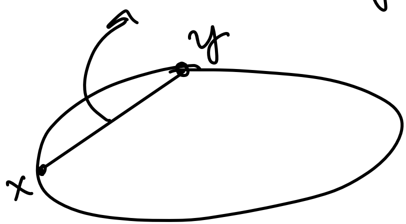
$$f(y) \geq l(y)$$

$$\Rightarrow \quad f(y) \geq f(x) + \nabla f(x)^T (y-x)$$

$$\Rightarrow \quad f(x) - f(y) \leq \nabla f(x)^T (x-y)$$

2. <u>Convex Set</u> : Set S is convex if for any $x, y \in S$
$\lambda \in [0,1]$

$$((1-\lambda)x + \lambda y) \in S$$



3. Gradient Descent Basic Update Rule

- Start at $x_{(0)}$
- For $i = 0, \ldots, T$:

$$x_{(i+1)} = x_i - \eta \nabla f(x_{(i)})$$

$\rightarrow$ $f(x+\eta v) - f(x) \approx \eta \nabla f(x)^T v$, small η

- we want $f(x+\eta v) - f(x) < 0$, then
- $\therefore$ setting $v = -\nabla f(x)$, we get

$$f(x+\eta v) - f(x) \approx -\eta \|\nabla f(x)\|^2$$

$\therefore$ update rule is negative gradient.

4. <u>Definitions</u>:

(a) <u>G-Lipschitz</u>: $\exists L < \infty$, $\forall x, y$

$$\| f(x) - f(y) \| \leq G \|x-y\|$$

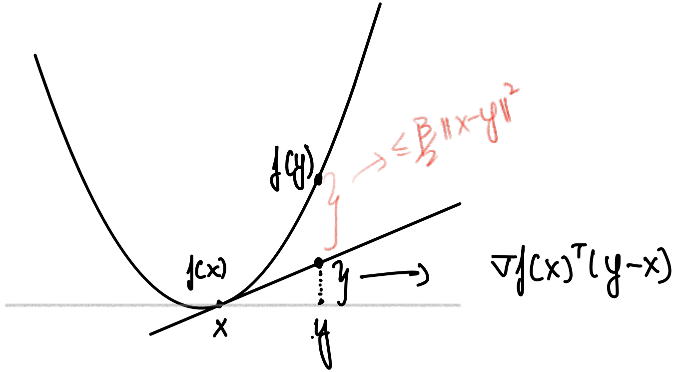
For differentiable functions

$$\| \nabla f(x) \| \leq G \quad , \quad \forall x$$

(b) <u>β-Smooth</u>: f is β smooth if $\forall x, y$

$$\| \nabla f(x) - \nabla f(y) \| \leq \beta \|x-y\|$$

$$\Rightarrow 0 \leq f(y) - f(x) - \nabla f(x)^T (y-x) \leq \frac{\beta}{2} \|x-y\|^2$$

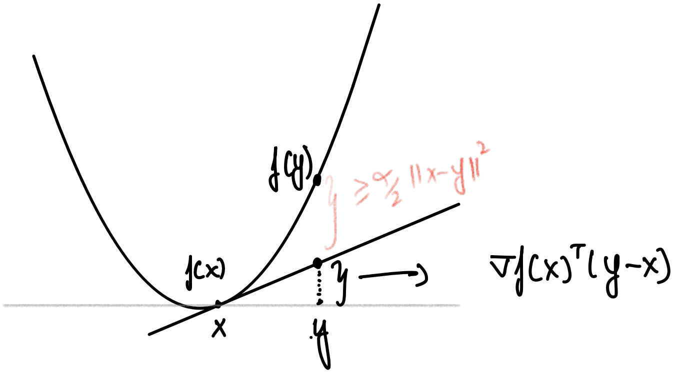


- $f(y) \leq f(x) + \nabla f(x)^T(y-x) + \frac{\beta}{2}\|x-y\|^2.$

- $\|\nabla^2 f(x)\| \leq \beta \quad , \forall x.$

(c) Def: α- Strongly Convex if $\forall\, x, y$

- $f(y) - f(x) - \nabla f(x)^T (y-x) \geq \frac{\alpha}{2} \|x-y\|^2$.

