Lecture 8

Min cut using gradient descent

- projection matrices (how they look like)
- gradient descent with equality constraints
- min-cut LP formulation, lemma about integral solutions
- min-cut with GD (vertex space)
    - error analysis

      +
    - running time

Projection

let S be a subspace of $\mathbb{R}^n$ spanned by $\{s_1, s_2, \dots, s_k\}$ vectors $s_i \in \mathbb{R}^n$, $\dim(S) = k$

(thus $s_1, \dots, s_k$ are linearly independent) we want to describe the projection operator $P: \mathbb{R}^n \to S$

s.t. $Pv = \arg\min_{u \in S} \|v-u\|_2$

some observations $Pv \in S$

$v \in S \Rightarrow Pv = v$

($P^2 = P$ if I apply the operator twice I stay at the same position)

step by step

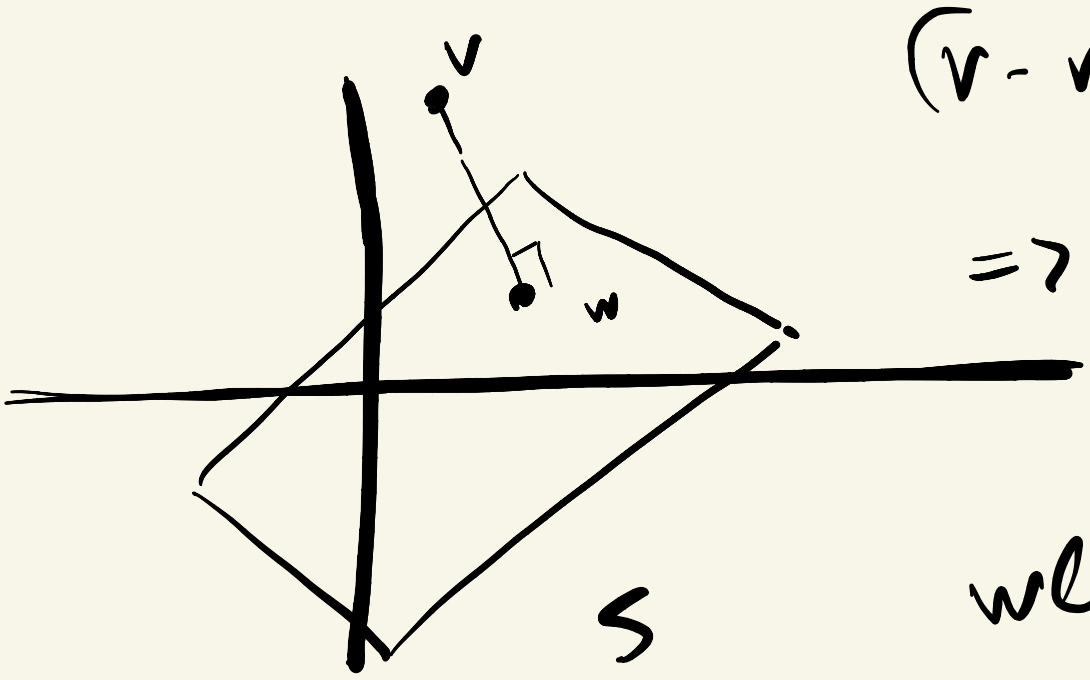
① How can we describe S?

$A = \begin{bmatrix} | & & | \\ s_1 & \dots & s_k \\ | & & | \end{bmatrix}$ S is spanned be $s_1, \dots, s_k$ means that

$$S = \{ x_1 \cdot \vec{s}_1 + x_2 \vec{s}_2 + \dots + x_k \vec{s}_k \mid x = \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix} \in \mathbb{R}^k \} =$$

$$= \underbrace{\{ A \cdot x \mid x \in \mathbb{R}^k \}}_{= \operatorname{Im}(A)}$$

② $w = P \cdot v$ what do we know about $v - w$?



