Advanced Graph Algorithms and Optimization Spring 2020

Effective Resistance, Gaussian Elimination as Optimization

Rasmus Kyng, Scribe: Hongjie Chen *Lecture 6 — Wednesday, March 25th*

1 What is a (Moore-Penrose) Pseudoinverse?

Recall that for a connected graph G with Laplacian $\boldsymbol{L}$, we have $\ker(\boldsymbol{L}) = \text{span}\{\boldsymbol{1}\}$, which means $\boldsymbol{L}$ is not invertible. However, we still want some matrix which behaves like a real inverse. To be more specific, given a Laplacian $\boldsymbol{L} \in \mathbb{R}^{V\times V}$, we want some matrix $\boldsymbol{L}^+ \in \mathbb{R}^{V\times V}$ s.t.

1) $(\boldsymbol{L}^+)^\top = \boldsymbol{L}^+$ (symmetric)
2) $\boldsymbol{L}^+\boldsymbol{1} = \boldsymbol{0}$, or more generally, $\boldsymbol{L}^+\boldsymbol{v} = \boldsymbol{0}$ for $\boldsymbol{v} \in \ker(\boldsymbol{L})$
3) $\boldsymbol{L}^+\boldsymbol{L}\boldsymbol{v} = \boldsymbol{L}\boldsymbol{L}^+\boldsymbol{v} = \boldsymbol{v}$ for $\boldsymbol{v} \perp \boldsymbol{1}$, or more generally, for $\boldsymbol{v} \in \ker(\boldsymbol{L})^\perp$

Under the above conditions, $\boldsymbol{L}^+$ is uniquely defined and we call it the pseudoinverse of $\boldsymbol{L}$. Note that there are many other equivalent definitions of the pseudoinverse of some matrix $\boldsymbol{A}$, and we can also generalize the concept to matrices that aren't symmetric or even square.

Let $\lambda_i, \boldsymbol{v}_i$ be the i-th pair of eigenvalue and eigenvector of $\boldsymbol{L}$, with $\{\boldsymbol{v}_i\}_{i=1}^n$ forming a orthogonal basis. Then by the spectral theorem,

$$\boldsymbol{L} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^\top = \sum_i \lambda_i \boldsymbol{v}_i\boldsymbol{v}_i^\top,$$

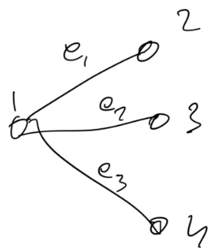
where $\boldsymbol{V} = \begin{bmatrix}\boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n\end{bmatrix}$ and $\boldsymbol{\Lambda} = \text{diag}\{\lambda_1, ..., \lambda_n\}$. And we can show that its pseudoinverse is exactly

$$\boldsymbol{L}^+ = \sum_{i,\lambda_i\neq 0} \lambda_i^{-1}\boldsymbol{v}_i\boldsymbol{v}_i^\top.$$

Checking conditions 1), 2), 3) is immediate.

2 Electrical Flows Again

Recall the incidence matrix $\boldsymbol{B} \in \mathbb{R}^{V\times E}$ of a graph $G = (V, E)$.



$$
\begin{array}{c|ccc}
 & e_1 & e_2 & e_3 \\
\hline
1 & +1 & +1 & +1 \\
2 & -1 & 0 & 0 \\
3 & 0 & -1 & 0 \\
4 & 0 & 0 & -1
\end{array}
$$

Figure 1: An example of a graph and its incidence matrix B.

In Lecture 1, we introduced the electrical flow routing demand $\boldsymbol{d} \in \mathbb{R}^V$. Let's call the electrical flow $\tilde{\boldsymbol{f}} \in \mathbb{R}^E$. The net flow constraint requires $\boldsymbol{B}\tilde{\boldsymbol{f}} = \boldsymbol{d}$. By Ohm's Law, $\tilde{\boldsymbol{f}} = \boldsymbol{R}^{-1}\boldsymbol{B}^\top \boldsymbol{x}$ for some voltage $x \in \mathbb{R}^V$ where $\mathbb{R} = \operatorname{diag}(\boldsymbol{r})$ and $\boldsymbol{r}(e) =$ resistance of edge e. We showed (in the exercises) that when $\boldsymbol{d} \perp \boldsymbol{1}$, there exists an voltage $\tilde{\boldsymbol{x}} \perp \boldsymbol{1}$ s.t. $\tilde{\boldsymbol{f}} = \boldsymbol{R}^{-1}\boldsymbol{B}^\top \tilde{\boldsymbol{x}}$ and $\boldsymbol{B}\tilde{\boldsymbol{f}} = \boldsymbol{d}$. This $\tilde{\boldsymbol{x}}$ solves $\boldsymbol{L}\boldsymbol{x} = \boldsymbol{d}$ where $\boldsymbol{L} = \boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{B}^\top$.

