Matrix Calculus

Omid Safarzadeh

December 24, 2021

Introduction

Matrix Multiplication [Bar09]

Definition 1.1

Let A be $m \times n$, and B be $n \times p$, and let the product AB be

$$C = AB$$

then C is a $m \times p$ matrix, with element (i, j) given by

$$c_{ij} = \Sigma_{k=1}^{n} a_{ik} b_{kj}$$

for all $i = 1, 2..., m$, $j = 1, 2, ..., p$.

Properties of determinants [Bar09]

We already know that for a 2x2 matrix, we have

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \Rightarrow \text{detA} = |A| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

- det I=1
- If you exchange two rows of a matrix, you reverse the sign of its determinant from positive to negative or from negative to positive.
- The determinant behaves like a linear function on the rows of the matrix:

$$\begin{vmatrix} ta + a' & tb + b' \\ c & d \end{vmatrix} = t \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a' & b' \\ c & d \end{vmatrix}$$

Properties of determinants

- If two rows of a matrix are equal, its determinant is zero.
- If $i \neq j$ subtracting t times row i from row j doesn't change the determinant

$$\begin{vmatrix} a & b \\ c - ta & d - tb \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

- If A has a row that is all zeros, then $\det A = 0$.
- The determinant of a triangular matrix is the product of the diagonal entries (pivots) $d_1, d_2, ..., d_n$.
- $\det AB = (\det A)(\det B)$
- $\det A^T = \det A$

Inverse Matrix[]

Definition 1.2

An $n \times n$ square matrix A is called invertible , if there exists an $n \times n$ square matrix B such that

$$AB = BA = I_n$$

where I_n denotes the $n \times n$ identity matrix and the multiplication used is ordinary matrix multiplication. If this is the case, then the matrix B is uniquely determined by A, and is called the inverse of A, denoted by A^{-1}

Matrix Transpose []

Definition 1.3

The transpose of a matrix is an operator which flips a matrix over its diagonal; that is, it switches the row and column indices of the matrix A by producing another matrix, often denoted by A^T. For example, consider matrix A as:

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

Transpose of matrix A will be defined as:

$$A^T = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$$

Matrix Transpose

Proposition 1.1

Let A be $m \times n$, and B be $n \times p$, and let the product AB be

$$C = AB$$

then

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

Proposition 1.2

Let A and B be $n \times n$, Let the product AB be given

$$C = AB$$

then

$$C^{-1} = B^{-1} A^{-1}$$

Properties of transpose[]

- $AB^T = A(B^T)$
- $(A^T)^T = A$
- $(A \pm B)^T = A^T \pm B^T$
- $(\kappa A)^T = \kappa A^T$
- $(AB)^T = B^T A^T$

Partioned Matrices [Bar09]

Definition 2.1

Let A be $m \times n$ and write:

$$A = \begin{bmatrix} B & C \\ D & E \end{bmatrix}$$

where B is $m_1 \times n_1$, E is $m_2 \times n_2$, C is $m_1 \times n_2$, D is $m_2 \times n_1$, $m_1 + m_2 = m$, and $n_1 + n_2 = n$. The above is said to be a partition of the matrix A.

Partioned Matrices

Proposition 2.1

Let A be a square, nonsingular matrix of order m. Partition A as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

so that A_{11} is a nonsingular matrix of order m_1, A_{22} is a nonsingular matrix of order m_2, and $m_1 + m_2 = m$. Then

$$A^{-1} = \begin{bmatrix} (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & -A_{11}A_{12}(A_{22} - A_{22}A_{11}^{-1}A_{12})^{-1} \\ -A_{22}^{-1}A_{21}(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \end{bmatrix}$$

Eigenvalues and Eigenvectors

To explain eigenvalues, we first explain eigenvectors. Almost all vectors change direction, when they are multiplied by A. Certain exceptional vectors x are in the same direction as Ax. Those are the "eigenvectors". Multiply an eigenvector by A, and the vector Ax is a number λ times the original x.

