ESI 6420 HOMEWORK 2 SOLUTIONS

MAKSIM LEVENTAL

Time spent: too much but time well spent is never wasted.

2.1 Claim: For $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{k \times n}$, if for all x it's the case that $Ax = 0$ implies $Bx = 0$ then $B = CA$ for some $C \in \mathbb{R}^{k \times m}$.

Proof. Assume without loss of generality that and $k \leq m$. Suppose $Ax = 0 \Rightarrow Bx = 0$. This is equivalent to $\text{null}(A) \subset \text{null}(B)$. By the fundamental theorem of Linear Algebra and the fact that $\text{row}(A) \subset \mathbb{R}^n$ and $\text{row}(B) \subset \mathbb{R}^n$ we have that $\text{null}(A) \cup \text{row}(A) = \text{null}(B) \cup \text{row}(B) = \mathbb{R}^n$ and hence

$$
\begin{gathered}
\text{null}(A) \subset \text{null}(B) \\
\Longleftrightarrow \\
\overline{\text{null}(A)} \subset \overline{\text{null}(B)} \\
\Longleftrightarrow \\
\text{row}(B) \subset \text{row}(A)
\end{gathered}
$$

Therefore for all $i = 1, \ldots, k$ there exist $c_i \in \mathbb{R}^m$ such that

$$A^T c_i = b_i^T$$

where $b_i^T = \left(B^T\right)_i$ the ith row of B. Collecting all of the c_i into a matrix C^T we get that

$$A^T C^T = B^T$$

or

$$CA = B$$

$\square$

2.2 Claim: For $\{v_1, \ldots, v_m\} \subset \mathbb{R}^n$ and $m \geq n+2$ there exist $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m) \neq \mathbf{0}$ such that

$$\sum_{i=1}^{m} \alpha_i v_i = 0$$

and

$$\sum_{i=1}^{m} \alpha_i = 0$$

Proof. Assume $m \geq n+2$. Let

$$w_i = \begin{pmatrix} v_i \\ 1 \end{pmatrix}$$

Note that $w_i \in \mathbb{R}^{(n+1)}$. Since $m \geq n+2$ it's the case that $m > n+1$ and hence $\{w_1, \ldots, w_m\}$ is linearly dependent (since the maximal number of linearly independent vectors in $\mathbb{R}^{(n+1)}$

Date: September 30, 2015.

is $n+1$). Therefore, with $W=(w_1,w_2,\dots,w_m)$ it must be case that there exists $\boldsymbol{\alpha}\in\mathbb{R}^m$ such that

$$\begin{aligned}
W\boldsymbol{\alpha} &= \begin{pmatrix} v_1 & \cdots & v_m \\ 1 & \cdots & 1 \end{pmatrix}\begin{pmatrix}\alpha_1\\ \vdots \\ \alpha_m\end{pmatrix} \\
&= \begin{pmatrix}\sum_{i=1}^m \alpha_i v_i \\ \sum_{i=1}^m \alpha_i\end{pmatrix} \\
&= \mathbf{0}
\end{aligned}$$

□

2.3 Claim: For $A:U\to V$ and $B:V\to W$ linear maps

$$\dim\left(R\left[A\right]\cap\ker\left[B\right]\right)=\dim\left(R\left[A\right]\right)-\dim\left(R\left[BA\right]\right)=\dim\left(\ker\left[BA\right]\right)-\dim\left(\ker\left[A\right]\right)$$

$$\begin{aligned}
\dim\left(R\left[BA\right]\right) &= \dim\left(R\left[A\right]\right)-\dim\left(R\left[A\right]\cap\ker\left[B\right]\right)\\
\dim\left(R\left[A\right]\right)+\dim\left(\ker\left[A\right]\right) &= \dim\left(\ker\left[BA\right]\right)+\dim\left(R\left[BA\right]\right)
\end{aligned}$$

where $R\left[X\right]$ is the column-space of X.

Proof. Let u,v,w be the dimensions of U,V,W respectively. Since $BA:U\to W$ the rank-nullity theorem states that

$$\operatorname{rank}\left(A\right)+\operatorname{nullity}\left(A\right)=u=\operatorname{rank}\left(BA\right)+\operatorname{nullity}\left(BA\right)$$

and since $\operatorname{rank}\left(X\right)=\dim\left(R\left[X\right]\right)$ and $\operatorname{nullity}\left(X\right)=\dim\left(\ker\left[X\right]\right)$ it's the case that

$$\dim\left(R\left[A\right]\right)+\dim\left(\ker\left[A\right]\right)=\dim\left(\ker\left[BA\right]\right)+\dim\left(R\left[BA\right]\right)$$

or

$$\dim\left(R\left[A\right]\right)-\dim\left(R\left[BA\right]\right)=\dim\left(\ker\left[BA\right]\right)-\dim\left(\ker\left[A\right]\right)$$