And we also made the following claim.

Claim 2.1.

$$\tilde{\boldsymbol{f}} = \arg\min_{\boldsymbol{B}\boldsymbol{f}=\boldsymbol{d}} \boldsymbol{f}^\top \boldsymbol{R}\boldsymbol{f} \ \textit{ where } \boldsymbol{f}^\top \boldsymbol{R}\boldsymbol{f} = \sum_e \boldsymbol{r}(e)\boldsymbol{f}(e)^2, \tag{1}$$

But we didn't prove this. Now, let's prove it.

Proof. Consider any $\boldsymbol{f} \in \mathbb{R}^E$ s.t. $\boldsymbol{B}\boldsymbol{f} = \boldsymbol{d}$. For any $\boldsymbol{x} \in \mathbb{R}^V$, we have

$$
\begin{aligned}
\frac{1}{2}\boldsymbol{f}^\top \boldsymbol{R}\boldsymbol{f} &= \frac{1}{2}\boldsymbol{f}^\top \boldsymbol{R}\boldsymbol{f} - \boldsymbol{x}^\top(\underbrace{\boldsymbol{B}\boldsymbol{f} - \boldsymbol{d}}_{\boldsymbol{0}}) \\
&\geq \min_{\boldsymbol{f}\in\mathbb{R}^E} \underbrace{\frac{1}{2}\boldsymbol{f}^\top \boldsymbol{R}\boldsymbol{f} - \boldsymbol{x}^\top \boldsymbol{B}\boldsymbol{f} + \boldsymbol{d}^\top \boldsymbol{x}}_{g(\boldsymbol{f})} \\
&= \boldsymbol{d}^\top \boldsymbol{x} - \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x}
\end{aligned}
$$

since $\nabla_{\boldsymbol{f}} g(\boldsymbol{f}) = \boldsymbol{0}$ gives us $\boldsymbol{f} = \boldsymbol{R}^{-1}\boldsymbol{B}^\top \boldsymbol{x}$. Thus, for all $\boldsymbol{f} \in \mathbb{R}^E$ s.t. $\boldsymbol{B}\boldsymbol{f} = \boldsymbol{d}$ and all $\boldsymbol{x} \in \mathbb{R}^V$,

$$\frac{1}{2}\boldsymbol{f}^\top \boldsymbol{R}\boldsymbol{f} \geq \boldsymbol{d}^\top \boldsymbol{x} - \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x}. \tag{2}$$

But for the electrical flow $\tilde{\boldsymbol{f}}$ and electrical voltage $\tilde{\boldsymbol{x}}$, we have $\tilde{\boldsymbol{f}} = \boldsymbol{R}^{-1}\boldsymbol{B}^\top \tilde{\boldsymbol{x}}$ and $\boldsymbol{L}\tilde{\boldsymbol{x}} = \boldsymbol{d}$. So

$$\tilde{\boldsymbol{f}}^\top \boldsymbol{R}\tilde{\boldsymbol{f}} = \left(\boldsymbol{R}^{-1}\boldsymbol{B}^\top \tilde{\boldsymbol{x}}\right)^\top \boldsymbol{R}\left(\boldsymbol{R}^{-1}\boldsymbol{B}^\top \tilde{\boldsymbol{x}}\right) = \tilde{\boldsymbol{x}}^\top \boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{B}^\top \tilde{\boldsymbol{x}} = \tilde{\boldsymbol{x}}^\top \boldsymbol{L}\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{x}}^\top \boldsymbol{d}.$$

