Linear Algebra I & II

Jiayi Li

Department of Statistics, UCLA

Operations and Properties

Basic Notations

- By $A \in \mathbb{R}^{m \times n}$ we denote a matrix with m rows and n columns.
- By $x \in \mathbb{R}^n$, we denote a vector with n entries. An n-dimensional vector is often thought of as a $n \times 1$ matrix, known as a column vector. If we want to explicitly represent a row vector, we typically write x^T.
- The i-th element of vector x is denoted as x_i, $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$
- Use notation a_{ij} to denote the entry of matrix A in the i-th row and j-th column.

Matrix Multiplication

The product of two matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$,

$$C = AB \in \mathbb{R}^{m \times p}, \;\; C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}$$

Given two vectors $x, y \in \mathbb{R}^n$, the inner product (dot product) is a real number given by

$$x^T y \in \mathbb{R} = \sum_{i=1}^{n} x_i y_i$$

Given two vectors $x \in \mathbb{R}^m, y \in \mathbb{R}^n$, the outer product is given by

$$xy^T \in \mathbb{R}^{m \times n} = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_m y_1 & x_m y_2 & \cdots & x_m y_n \end{bmatrix}$$

Matrix Multiplication

The product of two matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$,

$$AB = \begin{bmatrix} - & a_1^T & - \\ - & a_2^T & - \\ & \vdots & \\ - & a_m^T & - \end{bmatrix} \begin{bmatrix} | & & | \\ b_1 & \cdots & b_p \\ | & & | \end{bmatrix}$$

$$= \begin{bmatrix} a_1^T b_1 & a_1^T b_2 & \cdots & a_1^T b_p \\ a_2^T b_1 & a_2^T b_2 & \cdots & a_2^T b_p \\ \vdots & \vdots & \ddots & \vdots \\ a_m^T b_1 & a_m^T b_2 & \cdots & a_m^T b_p \end{bmatrix} = \sum_{i=1}^{n} a_i b_i^T$$

Matrix Multiplication

Associative:

$$(AB)C = A(BC)$$

$$((AB)C)_{ij} = \sum_{k=1}^{p}(AB)_{ik}C_{kj} = \sum_{k=1}^{p}(\sum_{l=1}^{n} A_{il}B_{lk})C_{kj}$$

$$= \sum_{k=1}^{p}(\sum_{l=1}^{n} A_{il}B_{lk}C_{kj}) = \sum_{l=1}^{n}(\sum_{k=1}^{p} A_{il}B_{lk}C_{kj}) = \sum_{l=1}^{n} A_{il}(\sum_{k=1}^{p} B_{lk}C_{kj})$$

$$= \sum_{l=1}^{n} A_{il}(BC)_{lj} = (A(BC))_{ij}$$

Distributive:

$$A(B + C) = AB + AC$$

Not commutative:

$$AB \neq BA \text{ in general}$$

Exercise 1

1.1 Find the inner product (if it has) and the outer product of the vectors

(a) $A = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, $B = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ (b) $A = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, $B = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$,

(c) $A = \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix}$, $B = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$

1.2 Find the product of the matrices

(a) $A = \begin{bmatrix} 5 & 2 \\ 1 & 3 \end{bmatrix}$, $B = \begin{bmatrix} 3 & 1 \\ 1 & 4 \end{bmatrix}$, $AB = ?$

(b) $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$, $B = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 4 & 1 \end{bmatrix}$, $AB = ?$

(c) $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$, $B = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 4 & 1 \end{bmatrix}$, $C = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$, $ABC = ?$

Exercise 1

1.3 True or False

(a) For $A, B \in \mathbb{R}^{n \times n}$, $A^2 - B^2 = (A+B)(A-B)$

(b) For $A, B \in \mathbb{R}^{n \times n}$, $(AB)^2 = A^2B^2$

1.4 Find the product of the matrices

(a) $A = \begin{bmatrix} -4 & -y \\ -2x & -4 \end{bmatrix}$, $B = \begin{bmatrix} -4x & 0 \\ 2y & -5 \end{bmatrix}$, $AB = ?$

(b)$A = \begin{bmatrix} -4 & -y \\ -2x & -4 \\ z & 1 \end{bmatrix}$, $B = \begin{bmatrix} -4x & 0 & 3z \\ 2y & -5 & -z \end{bmatrix}$, $AB = ?$

Identity and Diagonal Matrices

The identity matrix $I \in \mathbb{R}^{n \times n}$, is a square matrix with ones on the digonal and zeros everywhere else.

$$I_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

For all $A \in \mathbb{R}^{m \times n}$,

$$AI = A = IA$$

(Usually the dimension of I is not specified)
A diagonal matrix is a matrix where all non-digonal elements are 0, typically denoted as $D = \text{diag}(d_1, d_2, ..., d_n)$.
Clearly, $I = \text{diag}(1, 1, ..., 1)$.