Now let $\{u_1,\dots,u_k\}$ be a basis for $R\left[A\right]\cap\ker\left[B\right]$. The set $\{u_1,\dots,u_k\}$ is linearly independent in $R\left[A\right]$ and hence can be extended to a basis: let $\{u_1,\dots,u_k,v_1,\dots,v_n\}$ be a basis for $R\left[A\right]$. Finally let $\{y_1,\dots,y_m\}$ be a basis for $R\left[BA\right]$ and pick an arbitrary y_i. By definition there exists x_i such that

$$BAx_i=y_i$$

But Ax_i is in $R\left[A\right]$ and therefore there exist $\{\alpha_1,\dots,\alpha_k,\beta_1,\dots,\beta_n\}$ such that

$$Ax_i=\sum_{i=1}^k\alpha_i u_i+\sum_{i=1}^n\beta_i v_i$$

But

$$y_i=B\left(Ax_i\right)=B\left(\sum_{i=1}^k\alpha_i u_i+\sum_{i=1}^n\beta_i v_i\right)=\sum_{i=1}^k\alpha_i Bu_i+B\left(\sum_{i=1}^n\beta_i v_i\right)$$

and by $\{u_i\}$ being a basis for $\ker\left[B\right]$

$$y_i=\cancel{\sum_{i=1}^k\alpha_i Bu_i}+B\left(\sum_{i=1}^n\beta_i v_i\right)=\sum_{i=1}^n\beta_i Bv_i=\sum_{i=1}^n\beta_i w_i$$

where $w_i\coloneqq Bv_i$. Therefore

$$R\left[BA\right]=\operatorname{span}\left(\{y_1,\dots,y_m\}\right)=\operatorname{span}\left(\{w_1,\dots,w_n\}\right)$$

but note that $\{w_1,\dots,w_n\}$ is linearly independent: if there existed $\{\gamma_1,\dots,\gamma_n\}$ such that $\sum_{i=1}^n\gamma_i w_i=0$ then

$$\sum_{i=1}^n\gamma_i w_i=\sum_{i=1}^n\gamma_i Bv_i=B\left(\sum_{i=1}^n\gamma_i v_i\right)=0$$

and $\{v_1, \ldots, v_n\}$ were defined to be linearly independent of $\ker[B]$ (and hence no vector in their span could be in $\ker[B]$). Hence (finally)

$$\begin{aligned}\dim(R[BA]) &= \dim(\text{span}(\{y_1, \ldots, y_m\})) \\ &= \dim(\text{span}(\{w_1, \ldots, w_n\})) \\ &= \dim(\{u_1, \ldots, u_k, v_1, \ldots, v_n\} \setminus \{u_1, \ldots, u_k\})\end{aligned}$$

and $\{u_1, \ldots, u_k\}$ and $\{v_1, \ldots, v_n\}$ are linearly independent

$$\begin{aligned}\dim(R[BA]) &= \dim(\{u_1, \ldots, u_k, v_1, \ldots, v_n\} \setminus \{u_1, \ldots, u_k\}) \\ &= \dim(\{u_1, \ldots, u_k, v_1, \ldots, v_n\}) - \dim(\{u_1, \ldots, u_k\}) \\ &= \dim(R[A]) - \dim(R[A] \cap \ker[B])\end{aligned}$$

Corollary. *For D a linear operator*

$$\dim\left(\ker\left[D^{n+1}\right]\right) = \dim(\ker[D]) + \sum_{j=1}^{n} \dim\left(R\left[D^k\right] \cap \ker[D]\right)$$

Proof. By induction on n. Let $n = 1$ and by the (manipulated above) result with $A = B = D$

$$\dim(R[D] \cap \ker[D]) + \dim(\ker[D]) = \dim(\ker[DD])$$

or

$$\dim\left(\ker\left[D^{1+1}\right]\right) = \dim(\ker[D]) + \sum_{j=1}^{1} \dim\left(R\left[D^k\right] \cap \ker[D]\right)$$

Assume $n > 1$. Then by applying the above result again with $B = D^n$ and $A = D$, and applying the induction hypothesis

$$\begin{aligned}\dim\left(\ker\left[D^{n+1}\right]\right) &= \dim\left(\ker\left[D^1 D^n\right]\right) \\ &= \dim(R[D^n] \cap \ker[D]) + \dim(\ker[D]) \\ &= \dim(R[D^n] \cap \ker[D]) + \dim(\ker[D]) \\ &+ \sum_{j=1}^{n-1} \dim\left(R\left[D^k\right] \cap \ker[D]\right) \\ &= \dim(\ker[D]) + \sum_{j=1}^{n} \dim\left(R\left[D^k\right] \cap \ker[D]\right)\end{aligned}$$

□

□

2.5 Claim: For V and n-dimensional real vector space and U, W m-dimensional subspaces of V, if $u \perp W$ for all $u \in U \setminus \{0\}$ then there exists $w \in W \setminus \{0\}$ such that $w \perp U \setminus \{0\}$.