The basic equation is $Ax = \lambda x$. **The number** λ **is an eigenvalue of** A.

The eigenvalue λ tells whether the special vector x is stretched or shrunk or reversed or left unchanged—when it is multiplied by A.

Eigenvalues and Eigenvectors

Example 3.1

$$A = \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix} \qquad det(A-\lambda) = \begin{bmatrix} 0.8-\lambda & 0.3 \\ 0.2 & 0.7-\lambda \end{bmatrix} = \lambda^2 - \frac{3}{2}\lambda + \frac{1}{2}$$

$$= (\lambda - 1)(\lambda - \frac{1}{2}) = 0.$$

Matrix decomposition [Wik21]

- In the mathematical discipline of linear algebra, a **matrix decomposition** or **matrix factorization** is a factorization of a matrix into a product of matrices.
- Some decompositions related to solving systems of linear eqautions are listed below:
  1. LU decomposition
  2. Cholesky decomposition
  3. QR decomposition
- And some decompositions based on eigenvalues are:
  1. Singular value decomposition (SVD)

LU decomposition

- **lower–upper (LU) decomposition** factors a matrix as the product of a lower triangular matrix and an upper triangular matrix.
- Let A be a square matrix. Based on definition $A = LU$ where L is a lower triangular matrix and U is a upper triangular matrix

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} \ell_{11} & 0 & 0 \\ \ell_{21} & \ell_{22} & 0 \\ \ell_{31} & \ell_{32} & \ell_{33} \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}.$$

- For example, it is easy to verify (by expanding the matrix multiplication) that $a_{11} = \ell_{11} u_{11}$. If $a_{11} = 0$, then at least one of ℓ_{11} and u_{11} has to be zero, which implies that either L or U is singular.

LU decomposition

Example 4.1

$$\begin{bmatrix} 4 & 3 \\ 6 & 3 \end{bmatrix} = \begin{bmatrix} \ell_{11} & 0 \\ \ell_{21} & \ell_{22} \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{bmatrix}.$$

- Expanding the matrix multiplication gives

$$\ell_{11}.u_{11} + 0.0 = 4$$

$$\ell_{11}.u_{12} + 0.u_{22} = 3$$

$$\ell_{21}.u_{11} + \ell_{22}.0 = 6$$

$$\ell_{21}.u_{12} + \ell_{22}.u_{22} = 3.$$

Example 8.1 cont.

- Hence,

$$\ell_{11} = \ell_{22} = 1$$
$$\ell_{21} = 1.5$$
$$u_{11} = 14$$
$$u_{12} = 3$$
$$u_{22} = -1.5$$

So we have:

$$\begin{bmatrix} 4 & 3 \\ 6 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1.5 & 1 \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 0 & -1.5 \end{bmatrix}.$$

Cholesky decomposition

- The **Cholesky decomposition** is a decomposition of a Hermitian, positive-definite matrix into the product of a lower triangular matrix and its conjugate transpose, which is useful for efficient numerical solutions.
- The Cholesky decomposition of a Hermitian positive-definite matrix A, is a decomposition of the form

$$A = LL^T$$

where L is a lower triangular matrix with real and positive diagonal entries, and L^T denotes the conjugate transpose of L.

$$A = LL^T = \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \begin{pmatrix} L_{11} & L_{21} & L_{31} \\ 0 & L_{22} & L_{32} \\ 0 & 0 & L_{33} \end{pmatrix}$$

Cholesky decomposition

- Where

$$L_{j,j} = (\pm)\sqrt{A_{j,j} - \sum_{k=1}^{j-1} L_{j,k}^2}$$

and

$$L_{i,j} = \frac{1}{L_{j,j}}\left(A_{i,j} - \sum_{k=1}^{j-1} L_{i,k}L_{j,k}\right) \quad \text{for } i > j.$$

Cholesky decomposition

Example 4.2

$$\begin{pmatrix} 4 & 12 & -16 \\ 12 & 37 & -43 \\ -16 & -43 & 98 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 6 & 1 & 0 \\ -8 & 5 & 3 \end{pmatrix} \begin{pmatrix} 2 & 6 & -8 \\ 0 & 1 & 5 \\ 0 & 0 & 3 \end{pmatrix}$$

- **Calculating** $L_{j,j}$ **and** $L_{i,j}$: Exercise!

