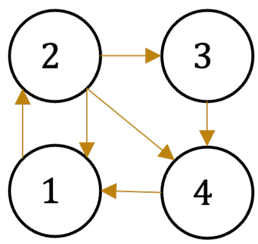
COMP/ENGN8535 Homework Assignment #2
Solutions

1. (20 points) Consider the network of 4 webpages shown below.



   (a) (5 points) What is the transition matrix P under the Random Surfer Model for this network?
   Solution:

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

   (b) (5 points) What is the Google matrix G under the Modified Surfer Model for this network with $\alpha = 0.85$?
   Solution: In this case, $S = P$ since there are no nodes with $N_i = 0$. Thus,

$$\begin{aligned} G &= \alpha S + \frac{1-\alpha}{n} \\ &= \alpha P + \frac{1-\alpha}{n} \\ &= 0.85P + 0.15/4 \\ &= \begin{bmatrix} 0.0375 & 0.8875 & 0.0375 & 0.0375 \\ 0.3208 & 0.0375 & 0.3208 & 0.3208 \\ 0.0375 & 0.0375 & 0.0375 & 0.8875 \\ 0.8875 & 0.0375 & 0.0375 & 0.0375 \end{bmatrix} \end{aligned}$$

   (c) (10 points) Using the Google matrix G and α above, rank the webpages with the (full) PageRank algorithm. Use the power method to approximate the stationary

distribution π_∞ with 5 iterations from $\pi_0 = \begin{bmatrix}0.25 & 0.25 & 0.25 & 0.25\end{bmatrix}^T$.
Solution:

$$\begin{aligned}
\pi_\infty &\approx G^T \pi_4 \\
\pi_4 &= G^T \pi_3 \\
\pi_3 &= G^T \pi_2 \\
\pi_2 &= G^T \pi_1 \\
\pi_1 &= G^T \pi_0
\end{aligned}$$

or alternatively

$$\begin{aligned}
\pi_\infty &\approx (G^T)^5 \pi_0 \\
&= \begin{bmatrix}0.3289 \\ 0.3135 \\ 0.1194 \\ 0.2383\end{bmatrix}
\end{aligned}$$

Ranking of pages (in descending order) is: Page 1 > Page 2 > Page 4 > Page 3.

2. (20 points) Consider the matrix

$$X = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}.$$

(a) (10 points) Find the SVD of X.

Solution: $XX^T \in \mathbb{R}^{2\times 2}$ is smaller than $X^TX \in \mathbb{R}^{3\times 3}$ so there are at most 2 non-zero singular values of X corresponding to the square-roots of the eigenvalues of XX^T. We determine the singular values of X by solving

$$\det(XX^T - \lambda I) = 0$$

giving $\lambda_1 = 25$ and $\lambda = 9$. Thus, the singular values are $\sigma_1 = 5$ and $\sigma_2 = 3$ and $\Sigma \in \mathbb{R}^{2\times 3}$ (which has the same dimensions as X) is given by

$$\Sigma = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix}.$$

The right-singular vectors $V \in \mathbb{R}^{3\times 3}$ are the (normalized) eigenvectors of X^TX. The eigenvalues of X^TX are found by solving

$$\det(X^TX - \lambda I) = 0$$

and are $\lambda_1 = 25, \lambda_2 = 9, \lambda_3 = 0$ (alternatively recall that X^TX and XX^T must have the same non-zero eigenvalues, with any additional eigenvalues being zero). The right-singular vectors $V \in \mathbb{R}^{3\times 3}$ are the eigenvectors of X^TX. The eigenvector corresponding to $\lambda_1 = 25$ normalised to have a norm of 1 is

$$v_1 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \\ 0 \end{bmatrix}$$

and the eigenvector corresponding to $\lambda_1 = 9$ normalised to have a norm of 1 is

$$v_2 = \begin{bmatrix} \frac{\sqrt{2}}{6} \\ -\frac{\sqrt{2}}{6} \\ \frac{2\sqrt{2}}{3} \end{bmatrix}$$

and the eigenvector corresponding to $\lambda_1 = 9$ normalised to have a norm of 1 is

$$v_3 = \begin{bmatrix} -\frac{2}{3} \\ \frac{2}{3} \\ \frac{1}{3} \end{bmatrix}$$

both found by solving the linear equations $(X^TX - \lambda I)v = 0$. Thus,

$$V = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{6} & -\frac{2}{3} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{6} & \frac{2}{3} \\ 0 & \frac{2\sqrt{2}}{3} & \frac{1}{3} \end{bmatrix}$$

