CS205 Homework #3 Solutions

Problem 1

Consider an $n \times n$ matrix $\mathbf{A}$.

1. Show that if $\mathbf{A}$ has distinct eigenvalues all the corresponding eigenvectors are linearly independent.

2. Show that if $\mathbf{A}$ has a full set of eigenvectors (i.e. any eigenvalue λ with multiplicity k has k corresponding linearly independent eigenvectors), it can be written as $\mathbf{A} = \mathbf{Q\Lambda Q}^{-1}$ where $\mathbf{\Lambda}$ is a diagonal matrix of $\mathbf{A}$'s eigenvalues and $\mathbf{Q}$'s columns are $\mathbf{A}$'s eigenvectors. Hint: show that $\mathbf{AQ} = \mathbf{Q\Lambda}$ and that $\mathbf{Q}$ is invertible.

3. If $\mathbf{A}$ is symmetric show that any two eigenvectors corresponding to different eigenvalues are orthogonal.

4. If $\mathbf{A}$ is symmetric show that it has a full set of eigenvectors. Hint: If $(\lambda, \mathbf{q})$ is an eigenvalue, eigenvector ($\mathbf{q}$ normalized) pair and λ is of multiplicity $k > 1$, show that $\mathbf{A} - \lambda\mathbf{q}\mathbf{q}^T$ has an eigenvalue of λ with multiplicity $k-1$. To show that consider the Householder matrix $\mathbf{H}$ such that $\mathbf{Hq} = \mathbf{e}_1$ and note that $\mathbf{HAH}^{-1} = \mathbf{HAH}$ and $\mathbf{A}$ are similar.

5. If $\mathbf{A}$ is symmetric show that it can be written as $\mathbf{A} = \mathbf{Q\Lambda Q}^T$ for an orthogonal matrix $\mathbf{Q}$. (You may use the result of (4) even if you didn't prove it)

Solution

1. Assume that the eigenvectors $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n$ are *not* linearly independent. Therefore, among these eigenvectors there is one (say $\mathbf{q}_1$) which can be written as a linear combination of some of the others (say $\mathbf{q}_2, \ldots, \mathbf{q}_k$) where $k \leq n$. Without loss of generality, we can assume that $\mathbf{q}_2, \ldots, \mathbf{q}_k$ are linearly independent (we can keep removing vectors from a linearly dependent set until it becomes independent), therefore the decomposition of $\mathbf{q}_1$ into a linear combination $\mathbf{q}_1 = \sum_{i=2}^{k} \alpha_k \mathbf{q}_k$ is unique. However

$$\mathbf{q}_1 = \sum_{i=2}^{k} \alpha_k \mathbf{q}_k \Rightarrow \mathbf{A}\mathbf{q}_1 = \mathbf{A}\left(\sum_{i=2}^{k} \alpha_k \mathbf{q}_k\right) \Rightarrow \lambda_1 \mathbf{q}_1 = \sum_{i=2}^{k} \lambda_k \alpha_k \mathbf{q}_k \Rightarrow \mathbf{q}_1 = \sum_{i=2}^{k} \frac{\lambda_k}{\lambda_1} \alpha_k \mathbf{q}_k$$

In the third equation above, we assumed that $\lambda_1 \neq 0$. This is valid, since otherwise the same equation would indicate that there is a linear combination of $\mathbf{q}_2, \ldots, \mathbf{q}_k$ that is equal to zero (which contradicts their linear independence). From the last equation above and the uniqueness of the decomposition of $\mathbf{q}_1$ into a linear combination of $\mathbf{q}_2, \ldots, \mathbf{q}_k$ we have $\frac{\lambda_k}{\lambda_1}\alpha_k = \alpha_k$. There must be an $\alpha_{k_0} \neq 0,\ 2 \leq k_0 \leq k$, otherwise we would have $\mathbf{q}_1 = \mathbf{0}$. Thus $\lambda_1 = \lambda_{k_0}$ which is a contradiction.