$(v-w) \perp S \Rightarrow$

$\Rightarrow (v-w) \perp \operatorname{Im}(A)$

whatever "guy" lives in $\operatorname{Im}(A) = S$ should be orthogonal to $v - w$.

let $x_1 s_1 + x_2 s_2 + \ldots + x_k s_k$ be an arbitrary "guy" ($x \in \mathbb{R}^k$) leaving in S. Then

$$(v-w)^T (x_1 s_1 + \ldots + x_k s_k) = 0 \quad \forall x \in \mathbb{R}^k$$

$\Downarrow$

$$\left.\begin{array}{ll} x = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \Rightarrow & (v-w)^T \cdot s_1 = 0 \\ x = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \Rightarrow & (v-w)^T s_2 = 0 \\ \vdots & \vdots \\ x = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} & (v-w)^T s_k = 0 \end{array}\right\} (v-w)^T A = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

(and of course if $(v-w)^T \cdot A = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$ then $(v-w)^T \cdot y = 0 \quad \forall y \in S$)

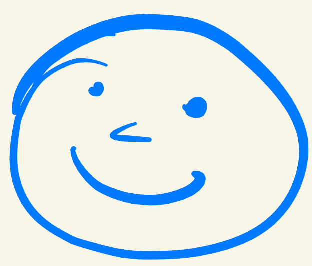
$$(v-w)^T A = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \Rightarrow v^T \cdot A = w^T A \Rightarrow$$

$$w \in S = \mathrm{Im}(A) \Rightarrow \exists x \in \mathbb{R}^k \text{ s.t. } w = Ax$$

$$\Longrightarrow v^T A = x^T A^T A \Rightarrow$$

$$\Rightarrow A^T A \cdot x = A^T v \Rightarrow x = (A^T A)^{-1} A^T v$$

$$\Rightarrow w = \underbrace{A\,(A^T A)^{-1} A^T}_{\parallel \atop P} \cdot v \qquad x$$



some "sanity checks":

① If $s_1, \dots, s_k$ form an orthonormal basis of S then $A^T A = I$ and $P = A \cdot A^T$

$$w = A \begin{bmatrix} - s_1^T - \\ \vdots \\ - s_k^T - \end{bmatrix} \cdot v = A \begin{bmatrix} s_1^T v \\ \vdots \\ s_k^T v \end{bmatrix} = \begin{bmatrix} | & & | \\ s_1 & \cdots & s_k \\ | & & | \end{bmatrix} \begin{bmatrix} s_1^T v \\ \vdots \\ s_k^T v \end{bmatrix} =$$

$$= \sum_{i=1}^{k} (s_i^T v) \cdot \vec{s}_i$$

"projection of v

② $P^2 = A \underbrace{(A^TA)^{-1} \underbrace{A^T A}}_{I} (A^TA)^{-1}A^T =$

$v_i v_i^T$

$= A(A^TA)^{-1}A^T = P$

---

what do you do when A^TA is not invertible?

let $Q = A^TA$ be a symmetric not full rank matrix.

(Def to remember
$\ker(Q) = \{x \mid Qx = 0\}$)

$Q = U \cdot \Lambda \cdot U^T$, U is unitary
Λ is diagonal

$Q = \sum_{i=1}^{k} \lambda_i u_i u_i^T$

Def. pseudoinverse

$Q^+ = \sum_{\substack{i=1 \\ \lambda_i \neq 0}}^{k} \frac{1}{\lambda_i} \cdot u_i u_i^T$

it is called pseudoinverse because as far as $v \notin \ker(Q)$ then

$Q^+ \cdot Q \cdot v = v$

we have a 1-1 mapping
$\forall v \perp \ker(Q)$

$v \perp \ker(Q) \Rightarrow v \in \mathrm{Im}(Q^T) = \mathrm{Im}(Q)$

$\exists x$ s.t. $v = Q \cdot x$

$Q^+ Q v = \sum_{\lambda_i \neq 0} u_i u_i^T \cdot v = \left(\sum_{\lambda_i \neq 0} u_i u_i^T\right)\left(\sum \lambda_i u_i u_i^T\right)x =$