Proof. Fix bases $\{u_1, \ldots, u_m\}, \{w_1, \ldots, w_m\}$ for for U, W. The hypothesis $u \perp W$ for all $u \in U \setminus \{0\}$ implies that $u_j \cdot w_i = 0$ for all i, j. Let

$$w = \sum_{i=1}^{m} w_i$$

Then $w \perp U \setminus \{0\}$ since for arbitrary $u \in U \setminus \{0\}$

$$u = \sum_{i=1}^{m} \alpha_i u_i$$

and

$$w \cdot u = \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i u_j \cdot w_i = 0$$

□

3.1 Claim: $\langle\langle I\rangle\rangle \geq 1$ for all matrix norms $\langle\langle\cdot\rangle\rangle$.

Proof. Since $II = I$ we have that

$$\langle\langle I\rangle\rangle = \langle\langle II\rangle\rangle \leq \langle\langle I\rangle\rangle \langle\langle I\rangle\rangle$$

cancelling $\langle\langle I\rangle\rangle$ from both sides of the inequality we have that

$$1 \leq \langle\langle I\rangle\rangle$$

□

3.2 Claim: The Frobenius norm

$$\langle\langle A\rangle\rangle_F = \sqrt{\operatorname{tr}(A^T A)} = \sqrt{\sum_i \sum_j |a_{ij}|^2}$$

is a matrix norm.

Proof. It is positively homogenous, subadditive, and positive definite because it is equivalent to the Euclidean norm on $\mathbb{R}^{n^2}$. Further it satisfies "submultiplicativity": first note that for A, B

$$(AB)_{ij} = \sum_k a_{ik} b_{kj}$$

Therefore

$$\begin{aligned}\langle\langle AB\rangle\rangle_F &= \sqrt{\sum_i \sum_j \left((AB)_{ij}\right)^2} \\ &= \sqrt{\sum_i \sum_j \left(\left|\sum_k a_{ik} b_{kj}\right|\right)^2}\end{aligned}$$

now by Cauchy-Schwarz

$$\begin{aligned}\langle\langle AB\rangle\rangle_F &= \sqrt{\sum_i \sum_j \left((AB)_{ij}\right)^2} \\ &= \sqrt{\sum_i \sum_j \left(\left|\sum_k a_{ik} b_{kj}\right|\right)^2} \\ &\leq \sqrt{\sum_i \sum_j \left(\sum_k |a_{ik}|^2 \sum_k |b_{kj}|^2\right)} \\ &= \sqrt{\left(\sum_i \sum_k |a_{ik}|^2\right)\left(\sum_j \sum_k |b_{kj}|^2\right)} \\ &= \langle\langle A\rangle\rangle_F \langle\langle B\rangle\rangle_F\end{aligned}$$

□

3.3 Let the induced matrix p-norm be

$$\langle\langle A\rangle\rangle_p = \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|Ax\|_p}{\|x\|_p}$$

(a) Claim: The induced matrix p-norm is equivalent to

$$\langle\langle A\rangle\rangle_p = \sup_{x \in \mathbb{R}^n, \|x\|_p = 1} \|Ax\|_p$$

Proof. Let y be such that

$$y = \operatorname{argmax}\left(\frac{\|Ax\|_p}{\|x\|_p}\right)$$

Then

$$\langle\langle A\rangle\rangle_p = \frac{\|Ay\|_p}{\|y\|_p}$$

Let $z \in \mathbb{R}^n, c \in \mathbb{R}$ be such that $cz = y$ and $\|z\|_p = 1$. Then

$$\frac{\|Ay\|_p}{\|y\|_p} = \frac{\|A(cz)\|_p}{\|cz\|_p}$$

Then by linearity of matrix multiplication and the p-norm

$$\begin{aligned}\frac{\|c\|_p \|Az\|_p}{\|c\|_p \|z\|_p} &= \frac{\|Az\|_p}{\|z\|_p} \\ &= \|Az\|_p\end{aligned}$$

Now suppose that there were some z' such that $\|z'\|_p = 1$ and $\|Az'\|_p > \|Az\|_p$. Then

$$\|Az'\|_p > \|Az\|_p = \frac{\|Az\|_p}{\|z\|_p} = \frac{\left\|A\frac{y}{c}\right\|_p}{\left\|\frac{y}{c}\right\|_p} = \frac{\left\|\frac{1}{c}\right\|_p \|Ay\|_p}{\left\|\frac{1}{c}\right\|_p \|y\|_p} = \frac{\|Ay\|_p}{\|y\|_p}$$

which is a contradiction (since y was defined to the argmax of $\|Ax\|_p / \|x\|_p$). Hence it suffices to define

$$\langle\langle A\rangle\rangle_p = \sup_{x\in\mathbb{R}^n, \|x\|_p=1} \|Ax\|_p$$

□

(b) Claim: the induced matrix p-norm is a legitimate matrix norm.