2. Since $\mathbf{A}$ has a full set of eigenvectors, we can write $\mathbf{A}\mathbf{q}_i = \lambda_i \mathbf{q}_i$, $i = 1, \ldots, n$. We can collect all these equations into the matrix equation $\mathbf{AQ} = \mathbf{Q\Lambda}$, where $\mathbf{Q}$ has the n eigenvectors as columns, and $\mathbf{\Lambda} = diag(\lambda_1, \ldots, \lambda_n)$. From (1) we have that the columns of $\mathbf{Q}$ are linearly independent, therefore it is invertible. Thus $\mathbf{AQ} = \mathbf{Q\Lambda} \Rightarrow \mathbf{A} = \mathbf{Q\Lambda Q}^{-1}$.

3. If λ_1, λ_2 are two distinct eigenvalues and $\mathbf{q}_1, \mathbf{q}_2$ are their corresponding eigenvectors (there could be several eigenvectors per eigenvalue, if they have multiplicity higher than 1), then

$$\left.\begin{aligned} \mathbf{q}_1^T \mathbf{A} \mathbf{q}_2 &= \mathbf{q}_1^T (\mathbf{A}\mathbf{q}_2) = \lambda_2 (\mathbf{q}_1^T \mathbf{q}_2) \\ \mathbf{q}_1^T \mathbf{A} \mathbf{q}_2 &= \mathbf{q}_1^T \mathbf{A}^T \mathbf{q}_2 = (\mathbf{A}\mathbf{q}_1)^T \mathbf{q}_2 = \lambda_1 (\mathbf{q}_1^T \mathbf{q}_2) \end{aligned}\right\} \Rightarrow \lambda_1 (\mathbf{q}_1^T \mathbf{q}_2) = \lambda_2 (\mathbf{q}_1^T \mathbf{q}_2)$$

$$\Rightarrow (\lambda_1 - \lambda_2)(\mathbf{q}_1^T \mathbf{q}_2) = 0 \overset{\lambda_1 \neq \lambda_2}{\Longrightarrow} \mathbf{q}_1^T \mathbf{q}_2 = 0$$

4. In the review session we saw that that for a symmetric matrix $\mathbf{A}$, if λ is a multiple eigenvalue and $\mathbf{q}$ one of its associated eigenvectors, then the matrix $\mathbf{A} - \lambda \mathbf{q}\mathbf{q}^T$ reduces the multiplicity of λ by 1, moving $\mathbf{q}$ from the space of eigenvectors of λ into the nullspace (or the space of eigenvectors of zero).

   Assume that the eigenvalue λ with multiplicity k has only l linearly independent eigenvectors, where $l < k$, say $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_l$. Then, using the previous result, the matrix $\mathbf{A} - \lambda \sum_{i=1}^{l} \mathbf{q}_i \mathbf{q}_i^T$ has λ as an eigenvalue with multiplicity $l - k$ and the vectors $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_l$ belong to the set of eigenvectors associated with the zero eigenvalue. Since λ is an eigenvalue of this matrix, there must be a vector $\hat{\mathbf{q}}$ such that $\left(\mathbf{A} - \lambda \sum_{i=1}^{l} \mathbf{q}_i \mathbf{q}_i^T\right) \hat{\mathbf{q}} = \lambda \hat{\mathbf{q}}$. Furthermore, we have that $\hat{\mathbf{q}}$ must be orthogonal to all vectors $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_l$ since they are eigenvectors of a different eigenvalue (the zero eigenvalue). Therefore

$$\left(\mathbf{A} - \lambda \sum_{i=1}^{l} \mathbf{q}_i \mathbf{q}_i^T\right) \hat{\mathbf{q}} = \lambda \hat{\mathbf{q}} \Rightarrow \mathbf{A}\hat{\mathbf{q}} - \lambda \sum_{i=1}^{l} \mathbf{q}_i \mathbf{q}_i^T \hat{\mathbf{q}} = \lambda \hat{\mathbf{q}} \Rightarrow \mathbf{A}\hat{\mathbf{q}} = \lambda \hat{\mathbf{q}}$$

   This means that $\hat{\mathbf{q}}$ is an eigenvector of the original matrix $\mathbf{A}$ associated with λ and is orthogonal to all $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_l$, which contradicts our hypothesis that there exist only $l < k$ linearly independent eigenvectors.