QR decomposition

- The QR decomposition (also called the QR factorization) of a matrix is a decomposition of the matrix into an orthogonal matrix and a triangular matrix. A QR decomposition of a real square matrix A is a decomposition of A as
$$A = QR$$
where Q is an orthogonal matrix (i.e. $Q^T Q = I$) and R is an upper triangular matrix.
- There are several methods for actually computing the QR decomposition. One of such method is the Gram-Schmidt process.

Gram-Schmidt process []

Consider the Gram-Schmidt procedure, with the vectors to be considered in the process as columns of the matrix A. That is,

$$A = \left[a_1 | a_2 | \ \dots \ | a_n\right].$$

Then,

$$u_1 = a_1, \quad e_1 = \frac{u_1}{||u_1||}$$

$$u_2 = a_2 - (a_2.e_1)e_1, \quad e_2 = \frac{u_2}{||u_2||}$$

$$u_{k+1} = a_{k+1} - (a_{k+1}.e_1)e_1 - \cdots - (a_{k+1}.e_k)e_k, \quad e_1 = \frac{u_{k+1}}{||u_{k+1}||}$$

- Note that $||.||$ is the L_2 norm.

QR decomposition

The resulting QR decomposition is

$$A = \left[a_1 | a_2 | \ \dots \ | a_n\right] = \left[e_1 | e_2 | \ \dots \ | e_n\right] \begin{bmatrix} a_1.e_1 & a_2.e_1 & \cdots & a_n.e_1 \\ 0 & a_2.e_2 & \cdots & a_n.e_2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_n.e_n \end{bmatrix} = QR.$$

- Note that once we find $e_1, \dots, e_n$, it is not hard to write the QR factorization.

Example 4.3

Consider the matrix

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix},$$

with the vectors $a_1 = (1,1,0)^T$, $a_2 = (1,0,1)^T$ and $a_3 = (0,1,1)^T$.
Performing the Gram-Schmidt procedure, we obtain:

$$u_1 = a_1 = (1,1,0)^t$$

$$e_1 = \frac{u_1}{||u_1||} = \frac{1}{\sqrt{2}}(1,1,0) = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0),$$

$$u_2 = a_2 - (a_2.e_1)e_1 = (1,0,1) - \frac{1}{\sqrt{2}}(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0) = (\frac{1}{2}, -\frac{1}{2}, 1)$$

$$e_2 = \frac{u_2}{||u_2||} = \frac{1}{\sqrt{3/2}}(\frac{1}{2}, -\frac{1}{2}, 1) = (\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}})$$

Example 8.3 cont.

$$u_3 = a_3 - (a_3.e_1)e_1 - (a_3.e_2)e_2$$

$$= (0,1,1) - \frac{1}{\sqrt{2}}(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0) - (\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}) = (-\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$$

$$e_3 = \frac{u_3}{||u_3||} = (-\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}).$$

Thus

$$Q = \left[e_1 | e_2 | \ \dots \ | e_n\right] = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}.$$

$$R = \begin{bmatrix} a_1.e_1 & a_2.e_1 & a_3.e_1 \\ 0 & a_2.e_2 & a_3.e_2 \\ 0 & 0 & a_3.e_3 \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & \frac{3}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{2}{\sqrt{3}} \end{bmatrix}.$$

SVD

- **singular value decomposition (SVD)** is a factorization of a real or complex matrix. It generalizes the eigendecomposition of a square normal matrix with an orthonormal eigenbasis to any $n \times p$:

$$A_{n\times p} = U_{n\times n} S_{n\times p} V_{p\times p}^T$$

Where

$$U^T U = I_{n\times n}$$

$$V^T V = I_{p\times p} \quad \text{(i.e. } U \text{ and } V \text{ are orthogonal)}$$

- The eigenvectors of $A^T A$ make up the columns of V.
- The eigenvectors of AA^T make up the columns of U.
- The singular values in S are square roots of eigenvalues from AA^T or $A^T A$.