Therefore,

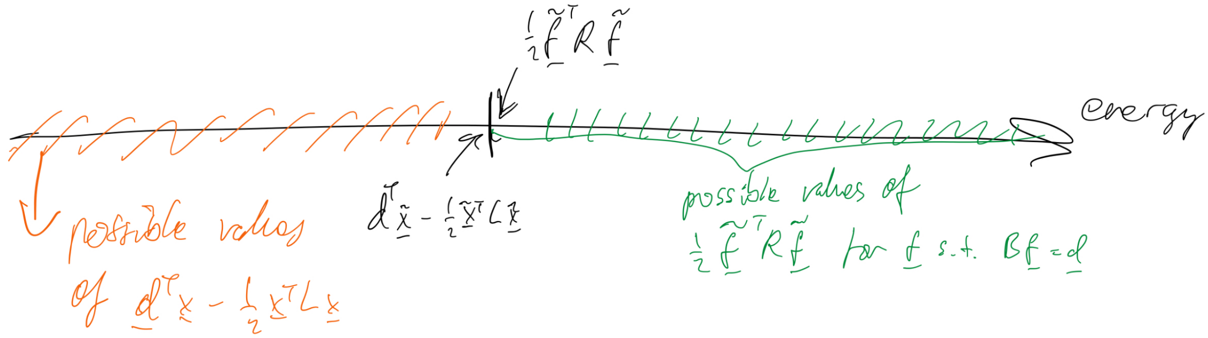
$$\frac{1}{2}\tilde{\boldsymbol{f}}^\top \boldsymbol{R}\tilde{\boldsymbol{f}} = \boldsymbol{d}^\top \tilde{\boldsymbol{x}} - \frac{1}{2}\tilde{\boldsymbol{x}}^\top \boldsymbol{L}\tilde{\boldsymbol{x}}. \tag{3}$$

By combining Equation (2) and Equation (3), we see that for all $\boldsymbol{f}$ s.t. $\boldsymbol{B}\boldsymbol{f} = \boldsymbol{d}$,

$$\frac{1}{2}\boldsymbol{f}^\top \boldsymbol{R}\boldsymbol{f} \geq \boldsymbol{d}^\top \tilde{\boldsymbol{x}} - \frac{1}{2}\tilde{\boldsymbol{x}}^\top \boldsymbol{L}\tilde{\boldsymbol{x}} = \frac{1}{2}\tilde{\boldsymbol{f}}^\top \boldsymbol{R}\tilde{\boldsymbol{f}}.$$

Thus $\tilde{\boldsymbol{f}}$ is the minimum electrical energy flow among all flows that route demand $\boldsymbol{d}$, proving Equation (1) holds.

The drawing below shows how the quantities line up:



$\square$

3 Effective Resistance

Given a graph $G = (V, E)$, for any pair of vertices $(a, b) \in V$, we want to compute the cost (or energy) of routing 1 unit of current from a to b. We call such cost the effective resistance between a and b, denoted by $R_{\text{eff}}(a, b)$. Recall for a single resistor $r(a, b)$,

$$\text{energy} = r(a, b) f^2(a, b) = r(a, b).$$

So when we have a graph consisting of just one edge (a, b), the effective resistance is just $R_{\text{eff}}(a, b) = r(a, b)$.

In a general graph, we can also consider the energy required to route one unit of current between two vertices. For any pair $a, b \in V$, we have

$$R_{\text{eff}}(a, b) = \min_{\boldsymbol{B}\boldsymbol{f} = \boldsymbol{e}_b - \boldsymbol{e}_a} \boldsymbol{f}^\top \boldsymbol{R}\boldsymbol{f}.$$

Note that the cost of routing F units of flow from a to b will be $R_{\text{eff}}(a, b) \cdot F^2$.

Since $(\boldsymbol{e}_b - \boldsymbol{e}_a)^\top \boldsymbol{1} = 0$, we know from the previous section that $R_{\text{eff}}(a, b) = \tilde{\boldsymbol{f}}^\top \boldsymbol{R}\tilde{\boldsymbol{f}}$ where $\tilde{\boldsymbol{f}}$ is the electrical flow. Now we can write $\boldsymbol{L}\tilde{\boldsymbol{x}} = \boldsymbol{e}_b - \boldsymbol{e}_a$ and $\tilde{\boldsymbol{x}} = \boldsymbol{L}^+(\boldsymbol{e}_b - \boldsymbol{e}_a)$ for the electrical voltages routing 1 unit of current from a to b. Now the energy of routing 1 unit of current from a to b is