5. From (4) we have that $\mathbf{A}$ has a full set of eigenvectors. We can assume that the eigenvectors of that set corresponding to the same eigenvalue are orthogonal to each other (we can use the Gram-Schmidt algorithm to make an orthogonal set from a linearly independent set). We also know from (3) that eigenvectors corresponding to different eigenvalues are orthogonal. Therefore, we can find a full, orthonormal set of eigenvectors for $\mathbf{A}$. Under that premise, the matrix $\mathbf{Q}$ containing all the eigenvectors as its columns is orthogonal, that is $\mathbf{Q}^T\mathbf{Q} = \mathbf{I} \Rightarrow \mathbf{Q}^{-1} = \mathbf{Q}^T$. Finally, using (2) we have

$$\mathbf{A} = \mathbf{Q\Lambda Q}^{-1} = \mathbf{Q\Lambda Q}^T$$

Problem 2

[Adapted from Heath 4.23] Let $\mathbf{A}$ be an $n \times n$ matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$. Recall that these are the roots of the characteristic polynomial of $\mathbf{A}$, defined as $f(\lambda) \equiv \det(\mathbf{A} - \lambda\mathbf{I})$. Also we define the multiplicity of an eigenvalue to be the degree of it as a root of the characteristic polynomial.

1. Show that the determinant of $\mathbf{A}$ is equal to the product of its eigenvalues, i.e. $\det(\mathbf{A}) = \prod_{j=1}^{n} \lambda_j$.

2. The *trace* of a matrix is defined to be the sum of its diagonal entries, i.e., $\text{trace}(\mathbf{A}) = \sum_{j=1}^{n} a_{jj}$. Show that the trace of $\mathbf{A}$ is equal to the sum of its eigenvalues, i.e. $\text{trace}(\mathbf{A}) = \sum_{j=1}^{n} \lambda_j$.

3. Recall a matrix $\mathbf{B}$ is similar to $\mathbf{A}$ if $\mathbf{B} = \mathbf{T}^{-1}\mathbf{A}\mathbf{T}$ for a non-singular matrix $\mathbf{T}$. Show that two similar matrices have the same trace and determinant.

4. Consider the matrix

$$\mathbf{A} = \begin{pmatrix} -\frac{a_{m-1}}{a_m} & -\frac{a_{m-2}}{a_m} & \cdots & -\frac{a_1}{a_m} & -\frac{a_0}{a_m} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

What is $\mathbf{A}$'s characteristic polynomial? Describe how you can use the power method find the largest root (in magnitude) of an arbitrary polynomial.

Solution

1. Since $(-1)^n$ is the highest order term coefficient and $\lambda_1, \ldots, \lambda_n$ are solutions to f we write $f(\lambda) = (-1)^n(\lambda - \lambda_1)\cdots(\lambda - \lambda_n)$. If we evaluate the characteristic polynomial at zero we get $f(0) = \det(\mathbf{A} - 0\mathbf{I}) = \det \mathbf{A}$. Also:

$$\det \mathbf{A} = f(0) = (-1)^n(0 - \lambda_1)\cdots(0 - \lambda_n) = (-1)^{2n} \prod_{i=1}^{n} \lambda_i = \prod_{i=1}^{n} \lambda_i$$

2. Consider the minor cofactor expansion of $\det(\mathbf{A} - \lambda I)$ which gives a sum of terms. Each term is a product of n factors comprising one entry from each row and each column. Consider the minor cofactor term containing members of the diagonal $(a_{11} - \lambda)(a_{22} - \lambda)\cdots(a_{nn} - \lambda)$. The coefficient for the λ^{n-1} term will be $(-1)^n(\sum_{i=1}^{n} -\lambda_i) = (-1)^{n+1}\sum_{i=1}^{n} \lambda_i$. Observe that this minor cofactor term is the only one that will contribute to the λ^{n-1} order terms (i.e. if you didn't use the diagonal for one row or column in the determinant you lose two λ's so you can't get up to λ^{n-1} order). Thus coefficient of the λ^{n-1} term is the trace of the matrix. We have $\text{trace}\mathbf{A} = f(\lambda) =$

$(-1)^n(\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n)$. The λ^{n-1} coeffcient from this is going to be $(-1)^n \sum_{i=1}^n -\lambda_i = (-1)^{2n} \sum_{i=1}^n \lambda_i = \sum_{i=1}^n \lambda_i$.