Matrix Transpose

The transpose of a matrix results from flipping the rows and columns. Given $A \in \mathbb{R}^{m \times n}$, its transpose $A^T \in \mathbb{R}^{n \times m}$, where

$$(A^T)_{ij} = A_{ji}$$

The transpose of a row (column) vector is a column (row) vector.

Tranposes have the following properties,

- $(A^T)^T = A$
- $(AB)^T = B^T A^T$
- $(A+B)^T = A^T + B^T$

A square matrix is symmetric if $A = A^T$.

Trace

The trace of a square matrix $A \in \mathbb{R}^{n \times n}$, denoted $\mathrm{tr}(A)$, is the sum of diagonal elements in the matrix:

$$\mathrm{tr}(A) = \sum_{i=1}^{n} A_{ii}$$

The trace has the following properties,

- For $A \in \mathbb{R}^{n \times n}$, $\mathrm{tr}(A) = \mathrm{tr}(A^T)$
- For $A, B \in \mathbb{R}^{n \times n}$, $\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B)$
- For $A \in \mathbb{R}^{n \times n}$, $t \in \mathbb{R}$, $\mathrm{tr}(tA) = t \cdot \mathrm{tr}(A)$
- For A, B such that AB is square, $\mathrm{tr}(AB) = \mathrm{tr}(BA)$
- For A, B, C such that ABC is square, $\mathrm{tr}(ABC) = \mathrm{tr}(BAC) = \mathrm{tr}(CAB)$, and so on for the product of more matrices.

Norms

Formally, a norm is any function $f : \mathbb{R}^n \to \mathbb{R}$ that satisfies 4 properties:

1. For all $x \in \mathbb{R}^n$, $f(x) \geq 0$ (non-negativity)
2. $f(x) = 0$ iff $x = 0$ (definiteness)
3. For all $x \in \mathbb{R}^n$, $t \in \mathbb{R}$, $f(tx) = |t|f(x)$ (homogeneity)
4. For all $x, y \in \mathbb{R}^n$, $f(x + y) \leq f(x) + f(y)$ (triangle inequality)

A norm of a vector x is informally a measure of "length" of the vector. e.g. the commonly used Euclidean or l_2 norm,

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$$

note that $\|x\|_2^2 = x^T x$.

Norms

By definition, l_p, $(p \geq 1)$ norm:

$$\|x\|_p = (\sum_{i=1}^{n} |x_i|^p)^{1/p}$$

l_1 norm: $\|x\|_1 = \sum_{i=1}^{n} |x_i|$
l_∞ norm: $\|x\|_\infty = \max_i |x_i|$
Norms can also be defined for matrices, such as the Frobenius norm,

$$\|x\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^2} = \sqrt{\operatorname{tr}(A^T A)}$$

Exercise 2

2.1 For $v \in \mathbb{R}^n$, prove that vv^T is a symmetric matrix.

2.2 For $v, w \in \mathbb{R}^n$, prove that $\text{tr}(vw^T) = v^T w$.

2.3 Calculate the following expressions, using the following matrices: $A = \begin{bmatrix} 2 & 3 \\ 5 & -1 \end{bmatrix}$, $B = \begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix}$, $v = \begin{bmatrix} 2 \\ -4 \end{bmatrix}$

(a) $AB^T + vv^T$
(b)$Av - 2v$
(c)$v^T B$
(d)$v^T v + v^T BA^T v$

Exercise 2

2.4 Let a and b be vectors in $\mathbb{R}^n$ such that

$$\|a\| = \|b\| = 1$$

and the inner product

$$a^T b = -\frac{1}{2}$$

determine $\|a - b\|$.

Rank

A set of non-zero vectors $\{x_1, x_2, \cdots, x_n\} \subset \mathbb{R}^m$ is said to be (linearly) independent if no vector can be represented as a linear combination of the remaining vectors. Conversely, if one vector belonging to the set can be represented as a linear combination of the remaining vectors, then the vectors are said to be (linearly) dependent. That is, if

$$x_n = \sum_{i=1}^{n-1} \alpha_i x_i$$

for some scalar values $\alpha_1, \cdots, \alpha_n \in \mathbb{R}$, then we say that the vectors $x_1, \cdots, x_n$ are linearly dependent, otherwise they are independent. E.g.

$$x_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix},\ x_2 = \begin{bmatrix} 4 \\ 1 \\ 5 \end{bmatrix},\ x_3 = \begin{bmatrix} 2 \\ -3 \\ -1 \end{bmatrix}$$