$$R_{\text{eff}}(a, b) = \tilde{\boldsymbol{f}}^\top \boldsymbol{R}\tilde{\boldsymbol{f}} = \tilde{\boldsymbol{x}}^\top \boldsymbol{L}\tilde{\boldsymbol{x}} = (\boldsymbol{e}_b - \boldsymbol{e}_a)^\top \boldsymbol{L}^+\boldsymbol{L}\boldsymbol{L}^+(\boldsymbol{e}_b - \boldsymbol{e}_a) = (\boldsymbol{e}_b - \boldsymbol{e}_a)^\top \boldsymbol{L}^+(\boldsymbol{e}_b - \boldsymbol{e}_a),$$

where the last equality is due to $\boldsymbol{L}^+\boldsymbol{L}\boldsymbol{L}^+ = \boldsymbol{L}^+$.

Remark 3.1. We have now seen several different expressions that all take on the same value: the energy of the electrical flow. It's useful to remind yourself what these are. Consider an electrical flow $\tilde{\boldsymbol{f}}$ routes demand $\boldsymbol{d}$, and associated electrical voltages $\tilde{\boldsymbol{x}}$. We know that $\boldsymbol{B}\tilde{\boldsymbol{f}} = \boldsymbol{d}$, and $\boldsymbol{f} = \boldsymbol{R}^{-1}\boldsymbol{B}^\top \tilde{\boldsymbol{x}}$, and $\boldsymbol{L}\tilde{\boldsymbol{x}} = \boldsymbol{d}$, where $\boldsymbol{L} = \boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{B}^\top$. And we have seen how to express the electrical energy using many different quantities:

$$\tilde{\boldsymbol{f}}^\top \boldsymbol{R}\tilde{\boldsymbol{f}} = \tilde{\boldsymbol{x}}^\top \boldsymbol{L}\tilde{\boldsymbol{x}} = \boldsymbol{d}^\top \boldsymbol{L}^+ \boldsymbol{d} = \boldsymbol{d}^\top \tilde{\boldsymbol{x}} = \tilde{\boldsymbol{f}}^\top \boldsymbol{B}^\top \tilde{\boldsymbol{x}}$$

Claim 3.2. *Any PSD matrix $\boldsymbol{A}$ has a PSD square root $\boldsymbol{A}^{1/2}$ s.t. $\boldsymbol{A}^{1/2}\boldsymbol{A}^{1/2} = \boldsymbol{A}$.*

Proof. By the spectral theorem, $AA = \sum_i \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^\top$ where $\{\boldsymbol{v}_i\}$ are orthonormal. Let $\boldsymbol{A}^{1/2} = \sum_i \lambda_i^{1/2} \boldsymbol{v}_i \boldsymbol{v}_i^\top$. Then

$$\begin{aligned}
\boldsymbol{A}^{1/2}\boldsymbol{A}^{1/2} &= \left(\sum_i \lambda_i^{1/2} \boldsymbol{v}_i \boldsymbol{v}_i^\top\right)^2 \\
&= \sum_i \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^\top \boldsymbol{v}_i \boldsymbol{v}_i^\top + \sum_{i \neq j} \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^\top \boldsymbol{v}_j \boldsymbol{v}_j^\top \\
&= \sum_i \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^\top
\end{aligned}$$

where the last equality is due to $\boldsymbol{v}_i^\top \boldsymbol{v}_j = \delta_{ij}$. It's easy to see that $\boldsymbol{A}^{1/2}$ is also PSD. □

Let $\boldsymbol{L}^{+/2}$ be the square root of $\boldsymbol{L}$. So

$$R_{\text{eff}}(a,b) = (\boldsymbol{e}_b - \boldsymbol{e}_a)^\top \boldsymbol{L}^+ (\boldsymbol{e}_b - \boldsymbol{e}_a) = \|\boldsymbol{L}^{+/2}(\boldsymbol{e}_b - \boldsymbol{e}_a)\|^2.$$

Example: Effective resistance in a path. Consider a path graph on vertices $V = \{1, 2, 3, \ldots, k+1\}$, with with resistances $\boldsymbol{r}(1), \boldsymbol{r}(2), \ldots, \boldsymbol{r}(k)$ on the edges of the path.

Figure 2: A path graph with k edges.