Proof. Using the equivalent definition from part (a) we prove

(i) Positive homogenous: let $c \in \mathbb{R}$ and then by the homogeneity of the p-norm

$$\begin{aligned}\langle\langle cA\rangle\rangle_p &= \sup_{x\in\mathbb{R}^n, \|x\|_p=1} \|(cA)x\|_p \\ &= \sup_{x\in\mathbb{R}^n, \|x\|_p=1} |c|\,\|Ax\|_p \\ &= |c|\left(\sup_{x\in\mathbb{R}^n, \|x\|_p=1} \|Ax\|_p\right) \\ &= |c|\,\langle\langle A\rangle\rangle_p\end{aligned}$$

(ii) Subadditivity: let $A, B \in M_n(\mathbb{R})$ and then by the subadditivity of the p-norm

$$\begin{aligned}\langle\langle A+B\rangle\rangle_p &= \sup_{x\in\mathbb{R}^n, \|x\|_p=1} \|(A+B)x\|_p \\ &\leq \sup_{x\in\mathbb{R}^n, \|x\|_p=1} \left(\|Ax\|_p + \|Bx\|_p\right) \\ &= \sup_{x\in\mathbb{R}^n, \|x\|_p=1} \|Ax\|_p + \sup_{x\in\mathbb{R}^n, \|x\|_p=1} \|Bx\|_p \\ &= \langle\langle A\rangle\rangle_p + \langle\langle B\rangle\rangle_p\end{aligned}$$

(iii) Positive-definiteness: let $A \neq 0$ and then firstly note that

$$\langle\langle -A\rangle\rangle_p = |-1|\,\langle\langle A\rangle\rangle_p = \langle\langle A\rangle\rangle_p$$

and by the subadditivity of the p-norm

$$\begin{aligned} 0 &= \langle\langle A + (-A) \rangle\rangle_p \\ &\leq \langle\langle A \rangle\rangle_p + \langle\langle -A \rangle\rangle_p \\ &= 2 \langle\langle A \rangle\rangle_p \end{aligned}$$

hence $\langle\langle A \rangle\rangle_p \geq 0$.

(iv) Submultiplicativity: first note that for arbitrary y, by definition of $\langle\langle A \rangle\rangle_p$

$$\|Ay\|_p \leq \langle\langle A \rangle\rangle_p \|y\|_p$$

Then let $y = Bx$ for $\|x\| = 1$ and

$$\begin{aligned} \langle\langle AB \rangle\rangle_p &= \sup_{x \in \mathbb{R}^n, \|x\|_p = 1} \|ABx\|_p \\ &= \sup_{x \in \mathbb{R}^n, \|x\|_p = 1} \|Ay\|_p \\ &\leq \sup_{x \in \mathbb{R}^n, \|x\|_p = 1} \langle\langle A \rangle\rangle_p \|y\|_p \\ &= \sup_{x \in \mathbb{R}^n, \|x\|_p = 1} \langle\langle A \rangle\rangle_p \|Bx\|_p \\ &= \langle\langle A \rangle\rangle_p \langle\langle B \rangle\rangle_p \end{aligned}$$

□

(c) Claim: The supremum of a set A is equal to the infimum of the set B of upperbounds on A.

Proof. By contradiction. Let $l = \sup A$. Then l is an upperbound on A and hence $l \in B$. Consider the set B' of lower bounds of the set B. Note that l is a lower bound on B, since if that were not the case there would exist some element $b \in B$ such that $b < l$, but l is the least upper bound. Suppose $l \neq \inf B$. Then there exists b such that b is a lower bound on B and $l < b$. But then b is not a lower bound on B since $l \in B$ and $l < b$. □

Corollary. *The induced matrix p-norm is equivalent to* $\langle\langle A \rangle\rangle_p = \inf \Lambda$ *where*

$$\Lambda = \left\{ \lambda \middle| \frac{\|Ax\|_p}{\|x\|_p} \leq \lambda, x \in \mathbb{R}^n, \lambda \in \mathbb{R} \right\}$$

Proof. By definition every $\lambda \in \Lambda$ is an upper bound on the set $C = \left\{ c \middle| c = \|Ax\|_p / \|x\|_p, x \in \mathbb{R}^n \right\}$. Since $\langle\langle A \rangle\rangle_p = \sup(C)$ by the above we have that

$$\langle\langle A \rangle\rangle_p = \sup C = \inf \Lambda$$

□

(d) Claim: $\langle\langle I \rangle\rangle = 1$.