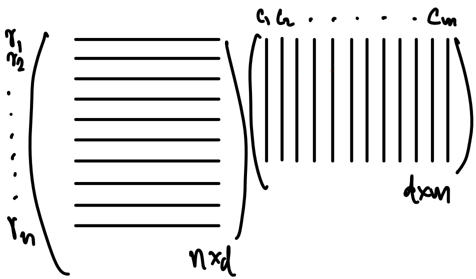


- $\|\nabla^2 f(x)\| \geq \alpha \quad \forall\, x$

$\Rightarrow$ For scalars: $\alpha \leq f''(x) \leq \beta$.

5. Multiplying $n \times d$ matrix with $d \times m$ matrix

$$A_{n\times d}\, B_{d\times m} = C_{n\times m}.$$



$$= \begin{pmatrix} \langle r_1, c_1\rangle, & \langle r_2, c_2\rangle, & \cdots, & \langle r_1, c_m\rangle \\ \langle r_2, c_1\rangle, & \langle r_2, c_2\rangle, & \cdots & \langle r_2, c_m\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle r_n, c_1\rangle, & \langle r_n, c_2\rangle, & \cdots & \langle r_n, c_m\rangle \end{pmatrix}$$

$$A \qquad B \qquad = \qquad C$$

- each $\langle r_i, c_j\rangle$ takes $O(d)$ operations.
- Compute $n \times m$ such inner products.
- Total time: $O(n \times m \times d)$.

6. Compute Gradient of Basic Functions : $\mathbb{R}^d \to \mathbb{R}$

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_d}(x) \end{bmatrix}$$

e.g. $f(x) = x_1^2 + x_2^2$, $d = 2$

$$\nabla f(x) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} .$$

7. <u>Condition Number</u>

For a matrix A:

$$\kappa(A) := \frac{\sigma_{max}(A)}{\sigma_{min}(A)}$$

$$= \frac{|\lambda_{max}(A)|}{|\lambda_{min}(A)|} \rightarrow \text{for symmetric } A.$$

- For a function g:

$$\kappa(g) := \frac{|\lambda_{max}(\nabla^2 g)|}{|\lambda_{min}(\nabla^2 g)|}$$

8.

$$f(x) = \frac{1}{2}\|Ax - b\|^2$$

$$f(x) = \frac{1}{2} \sum_{i=1}^{n} \left(\langle a_i, x \rangle - b_i \right)^2$$

$$\frac{\partial f(x)}{\partial x_k} = \frac{1}{2} \sum_{i=1}^{n} 2 \left(\langle a_i, x \rangle - b_i \right) \cdot a_{i,k}$$

$$\frac{\partial^2 f(x)}{\partial x_k \partial x_j} = \frac{1}{2} \sum_{i=1}^{n} 2 \cdot a_{ik} \, a_{ij}$$

$$\therefore \nabla f(x) = A^T \cdot (Ax - b)$$

$$\nabla^2 f(x) = A^T A$$

$\therefore$ Applying gradient descent, step size $\lambda_{max}(A^TA) := \lambda_{max}$

$$x_{t+1} = x_t - \frac{1}{\lambda_{max}} A^TA x_t + \frac{1}{\lambda_{max}} A^T b$$

Note that $A^T(Ax^* - b) = 0 \Rightarrow A^Tb = A^TAx^*$

$$x_{t+1} = x_t - \frac{1}{\lambda_{max}} A^TA x_t + \frac{1}{\lambda_{max}} A^TA x^*$$

$$x_{t+1} - x^* = \left(I - \frac{1}{\lambda_{max}} A^TA\right)(x_t - x^*)$$

Applying iterations, we get

$$(x_T - x^*) = \left(I - \frac{1}{\lambda_{max}} A^T A\right)^T (x_T - x^*)$$

- this converges quickly if max eigenvalue of $\left(I - \frac{1}{\lambda_{max}} A^T A\right)$ is small.