The effective resistance between the endpoints is

$$R_{\text{eff}}(1, k+1) = \sum_{i=1}^{k} \boldsymbol{r}(i)$$

To see this, observe that to have 1 unit of flow going from vertex 1 to vertex $k+1$, we must have one unit flowing across each edge i. Let $\boldsymbol{\Delta}(i)$ be the voltage difference across edge i, and $\boldsymbol{f}(i)$ the flow on the edge. Then $1 = \boldsymbol{f}(i) = \frac{\boldsymbol{\Delta}(i)}{\boldsymbol{r}(i)}$, so that $\boldsymbol{\Delta}(i) = \boldsymbol{r}(i)$. The electrical voltages are then $\tilde{\boldsymbol{x}} \in \mathbb{R}^V$ where $\tilde{\boldsymbol{x}}(i) = \tilde{\boldsymbol{x}}(1) + \sum_{j<i} \boldsymbol{\Delta}(j)$. Hence the effective resistance is

$$R_{\text{eff}}(1, k+1) = \boldsymbol{d}^\top \tilde{\boldsymbol{x}} = (\boldsymbol{e}_{k+1} - \boldsymbol{e}_1)^\top \tilde{\boldsymbol{x}} = \tilde{\boldsymbol{x}}(k+1) - \tilde{\boldsymbol{x}}(1) = \sum_{i=1}^{k} \boldsymbol{r}(i).$$

This behavior is sometimes known as the fact that the resistance of resistors adds up when they are connected in series.

Example: Effective resistance of parallel edges. So far, we have only considered graphs with at most one edge between any two vertices. But that math also works if we allow a pair of vertices to have multiple distinct edges connecting them. We refer to this as *multi-edges*. Suppose we have a graph on just two vertices, $V = \{1, 2\}$, and these are connected by k parallel multi-edges with resistances $\boldsymbol{r}(1), \boldsymbol{r}(2), \ldots, \boldsymbol{r}(k)$.

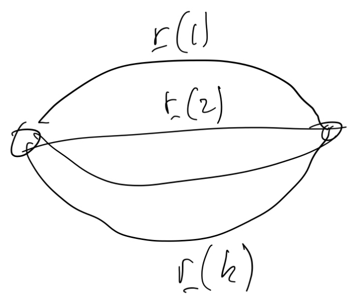


Figure 3: A graph on just two vertices with k parallel multiedges.

The effective resistance between the endpoints is

$$R_{\text{eff}}(1, 2) = \frac{1}{\sum_{i=1}^{k} 1/\boldsymbol{r}(i)}.$$

Let's see why. Our electrical voltages $\tilde{\boldsymbol{x}} \in \mathbb{R}^V$ can be described by just the voltage difference $\Delta \in \mathbb{R}$ between vertex 1 and vertex 2, i.e. $\tilde{\boldsymbol{x}}(2) - \tilde{\boldsymbol{x}}(1) = \Delta$. which creates a flow on edge i of $\tilde{\boldsymbol{f}}(i) = \Delta/\boldsymbol{r}(i)$. Thus the total flow from vertex 1 to vertex 2 is $1 = \sum_i \Delta/\boldsymbol{r}(i)$, so that $\Delta = \frac{1}{\sum_{i=1}^{k} 1/\boldsymbol{r}(i)}$. Meanwhile, the effective resistance is also

$$R_{\text{eff}}(1, 2) = (\boldsymbol{e}_2 - \boldsymbol{e}_1)^\top \tilde{\boldsymbol{x}} = \Delta = \frac{1}{\sum_{i=1}^{k} 1/\boldsymbol{r}(i)}$$

3.1 Effective Resistance is a Distance

Definition 3.3. Consider a weighted undirected graph G with vertex set V. We say function $d : V \times V \to \mathbb{R}$, which takes a pair of vertices and returns a real number, is a *distance* if it satisfies

1. $d(a,a) = 0$ for all $a \in V$
2. $d(a,b) \geq 0$ for all $a,b \in V$.
3. $d(a,b) = d(b,a)$ for all $a,b \in V$.
4. $d(a,b) \leq d(a,c) + d(c,b)$ for all $a,b,c \in V$.

Lemma 3.4. *R_{eff} is a distance.*

Before proving this lemma, let's see a claim that will help us finish the proof.