Proof. To wit

$$\langle\langle I \rangle\rangle_p = \sup_{x \in \mathbb{R}^n, \|x\|_p = 1} \|Ix\|_p = \sup_{x \in \mathbb{R}^n, \|x\|_p = 1} \|x\|_p = 1$$

□

(e) Claim: For $\langle\langle A \rangle\rangle_2$ it's the case that

$$\langle\langle A \rangle\rangle_2 = \sqrt{\lambda_{\max}(A^T A)}$$

Proof. Firstly

$$\begin{aligned}\langle\langle A\rangle\rangle_2 &= \sup_{x\in\mathbb{R}^n,\|x\|_2=1} \|Ax\|_2 \\ &= \sup_{x\in\mathbb{R}^n,\|x\|_2=1} (Ax)^T (Ax) \\ &= \sup_{x\in\mathbb{R}^n,\|x\|_2=1} x^T \left(A^T A\right) x\end{aligned}$$

Since $A^T A$ is symmetric there exist P, Σ such that $A^T A = P\Sigma P^T$ and so

$$\begin{aligned}\langle\langle A\rangle\rangle_2 &= \sup_{x\in\mathbb{R}^n,\|x\|_2=1} x^T \left(P\Sigma P^T\right) x \\ &= \sup_{x\in\mathbb{R}^n,\|x\|_2=1} \left(P^T x\right)^T \Sigma \left(P^T x\right)\end{aligned}$$

Let $y = P^T x$ and note that $\|y\|_2 = 1$ since P^T is an isometry. Further note that $A^T A$ is positive semi-definite since $x^T \left(A^T A\right) x = \|Ax\| \geq 0$ and therefore $\Sigma_{ii} \geq 0$. Let $\left(\sqrt{\Sigma}\right)_{ij} = \sqrt{\Sigma_{ij}}$ and then

$$\begin{aligned}\langle\langle A\rangle\rangle_2 &= \sup_{y\in\mathbb{R}^n,\|y\|_2=1} \left(y^T\sqrt{\Sigma}\right)\left(\sqrt{\Sigma}y\right) \\ &= \sup_{y\in\mathbb{R}^n,\|y\|_2=1} \left(\sqrt{\Sigma}y\right)^T\left(\sqrt{\Sigma}y\right)\end{aligned}$$

since $\sqrt{\Sigma}$ is symmetric. Finally

$$\langle\langle A\rangle\rangle_2 = \sup_{y\in\mathbb{R}^n,\|y\|_2=1} \left\|\sqrt{\Sigma}y\right\|_2$$

and under the constraint $\|y\|_2 = 1$ it's the case that $\left\|\sqrt{\Sigma}y\right\|_2$ is maximized at $y = e_j$ where j corresponds to the $\left(\sqrt{\Sigma}\right)_{jj}$ is maximum, i.e. the square root of the maximal eigenvalue of $A^T A$, and furthermore $\left\|\sqrt{\Sigma}e_j\right\|_2 = \sqrt{\lambda_j}$. □

(f) Claim:$\langle\langle A\rangle\rangle_F$ is not an induced p-norm.

Proof. By counterexample: take

$$A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Then

$$A^T A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and

$$\langle\langle A\rangle\rangle_F = \sqrt{1+1} = \sqrt{2}$$

but

$$
\begin{aligned}
\langle\langle A\rangle\rangle_p &= \sup_{x\in\mathbb{R}^n,\|x\|_p=1}\left\|\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}x\right\|_p\\
&= \sup_{x\in\mathbb{R}^n,\|x\|_p=1}\left\|\begin{pmatrix}x_1\\ -x_2\end{pmatrix}\right\|_p\\
&= \sup_{x\in\mathbb{R}^n,\|x\|_p=1}\left(|x_1|^p+|-x_2|^p\right)^{1/p}\\
&= \sup_{x\in\mathbb{R}^n,\|x\|_p=1}\left(|x_1|^p+|x_2|^p\right)^{1/p}\\
&= \sup_{x\in\mathbb{R}^n,\|x\|_p=1}\|x\|_p\\
&= 1
\end{aligned}
$$

