Lecture 9

Min-cut using gradient descent continued

- Recap of last lecture results
- Min-cut using gradient descent (edges space)

Recap

① Let $S = \text{span}\{s_1, s_2, \ldots, s_k\}$ then the projection operator

$$P : \mathbb{R}^n \to S \;, \quad Pv = \underset{u \in S}{\arg\min} \|v - u\|$$

is a matrix $P = B(B^T B)^+ B^T$

where $B = \begin{bmatrix} | & | & & | \\ s_1 & s_2 & \cdots & s_k \\ | & | & & | \end{bmatrix}$

useful properties

$\ker(P) = \text{span}\{I-P\}$

$P^2 = P \Rightarrow P(I-P) = 0$

(2) Gradient descent for a problem with equality constraints $Px = b$ where P is a projection matrix.

$$\min f(x) \quad \text{s.t } Px = b \implies$$

start with feasible x_0 ($Px_0 = b$)

for $t = 0 \ldots T-1$

$$x_{t+1} = x_t - \frac{1}{\beta}(I-P)\nabla f(x_t)$$

Lipschitz parameter of $(I-P)\nabla f(x)$

to get an ε-additive error we need

$$T = O\left(\frac{\beta}{\varepsilon}\|x^* - x_0\|^2\right)$$

If we use instead Nesterov accelerated method we only need $T = O\left(\sqrt{\frac{\beta}{\varepsilon}}\|x^* - x_0\|\right)$ iterations

(3) formulated min-cut as

$$\min \sum_{(u,v)\in E} \sqrt{(x_u - x_v)^2 + \mu^2}$$

$$\text{s.t } x_s - x_t = 1$$

complexity per iteration

number of iterations needed

nesterov $\implies$ $(1+\varepsilon)$-multiplicative approximation with running time $O\left(|E| \cdot \frac{\sqrt{|E|}\cdot\sqrt{|V|}}{\varepsilon \cdot F}\right)$

A closer look:

$$O\left(|E| \cdot \frac{\sqrt{|E|}\sqrt{|V|}}{\varepsilon F}\right)$$

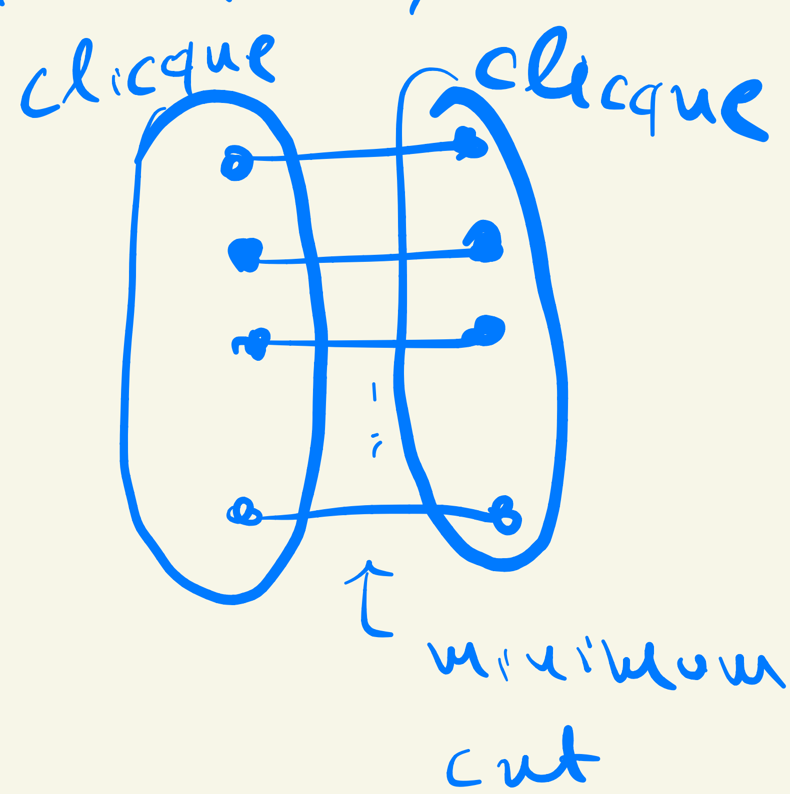
the $\sqrt{v}$ factor comes from

$\|x^* - x_0\| \le \sqrt{v}$ and (probably)

