

Matrix Theory: Assignment 12 Solutions

General Instructions

The following document contains the solutions to the theory-based questions for Assignment 12. Please note that the solutions provided may not be the only possible way to solve the questions. They indicate only one of the many (possibly) valid solutions. The solutions provided are relatively crisp and do not include all the steps that you must have. Your solution should be logical and contain all supporting arguments. Feel free to contact us via email in case you find any discrepancy in the solutions provided.

Question 1

Given $\mathbf{G} = \mathbf{A}^T \mathbf{A}$ where $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $m \geq n$. Now

$$\begin{aligned}\mathbf{G}^T &= (\mathbf{A}^T \mathbf{A})^T \\ &= \mathbf{A}^T \mathbf{A} \\ &= \mathbf{G}\end{aligned}$$

i.e., \mathbf{G} is a symmetric matrix $\in \mathbb{R}^{n \times n}$. Let λ be an eigenvalue of \mathbf{G} for the non-zero eigen vector \mathbf{x} .

$$\begin{aligned}\mathbf{G}\mathbf{x} &= \lambda\mathbf{x} \\ \implies \mathbf{A}^T \mathbf{A}\mathbf{x} &= \lambda\mathbf{x} \\ \implies \mathbf{x}^T \mathbf{A}^T \mathbf{A}\mathbf{x} &= \mathbf{x}^T \lambda\mathbf{x} \\ \implies \|\mathbf{A}\mathbf{x}\|_2^2 &= \lambda \|\mathbf{x}\|_2^2 \\ \implies \lambda &= \frac{\|\mathbf{A}\mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2}\end{aligned}$$

λ must be both real and non-negative to satisfy the above equation with $\lambda = 0$ iff $\|\mathbf{x}\|_2^2 = 0$. But this contradicts the assumption that \mathbf{x} is a non-zero vector. Thus the eigenvalues of \mathbf{G} are real and positive.

Question 2

Let any matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ with rank r . Let $\mathbf{v}_1 \in \mathbb{R}^n$ be an unit vector in the row space of \mathbf{A} . Corresponding to \mathbf{v}_1 we can find a vector \mathbf{u}_1 which is an unit vector such that

$$\mathbf{A}\mathbf{v}_1 = \sigma_1 \mathbf{u}_1$$

where σ_1 will handle the scaling of the vector and \mathbf{u}_1 is the direction of the vector.

As we know that $r = \dim(C(A)) = \dim(R(A))$, so taking an orthonormal basis of $R(A)$ i.e. $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$, we can find vectors $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ such that

$$\mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i \quad \forall i = 1, \dots, r$$

Similarly finding the orthonormal basis of $N(A)$ i.e. $\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$, we can find vectors $\{\mathbf{u}_{r+1}, \dots, \mathbf{u}_n\}$ such that

$$\mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i \quad \forall i = r+1, \dots, n$$

Now, we can write the SVD of \mathbf{A} as

$$\begin{aligned} \mathbf{A} &= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \\ &= [\mathbf{u}_1 \quad \dots \quad \mathbf{u}_r \quad \mathbf{u}_{r+1} \quad \dots \quad \mathbf{u}_n] \begin{bmatrix} \sigma_1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_r & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_r^T \\ \mathbf{v}_{r+1}^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix} \\ &= [\mathbf{u}_1 \quad \dots \quad \mathbf{u}_r \quad \mathbf{u}_{r+1} \quad \dots \quad \mathbf{u}_n] \begin{bmatrix} \sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_r \\ 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_r^T \end{bmatrix} \\ &= [\mathbf{u}_1 \quad \dots \quad \mathbf{u}_r] \begin{bmatrix} \sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_r \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_r^T \end{bmatrix} \\ &= \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T \end{aligned}$$

Thus, we can say that $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ and $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ are the left and the right singular matrices respectively.

Claim: We can choose the set $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ such that they are the eigenvectors of $\mathbf{A}^T \mathbf{A}$.

Proof: As $\mathbf{A}^T \mathbf{A}$ is a symmetric matrix we can use spectral theorem and say that $\mathbf{A}^T \mathbf{A}$ is diagonalizable, and we can write it as

$$\mathbf{A}^T \mathbf{A} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T$$

where \mathbf{V} is the matrix of eigenvectors of $\mathbf{A}^T \mathbf{A}$ and $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues of $\mathbf{A}^T \mathbf{A}$ and \mathbf{V} is an orthogonal matrix.