Claim 3.5. *Let $\boldsymbol{L}\tilde{\boldsymbol{x}} = \boldsymbol{e}_b - \boldsymbol{e}_a$. Then for all $\boldsymbol{c} \in V$, we have $\tilde{\boldsymbol{x}}(b) \geq \tilde{\boldsymbol{x}}(c) \geq \tilde{\boldsymbol{x}}(a)$.*

We only sketch a proof of this claim:

Proof sketch. Consider any $c \in V$, where $c \neq a, b$. Now $(\boldsymbol{L}\tilde{\boldsymbol{x}})(c) = 0$, i.e.

$$\left(\sum_{(u,c)} w(u,c)\right) \tilde{\boldsymbol{x}}(c) - \left(\sum_{(u,c)} w(u,c)\tilde{\boldsymbol{x}}(u)\right) = 0$$

Rearranging $\tilde{\boldsymbol{x}}(c) = \frac{\sum_{(u,c)} w(u,c)\tilde{\boldsymbol{x}}(u)}{\sum_{(u,c)} w(u,c)}$. This tells us that $\tilde{\boldsymbol{x}}(c)$ is a weighted average of the voltages of its neighbors. From this, we can show that $\tilde{\boldsymbol{x}}(a)$ and $\tilde{\boldsymbol{x}}(b)$ are the extreme values. $\square$

Proof. It is easy to check that conditions 1, 2, and 3 of Definition 3.3 are satisfied by R_{eff}. Let us confirm condition 4.

For any u, v, let $\tilde{\boldsymbol{x}}_{u,v} = \boldsymbol{L}^{+}(-\boldsymbol{e}_u + \boldsymbol{e}_v)$. Then

$$\tilde{\boldsymbol{x}}_{a,b} = \boldsymbol{L}^{+}(-\boldsymbol{e}_a + \boldsymbol{e}_b) = \boldsymbol{L}^{+}(-\boldsymbol{e}_a + \boldsymbol{e}_c - \boldsymbol{e}_c + \boldsymbol{e}_b) = \tilde{\boldsymbol{x}}_{a,c} + \tilde{\boldsymbol{x}}_{c,b}.$$

Thus,

$$\begin{aligned} R_{\text{eff}}(a,b) = (-\boldsymbol{e}_a + \boldsymbol{e}_b)^{\top} \tilde{\boldsymbol{x}}_{a,b} &= (-\boldsymbol{e}_a + \boldsymbol{e}_b)^{\top} (\tilde{\boldsymbol{x}}_{a,c} + \tilde{\boldsymbol{x}}_{c,b}) \\ &= -\tilde{\boldsymbol{x}}_{a,c}(a) + \tilde{\boldsymbol{x}}_{a,c}(b) - \tilde{\boldsymbol{x}}_{c,b}(a) + \tilde{\boldsymbol{x}}_{c,b}(b) \\ &= -\tilde{\boldsymbol{x}}_{a,c}(a) + \tilde{\boldsymbol{x}}_{a,c}(c) - \tilde{\boldsymbol{x}}_{c,b}(c) + \tilde{\boldsymbol{x}}_{c,b}(b). \end{aligned}$$

where in the last line we applied Lemma 3.5 to show that $\tilde{\boldsymbol{x}}_{a,c}(b) \leq \tilde{\boldsymbol{x}}_{a,c}(c)$ and $-\tilde{\boldsymbol{x}}_{c,b}(a) \leq -\tilde{\boldsymbol{x}}_{c,b}(c)$. $\square$

4 An Optimization Perspective on Gaussian Elimination for Laplacians

In this section, we will explore how to exactly minimize a Laplacian quadratic form by minimizing over one variable at a time. It turns out that this is in fact Gaussian Elimination in disguise – or, more precisely, the variant of Gaussian elimination that we tend to use on symmetric matrices, which is called Cholesky factorization.

Consider a Laplacian $\boldsymbol{L}$ of a connected graph $G = (V, E, \boldsymbol{w})$, where $\boldsymbol{w} \in \mathbb{R}^E$ is a vector of positive edge weights. Let $\boldsymbol{W} \in \mathbb{R}^{E \times E}$ be the diagonal matrix with the edge weights on the diagonal, i.e. $\boldsymbol{W} = \operatorname{diag}(\boldsymbol{w})$ and $\boldsymbol{L} = \boldsymbol{B}\boldsymbol{W}\boldsymbol{B}^\top$. Let $\boldsymbol{d} \in \mathbb{R}^V$ be a demand vector s.t. $\boldsymbol{d} \perp \boldsymbol{1}$.