we cannot hope to get easily

something better than that, as

$\|x^*\| = O(\sqrt{v})$



Idea: change space

vertex space $\longrightarrow$ edge space

assume that instead of having as optimization variable $x \in \mathbb{R}^{|V|}$ we had $y \in \mathbb{R}^{|E|}$ then $\|y^*\| \le \sqrt{F}$ and we may hope to get $\|y_0 - y^*\| \le \sqrt{F}$

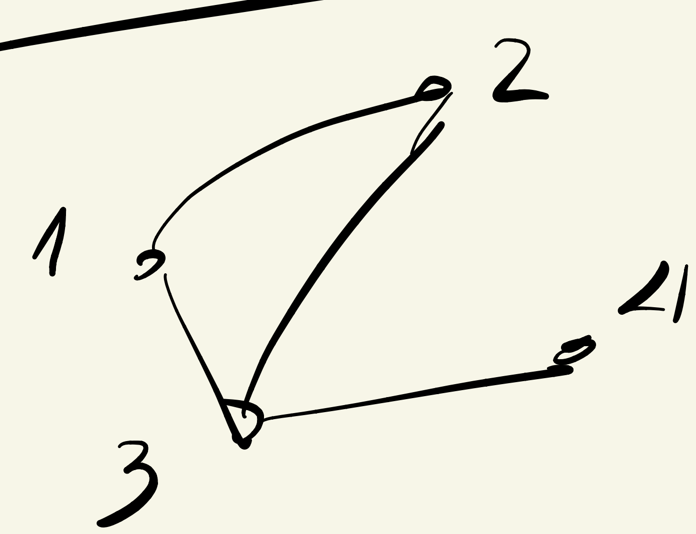
Min-cut using gradient descent on the edge space

Initial LP (vertex space)

$$\min \sum_{e=(u,v)\in E} |x_u - x_v|$$
$$\text{s.t} \quad x_s - x_t = 1$$

let $B \in \mathbb{R}^{|E|\times|V|}$ the matrix whose i-th row, which corresponds to (let's say) edge (u,v), is full of zeros except from the coloumns which correspond to u ⇝ which has a 1 and v ⇝ which has a -1

easier with an example



$B =$

	1	2	3	4
(1,2)	1	-1	0	0
(1,3)	1	0	-1	0
(2,3)	0	1	-1	0
(3,4)	0	0	1	-1

note that $\sum_{(u,v) \in E} |x_u - x_v| = \|B \cdot x\|_1$

let's define $y = Bx \in \mathbb{R}^{|E|}$ the vector which describes the edge space =>

=> $\min \sum_{(u,v)\in E} |x_u - x_v|$ s.t $x_s - x_t = 1$ => $\min \|y\|_1$ s.t $y = Bx$, $x_s - x_t = 1$

now our goal is to eliminate the x variables from the new formulation

(and hopefully get an equality constraint which is associated with a projection matrix... we didn't do so much work for nothing :P

to that end we will introduce the Laplacian matrix

Laplacian matrix

$L := B^T B$ is called the Laplacian matrix of a graph.

some observations:

① let $\tilde{x}_u \in \mathbb{R}^{|V|}$, $\tilde{x}_u = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ ← row which corresponds to vertex u

then $L = \sum_{(u,v) \in E} (\tilde{x}_u - \tilde{x}_v) \cdot (\tilde{x}_u - \tilde{x}_v)^T$

② $L = D - A$, $D \in \mathbb{R}^{|V| \times |V|}$ ⇝ diagonal matrix where $D_{uu} = \deg(u)$

↳ $A \in \mathbb{R}^{|V| \times |V|}$ is the adjacency matrix of the graph

③ the all one vector $\mathbb{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \in \ker(L)$