As $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}^T \mathbf{A})$, we take the eigenvectors corresponding to the non-zero eigenvalues of $\mathbf{A}^T \mathbf{A}$ and normalize them to get the set $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$.

□

Now, we have defined our right singular values as eigenvectors of $\mathbf{A}^T \mathbf{A}$, which are orthogonal to each other. Thus, we can say that $\mathbf{v}_i^T \mathbf{v}_j = 0$ for $i \neq j$.

Finding the dot product of left singular vectors

$$\begin{aligned} \langle \mathbf{u}_i, \mathbf{u}_j \rangle &= \left\langle \left(\frac{\mathbf{A} \mathbf{v}_i}{\sigma_i} \right), \left(\frac{\mathbf{A} \mathbf{v}_j}{\sigma_j} \right) \right\rangle \\ &= \frac{1}{\sigma_i \sigma_j} \mathbf{v}_i^T \mathbf{A}^T \mathbf{A} \mathbf{v}_j \\ &= \frac{1}{\sigma_i \sigma_j} \mathbf{v}_i^T \lambda_j \mathbf{v}_j \\ &= \frac{1}{\sigma_i \sigma_j} \lambda_j \mathbf{v}_j^T \mathbf{v}_i \\ &= \begin{cases} 0 & i \neq j \\ \frac{\lambda_j}{\sigma_i \sigma_j} & i = j \end{cases} \end{aligned}$$

Thus, the left singular vectors are also orthogonal to each other.

□

Question 3

a) Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{x} \in \mathbb{R}^n$

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$\mathbf{Ax} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n a_{1i}x_i \\ \sum_{i=1}^n a_{2i}x_i \\ \vdots \\ \sum_{i=1}^n a_{mi}x_i \end{bmatrix}$$

Notice, each entry of \mathbf{Ax} is a scalar. Let each of its entries be y_i . Thus, we have:

$$\mathbf{Ax} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

$$\frac{\partial \mathbf{Ax}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial \mathbf{x}} \\ \frac{\partial y_2}{\partial \mathbf{x}} \\ \vdots \\ \frac{\partial y_m}{\partial \mathbf{x}} \end{bmatrix}$$

$$\frac{\partial y_i}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_i}{\partial x_1} & \frac{\partial y_i}{\partial x_2} & \cdots & \frac{\partial y_i}{\partial x_n} \end{bmatrix}$$

Therefore,

$$\frac{\partial \mathbf{Ax}}{\partial \mathbf{x}} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \mathbf{A}$$

b)

$$\mathbf{x}^T \mathbf{Ax} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n a_{1i} x_i \\ \sum_{i=1}^n a_{2i} x_i \\ \vdots \\ \sum_{i=1}^n a_{mi} x_i \end{bmatrix}$$

Notice the issue with dimensionality here. For this product to exist, $m = n$ is a necessity. Now, we make this adjustment in our product to get:

$$\begin{aligned} \mathbf{x}^T \mathbf{Ax} &= \langle \mathbf{x}, \mathbf{Ax} \rangle = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j \\ &= \sum_{i=1}^n a_{ii} x_i^2 + \sum_{i=1}^n \sum_{j=1, j \neq i}^n a_{ij} x_i x_j \end{aligned}$$

Which is a scalar. Therefore we have:

$$\frac{\partial \mathbf{x}^T \mathbf{Ax}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial \mathbf{x}^T \mathbf{Ax}}{\partial x_1} & \frac{\partial \mathbf{x}^T \mathbf{Ax}}{\partial x_2} & \cdots & \frac{\partial \mathbf{x}^T \mathbf{Ax}}{\partial x_n} \end{bmatrix}$$