- Eigenvalues of this are:

$$1 - \frac{\lambda_1(A^T A)}{\lambda_{max}(A^T A)}, \quad 1 - \frac{\lambda_2(A^T A)}{\lambda_{max}(A^T A)}, \quad \cdots\cdots \quad 1 - \frac{\lambda_{min}(A^T A)}{\lambda_{max}(A^T A)}$$

$\therefore$ Largest eigenvalue is $\left(1 - \frac{1}{K}\right)$

$$\therefore \quad \|x^T - x^*\|^2 \leq \left\| \left(I - \frac{1}{\lambda_{max}} A^T A\right)^T (x_0 - x^*) \right\|^2$$

$$\leq \left(1 - \frac{1}{K}\right)^{2T} \|x_0 - x^*\|^2$$

$$\leq \left(1 - \frac{1}{K}\right)^{\frac{2T}{K} K} \|x_0 - x^*\|^2$$

$$= e^{-2T/K} \|x_0 - x^*\|^2.$$

$\therefore$ $1/K$ decides the rate of convergence to optimal.

($\therefore$ K the smaller the better)

4. <u>Positive Semi-Definite</u> (PSD), $H \succeq 0$.

- A square, symmetric matrix $H \in \mathbb{R}^{d \times d}$ is PSD if for any vector $y \in \mathbb{R}^d$

$$y^T H y \geq 0$$

- Equivalently, all eigenvalues of H, $\lambda_i \geq 0$, $\forall i \in [d]$.

Because: $H = V^T \Sigma V$ → eigendecomposition.

1. Take $y = v_i$ → eigenvector

$$v_i^T H v_i = \lambda_i v_i^T v_i = \lambda_i \|v_i\|^2 \geq 0 \Rightarrow \lambda_i \geq 0.$$

2. $H = V^T \Sigma V$ & $\Sigma = \mathrm{diag}(\overbrace{\lambda_1, \dots, \lambda_d}^{\geq 0})$

$$\Rightarrow \quad y^T V^T \Sigma V y = (\Sigma^{\frac{1}{2}} V y)^T (\Sigma^{\frac{1}{2}} V y)$$
$$= \|\Sigma^{\frac{1}{2}} V y\|^2 \geq 0.$$

- Is $A^T A$ PSD for any $A \in \mathbb{R}^{n \times d}$?

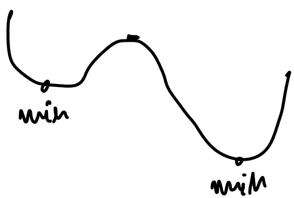
$\forall y$,
$$y^T A^T A y = (Ay)^T (Ay)$$
$$= \|Ay\|^2 \geq 0.$$

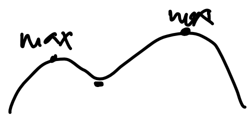
10. Stationary Points

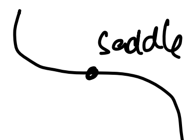
For a differentiable f, a stationary point is any x with

$$\nabla f(x) = 0$$

Stationary points can be:
- Global min/max
- Local min/max
- Saddle point







11. <u>Center of Gravity Method</u>:

- f bdd. between $[-B, B]$, on cvx set S.
- Center of gravity o/p $\hat{x}$ s.t. $f(\hat{x}) \le f(x^*) + \varepsilon$ using $O(d \log(B/\varepsilon))$ calls to gradient oracle for cvx. f.

$\Rightarrow$ Need a "representation" of S & not just proj. oracle.

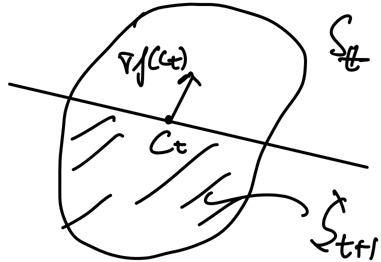
- Center of gravity of S is defined as

$$c = \frac{\int_{x \in S} x \, dx}{\mathrm{Vol}(S)} = \frac{\int_{x \in S} x \, dx}{\int_{x \in S} 1 \, dx}$$

· $S_1 = \mathcal{S}$. For each $t = 1, \ldots, T$, let c_t : center of grav($\mathcal{S}_t$)

$$H := \{ x \in \mathbb{R}^d : \langle x - c_t, \nabla f(c_t) \rangle \le 0 \}$$

$$S_{t+1} = S_t \cap H$$



$\rightarrow$ $$f(y) \geq f(c_t) + \nabla f(c_t)\,(y - c_t)$$

if $\nabla f(c_t)(y - c_t) \geq 0$ then $f(y) \geq f(c_t)$

$\therefore$ we the rule is OK.