□

4.1 The Laplacian is $L = D - A$ where D is the degree matrix and A is the adjacency matrix, hence

$$
L = D - A = \begin{pmatrix}2&0&0&0&0&0\\0&3&0&0&0&0\\0&0&2&0&0&0\\0&0&0&3&0&0\\0&0&0&0&3&0\\0&0&0&0&0&1\end{pmatrix} - \begin{pmatrix}0&1&0&0&1&0\\1&0&1&0&1&0\\0&1&0&1&0&0\\0&0&1&0&1&1\\1&1&0&1&0&0\\0&0&0&1&0&0\end{pmatrix} = \begin{pmatrix}2&-1&0&0&-1&0\\-1&3&-1&0&-1&0\\0&-1&2&-1&0&0\\0&0&-1&3&-1&-1\\-1&-1&0&-1&3&0\\0&0&0&-1&0&1\end{pmatrix}
$$

where

4.2 Claim: $L^T = L$.

Proof. For an undirected graph the adjacency matrix A is symmetric since $A_{ij} = 1$ iff $(i,j) \in E$ but by undirected $(i,j) \in E$ implies $(j,i) \in E$ and so $A_{ij} = A_{ji}$. D is symmetric since it is diagonal. Therefore

$$
L^T = (D-A)^T = D^T - A^T = D - A = L
$$

□

4.3 Claim: L is PSD.

Proof. If for L we can write

$$
u^T L u = \frac{1}{2}\sum_{i,j\,\middle|\,(i,j)\in E}(u_i-u_j)^2
$$

for arbitrary u then L is PSD, since

$$
\sum_{i,j\,\middle|\,(i,j)\in E}(u_i-u_j)^2 \geq 0
$$

Let $u = (u_1,\ldots,u_n)$ and using $L = D - A$ we have

$$
\begin{aligned}
u^T(D-A)u &= u^T D u - u^T A u\\
&= \sum_{i=1}^n d(i)\,u_i^2 - u^T A u\\
&= \sum_{i=1}^n d(i)\,u_i^2 - \sum_{i,j} u_i u_j A_{ij}
\end{aligned}
$$

Since $A_{ij} = \delta_{\{(i,j)\in E\}}$

$$u^T (D - A) u = \sum_{i=1}^{n} d(i) u_i^2 - \sum_{i,j \middle| (i,j)\in E} u_i u_j$$

Now $d(i) = \sum_{j=1}^{n} \delta_{(i,j)\in E}$ (self edges are not in E, i.e. there does not exist i such that $(i,i) \in E$) and since the graph is undirected for all i, j it's the case that $\delta_{\{(i,j)\in E\}} = 1 \iff \delta_{\{(j,i)\in E\}} = 1$. Essentially there is double counting in the $d(i)$. Therefore

$$\begin{aligned}
u^T (D - A) u &= \sum_{i=1}^{n} d(i) u_i^2 - \sum_{i,j \middle| (i,j)\in E} u_i u_j \\
&== \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \delta_{(i,j)\in E} \right) u_i^2 - \sum_{i,j \middle| (i,j)\in E} u_i u_j \\
&= \sum_{i,j \middle| (i,j)\in E}^{n} 2\left(u_i^2 + u_j^2\right) - \sum_{i,j \middle| (i,j)\in E} u_i u_j \\
&= \frac{1}{2} \left(\sum_{i,j \middle| (i,j)\in E}^{n} \left(u_i^2 + u_j^2\right) - 2u_i u_j \right) \\
&= \frac{1}{2} \left(\sum_{i,j \middle| (i,j)\in E}^{n} u_i^2 - 2u_i u_j + u_j^2 \right) \\
&= \frac{1}{2} \sum_{i,j \middle| (i,j)\in E}^{n} (u_i - u_j)^2
\end{aligned}$$

□

4.4 Claim: 0 is always an eigenvalue of L

Proof. If a matrix A is diagonalizable then 0 is an eigenvalue iff A is singular, since

$$\begin{aligned}
\det(A) &= \det\left(PDP^T\right) \\
&= \det\left(P^T PD\right) \\
&= \det(D) \\
&= \prod_i \lambda_i
\end{aligned}$$

Therefore if the nullspace of A is non-trivial then A has an eigenvalue equal to zero. Note that summing across any row of a Laplacian matrix L, for an undirected graph, we get zero. Therefore let $u = (1, \ldots, 1)$ implies $Lu = 0$. Hence 0 is an eigenvalue. □

4.5 Claim: If $G = (V, E)$ is path connected then the multiplicity of the zero eigenvalue is 1.

Proof. We prove a strong fact: the dimension of the eigenspace associated with the eigenvalue 0 is the number of connected components of G. Indeed the number of connected components of the graph corresponds to the block structure of the Laplacian. If there exists more than one block, e.g.