Rank

The column rank of a matrix $A \in \mathbb{R}^{m\times n}$ is the size of the largest subset of columns of A that constitute a linearly independent set. In the same way, the row rank is the largest number of rows of A that constitute a linearly independent set. For any matrix $A \in \mathbb{R}^{m\times n}$, it turns out that the column rank of A is equal to the row rank of A, and so both quantities are referred to collectively as the rank of A, denoted as rank(A). The following are some basic properties of the rank:

- For $A \in \mathbb{R}^{m\times n}$, $\text{rank}(A) \leq \min(m, n)$. If $\text{rank}(A) = \min(m, n)$, A is said to be full rank.
- For $A \in \mathbb{R}^{m\times n}$, $\text{rank}(A) = \text{rank}(A^T)$.
- For $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times p}$, $\text{rank}(AB) \leq \min(\text{rank}(A), \text{rank}(B))$.
- For $A \in \mathbb{R}^{m\times n}$, $\text{rank}(A + B) \leq \text{rank}(A) + \text{rank}(B)$.

Inverse

The inverse of a square matrix $A \in \mathbb{R}^{n\times n}$ is denoted A^{-1}, and is the unique matrix such that

$$A^{-1}A = I = AA^{-1}.$$

Non-square matrices do not have inverses by definition. For some square matrices A, A^{-1} may not exist. In particular, we say that A is invertible or non-singular if A^{-1} exists and non-invertible or singular otherwise.

In order for a square matrix A to have an inverse, then A must be full rank.

The following are properties of the inverse; all assume that $A, B \in \mathbb{R}^{n\times n}$ are non-singular:

- $(A^{-1})^{-1} = A$
- $(AB)^{-1} = B^{-1}A^{-1}$
- $(A^{-1})^T = (A^T)^{-1} = A^{-T}$

Orthogonal Matrices

Two vectors $x, y \in \mathbb{R}^n$ are orthogonal if $x^T y = 0$. A vector $x \in \mathbb{R}^n$ is normalized if $\|x\|_2 = 1$. A square matrix $U \in \mathbb{R}^{n \times n}$ is orthogonal if all its columns are orthogonal to each other and are normalized (the columns are then referred to as being orthonormal), that is

$$U^T U = I = UU^T$$

In other words, the inverse of an orthogonal matrix is its transpose.
Operating on a vector with an orthogonal matrix will not change its Euclidean norm, i.e.

$$\|Ux\| = \|x\|_2$$

for any orthogonal $U \in \mathbb{R}^{n \times n}$, $x \in \mathbb{R}^n$

Determinant

The determinant of a square matrix $A \in \mathbb{R}^{n \times n}$, is a function $\det : \mathbb{R}^{n \times n} \to \mathbb{R}$, and is denoted $|A|$ or $\det A$.
Several properties of determinant:

- For $A \in \mathbb{R}^{n \times n}$, $|A| = |A^T|$.
- For $A, B \in \mathbb{R}^{n \times n}$ $|AB| = |A||B|$.
- For $A \in \mathbb{R}^{n \times n}$, $|A| = 0$ iff A is singular.
- For $A \in \mathbb{R}^{n \times n}$ and A non-singular, $|A^{-1}| = 1/|A|$.

Define for $A \in \mathbb{R}^{n \times n}$, $A_{\setminus i, \setminus j} \in \mathbb{R}^{(n-1)\times(n-1)}$ to be the matrix that results from deleting the i-th row and j-th column from A, then

$$|A| = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} |A_{\setminus i, \setminus j}| = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} |A_{\setminus i, \setminus j}|$$

Determinant

$$\begin{vmatrix} a_{11} \end{vmatrix} = a_{11}$$

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32}$$

$$-a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}$$

Exercise 3

3.1 Determine the rank of A and A^TA when

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

3.2 Let v be a nonzero vector in $\mathbb{R}^n$. Then the dot product $v \cdot v = v^Tv \neq 0$.
Set $a := \frac{2}{v^Tv}$ and define the $n \times n$ matrix A by

$$A = I - avv^T$$

where I is the $n \times n$ identity matrix.
Prove that A is a symmetric matrix and orthogonal.