$\underset{U \text{ is unitary}}{=} \left(\sum \lambda_i u_i u_i^T\right)x = Qx = v$

overall:

when A^TA is not invertible then we need to solve

$A^TA\,x = A^Tv$ which is solvable

assuming $x \in \mathrm{Im}(A^T) \perp \ker(A) =$
$= \ker(A^TA)$

we get $x = (A^TA)^+ A^T v$

and $w = A \cdot x = A \underbrace{(A^TA)^+ A^T}_{= P} \cdot v$

by now it should be clear how to define the projection matrix for a set S which is spanned by $\{s_1, s_2, \dots, s_k\}$

Gradient descent with equality constraints

Goal: solve using GD

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

$A \in \mathbb{R}^{m \times n}$
$x \in \mathbb{R}^n$
$b \in \mathbb{R}^m$
$m \leq n$

$\{x \mid Ax = b\} = x_0 + \ker(A)$, x_0 feasible solution

let $\{\mu_1, \dots, \mu_k\}$ be a basis of $\ker(A)$

$U = \begin{bmatrix} | & & | \\ \mu_1 & \cdots & \mu_k \\ | & & | \end{bmatrix} \Rightarrow \ker(A) = \{U \cdot z \mid z \in \mathbb{R}^k\}$

thus $x_0 + \ker(A) = \{x_0 + Uz \mid z \in \mathbb{R}^k\}$

$$\begin{aligned} &\min f(x) \\ &\text{s.t } Ax = b \end{aligned} \Longrightarrow \min f(x_0 + Uz), \; z \in \mathbb{R}^k$$

constrained

unconstrained

$$\nabla_z f(x_0+Uz) = U \cdot \nabla_{x_0+U_z} f(x_0+Uz)$$

G.D

start with $z_0 \in \mathbb{R}^k$, $t=0$

update step $z_{t+1} = z_t - \eta_t U \nabla f(x_0+Uz_t)$

$t \leftarrow t+1$

repeat until stopping condition

---

now we will focus to the case where the equality constraint is of the form $P \cdot x = b$ where P is a projection matrix.

observations

subspace to which P projects onto

- $Pb = b$ (o.w the optimization problem is not feasible, since
  - $Pb \neq b \Rightarrow b \notin S$
  - $Px \in S \;\; \forall x$

  so $\overset{\in S}{Px} = \overset{\notin S}{b}$ can never be true

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

$P \cdot (I-P) = P - P^2$ $\underline{\underline{P=P^2}}$, since P is a projection matrix

$= 0$

$$\mathcal{U} \cdot \nabla_{x_0+\mathcal{U}z} f(x_0+\mathcal{U}z)$$

$Pb = b$
$\Rightarrow b$ is feasible

$\ker(P) = \mathrm{span}\{I-P\}$
$\Rightarrow \mathcal{U} = I - P$

$$(I-P)\nabla f(b + (I-P)z)$$

let z_t be feasible
$(Pz_t = b)$

$$(I-P)\nabla f(b + z_t - \underbrace{P \cdot z_t}_{= b}) =$$

$$= (I-P)\nabla f(z_t)$$

moreover $z_{t+1} = z_t - \eta \cdot (I-P)\nabla f(z_t)$

is still feasible

$$P z_{t+1} = P z_t - \eta \underbrace{P(I-P)}_{0}\nabla f(z_t) =$$

$$= P z_t \overset{z_t \text{ is feasible}}{=} b$$

If I start from a feasible point
I will remain feasible and consequently
I will have this simplification

Let β be the Lipschitz parameter of $(I-P)\nabla f(z)$ $\left(\|(I-P)\nabla f(z) - (I-P)\nabla f(z')\| \le \beta\|z-z'\|\right)$

then GD is

- is feasible ($Pz_0 = b$), $t=0$
- $z_{t+1} = z_t - \frac{1}{\beta}(I-P)\nabla f(z_t)$

$t \leftarrow t+1$

repeat if stopping condition is not satisfied

unconstrained!!!