④ if the graph is connected then $\ker(L) = \text{span}\{\mathbb{1}\}$

↓

in that case the pseudoinverse L^+ can invert any vector perpendicular to $\mathbb{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$

Theorem (informally)

Laplacian of a graph

$\mathcal{L} x = b$ can be solved in time $\tilde{O}(|E|)$

(the running time also depends on the error that we want to achieve, but forget that for now)

Now we are ready to eliminate x from our equations!!!

$$y = B \cdot x \iff y \in \mathrm{Im}(B) \Rightarrow$$

$$\iff \Pi y = y \iff (I - \Pi) \cdot y = 0$$

projection onto the set $\mathrm{Im}(B)$

therefore we can rewrite

$y = Bx$

$x_s - x_t = 1$

$$\begin{cases} (I - \Pi) y = 0 \\ \Pi = B(B^T B)^+ B^T = B \mathcal{L}^+ B^T \\ x_s - x_t = 1 \end{cases}$$

let $\tilde{x}_s = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ ← row which corresponds to vertex s

$\tilde{x}_t = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ ← row which correspond to vertex t

so x_s, x_t are coefficients of the vertex space vector x (in our first min-cut formulation)

while $\tilde{x}_s, \tilde{x}_t$ are vectors with a 1 in the position which corresponds to either vertex s or t.

$\tilde{x}_s - \tilde{x}_t \perp \mathbb{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix}$

$x_s - x_t = 1 \Rightarrow x^T (\tilde{x}_s - \tilde{x}_t) = 1 \Rightarrow$

$\Rightarrow x^T (B^T B)(B^T B)^+ (\tilde{x}_s - \tilde{x}_t) = 1 \Rightarrow$

$\Rightarrow x^T B^T B \mathcal{L}^+ (\tilde{x}_s - \tilde{x}_t) = 1 \Rightarrow$

$\Rightarrow (Bx)^T \cdot B \mathcal{L}^+ (\tilde{x}_s - \tilde{x}_t) = 1 \Rightarrow$

$$\Rightarrow (Bx)^T \cdot B\mathcal{L}^+(\tilde{x}_s - \tilde{x}_t) = 1 \xrightarrow[\text{let } y_{st} = B\mathcal{L}^+(\tilde{x}_s - \tilde{x}_t)]{y = Bx \text{ (from the first constraint)}}$$

$$= y^T y_{st} = 1 \quad !!!$$

new formulation

$$\min \|y\|_1$$
$$\text{s.t. } (I - \Pi)y = 0$$
$$y^T y_{st} = 1$$

where $\Pi = B\mathcal{L}^+B^T$ are

$y_{st} = B\mathcal{L}^+(\tilde{x}_s - \tilde{x}_t)$

constant

now let's try to simplify the formulation even more.

Our goal is to describe $(I - \Pi)y = 0$, $y^T y_{st} = 1$

as a single constraint of the form

$Py = b$, where P is a projection matrix

$$(I-\Pi)\cdot y = 0$$

$$y^T y_{st} = 1$$

Note that (1) $(I-\Pi)\, y_{st} =$

$$= (I - B L^+ B^T) \cdot B L^+ (\tilde{x}_s - \tilde{x}_t) =$$

$$= B L^+ (\tilde{x}_s - \tilde{x}_t) - B L^+ \underbrace{\underbrace{B^T B}_{L} L^+ (\tilde{x}_s - \tilde{x}_t)}_{\tilde{x}_s - \tilde{x}_t} =$$

$$= B L^+ (\tilde{x}_s - \tilde{x}_t) - B L^+ (\tilde{x}_s - \tilde{x}_t) =$$

$$= 0$$

(2) solutions to $y^T y_{st} = 1$

are of the form

$y_{feasible} + z$, where

$y^T_{feasible} \cdot y_{st} = 1$ and $z^T \cdot y_{st} = 0$

If we set $y_{feasible} = \frac{y_{st}}{\|y_{st}\|^2}$ then

solutions are of the form

$$y_{st} / \|y_{st}\|^2 + z \;,\; z \perp y_{st}$$

(3) solutions of $(I-\Pi).y=0$ are of the form $y_{feasible}+z$ where

$(I-\Pi)y_{feasible}=0$ and $z\in\ker(I-\Pi)$

(In that case since the RHS is 0 both $y_{feasible}$ and z belong to $\ker(I-\Pi)$)