Example 4.4

To understand how to solve for SVD, let's take the example of the matrix:

$$A = \begin{bmatrix} 2 & 4 \\ 1 & 3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

and

$$AA^T = \begin{bmatrix} 2 & 4 \\ 1 & 3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 4 & 0 & 0 \\ 1 & 3 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 20 & 14 & 0 & 0 \\ 14 & 10 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = W$$

Since, $W\mathbf{x} = \lambda\mathbf{x}$ then $(W - \lambda I)\mathbf{x} = 0$

Example 8.4 cont.

$$\begin{bmatrix} 20-\lambda & 14 & 0 & 0 \\ 14 & 10-\lambda & 0 & 0 \\ 0 & 0 & -\lambda & 0 \\ 0 & 0 & 0 & -\lambda \end{bmatrix} \mathbf{x} = (W - \lambda I)\mathbf{x} = 0$$

By solving the above equation, we have:

$$\lambda_1 = \lambda_2 = 0$$

$$\lambda_3 = 15 + \sqrt{221} \sim 29 \qquad \lambda_4 = 15 - \sqrt{221} \sim 0.117$$

This value can be used to determine the eigenvector that can be placed in the columns of U. Thus we obtain the following equations:

$$19.883x_1 + 14x_2 = 0$$

$$14x_1 + 9.883x_2 = 0$$

$$x_3 = x_4 = 0$$

Example 8.4 cont.

Upon simplifying the first two equations we obtain a ratio which relates the value of x_1 to x_2. The values of x_1 and x_2 are chosen such that the elements of the S are the square roots of the eigenvalues. Thus a solution that satisfies the above equation $x_1 = -0.58$ and $x_2 = 0.82$ and $x_3 = x4 = 0$ Substituting the other eigenvalue we obtain:

$$-9.883x_1 + 14x_2 = 0$$

$$14x_1 - 19.883x_2 = 0$$

$$x_3 = x_4 = 0$$

Thus a solution that satisfies this set of equations is $x_1 = 0.82$ and $x_2 = -0.58$ and $x_3 = x_4 = 0$ (this is the first column of the U matrix). Combining these we obtain:

$$U = \begin{bmatrix} 0.82 & -0.58 & 0 & 0 \\ 0.58 & 0.82 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Example 8.4 cont.

Similarly $A^T A$ makes up the columns of V so we can do a similar analysis to find the value of V.

$$A^T.A = \begin{bmatrix} 2 & 4 \\ 1 & 3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 4 & 0 & 0 \\ 1 & 3 & 0 & 0 \end{bmatrix}$$

and similarly we obtain the expression:

$$V = \begin{bmatrix} 0.4 & -0.91 \\ 0.91 & 0.4 \end{bmatrix}$$

Finally as mentioned previously the S is the square root of the eigenvalues from AA^T or $A^T A$. and can be obtained directly giving us:

$$S = \begin{bmatrix} 5.47 & 0 \\ 0 & 0.37 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Matrix Differentiation [Bar09]

Let $\mathbf{y} = \psi(\mathbf{x})$, where $\mathbf{y}$ is an m-element vector, and $\mathbf{x}$ is an n-element vector. The symbol

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \dots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \dots & \frac{\partial y_2}{\partial x_n} \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \dots & \frac{\partial y_m}{\partial m_n} \end{bmatrix}$$

will denote the $m \times n$ matrix of first-order partial derivatives of the transformation from $\mathbf{x}$ to $\mathbf{y}$. Such a matrix is called the Jacobian matrix of the transformation $\psi()$.