Finding the k^{th} entry of the above vector:

$$\begin{aligned}
\frac{\partial \mathbf{x}^T \mathbf{A} \mathbf{x}}{\partial x_k} &= \frac{\partial}{\partial x_k} \left(\sum_{i=1}^n a_{ii} x_i^2 + \sum_{i=1}^n \sum_{j=1, j \neq i}^n a_{ij} x_i x_j \right) \\
&= \frac{\partial}{\partial x_k} \left(\sum_{i=1}^n a_{ii} x_i^2 \right) + \frac{\partial}{\partial x_k} \left(\sum_{i=1}^n \sum_{j=1, j \neq i}^n a_{ij} x_i x_j \right) \\
&= 2a_{kk} x_k + \sum_{j=1, j \neq k}^n a_{kj} x_j + \sum_{j=1, j \neq k}^n a_{jk} x_j \\
&= 2a_{kk} x_k + \sum_{j=1, j \neq k}^n (a_{kj} + a_{jk}) x_j \\
&= \sum_{j=1}^n (a_{kj} + a_{jk}) x_j
\end{aligned}$$

Writing the above in vector form:

$$\begin{aligned}
\frac{\partial \mathbf{x}^T \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} &= \begin{bmatrix} \frac{\partial \mathbf{x}^T \mathbf{A} \mathbf{x}}{\partial x_1} & \frac{\partial \mathbf{x}^T \mathbf{A} \mathbf{x}}{\partial x_2} & \dots & \frac{\partial \mathbf{x}^T \mathbf{A} \mathbf{x}}{\partial x_n} \end{bmatrix} \\
&= \begin{bmatrix} \sum_{j=1}^n (a_{1j} + a_{1k}) x_1 \\ \sum_{j=1}^n (a_{2j} + a_{2k}) x_2 \\ \vdots \\ \sum_{j=1}^n (a_{nj} + a_{nk}) x_n \end{bmatrix}^T \\
&= \left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \begin{bmatrix} a_{11} + a_{11} & a_{12} + a_{21} & \dots & a_{1n} + a_{n1} \\ a_{21} + a_{12} & \dots & \dots & a_{2n} + a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} + a_{1n} & \dots & \dots & a_{nn} + a_{nn} \end{bmatrix} \right)^T
\end{aligned}$$

This is nothing but $\mathbf{x}^T (\mathbf{A} + \mathbf{A}^T)$. Another form, $(\mathbf{A} + \mathbf{A}^T) \mathbf{x}$ is also often used. The difference in representation indicates either the row or column form of the output. While solving an equation, however, we must be careful as to what format we are using and stay consistent with it throughout.

c) Yes, this is possible. When $\mathbf{A} \mathbf{x} = \mathbf{b}$ does not have a solution, we try to minimize the 2-norm $\|\mathbf{A} \mathbf{x} - \mathbf{b}\|$. A slightly simpler calculation would be to minimize the square of this term. Thus, we try to minimize $\|\mathbf{A} \mathbf{x} - \mathbf{b}\|^2$.

$$\begin{aligned}
\|\mathbf{Ax} - \mathbf{b}\|^2 &= \langle \mathbf{Ax} - \mathbf{b} \rangle \langle \mathbf{Ax} - \mathbf{b} \rangle \\
&= (\mathbf{Ax} - \mathbf{b})^T (\mathbf{Ax} - \mathbf{b}) \\
&= (\mathbf{Ax})^T \mathbf{Ax} + (\mathbf{Ax})^T (-\mathbf{b}) + (-\mathbf{b})^T (\mathbf{Ax}) + (-\mathbf{b})^T (-\mathbf{b}) \\
&= \mathbf{x}^T \mathbf{A}^T \mathbf{Ax} - \mathbf{x}^T \mathbf{A}^T \mathbf{b} - \mathbf{b}^T \mathbf{Ax} + \mathbf{b}^T \mathbf{b}
\end{aligned}$$

Differentiating w.r.t \mathbf{x} , and equating to 0, we get:

$$\begin{aligned}
&(\mathbf{A}^T \mathbf{A} + (\mathbf{A}^T \mathbf{A})^T) \mathbf{x} - \mathbf{A}^T \mathbf{b} - (\mathbf{b}^T \mathbf{A})^T = 0 \\
\implies \mathbf{x} &= (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}
\end{aligned}$$

The same result can also be obtained from the following differentiation:

$$\begin{aligned}
&\mathbf{x}^T (\mathbf{A}^T \mathbf{A} + (\mathbf{A}^T \mathbf{A})^T) - (\mathbf{A}^T \mathbf{b})^T - \mathbf{b}^T \mathbf{A} = 0 \\
\implies \mathbf{x}^T &= \mathbf{b}^T \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1} \\
\implies \mathbf{x} &= (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}
\end{aligned}$$