Let us define an energy

$$\mathcal{E}(\boldsymbol{x}) = -\boldsymbol{d}^\top \boldsymbol{x} + \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x}$$

Note that this function is convex and is minimized at $\boldsymbol{x}$ s.t. $\boldsymbol{L}\boldsymbol{x} = \boldsymbol{d}$.

We will now explore an approach to solving the minimization problem

$$\min_{\boldsymbol{x} \in \mathbb{R}^V} \mathcal{E}(\boldsymbol{x})$$

Let $\boldsymbol{x} = \begin{pmatrix} y \\ \boldsymbol{z} \end{pmatrix}$ where $y \in \mathbb{R}$ and $\boldsymbol{z} \in \mathbb{R}^{V \setminus \{1\}}$.

We will now explore how to minimize over y, given any $\boldsymbol{z}$. Once we find an expression for y in terms of $\boldsymbol{z}$, we will be able to reduce will result in new quadratic minimization problem in $\boldsymbol{z}$,

$$\mathcal{E}'(\boldsymbol{z}) = -\boldsymbol{d}'^\top \boldsymbol{z} + \frac{1}{2}\boldsymbol{z}^\top \boldsymbol{L}'\boldsymbol{z}$$

where $\boldsymbol{d}'$ is a demand vector on the remaining vertices, with $\boldsymbol{d} \perp \boldsymbol{1}$ and $\boldsymbol{L}'$ is a Laplacian of a graph on the remaining vertices $V' = V \setminus \{1\}$. We can then repeat the procedure to eliminate another variable and so on. Eventually, we can then find all the solution to our original minimization problem.

To help us understand how to mimize over the first variable, we introduce some notation for the first row and column of the Laplacian:

$$\boldsymbol{L} = \begin{pmatrix} W & -\boldsymbol{a}^\top \\ -\boldsymbol{a} & \operatorname{diag}(\boldsymbol{a}) + \boldsymbol{L}_{-1} \end{pmatrix} \tag{4}$$

Note that W is the weighted degree of vertex 1, and that

$$\begin{pmatrix} W & -\boldsymbol{a}^\top \\ -\boldsymbol{a} & \operatorname{diag}(\boldsymbol{a}) \end{pmatrix} \tag{5}$$

is the Laplacian of the subgraph of G containing only the edges incident on vertex 1, while $\boldsymbol{L}_{-1}$ is the Laplacian of the subgraph of G containing all edges *not* incident on vertex 1.

Let us also write $\boldsymbol{d} = \begin{pmatrix} b \\ \boldsymbol{c} \end{pmatrix}$ where $y \in \mathbb{R}$ and $\boldsymbol{c} \in \mathbb{R}^{V \setminus \{1\}}$.

Now,

$$\mathcal{E}(\boldsymbol{x}) = -\boldsymbol{d}^\top \boldsymbol{x} + \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x} = -\begin{pmatrix} b \\ \boldsymbol{c} \end{pmatrix}^\top \begin{pmatrix} y \\ \boldsymbol{z} \end{pmatrix} + \frac{1}{2}\begin{pmatrix} y \\ \boldsymbol{z} \end{pmatrix}^\top \begin{pmatrix} W & -\boldsymbol{a}^\top \\ -\boldsymbol{a} & \operatorname{diag}(\boldsymbol{a}) + \boldsymbol{L}_{-1} \end{pmatrix} \begin{pmatrix} y \\ \boldsymbol{z} \end{pmatrix}$$

$$= -by - \boldsymbol{c}^\top \boldsymbol{z} + \frac{1}{2}\left(y^2 W - 2y\boldsymbol{a}^\top \boldsymbol{z} + \boldsymbol{z}^\top \operatorname{diag}(\boldsymbol{a})\boldsymbol{z} + \boldsymbol{z}^\top \boldsymbol{L}_{-1}\boldsymbol{z}\right)$$