from (1) we know that $(I-\Pi)\dfrac{y_{s+}}{\|y_{s+}\|^2}=0$

thus the solutions are of the form $\dfrac{y_{s+}}{\|y_{s+}\|^2}+z$, $z\in\ker(I-\Pi)$

From (2) and (3) we get that:

$$\left.\begin{array}{l}(I-\Pi).y=0\\ y^T y_{s+}=1\end{array}\right\}\quad y=\frac{y_{s+}}{\|y_{s+}\|^2}+z,\quad \begin{array}{l}z\perp y_{s+}\\ z\perp \mathrm{span}(I-\Pi)\end{array}$$

$\Rightarrow$

$$y = \frac{y_{st}}{\|y_{st}\|^2} + z \quad , z \perp y_{st} \qquad \Rightarrow \quad Py = b$$

$$z \perp \operatorname{span}(I - \Pi)$$

where

$$P = (I - \Pi) + \frac{y_{st} y_{st}^T}{\|y_{st}\|^2}$$

$$b = \frac{y_{st}}{\|y_{st}\|^2}$$

(because P is symmetric and $P^2 = P$, P is a projection matrix)

(almost) final LP

$$\min \ \|y\|_1$$
$$\text{s.t.} \ Py = b$$

, where

$$P = (I - \Pi) + \frac{y_{st} y_{st}^T}{\|y_{st}\|^2}$$

$$b = \frac{y_{st}}{\|y_{st}\|^2}$$

$$\Pi = B L^+ B^T$$

$$y_{st} = B L^+ (\tilde{x}_s - \tilde{x}_t)$$

$$\Downarrow$$

final LP

$$\min \sum_{e \in E} \sqrt{y_e^2 + \mu^2}$$
$$\text{s.t.} \ Py = b$$

final LP

$$\min \sum_{e \in E} \sqrt{y_e^2 + \mu^2}$$
$$\text{s.t.} \quad Py = b$$

$\Longrightarrow$

$$y_0 = \frac{y_{st}}{\|y_{st}\|^2}$$

$$g(y) = \sum_e \sqrt{y_e^2 + \mu^2}$$

for $t = 0$ to $T-1$

$$y_{t+1} = y_t - \frac{1}{\beta}(I - \Pi)\nabla g(y)$$

Lipschitz parameter of $(I-\Pi)\nabla g(y)$

Running time per iteration

$$I - \Pi = I - (I - P) - \frac{y_{st} y_{st}^T}{\|y_{st}\|^2} =$$

$$= P - \frac{y_{st} y_{st}^T}{\|y_{st}\|^2} = B L^+ B^T - \frac{y_{st} y_{st}^T}{\|y_{st}\|^2}$$

and we want to compute $\left(B L^+ B^T - \frac{y_{st} y_{st}^T}{\|y_{st}\|^2}\right) \nabla g(y)$

- $\frac{y_{st} y_{st}^T \nabla g(y)}{\|y_{st}\|^2} \rightsquigarrow O(|E|)$
- $B L^+ B^T \nabla g(y)$ — $O(|E|)$ (because B has only $2|E|$ non-zero entries) $+ \nabla g(y) \rightsquigarrow O(|E|)$
- $L^+(B^T \nabla g(y)) \leftarrow O(|E|)$ (by solving $Lx = B^T \nabla f(y)$)
- $B(L^+ B^T \nabla g(y)) \leftarrow O(|E|)$ (because B has only $2|E|$ non-zero entries)

all together $\tilde{O}(|E|)$

Now to do the complete analysis, we calculate the Lipschitz constant β of $(I-\Pi)\nabla g(y)$