3. Two matrices that are similar have the same eigenvalues. Thus the formulas given above for the trace and the determinant that depend only on eigenvalues yield the same value. Alternatively recall that $\det(\mathbf{AB}) = \det \mathbf{A} \det \mathbf{B}$ so $\det(\mathbf{B^{-1}AB}) = (\det \mathbf{B}^{-1})(\det \mathbf{A})(\det \mathbf{B}) = (\det \mathbf{A})(\det \mathbf{B}^{-1}\mathbf{B}) = \det \mathbf{AI} = \det \mathbf{A}$. And $\text{trace}(\mathbf{AB}) = \sum_{i=1}^n \sum_{j=1}^n a_{ij}b_{ji} = \sum_{i=1}^n \sum_{j=1}^n b_{ij}a_{ji} = \text{trace}(\mathbf{BA})$. Thus we have

$$\text{trace}((\mathbf{B}^{-1}\mathbf{A})\mathbf{B}) = \text{trace}(\mathbf{BB}^{-1}\mathbf{A}) = \text{trace}\mathbf{A}$$

4. We can use the minor cofactor expansion along the top row. Expanding about $i = 1, j = 1$ we get $(a_{m-1}/a_m - \lambda)\lambda^{m-1} = -\lambda^m - a_{m-1}/a_m\lambda^{m-1}$ because the submatrix $A(2:m, 2:m)$ is triangular with $-\lambda$ on the diagonal. Then for $i = 1, j = k$ we get $-a_{m-k}/a_m\lambda^{m-k}$ because we get a triangular sub matrix with $m-k$ λ's on the diagonal and $k-1$ 1's. Thus we get the equation

$$\det(\mathbf{A} - \lambda\mathbf{I}) = -\lambda^m - \frac{a_{m-1}}{a_m}\lambda^{m-1} - \sum_{k=2}^{m} \frac{a_{m-k}}{a_m}\lambda^{m-k} = 0$$

The roots of the above polynomial are the same as if we multiply by $-a_m$ and so

$$a_m\lambda^m + a_{m-1}\lambda^{m-1} + \cdots + a_1\lambda + a_0 = 0$$

To solve a polynomial equation simply form the matrix given in the homework. The eigenvalues of this matrix are the roots of your polynomial. Use the power method to get the first eigenvalue which is also the first root. Then use deflation to form a new matrix and use the power method again to extract the second eigenvalue (and root). After m runs of the power method all roots are recovered.

Problem 3

Let $\mathbf{A}$ be a $m \times n$ matrix and $\mathbf{A} = \mathbf{U\Sigma V}^T$ its singular value decomposition.

1. Show that $\|\mathbf{A}\|_2 = \|\mathbf{\Sigma}\|_2$

2. Show that if $m \geq n$ then for all $\mathbf{x} \in R^n$

$$\sigma_{\min} \leq \frac{\|\mathbf{Ax}\|_2}{\|\mathbf{x}\|_2} \leq \sigma_{\max}$$

where $\sigma_{\min}, \sigma_{\max}$ are the smallest and largest singular values, respectively.

3. The Frobenius norm of a matrix $\mathbf{A}$ is defined as $\|A\|_F = \sqrt{\sum_{i,j} a_{ij}^2}$. Show that $\|A\|_F = \sqrt{\sum_{i=1}^p \sigma_i^2}$ where $p = \min\{m, n\}$.

4. If $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{b}$ is an m-vector prove that the solution $\mathbf{x}$ of minimum Euclidean norm to the least squares problem $\mathbf{A}\mathbf{x} \cong \mathbf{b}$ is given by

$$\mathbf{x} = \sum_{\sigma_i \neq 0} \frac{\mathbf{u}_i^T \mathbf{b}}{\sigma_i} \mathbf{v}_i$$

where $\mathbf{u}_i$, $\mathbf{v}_i$ are the columns of $\mathbf{U}$ and $\mathbf{V}$, respectively.

5. Show that the columns of $\mathbf{U}$ corresponding to non-zero singular values form an orthogonal basis of the column space of $\mathbf{A}$. What space do the columns of $\mathbf{U}$ corresponding to zero singular values span?