12. <u>Grunbaum Theorem</u> :

For any cvx set S with center of grav. c & any half-sp $Z = \{x \mid \langle a, x-c \rangle \geq 0\}$, then

$$\frac{\mathrm{Vol}(S \cap Z)}{\mathrm{Vol}(S)} \geq \frac{1}{e}$$

· Let Z be complement of H in center of grav. method, then, we cut off at least $\frac{1}{e}$ fraction of cvx. body.

$$\therefore \ \mathrm{Vol}(S_t) \leq \left(1 - \frac{1}{e}\right)^t \mathrm{vol}(S).$$

$\Rightarrow$ Cvg. rate.

→ We don't use center of gravity in practice because:

Computing centroid is hard #P-Hard even if S is intersection of half-sp.

11. __Linear Program__ :

Linear constraints; Linear objective.

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Ax \geq b \end{aligned}$$

eg.

$$\begin{aligned} \min \quad & c_1 x_1 + \cdots + c_n x_n \\ \text{s.t.} \quad & \\ & a_1 x_1 + \cdots + a_n x_n \geq b_1 \\ & c_1 x_1 + \cdots + c_n x_n \geq b_2 \\ & d_1 x_1 + \cdots + d_n x_n \geq b_3 \\ & \vdots \\ & x_1, x_2 \geq 0 \end{aligned}$$

eg.

$$\begin{aligned} \min_x \quad & c \\ \text{s.t.} \quad & Ax \geq b \end{aligned} \Rightarrow \text{constraint satisfaction.}$$

12. <u>Projection Oracle</u>:

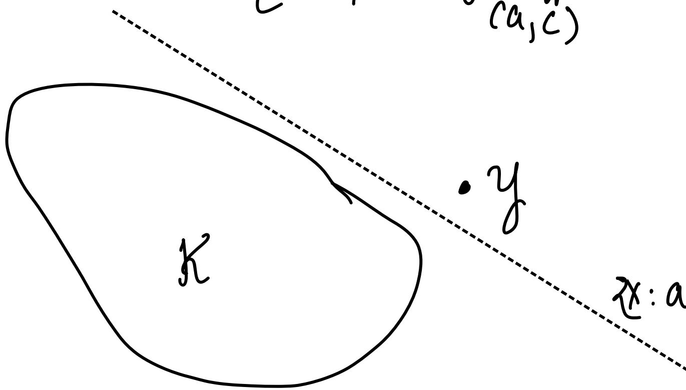
Given set $\mathcal{S}$, let x be any point, then

$$P_{\mathcal{S}}(x) = \underset{y \in \mathcal{S}}{\operatorname{argmin}} \|y - x\|_2 .$$

13. <u>Separation Oracle</u>

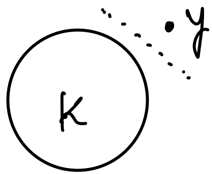
For a convex set K. Given any point y,

separation oracle $\longrightarrow$ $S_K(y) = \begin{cases} \emptyset & \text{if } y \in K \\ \text{separating hyperplane } (a,c), & \text{if } y \notin K \end{cases}$



Separation oracle for simple convex set.

e.g. ℓ_2 ball



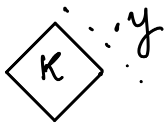
$\Rightarrow$ If $\|y\| \leq 1$, then return ϕ

else, return $\left(\frac{y}{\|y\|}, 1\right)$

For $x \in K$, $\frac{y^T}{\|y\|} x \leq \|x\| \leq 1$

For $y \notin K$, $\frac{y^T y}{\|y\|} = \|y\| > 1$

e.g. ℓ_1 ball



- $\|x\|_1 \leq 1$
- Basically, use y to give a separation oracle.

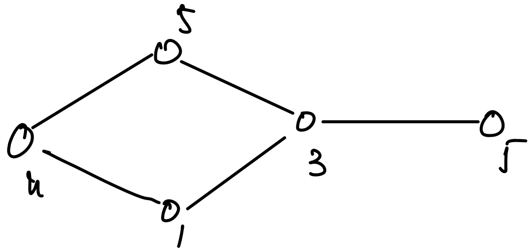
consider $v = (v_1, \dots, v_n)$, where $v_i = \text{sign}(y_i)$
then

if $x \in K$ $\quad v^T x \leq 1$

if $y \notin K$ $\quad v^T y = \|y\|_1 > 1$.