$$L = \begin{pmatrix} L_1 & 0 \\ 0 & L_2 \end{pmatrix}$$

that means there's no path from connected component L_1 to L_2 (since, up to relabeling[1], there are no edges from any vertex in L_1 to any in L_2). Furthermore this clearly corresponds to the dimension of the nullspace, i.e. the degenerate eigenspace associated with eigenvalue $\lambda = 0$. For example for the L above, if $L_1 \in \mathbb{R}^{m\times m}$ and $L_2 \in \mathbb{R}^{n\times n}$, then both

$$u_1 = \left(\underbrace{1,\ldots,1}_{m},\underbrace{0,\ldots,0}_{n}\right)$$

and

$$u_2 = \left(\underbrace{0,\ldots,0}_{m},\underbrace{1,\ldots,1}_{n}\right)$$

are eigenvectors corresponding to the eigenvalue 0 and $u_1 \cdot u_2 = 0$. □

5.1 Claim: $A^T A$ is symmetric and PSD.

Proof. Firstly

$$\left(A^T A\right)^T = A^T \left(A^T\right)^T = A^T A$$

Then consider

$$x^T \left(A^T A\right) x = (Ax)^T (Ax) = |Ax| \geq 0$$

□

5.2 Claim: Let v_j be orthonormal eigenbasis of $A^T A$, λ_j eigenvalue associated with v_j, $\sigma_j = \sqrt{\lambda_j}$, $w_j \in \operatorname{span}\{u_1,\ldots,u_{j-1}\}^{\perp}$ with $|w_j| = 1$, and

$$u_j = \begin{cases} \frac{1}{\sigma_j} A v_j & \text{if } \sigma_j \neq 0 \\ w_j & \text{otherwise} \end{cases}$$

Then $u_j^T \cdot u_i = \delta_{ij}$.

Proof. If $\sigma_i \neq 0$ and $\sigma_j \neq 0$ consider

$$\begin{aligned} u_j^T \cdot u_i &= \left(\frac{1}{\sigma_j} A v_j\right)^T \cdot \left(\frac{1}{\sigma_i} A v_i\right) \\ &= \frac{1}{\sigma_j \sigma_i} v_j^T \left(A^T A\right) v_i \\ &= \frac{1}{\sigma_j \sigma_i} v_j^T \lambda_i v_i \\ &= \frac{\lambda_i}{\sigma_j \sigma_i} v_j^T \cdot v_i \\ &= \frac{\lambda_i}{\sigma_j \sigma_i} \delta_{ij} \end{aligned}$$

since v_i, v_j are orthonormal eigenvectors of $A^T A$. Then

$$\begin{aligned} u_j^T \cdot u_i &= \frac{\lambda_i}{\sigma_j \sigma_i} \delta_{ij} \\ &= \frac{\lambda_i}{(\sigma_i)^2} \delta_{ij} \\ &= \frac{\lambda_i}{\left(\sqrt{\lambda_i}\right)^2} \delta_{ij} \\ &= \delta_{ij} \end{aligned}$$

[1] Simply write down the Laplacian for each connected component $L_1, \ldots, L_k$ (with distinct labels for vertices in distinct connected components) and take the direct sum of the matrices $\oplus_{i=1}^k L_i = \operatorname{diag}(L_1, \ldots, L_k)$.

Without loss of generality assume $0 = \sigma_j \neq \sigma_i$ and $i < j$. Then $w_j \in \operatorname{span}\{u_1, \ldots, u_{j-1}\}^{\perp}$. Then immediately

$$u_j^T \cdot u_i = (w_j)^T \cdot u_i = 0$$

Assume $0 = \sigma_j = \sigma_i$ and $i < j$, then by $w_j \in \operatorname{span}\{u_1, \ldots, u_{j-1}\}^{\perp}$

$$u_j^T \cdot u_i = (w_j)^T \cdot w_i = 0$$

□

5.3 Reduced SVD

(a) Claim: There exist U, Σ, V such that $A = U\Sigma V^T$ and Σ is diagonal and

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_n \end{pmatrix}$$

Proof. Let σ_j for $j = 1, \ldots, m$, with $m \leq n$ be the non-zero square roots of the $A^T A$ and v_j for $j =$. The by 5.2 we have that

$$Av_j = u_j \sigma_j$$

Succintly this is

$$A\left[v_1\middle|\cdots\middle|v_m\right] = \left[u_1\middle|\cdots\middle|u_m\right] \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_n \end{pmatrix}$$

i.e. $AV = U\Sigma$. Then since V is an orthonormal matrix (since the v_j are othonormal eigenvectors of $A^T A$)

$$AVV^T = A = U\Sigma V^T$$

□

(b) Claim: $VV^T = V^T V = I$

Proof. As mentioned in part (a), since the v_j are othonormal eigenvectors of $A^T A$ we have that

$$v_j^T \cdot v_i = v_j^T \cdot v_i = \delta_{ij}$$

and so $VV^T = V^T V = I$. □

(c) Claim: $U^T U = I$

Proof. By 5.2 $u_j^T \cdot u_i = \delta_{ij}$ and hence $U^T U = I$ □

5.4 For r being the largest i such that $\sigma_i \neq 0$ (note that $\sigma_i \geq \sigma_{i+1}$)

(a) Claim: $\{u_1, \ldots, u_r\}$ is a basis for $R[A]$.