Matrix Differentiation

- Notice that if $\mathbf{x}$ is actually a scalar, then the resulting Jacobian matrix is a $m \times 1$ matrix; that is, a single column (a vector). On the other hand, if $\mathbf{y}$ is actually a scalar, then the resulting Jacobian matrix is a $1 \times n$ matrix; that is, a single row (the transpose of a vector).
- Let $\mathbf{y} = \mathbf{A}\mathbf{x}$, where $\mathbf{y}$ is $m \times 1$, $\mathbf{x}$ is $n \times 1$, $\mathbf{A}$ is $m \times n$, and $\mathbf{A}$ does not depend on $\mathbf{x}$, then
$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{A}$$
- Let $\mathbf{y} = \mathbf{A}\mathbf{x}$, where $\mathbf{y}$ is $m \times 1$, $\mathbf{x}$ is $n \times 1$, $\mathbf{A}$ is $m \times n$, and $\mathbf{A}$ does not depend on $\mathbf{x}$. Suppose that $\mathbf{x}$ is a function of the vector $\mathbf{z}$, while $\mathbf{A}$ is independent of $\mathbf{z}$. Then
$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{A}\frac{\partial \mathbf{x}}{\partial \mathbf{z}}$$

Matrix Differentiation

- Let the scalar α be defined by

$$\alpha = \mathbf{y}^T \mathbf{A} \mathbf{x}$$

where $\mathbf{y}$ is $m \times 1$, $\mathbf{x}$ is $n \times 1$, $\mathbf{A}$ is $m \times n$, and $\mathbf{A}$ is independent of $\mathbf{x}$ and $\mathbf{y}$, then

$$\frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{y}^T \mathbf{A}$$

and

$$\frac{\partial \alpha}{\partial \mathbf{y}} = \mathbf{x}^T \mathbf{A}^T$$

- For the special case in which the scalar α is given by the quadratic form

$$\alpha = \mathbf{x}^T \mathbf{A}^T \mathbf{x}$$

where $\mathbf{x}$ is $n \times 1$, $\mathbf{A}$ is $n \times n$, and $\mathbf{A}$ does not depend on $\mathbf{x}$, then

$$\frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{x}^T (\mathbf{A} + \mathbf{A}^T).$$

Matrix Differentiation

- For the special case where **A** is a symmetric matrix and

$$\alpha = \mathbf{x}^T \mathbf{A} \mathbf{x}$$

  where $\mathbf{x}$ is $n \times 1$, $\mathbf{A}$ is $n \times n$, and $\mathbf{A}$ does not depend of $\mathbf{x}$, then

$$\frac{\partial \alpha}{\partial \mathbf{x}} = 2\mathbf{x}^T \mathbf{A}$$

- Let the scalar α defined by

$$\alpha = \mathbf{y}^T \mathbf{x}$$

  where $\mathbf{y}$ is $n \times 1$, $\mathbf{x}$ is $n \times 1$, and both $\mathbf{x}$ and $\mathbf{y}$ are functions of the vector $\mathbf{z}$. Then

$$\frac{\partial \alpha}{\partial \mathbf{z}} = \mathbf{x}^T \frac{\partial \mathbf{y}}{\partial \mathbf{z}} + \mathbf{y}^T \frac{\partial \mathbf{x}}{\partial \mathbf{z}}$$

Matrix Differentiation

- Let the scalar α defined by

$$\alpha = \mathbf{x}^T \mathbf{x}$$

where $\mathbf{x}$ is $n \times 1$, and $\mathbf{x}$ is a function of the vector $\mathbf{z}$, then

$$\frac{\partial \alpha}{\partial \mathbf{z}} = 2\mathbf{x}^T \frac{\partial \mathbf{x}}{\partial \mathbf{z}}$$