14. Relax & Round Approach:

Vertex Cover:



select nodes with min wt. s.t. they cover all edges.

- edge (i,j), vertex i or vertex j

each vertex variable x_i , $i \in [V]$

$x_e = 0$ or 1 $\Rightarrow$ denoting if we choose that vertex or not

$e=(i,j)$ $x_i = 1$ or $x_j = 1$ or both.

$\therefore$ minimize wt.

s.t. all edges covered

$$\min_{x_i \in \{0,1\}} \sum_i x_i w_i$$

s.t. $\quad x_i + x_j \geq 1 \quad , \forall (i,j) \in E$

$\quad x_i \in \{0,1\} \quad , \forall i \in V$

not a linear program

why?

$\downarrow$ Relax

$$\min_{x_i \in \{0,1\}} \quad \sum_i x_i w_i$$

$$\text{s.t.} \quad x_i + x_j \geq 1 \quad \forall (i,j) \in E$$

$$x_i \geq 0 \quad \forall i \in V$$

Suppose we solve this LP, x^*. We can get fractional solutions. Now, we will round it to integral solutions: $\tilde{x}$

idea: If $x_i^* \geq \frac{1}{2}$, $\tilde{x}_i = 1$

if $x_i^* < \frac{1}{2}$, $\tilde{x}_i = 0$

<u>Claim</u>: $\tilde{x}$ is a valid solution: ie satisfies the constraints.

for any $(i,j) \in E$, $x_i^* + x_j^* \geq 1$ and $x_i^* \geq 0$

$\therefore$ at least one of x_i^* or $x_j^* \geq 0.5$

$\forall (i,j) \in E$, $\tilde{x}_i + \tilde{x}_j \geq 1$.

<u>Claim</u>: $\sum_i w_i \tilde{x}_i \leq 2 \cdot \sum_i w_i x_i^*$.

because $\sum_i w_i \tilde{x}_i = \sum_{i : x_i^* \geq 1/2} w_i$

$\leq \sum_{i : x_i^* \geq 1/2} 2 x_i^* w_i \leq \sum_i 2 x_i^* w_i$.

- Let OPT denote the optimal cost of the best integral solution.

$\Rightarrow$ $\sum_i w_i \tilde{x}_i \Rightarrow \text{cost}(\tilde{x})$: Cost of our algo

$\sum_i w_i x_i^* \Rightarrow \text{cost}(x^*)$: Cost of optimal fractional solu.

$\therefore$ $\text{cost}(x^*) \leq \text{OPT}$

$\text{cost}(\tilde{x}) \leq 2\,\text{cost}(x^*)$

$\Downarrow$

$\text{cost}(\tilde{x}) \leq 2\,\text{cost}(x^*) \leq 2\,\text{OPT}$.

Spectral Graph Theory

1. Graph $G = (V, E, w)$.

- Adjacency Matrix $A \Rightarrow$ $A_{ij} = A_{ji} = w_e$ if $(i,j) \in e$

  $A_{ij} = A_{ji} = 0$ if $(i,j) \notin e$

- Laplacian Matrix $L = D - A$

  $D \Rightarrow$ degree matrix

- Normalized Adjacency & Normalized Laplacian

$\bar{A} = D^{-1/2} A D^{-1/2}$

$\bar{L} = D^{-1/2} L D^{-1/2}$

$\bar{L} = I - D^{-1/2} A D^{-1/2}$

$= I - \bar{A}.$

- Edge - Incidence Matrix. Assign arbit sign to edge (v_i, v_j) $+1, -1$

$$B = \#\text{edges} \begin{bmatrix} 1 & -1 & & \\ & 1 & -1 & \\ & & & \\ -1 & 1 & & \\ & & & \end{bmatrix} \begin{array}{l} \\ \rightarrow b_i \\ \\ \rightarrow b_j \\ \\ \end{array}$$

n: # nodes

each row corresponds to an edge with a sign.

$L = B^T B \Rightarrow \sum_k b_k b_k^T$, $b_k \in$ n dimn vec.