Now, to minimize over y, we set $\frac{\partial}{\partial y}\mathcal{E}(\boldsymbol{x}) = 0$ and get

$$-b + yW - \boldsymbol{a}^\top \boldsymbol{z} = 0.$$

Solving for y, we get that the minimizing y is

$$y = \frac{1}{W}(b + \boldsymbol{a}^\top \boldsymbol{z}). \tag{6}$$

Observe that

$$\mathcal{E}(\boldsymbol{x}) = -by - \boldsymbol{c}^\top \boldsymbol{z} + \frac{1}{2}\left(y^2 W - 2y\boldsymbol{a}^\top \boldsymbol{z} + \boldsymbol{z}^\top \operatorname{diag}(\boldsymbol{a})\boldsymbol{z} + \boldsymbol{z}^\top \boldsymbol{L}_{-1}\boldsymbol{z}\right)$$

$$= -by - \boldsymbol{c}^\top \boldsymbol{z} + \frac{1}{2}\left(\frac{1}{W}(yW - \boldsymbol{a}^\top \boldsymbol{z})^2 \underbrace{- \frac{1}{W}\boldsymbol{z}^\top \boldsymbol{a}\boldsymbol{a}^\top \boldsymbol{z} + \boldsymbol{z}^\top \operatorname{diag}(\boldsymbol{a})\boldsymbol{z} + \boldsymbol{z}^\top \boldsymbol{L}_{-1}\boldsymbol{z}}_{\text{Let } \boldsymbol{S} = \operatorname{diag}(\boldsymbol{a}) - \frac{1}{W}\boldsymbol{a}\boldsymbol{a}^\top + \boldsymbol{L}_{-1}}\right)$$

$$= -by - \boldsymbol{c}^\top \boldsymbol{z} + \frac{1}{2}\left(\frac{1}{W}(yW - \boldsymbol{a}^\top \boldsymbol{z})^2 + \boldsymbol{z}^\top \boldsymbol{S}\boldsymbol{z}\right),$$

where we simplified the expression by defining $\boldsymbol{S} = \operatorname{diag}(\boldsymbol{a}) - \frac{1}{W}\boldsymbol{a}\boldsymbol{a}^\top + \boldsymbol{L}_{-1}$. Plugging in $y = \frac{1}{W}(b + \boldsymbol{a}^\top \boldsymbol{z})$, we get

$$\min_y \mathcal{E}\begin{pmatrix} y \\ \boldsymbol{z} \end{pmatrix} = -\left(\boldsymbol{c} + b\frac{1}{W}\boldsymbol{a}\right)^\top \boldsymbol{z} - \frac{b^2}{2W} + \frac{1}{2}\boldsymbol{z}^\top \boldsymbol{S}\boldsymbol{z}.$$

Now, we define $\boldsymbol{L}' = \boldsymbol{S}$ and $\boldsymbol{d}' = \boldsymbol{c} + b\frac{1}{W}\boldsymbol{a}$, and $\mathcal{E}'(\boldsymbol{z}) = -\boldsymbol{d}'^\top \boldsymbol{z} + \frac{1}{2}\boldsymbol{z}^\top \boldsymbol{L}'\boldsymbol{z}$. And, we can see that

$$\arg\min_{\boldsymbol{z}} \min_y \mathcal{E}\begin{pmatrix} y \\ \boldsymbol{z} \end{pmatrix} = \arg\min_{\boldsymbol{z}} \mathcal{E}'(\boldsymbol{z}),$$

since dropping the constant term $-\frac{b^2}{2W}$ does not change what the minimizing $\boldsymbol{z}$ values are.

Claim 4.1.

1. $\boldsymbol{d}' \perp \boldsymbol{1}$
2. $\boldsymbol{L}' = \boldsymbol{S} = \operatorname{diag}(\boldsymbol{a}) - \frac{1}{W}\boldsymbol{a}\boldsymbol{a}^\top + \boldsymbol{L}_{-1}$ *is a Laplacian of a graph on the vertex set* $V \setminus \{1\}$.

We will prove Claim 4.1 in the next lecture. From the Claim, we see that the problem of finding $\arg\min_{\boldsymbol{z}} \mathcal{E}'(\boldsymbol{z})$, is exactly of the same form as finding $\arg\min_{\boldsymbol{x}} \mathcal{E}(\boldsymbol{x})$, but with one fewer variables.

We can get a minimizing $\boldsymbol{x}$ that solves $\arg\min_{\boldsymbol{x}} \mathcal{E}(\boldsymbol{x})$ by repeating the variable elimination procedure until we get down to a single variable and finding its value. We then have to work back up to getting a solution for $\boldsymbol{z}$, and then substitute that into Equation (6) to get the value for y.

Remark 4.2. In fact, this perspective on Gaussian elimination also makes sense for any positive definite matrix. In this setting, minimizing over one variable will leave us with another positive definite quadratic minimization problem.