The left-singular vectors $U \in \mathbb{R}^{2\times 2}$ are the eigenvectors of XX^T, or can be found via the SVD equation $X = U\Sigma V^T \implies XV = U\Sigma$ which implies that

$$X \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} = \begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix}$$

which is equal to

$$\begin{bmatrix} Xv_1 & Xv_2 & Xv_3 \end{bmatrix} = \begin{bmatrix} 5u_1 & 3u_2 & 0 \end{bmatrix}$$

so equating components in this matrix equation gives that

$$u_1 = \frac{1}{5} X v_1 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}$$

and

$$u_2 = \frac{1}{3} X v_2 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix}$$

Thus,

$$U = \begin{bmatrix} u_1 & u_2 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix}$$

SVD is thus

$$X = U\Sigma V^T$$

where

$$\Sigma = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} \quad U = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix} \quad V = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{6} & -\frac{2}{3} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{6} & \frac{2}{3} \\ 0 & \frac{2\sqrt{2}}{3} & \frac{1}{3} \end{bmatrix}$$

Note depending on sign of vectors in U and V other SVDs are possible including

$$X = U\Sigma V^T$$

where

$$\Sigma = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} \quad U = \begin{bmatrix} -\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \quad V = \begin{bmatrix} -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{6} & -\frac{2}{3} \\ -\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{6} & \frac{2}{3} \\ 0 & \frac{2\sqrt{2}}{3} & \frac{1}{3} \end{bmatrix}$$

(b) (5 points) What is the best rank-1 approximation of X in the sense of minimising the approximation error under the Frobenius norm?

Solution: Best rank-1 approximation involves using only first singular value and corresponding singular vectors to form

$$\hat{X}_1 = \sigma_1 u_1 v_1^T = \begin{bmatrix} 2.5 & 2.5 & 0 \\ 2.5 & 2.5 & 0 \end{bmatrix}.$$

(c) (5 points) What is the approximation error associated with the best rank-1 approximation of X under the Frobenius norm?

Solution: Approximation error is the value of largest singular value not used in the approximation, so it is 3.

3. (20 points) Consider the following dataset, containing 5 data points:

$$\mathbf{x}_1 = \begin{bmatrix} 4 \\ 3 \end{bmatrix}, \ \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \ \mathbf{x}_3 = \begin{bmatrix} -2 \\ 2 \end{bmatrix}, \ \mathbf{x}_4 = \begin{bmatrix} 4 \\ -1 \end{bmatrix}, \ \mathbf{x}_5 = \begin{bmatrix} 2 \\ 0 \end{bmatrix},$$

(a) (10 points) What is the first principal component of the data?
Solution: Mean vector

$$\mu = \begin{bmatrix} 1.6 \\ 1 \end{bmatrix}.$$

Mean-subtracted data matrix:

$$X = \begin{bmatrix} 2.4000 & -1.6000 & -3.6000 & 2.4000 & 0.4000 \\ 2.0000 & 0 & 1.0000 & -2.0000 & -1.0000 \end{bmatrix}.$$

Covariance matrix

$$C = \frac{1}{n} X X^T = \frac{1}{5} \begin{bmatrix} 5.4400 & -0.8000 \\ -0.8000 & 2.0000 \end{bmatrix}.$$

Eigenvalues of C are $\lambda_1 = 5.617$ and $\lambda_2 = 1.823$ with corresponding eigenvectors

$$u_1 = \begin{bmatrix} 0.9764 \\ -0.216 \end{bmatrix} \text{ and } u_2 = \begin{bmatrix} 0.2160 \\ 0.9764 \end{bmatrix}$$