$(i,j) \in e$

$$b_k b_k^T = \begin{array}{c} \\ i \rightarrow \\ j \rightarrow \end{array} \begin{array}{c} \begin{array}{cc} i & j \\ \downarrow & \downarrow \end{array} \\ \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \end{array} \Rightarrow \begin{array}{l} +1 \text{ on } (i,i), (j,j) \\ -1 \text{ on } (i,j), (j,i) \end{array}$$

2. $\therefore \quad L = \sum_{k} b_k b_k^T = b_1 b_1^T + \cdots + b_m b_m^T$

$\therefore \quad L = B^T B.$

$\therefore \quad L \succeq 0 \quad \because \quad x^T B^T B x = \|Bx\|^2 \geq 0$

- $x \in \mathbb{R}^n$

$$x^T L x = \sum_{(i,j)\in E} \left(x(i) - x(j)\right)^2$$

because $\quad Bx = \begin{pmatrix} \vdots \\ x(i) - x(j) \\ \vdots \end{pmatrix}_{(i,j)\in E}$

$$\therefore \quad x^T B^T B x = \sum_{(i,j)\in E} \left(x(i) - x(j)\right)^2 .$$

3. "Linear Algebraic Way" to compute cut.

$c \in \{1, -1\}^n$, cut vector.

$$c^T L c = \sum_{(i,j) \in E} (c(i) - c(j))^2$$

If $c(i) \neq c(j)$, then we get 4
& $c(i) = c(j)$, then we get 0

$$\therefore \quad c^T L c = 4 \, \text{cut}(S, S^c)$$

Also: $c^T 1 = |S| - |S^c|$

Balanced cut: Minimize both $c^T L c$ & $c^T 1$

cut

imbalance

$\therefore$ Balanced cut:

$$\min_{c \in \left\{-\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\right\}^n} \quad c^T L c \quad \text{s.t.} \quad c^T 1 = 0$$

$\rightarrow$ Relaxing $c \in \left\{-\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\right\}$ to $\|c\|_2 = 1$, we get

Then, we get

$$\min_{\|c\|_2 = 1} \quad c^T L x \quad \text{s.t.} \quad c^T 1 = 0$$

$\Rightarrow$ Minimized exactly by the smallest second eigenvalue.

- $\min_{\|v\|=1} v^T L v$ = smallest eigenvalue.

- Fact 1: $L = B^T B \Rightarrow$ PSD matrix

Let $v = \frac{1}{\sqrt{n}} \mathbb{1}$

$$v^T L v = \sum_{(i,j)\in E} (v(i) - v(j))^2$$
$$= \sum_{(i,j)\in E} (1-1)^2 \frac{1}{n} = 0$$

$\therefore$ $v_n = \frac{1}{\sqrt{n}} \mathbb{1}$ with eigenvalue 0

$\therefore$ $v_{n-1} = \underset{\|c\|=1}{\operatorname{argmin}} \; c^T L c$ s.t. $c^T \mathbb{1} = 0$.

Relax and Round

- $v_{n-1} = \underset{\|c\|=1}{\operatorname{argmin}} \; c^T L c \quad \text{s.t.} \quad c^T \mathbb{1} = 0$

- Round:

  Let S be all nodes with $v_{n-1}(i) < 0$

  Let S^c be all nodes with $v_{n-1}(i) \geq 0$

  $c = \operatorname{sign}(v_{n-1})$.

5. SBM, Planted Clique, Random Graph

SBM: $G_n(p,q)$ ~ dist over graphs on n nodes

- Split equally into 2 grps. B & C
  $n/2$ $n/2$ nodes

- Any 2 nodes in same group are conn. with prob p (with self-loop)

- Any 2 nodes in diff. group are conn. with prob. $q < p$.

- <u>Random Graph</u> $G(n,p)$
  - connect any 2 nodes with prob. p.
- <u>Planted Clique</u>
  1. Sample $G' = (V, E)$ from $G(n,p)$ (generally $p = \frac{1}{2}$)
  2. Choose $S \subseteq V$, k vertices unif at random.
  3. Construct G by planting clique on S.