$$(\nabla g(y))_e = \frac{\partial g(y)}{\partial y_e} = \frac{\partial}{\partial y_e}\sqrt{y_e^2+\mu^2} = \frac{y_e}{\sqrt{y_e^2+\mu^2}}$$

$$(\nabla^2 g(y))_{ee'} = \begin{cases} 0, & e \neq e' \\ \dfrac{\partial^2 g(y)}{\partial y_e^2} = \dfrac{1}{\sqrt{y_e^2+\mu^2}} - \dfrac{y_e^2}{(y_e^2+\mu^2)^{3/2}} = \end{cases}$$

$$= \frac{\mu^2}{(y_e^2+\mu^2)^{3/2}} \leq \frac{\mu^2}{\mu^{2\cdot 3/2}} =$$

$$= 1/\mu$$

$$\Rightarrow \|\nabla g(y) - \nabla g(y')\| \leq \frac{1}{\mu}\|y-y'\|$$

and because $I-\Pi$ is a projection matrix $((I-\Pi)^2 = (I-\Pi))$ we have

$$\|(I-\Pi)\nabla g(y) - (I-\Pi)\nabla g(y')\| \leq \|\nabla g(y) - \nabla g(y')\|$$

thus $(I-\Pi)\nabla g(y)$ is β-Lipschitz with

$$\beta = 1/\mu$$

iterations

As before, we assume that we use Nesterov accelerated method.

$T = O\left(\sqrt{\frac{\beta}{\delta}}\,\|y_0 - y^*\|_2\right)$ to get a

δ-additive error on the objective value of the final optimization program.

Thus, we want $sol \leq F + \mu|E| + \delta =$

$= (1+\varepsilon) F$

$$\mu = \frac{\varepsilon F}{2|E|}, \quad \delta = \frac{\varepsilon F}{2}$$

$$\sqrt{\frac{\beta}{\delta}} = \sqrt{\frac{1}{\mu \cdot \delta}} = O\left(\frac{\sqrt{|E|}}{\varepsilon F}\right)$$

$$\|y_0 - y^*\|_2 = \left\|\frac{y_{st}}{\|y_{st}\|^2} - y^*\right\| = \|proj(0) - proj(y^*)\|_2$$

$$\leq \|0 - y^*\| = \|y^*\| =$$

$$= \sqrt{F}$$

solution with smaller norm

(general solution was

$\frac{y_{st}}{\|y_{st}\|^2} + z$, $z \perp y_{st}$, $z \perp Im(I-P)$)

so # iterations $= O\left(\frac{\sqrt{E}}{\varepsilon\sqrt{F}}\right)$

overall running time to get a $(1+\varepsilon)$ multiplicative approximation is

$$\tilde{O}\left(|E| \cdot \frac{\sqrt{|E|}}{\varepsilon\sqrt{F}}\right) = \tilde{O}\left(\frac{|E|^{3/2}}{\varepsilon\sqrt{F}}\right)$$

(huge improvement over $O\left(\frac{|E|^{3/2}\sqrt{V}}{\varepsilon F}\right)$ for a large range of cut sizes

last (but not least) improvement

$\tilde{O}\left(\frac{|E|^{3/2}}{\varepsilon\sqrt{F}}\right)$ is extremely good when the cut sizes is big. what about balancing it with an algorithm which performs great when the cut size is small (like classic algorithms)

Karger and Levine $\quad \tilde{O}(|E|+|V|F)$

LRS $\quad \tilde{O}\left(\frac{|E|^{3/2}}{\varepsilon\sqrt{F}}\right)$

$$F \begin{cases} > \frac{|E|}{(|V|\varepsilon)^{2/3}} \Rightarrow \text{LRS} \\ \leq \frac{|E|}{(|V|\varepsilon)^{2/3}} \Rightarrow \text{Karger and Levine} \end{cases} \Rightarrow O\left(\frac{|E| \cdot |V|^{1/3}}{\varepsilon^{2/3}}\right)$$