(or the negatives of these since sign of the eigenvectors is arbitrary). The first principle component is the eigenvector of C corresponding to the largest eigenvalue, that is,

$$u_1 = \begin{bmatrix} 0.9764 \\ -0.216 \end{bmatrix}.$$

(b) (10 points) Use PCA to compute lower-dimensional 1-d representations $\{\mathbf{y}_j \in \mathbb{R} : 1 \leq j \leq 5\}$ of the original 2-dimensional points $\{\mathbf{x}_j \in \mathbb{R}^2 : 1 \leq j \leq 5\}$.
Solution: 1-d representations given by $\mathbf{y}_j = u_1^T x_j$ for $1 \leq j \leq 5$. Thus,

$$\begin{aligned} y_1 &= 1.9114 \\ y_2 &= -1.5622 \\ y_3 &= -3.7310 \\ y_4 &= 2.7753 \\ y_5 &= 0.6065 \end{aligned}$$

or

$$\begin{aligned} y_1 &= -1.9114 \\ y_2 &= 1.5622 \\ y_3 &= 3.7310 \\ y_4 &= -2.7753 \\ y_5 &= -0.6065 \end{aligned}$$

if the sign of u_1 is reversed above in (a).

4. (20 points) Given a real symmetric matrix $C \in \mathbb{R}^{d \times d}$, show that the maximum value of the constrained optimisation problem

$$\max_{\{u \in \mathbb{R}^d : \|u\|_2 = 1\}} u^T C u$$

is the largest eigenvalue λ_1 of C, that is, show that

$$\lambda_1 = \max_{\{u \in \mathbb{R}^d : \|u\|_2 = 1\}} u^T C u.$$

Similarly, show that the eigenvector $u_1 \in \mathbb{R}^d$ corresponding to the largest eigenvalue λ_1 of C is the maximising argument of the same optimisation problem, that is, show that

$$u_1 = \arg\max_{\{u \in \mathbb{R}^d : \|u\|_2 = 1\}} u^T C u.$$

Solution: First note that the constraint $\|u\|_2 = 1$ is the same as the constraint $\|u\|_2^2 = 1$ so we consider the optimisation problem

$$\max_{\{u \in \mathbb{R}^d : \|u\|_2^2 = 1\}} u^T C u$$

This is an equality constrained optimisation problem so we can use the Lagrange multiplier method. Construct the Lagrangian

$$L(u, \lambda) = u^T C u + \lambda(1 - \|u\|_2^2).$$

The gradient of the Lagrangian with respect to the vector u is

$$\nabla_u L(u, \lambda) = 2Cu - 2\lambda u.$$

Solutions u^* and corresponding Lagrange multipliers λ^* must satisfy

$$\begin{aligned} 0 &= \nabla_u L(u^*, \lambda^*) \\ &= 2Cu^* - 2\lambda^* u^*. \end{aligned}$$

Rearranging this equation we have that solutions the optimisation problem must satisfy

$$\lambda^* u^* = C u^*$$

and so solutions u^* and their corresponding Lagrange multipliers λ^* must be eigenvector-eigenvalue pairs. We therefore have that

$$\begin{aligned} \max_{\{u \in \mathbb{R}^d : \|u\|_2 = 1\}} u^T C u &= u^{*T} C u^* \\ &= u^{*T} \lambda^* u^* \\ &= \lambda^* u^{*T} u^* \\ &= \lambda^* \|u^*\|_2^2 \\ &= \lambda^* \end{aligned}$$

where the last line follows because of the constraint that $\|u\|_2 = 1$. The maximum value of the constrained optimisation problem is therefore the largest eigenvalue of C, proving that

$$\lambda_1 = \max_{\{u \in \mathbb{R}^d : \|u\|_2 = 1\}} u^T C u.$$

Since we also have that optimising arguments u^* must be the eigenvector of C corresponding to the eigenvalue of λ^* in the sense that

$$\lambda^* u^* = C u^*,$$

it follows that the eigenvector u_1 associated with λ_1 satisfies

$$u_1 = \arg\max_{\{u \in \mathbb{R}^d : \|u\|_2 = 1\}} u^T C u.$$

Alternative proof via eigendecomposition: Since C is a real-symmetric matrix, its eigen-decomposition has the form $C = Q \Lambda Q^T$, where Q is an orthogonal matrix whose columns are the eigenvectors of C and Λ is a diagonal matrix whose diagonal entries are the eigenvalues of C. Since $\|u^T Q\|_2 = \|Q^T u\|_2 = \|u\|_2$ for all u because Q is orthogonal, the optimisation problem