$O\left(\frac{\beta}{\varepsilon}\|z^*-z_0\|^2\right)$ iterations

to get an ε-error

addition

when using Nesterov accelerated methods we get

here is no 2 here!!!

$O\left(\sqrt{\frac{\beta}{\varepsilon}}\|z^*-z_0\|_2\right)$ iterations

Min-cut using LP (finally)

to make it a classic LP just note that
$|x| = \max\{x, -x\}$
$= \min y$ s.t $y \geq x$, $y \geq -x$

step 1

formulate the problem as an LP

$$\min \sum_{(u,v) \in E} |x_u - x_v|$$

s.t $x_s = 0$, $x_t = 1$

$x_u \in \{0,1\}\ \forall u \in V$

$ILP \Rightarrow LP$ $\Longrightarrow$

$$\min \sum_{(u,v) \in E} |x_u - x_v|$$

s.t $x_s - x_t = 1$

$0 \leq x_u \leq 1\ \forall u \in V$

(not a classic LP in this form but you can make it so)

claim

let's assume for simplicity that $x_u \in [0,1]\ \forall u \in V$

let x be a solution with objective value p. Then $\exists$ an integral solution with at most $\lfloor p \rfloor$ edges.

proof

let $S_\ell = \{v \mid x_v \geq \ell\}$ and $\delta(S_\ell)$ the set of edges associated with the cut $(S_\ell, V - S_\ell)$. Note that

$$p = \sum_{(u,v) \in E} |x_u - x_v| \geq \int_0^1 |\delta(S_\ell)|\, d\ell \Rightarrow$$

$$\exists \ell : |\delta(S_\ell)| \leq p$$

we can find it in $\min(|E|, |V| \log |V|)$

step 1 result: If we have a fractional solution of value p, we can easily find an integral solution of value $\lfloor p \rfloor$.

So let's concentrate in finding a fractional solution.

step 2: change the LP to something GD is good at.

make the function differentiable and smooth

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

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

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

Error analysis

$$|x_u - x_v| < \sqrt{(x_u-x_v)^2 + \mu^2} \leq |x_u - x_v| + \mu$$

let F be the number of edges in a minimum cut, and x^* the description of the cut.

$\Rightarrow$ the optimal solution of our optimization problem has value at most $\sum_{(u,v)\in E} \left(|x^*_u - x^*_v| + \mu \right) =$

$$= F + \mu|E|$$

now assume that using the accelerated Nesterov method we find a solution of value at most $F + \mu|E| + \delta$

parameter that we add to make our function differentiable

additive error

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

$\Rightarrow (1+\varepsilon)F$

Running time estimate

Nesterov requires $O\left(\sqrt{\frac{\beta}{\delta}}\ \|x_0 - x^*\|_2\right)$ iterations where $\beta \simeq 1/\mu$ is the smoothness parameter

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

$$\Rightarrow \sqrt{\frac{\beta}{\delta}} = 2\sqrt{\frac{|E|}{F^2}} = O\left(\frac{\sqrt{|E|}}{F}\right)$$

$$\|x_0 - x^*\|_2 \leq \sqrt{V}$$

and in each iteration the gradient computation requires $O(E)$

$$\rightsquigarrow O\left(\frac{|E|^{3/2} \cdot \sqrt{V}}{\varepsilon F}\right) \text{ vs } O(|E|^{3/2})$$

combinatorial algorithm which solves the problem exactly

observations

1. combinatorial is better even when $F = \sqrt{V}$

2. the running time get's better as the cut increases (something that is not true in the augmenting paths algorithms)

3. $\| x_0 - x^* \|_2 \leq \sqrt{V}$ is a very bad initial bound
   Can we ameliorate it?