6. <u>SBM</u> :

$$\mathbb{E}A = \begin{bmatrix} p\,p\,p\,p & q\,q\,q\,q \\ p\,p\,p\,p & q\,q\,q\,q \\ p\,p\,p\,p & q\,q\,q\,q \\ p\,p\,p\,p & q\,q\,q\,q \\ q\,q\,q\,q & p\,p\,p\,p \\ q\,q\,q\,q & p\,p\,p\,p \\ q\,q\,q\,q & p\,p\,p\,p \\ q\,q\,q\,q & p\,p\,p\,p \end{bmatrix}$$

$$\mathbb{E}D = (p+q)\frac{n}{2} I$$

$\mathbb{E}L = \mathbb{E}D - \mathbb{E}A \Rightarrow$ eigenvec. of A = eigenvec. of L

- smallest eigenvec : $\mathbb{1}$, eigvalue $= 0$

· Largest eigenvec

$$\begin{bmatrix} p&p&p&p&q&q&q&q \\ p&p&p&p&q&q&q&q \\ p&p&p&p&q&q&q&q \\ p&p&p&p&q&q&q&q \\ q&q&q&q&p&p&p&p \\ q&q&q&q&p&p&p&p \\ q&q&q&q&p&p&p&p \\ q&q&q&q&p&p&p&p \end{bmatrix} \begin{bmatrix} 1\\1\\1\\1\\1\\1\\1\\1 \end{bmatrix} = \frac{(p+q)}{2} \begin{bmatrix} 1\\1\\1\\1\\1\\1\\1\\1 \end{bmatrix}$$

· 2^{nd} Largest Eigenvec

$$\begin{bmatrix} p&p&p&p&q&q&q&q \\ p&p&p&p&q&q&q&q \\ p&p&p&p&q&q&q&q \\ p&p&p&p&q&q&q&q \\ q&q&q&q&p&p&p&p \\ q&q&q&q&p&p&p&p \\ q&q&q&q&p&p&p&p \\ q&q&q&q&p&p&p&p \end{bmatrix} \begin{bmatrix} 1\\1\\1\\1\\-1\\-1\\-1\\-1 \end{bmatrix} = \frac{(p-q)}{2} \begin{bmatrix} 1\\1\\1\\1\\-1\\-1\\-1\\-1 \end{bmatrix}, \quad \lambda_i = 0 \ \forall_{i=3,\dots,n}$$

* Matrix Concentration (understanding)

- $A \sim G(n,p)$.

$$\| A - \mathbb{E}A \|_2 \leq O(\sqrt{pn}) \quad \text{for } p \geq O\left(\frac{\log^4 n}{n}\right)$$

- Davis Kahan

- $A, \bar{A}$ s.t. $\|A - \bar{A}\| \leq \varepsilon$

$v_1, \ldots, v_n$ $\quad$ $\bar{v}_1, \ldots, \bar{v}_n$

$\theta(v_i, \bar{v}_i)$, angle b/w eigenvec:

$$\sin \theta(v_i, \bar{v}_i) \leq \frac{\varepsilon}{\min_{j \neq i} |\lambda_i - \lambda_j|}$$

$\lambda_1, \ldots, \lambda_n$ eigenval of $\bar{A}$.

- <u>In SBM:</u>

$$\min_{i \neq j} |\lambda_i - \lambda_j| = \min\left(qn, \frac{n}{2}(p-q)\right)$$

$$\varepsilon = \mathcal{O}(\sqrt{pn})$$

We get:

$$\sin\Theta(v_2, \bar{v}_2) \leq \frac{\mathcal{O}(\sqrt{pn})}{\min_{j \neq i} |\lambda_i - \lambda_j|}$$

$$= \frac{\mathcal{O}(\sqrt{pn})}{(p-q)\,n} = \mathcal{O}\left(\frac{\sqrt{p}}{(p-q)\sqrt{n}}\right)$$

$$\Rightarrow \|v_2 - \bar{v}_2\|^2 \leq O\left(\frac{p}{(p-q)^2 n}\right)$$

- $\text{sign}(v_2)$ & $\bar{v}_2$ are pretty much the same.

- Entry i contributes $\frac{1}{n}$, we get $\text{sign}(v_2)$, $\bar{v}_2$ differ by at most

$$O\left(\frac{p}{(p-q)^2}\right).$$