Proof. Consider for $x \in \mathbb{R}^n$

$$\begin{aligned}
Ax &= \left[u_1 \middle| \cdots \middle| u_m\right] \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_n \end{pmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix} x \\
&= \left[u_1 \middle| \cdots \middle| u_m\right] \begin{bmatrix} \sigma_1 v_1^T \\ \vdots \\ \sigma_r v_r^T \\ 0 \\ \vdots \\ 0 \end{bmatrix} x \\
&= \left[u_1 \middle| \cdots \middle| u_m\right] \begin{pmatrix} \sigma_1 v_{11} & \cdots & \sigma_1 v_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_r v_{r1} & \cdots & \sigma_1 v_{1n} \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix} x \\
&= \left[\sum_{i=1}^{r} u_i \sigma_i v_{i1} \middle| \cdots \middle| \sum_{i=1}^{r} u_i \sigma_i v_{in}\right] \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \\
&= \sum_{j=1}^{n} \sum_{i=1}^{r} u_i \sigma_i v_{ij} x_j \\
&= \sum_{i=1}^{r} u_i \left(\sum_{j=1}^{n} \sigma_i v_{ij} x_j \right)
\end{aligned}$$

Let $k_i = \left(\sum_{j=1}^{n} \sigma_i v_{ij} x_j\right)$ and then $Ax = \sum_{i=1}^{r} u_i k_i$ and the image of x is a linear combination of $\{u_1, \ldots, u_r\}$. Therefore the range (column-space) of A is spanned by $\{u_1, \ldots, u_r\}$ and since by 5.2 $u_j^T \cdot u_i = \delta_{ij}$ it's in fact a basis. □

(b) Claim: $\{u_{r+1}, \ldots, u_m\}$ is a basis for $\ker\left[A^T\right]$.

Proof. Firstly $\{u_1, \ldots, u_m\}$ spans $\mathbb{R}^m$. By the fundamental theorem of linear algebra $R[A] \perp \ker\left[A^T\right]$ and since by part (a) $\{u_1, \ldots, u_r\}$ is a basis for $R[A]$ we have that $\{u_1, \ldots, u_r\} \perp \ker\left[A^T\right]$. But by definition $\{u_{r+1}, \ldots, u_m\} \perp \{u_1, \ldots, u_r\}$ and furthermore for $j = r+2, \ldots, m$ it's the case that $u_j \perp \{u_{r+1}, \ldots, u_{j-1}\}$ and so $\{u_{r+1}, \ldots, u_m\}$ are pairwise orthogonal. Hence $\{u_{r+1}, \ldots, u_m\}$ is a suitable basis for $\ker\left[A^T\right]$. □

(c) Claim: $\{v_1, \ldots, v_r\}$ is a basis for $R\left[A^T\right]$.

Proof. Note that $A^T = \left(U\Sigma V^T\right) = V\Sigma U^T$, since Σ is diagonal. Then by similar reasoning as in (a) we have that $\{v_1, \ldots, v_r\}$ is a basis for $R\left[A^T\right]$. □

(d) Claim: $\{v_{r+1}, \ldots, v_n\}$ is a basis for $\ker[A]$.

Proof. By similar reasoning as in (b). □

5.5 Claim: For $1 \leq k \leq \operatorname{rank}(A)$

$$\min_{X \middle| \operatorname{rank}(X)=k} \langle\langle A - X \rangle\rangle_2 = \sigma_{k+1}$$

and

$$\operatorname{argmin} \langle\langle A - X \rangle\rangle_2 = X_k = \sum_{j=1}^{k} \sigma_j u_j v_j^T$$

Proof. In 3 parts: □

(a) Lemma: For X such that $\operatorname{rank}(X) = k$ there exists $z \in \ker[X] \cap \operatorname{span}\{v_1, \ldots, v_{k+1}\}$ such that $\|z\|_2 = 1$.

Proof. sdf □

(b) Lemma: $\langle\langle A - X \rangle\rangle_2 \geq \sigma_{k+1}$.

Proof. Using z from part (a)

$$\langle\langle A - X \rangle\rangle_2 = \sup_{x \in \mathbb{R}^n \middle| \|x\|=1}$$

□