$$\max_{\{u \in \mathbb{R}^d : \|u\|_2 = 1\}} u^T C u$$

is equivalent to

$$\begin{aligned} \max_{\{u \in \mathbb{R}^d : \|u\|_2 = 1\}} u^T C u &= \max_{\{u \in \mathbb{R}^d : \|u\|_2 = 1\}} u^T Q \Lambda Q^T u \\ &= \max_{\{u \in \mathbb{R}^d : \|u\|_2 = 1\}} u^T \Lambda u \end{aligned}$$

which is to maximise $\sum_{i=1}^d \lambda_i u(i)^2$, given $\sum_{i=1}^d u(i)^2 = 1$, giving the largest eigenvalue λ_1, proving

$$\lambda_1 = \max_{\{u \in \mathbb{R}^d : \|u\|_2 = 1\}} u^T C u.$$

Now, let u^* be the maximising argument. Since $u^{*T} C u^* = \lambda_1$ we have that $\lambda_1 = u^{*T} Q \Lambda Q^T u^*$ and so we must have that $u^{*T} Q = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}$ and $Q^T u^* = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T$. Recalling that the columns of Q are the eigenvectors of C, the equality $u^{*T} Q = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}$ implies that

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} = u^{*T} \begin{bmatrix} u_1 & u_2 & \cdots & u_d \end{bmatrix}$$

and so u^* is parallel to only the first eigenvector u_1, and since both have norm 1 they must be equivalent, proving

$$u_1 = \arg\max_{\{u \in \mathbb{R}^d : \|u\|_2 = 1\}} u^T C u.$$

5. (20 points) Consider a data matrix $X \in \mathbb{R}^{d \times n}$ whose columns are d-dimensional data points that have already had their mean subtracted (the mean vector of the data points in X is $\mu = 0$). In PCA, the dimensionality of each data point (d) is reduced to a lower dimension ($k \leq d$) through an eigendecomposition of the data covariance matrix $C = \frac{1}{n} X X^T$. Show how singular value decomposition (SVD) of the data matrix X can be used to perform PCA.

   Solution: Recall that the SVD of X is $X = U \Sigma V^T$ where U and V are orthogonal matrices and Σ is a diagonal matrix with its diagonal elements arranged in descending order along the diagonal. The property of the transpose that $(ABC)^T = (C^T B^T A^T)$ implies that $X^T = (U \Sigma V)^T = V \Sigma^T U^T = V \Sigma U^T$. Substituting $X = U \Sigma V^T$ and $X^T = V \Sigma U^T$ into the definition of the covariance matrix C used in PCA gives

$$
\begin{aligned}
C &= \frac{1}{n} X X^T \\
&= \frac{1}{n} (U \Sigma V^T)(V \Sigma U^T) \\
&= \frac{1}{n} U \Sigma^2 U^T
\end{aligned}
$$

   where the last line follows because V is orthogonal so $VV^T = I$ and Σ is diagonal so $\Sigma\Sigma = \Sigma^2$.

   Now, recall that in PCA, C is a real-symmetric matrix so it has an eigendecomposition

$$C = UDU^T$$

   where U is an orthogonal matrix with the eigenvectors of C in its columns, and D is a diagonal matrix of eigenvalues of C. Taking $D = \frac{1}{n} \Sigma^2$ and U as the matrix of left-singular vectors of X above, we can see that this expression for C matches the expression for C above in terms of the SVD of X. Thus, the left-singular vectors of X (i.e., the columns of U) are the eigenvectors of C, and so the columns of U are the principle components of the data X.

   We can therefore perform PCA of X by simply computing the SVD of X to find the principle components U – the first column of U is the first principle component, the second principle component is the second column of U and so on.

   As noted in the lectures, some sources will also describe $U\Sigma$ as the principle components with U being the principle axes — we will call U the principle components.

END.