Solution

1. If $\mathbf{Q}$ is an orthogonal matrix, then

$$\|\mathbf{Q}\mathbf{x}\|_2^2 = (\mathbf{Q}\mathbf{x})^T(\mathbf{Q}\mathbf{x}) = \mathbf{x}^T\mathbf{Q}^T\mathbf{Q}\mathbf{x} = \mathbf{x}^T\mathbf{x} = \|\mathbf{x}\|_2^2 \Rightarrow \|\mathbf{Q}\mathbf{x}\|_2 = \|\mathbf{x}\|_2$$

Consequently

$$\frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \frac{\|\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \frac{\|\mathbf{\Sigma}\mathbf{V}^T\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \frac{\|\mathbf{\Sigma}\mathbf{V}^T\mathbf{x}\|_2}{\|\mathbf{V}^T\mathbf{x}\|_2} = \frac{\|\mathbf{\Sigma}\mathbf{y}\|_2}{\|\mathbf{y}\|_2} \tag{1}$$

where $\mathbf{y} = \mathbf{V}^T\mathbf{x}$. Since $\mathbf{V}^T$ is a nonsingular matrix, as $\mathbf{x}$ ranges over the entire space $R^n \setminus \{\mathbf{0}\}$, so does $\mathbf{y}$, too. Therefore

$$\max_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \max_{\mathbf{y}\neq\mathbf{0}} \frac{\|\mathbf{\Sigma}\mathbf{y}\|_2}{\|\mathbf{y}\|_2} \Rightarrow \|\mathbf{A}\|_2 = \|\mathbf{\Sigma}\|_2$$

2. Based on (1) we also have

$$\min_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \min_{\mathbf{y}\neq\mathbf{0}} \frac{\|\mathbf{\Sigma}\mathbf{y}\|_2}{\|\mathbf{y}\|_2}$$

Therefore

$$\min_{\mathbf{y}\neq\mathbf{0}} \frac{\|\mathbf{\Sigma}\mathbf{y}\|_2}{\|\mathbf{y}\|_2} \leq \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \max_{\mathbf{y}\neq\mathbf{0}} \frac{\|\mathbf{\Sigma}\mathbf{y}\|_2}{\|\mathbf{y}\|_2}$$

However

$$\frac{\|\mathbf{\Sigma}\mathbf{y}\|_2}{\|\mathbf{y}\|_2} = \sqrt{\frac{\sum_{i=1}^n \sigma_i^2 x_i^2}{\sum_{i=1}^n x_i^2}} \leq \sqrt{\frac{\sum_{i=1}^n \sigma_{\max}^2 x_i^2}{\sum_{i=1}^n x_i^2}} = \sigma_{\max}$$

$$\frac{\|\mathbf{\Sigma}\mathbf{y}\|_2}{\|\mathbf{y}\|_2} = \sqrt{\frac{\sum_{i=1}^n \sigma_i^2 x_i^2}{\sum_{i=1}^n x_i^2}} \geq \sqrt{\frac{\sum_{i=1}^n \sigma_{\min}^2 x_i^2}{\sum_{i=1}^n x_i^2}} = \sigma_{\min}$$

Furthermore, if $\sigma_{\min} = \sigma_i, \sigma_{\max} = \sigma_j$, then

$$\frac{\|\mathbf{\Sigma e}_i\|_2}{\|\mathbf{e}_i\|_2} = \sigma_i = \sigma_{\min} \qquad \frac{\|\mathbf{\Sigma e}_j\|_2}{\|\mathbf{e}_j\|_2} = \sigma_j = \sigma_{\max}$$

Therefore

$$\sigma_{\min} \leq \frac{\|\mathbf{Ax}\|_2}{\|\mathbf{x}\|_2} \leq \sigma_{\max}$$

and equality is actually attainable on either side.