- Let the scalar α defined by

$$\alpha = \mathbf{y}^T \mathbf{A} \mathbf{x}$$

where $\mathbf{y}$ is $m \times 1$, $\mathbf{x}$ is $n \times 1$, $\mathbf{A}$ is $m \times n$, and both $\mathbf{x}$ and $\mathbf{y}$ are functions of the vector $\mathbf{z}$, while $\mathbf{A}$ does not depend on $\mathbf{z}$. Then

$$\frac{\partial \alpha}{\partial \mathbf{z}} = \mathbf{x}^T \mathbf{A}^T \frac{\partial \mathbf{y}}{\partial \mathbf{z}} + \mathbf{y}^T \mathbf{A} \frac{\partial \mathbf{x}}{\partial \mathbf{z}}$$

Matrix Differentiation

- Let the scalar α defined by the quadratic form

$$\alpha = \mathbf{x}^T \mathbf{A} \mathbf{x}$$

where $\mathbf{x}$ is $n \times 1$, $\mathbf{A}$ is $n \times n$, and $\mathbf{x}$ is a function of the vector $\mathbf{z}$, while $\mathbf{A}$ does not depend on $\mathbf{z}$. Then

$$\frac{\partial \alpha}{\partial \mathbf{z}} = \mathbf{x}^T (\mathbf{A} + \mathbf{A}^T) \frac{\partial \mathbf{x}}{\partial \mathbf{z}}$$

- For the special case where $\mathbf{A}$ is a symmetric matrix and

$$\alpha = \mathbf{x}^T \mathbf{A} \mathbf{x}$$

where $\mathbf{x}$ is $n \times 1$, $\mathbf{A}$ is $n \times n$, and $\mathbf{x}$ is a function of the vector $\mathbf{z}$, while $\mathbf{A}$ does not depend on $\mathbf{z}$. Then

$$\frac{\partial \alpha}{\partial \mathbf{z}} = 2\mathbf{x}^T \mathbf{A} \frac{\partial \mathbf{x}}{\partial \mathbf{z}}$$

Matrix Differentiation

Lemma 5.1

Let A be a $m \times n$ matrix whose elements are functions of the scalar parameter α. Then the derivative of the matrix A with respect to the scalar parameter α is the $m \times n$ matrix of element-by-element derivatives:

$$\frac{\partial A^{-1}}{\partial \alpha} = \begin{bmatrix} \frac{\partial \alpha_{11}}{\partial \alpha} & \frac{\partial \alpha_{12}}{\partial \alpha} & \cdots & \frac{\partial \alpha_{1n}}{\partial \alpha} \\ \frac{\partial \alpha_{21}}{\partial \alpha} & \frac{\partial \alpha_{22}}{\partial \alpha} & \cdots & \frac{\partial \alpha_{2n}}{\partial \alpha} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \alpha_{m1}}{\partial \alpha} & \frac{\partial \alpha_{m2}}{\partial \alpha} & \cdots & \frac{\partial \alpha_{mn}}{\partial \alpha} \end{bmatrix}$$

- Let $\mathbf{A}$ be a nonsingular, $m \times m$ matrix whose elements are functions of the scalar parameter α. Then

$$\frac{\partial \mathbf{A}^{-1}}{\partial \alpha} = \mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \alpha} \mathbf{A}^{-1}$$

References

[] *APPLIED NUMERICAL METHODS II.* https://www.math.ucla.edu/ yanovsky/Teaching/Math151B/handouts/Gra Accessed: 2010-09-30.

[] *Properties of Transpose and Inverse of a matrix.* https://math.ryerson.ca/ danziger/professor/MTH141/Handouts/elementa Accessed: 2010-09-30.

[Bar09] Randal J. Barnes. "Matrix Differentiation (and some other stuff)". In: 2009.

[Wik21] Wikipedia contributors. *Matrix decomposition — Wikipedia, The Free Encyclopedia.* [Online; accessed 23-December-2021]. 2021. URL: https://en.wikipedia.org/w/index.php?title=Matrix_decomposition&oldid=1055701521.