3. We have that $a_{ij} = [\mathbf{A}]_{ij} = [\mathbf{U\Sigma V}^T]_{ij} = \sum_k \sigma_k u_{ik} v_{jk}$. Therefore

$$\begin{aligned} \|A\|_F &= \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(\sum_{k=1}^{p}\sigma_k u_{ik} v_{jk}\right)\left(\sum_{l=1}^{p}\sigma_l u_{il} v_{jl}\right) \\ &= \sum_{k,l=1}^{p}\sigma_k\sigma_l\left(\sum_{i=1}^{m} u_{ik}u_{il}\right)\left(\sum_{j=1}^{n} v_{jk}v_{jl}\right) \end{aligned}$$

However $\sum_{i=1}^{m} u_{ik}u_{il}$ is the dot product between the k-th and l-th columns of $\mathbf{U}$ and therefore equals $+1$ when $k = l$ and 0 otherwise. Similarly, $\sum_{j=1}^{n} v_{jk}v_{jl}$ is the dot product between the k-th and l-th columns of $\mathbf{V}$ and therefore equals $+1$ when $k = l$ and 0 otherwise. Thus

$$\|A\|_F = \sum_{k,l=1}^{p}\sigma_k\sigma_l\delta_{kl}\delta_{kl} = \sum_{k=1}^{p}\sigma_k^2$$

where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.

4. The least squares solution must satisfy the normal equations. It might not be unique though if $\mathbf{A}$ does not have full column rank, in which case we will show that the formula given just provides one of the least squares solutions (they all lead to the same minimum for the residual). We can rewrite the normal equations as

$$\mathbf{A}^T\mathbf{A}\mathbf{x_0} = \mathbf{A}^T\mathbf{b} \Leftrightarrow \mathbf{V\Sigma U}^T\mathbf{U\Sigma V}^T\mathbf{x_0} = \mathbf{V\Sigma U}^T\mathbf{b} \Leftrightarrow \mathbf{\Sigma^2 V}^T\mathbf{x_0} = \mathbf{\Sigma U}^T\mathbf{b}$$

We can show that $\mathbf{x} = \sum_{\sigma_i \neq 0} \frac{\mathbf{u}_i^T\mathbf{b}}{\sigma_i}\mathbf{v}_i$ satisfies the last equality as follows

$$\begin{aligned} \mathbf{\Sigma^2 V}^T\mathbf{x} &= \mathbf{\Sigma^2 V}^T\left(\sum_{\sigma_i\neq 0}\frac{\mathbf{u}_i^T\mathbf{b}}{\sigma_i}\mathbf{v}_i\right) = \left(\sum_{\sigma_i\neq 0}\frac{\mathbf{u}_i^T\mathbf{b}}{\sigma_i}\left(\mathbf{\Sigma^2 V}^T\mathbf{v}_i\right)\right) \\ &= \left(\sum_{\sigma_i\neq 0}\frac{\mathbf{u}_i^T\mathbf{b}}{\sigma_i}\left(\sigma_i^2\mathbf{e}_i\right)\right) = \left(\sum_{\sigma_i\neq 0}\sigma_i\left(\mathbf{u}_i^T\mathbf{b}\right)\mathbf{e}_i\right) \\ &= \left(\sum_{\textbf{all}\ \sigma_i}\sigma_i\left(\mathbf{u}_i^T\mathbf{b}\right)\mathbf{e}_i\right) = \mathbf{\Sigma U}^T\mathbf{b} \end{aligned}$$

5. Using the singular value decomposition we can write

$$\mathbf{A} = \sum_{i=1}^{p} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$$

where $\sigma_1, \sigma_2, \ldots, \sigma_p,\ p \leq \min(m, n)$ are the nonzero singular values of $\mathbf{A}$. Let $\mathbf{x} \in R^n$ be an arbitrary vector. Then

$$\mathbf{A}\mathbf{x} = \left(\sum_{i=1}^{p} \sigma_i \mathbf{u}_i \mathbf{v}_i^T \right) \mathbf{x} = \sum_{i=1}^{p} \left(\sigma_i \mathbf{v}_i^T \mathbf{x} \right) \mathbf{u}_i$$

Therefore, any vector in the column space of $\mathbf{A}$ is contained in $span\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$. Additionally, any vector of $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p$ is in the column space of A since $\mathbf{u}_k = \frac{1}{\sigma_k} \mathbf{A}\mathbf{v}_k$. Therefore, those two vector spaces coincide.

The columns of $\mathbf{U}$ corresponding to zero singular values span the normal complement of the space spanned by $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p$. That is, they span the nullspace of $\mathbf{A}^T$.