3.3 Prove that $(A^T)^{-1} = (A^{-1})^T$.

Exercise 3

3.4 Find the inverse matrix of the matrix

$$A = \begin{bmatrix} \frac{2}{7} & \frac{3}{7} & \frac{6}{7} \\ \frac{6}{7} & \frac{2}{7} & -\frac{3}{7} \\ -\frac{3}{7} & \frac{6}{7} & -\frac{2}{7} \end{bmatrix}$$

(hint: show that A is orthogonal)

3.5 Let A, B, C be the following matrices,

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 3 & 0 \\ 1 & 0 & 5 \end{bmatrix}, C = \begin{bmatrix} -1 & 0 & 1 \\ 0 & 5 & 6 \\ 3 & 0 & 1 \end{bmatrix}$$

Compute and simplify the following expression,

$$(A^T - B)^T + C(B^{-1}C)^{-1}$$

Quadratic Forms and Positive Semidefinite Matrices

Given a square matrix $A \in \mathbb{R}^{n \times n}$ and a vector $x \in \mathbb{R}^n$, the scalar value $x^T Ax$ is called a quadratic form. Written explicitly, we see that

$$x^T Ax = \sum_{i=1}^{n} x_i (Ax)_i = \sum_{i=1}^{n} x_i \Big(\sum_{j=1}^{n} A_{ij} x_j \Big) = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j$$

Note that

$$x^T Ax = (x^T Ax)^T = x^T A^T x = \frac{1}{2} x^T (A + A^T) x$$

The matrices appearing in a quadratic form are symmetric.

Quadratic Forms and Positive Semidefinite Matrices

We give the following definitions:

- A symmetric matrix $A \in \mathbb{S}^n$ is positive definite (PD) if for all non-zero vectors $x \in \mathbb{R}^n$, $x^T Ax > 0$. This is usually denoted $A \succ 0$ (or just $A > 0$), and the set of all positive definite matrices is denoted $\mathbb{S}^n_{++}$.
- A symmetric matrix $A \in \mathbb{S}^n$ is positive semidefinite (PSD) if for all vectors $x \in \mathbb{R}^n$, $x^T Ax \geq 0$. This is usually denoted $A \succeq 0$ (or just $A \geq 0$), and the set of all positive definite matrices is denoted $\mathbb{S}^n_{+}$.
- Likewise, we define negative definite (ND) and negative semidefinite (NSD).

One important property of positive definite and negative definite matrices is that they are always full rank, and hence, invertible.

Eigenvalues and Eigenvectors

Given a square matrix $A \in \mathbb{R}^{n\times n}$, we say that $\lambda \in \mathbb{C}$ is an eigenvalue of A and $x \in \mathbb{C}^n$ is the corresponding eigenvector if

$$Ax = \lambda x, \;\; x \neq 0$$

note that for any eigenvector $x \in \mathbb{C}^n$, and scalar $t \in \mathbb{C}$, $A(tx) = tAx = t\lambda x = \lambda(tx)$, so tx is also an eigenvector. For this reason we usually assume that the eigenvector is normalized to have length 1.
We can rewrite the equation above to state that (λ, x) is an eigenvalue-eigenvector pair of A if

$$(\lambda I - A)x = 0, \;\; x \neq 0$$

But $(\lambda I - A)x = 0$ has a non-zero solution to x iff $(\lambda I - A)$ has a non-empty nullspace, which is only the case if $(\lambda I - A)$ is singular, i.e. $|\lambda I - A| = 0$.

Eigenvalues and Eigenvectors

We expand this expression into a polynomial in λ, where λ will have maximum degree n. We then find the n (possibly complex) roots of this polynomial to find the n eigenvalues $\lambda_1, \cdots, \lambda_n$. To find the eigenvector corresponding to the eigenvalue λ_i, we simply solve the linear equation $(\lambda_i I - A)x = 0$ and get $x_1, \cdots, x_n$.

- The trace of A is equal to the sum of its eigenvalues, $\mathrm{tr}A = \sum_{i=1}^n \lambda_i$.
- The determinant of A is equal to the product of its eigenvalues, $|A| = \prod_{i=1}^n \lambda_i$.
- The rank of A is equal to the number of non-zero eigenvalues of A.
- If A is non-singular then $1/\lambda_i$ is an eigenvalue of A^{-1} with associated eigenvector x_i, i.e., $A^{-1}x_i = (1/\lambda_i)x_i$.

We can write all the eigenvector equations simultaneously as

$$AX = X\Lambda,\ X = (x_1, \cdots, x_n), \Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$$

Eigenvalues and Eigenvectors

Two remarkable properties come about when we look at the eigenvalues and eigenvectors of a symmetric matrix $A \in \mathbb{S}^n$. First, all the eigenvalues of A are real. Secondly, the eigenvectors of A are orthonormal.
We can therefore represent A as $A = U\Lambda U^T$, using this, we can show that the definiteness of a matrix depends entirely on the sign of its eigenvalues. Suppose $A \in \mathbb{S}^n = U\Lambda U^T$. Then

$$x^T A x = x^T U\Lambda U^T x = y^T \Lambda y = \sum_{i=1} \lambda_i y_i^2$$

where $y = U^T x$ (since U is full rank, any vector $y \in \mathbb{R}^n$ can be represented in this form).

Matrix Calculus

Gradient

Suppose that $f : \mathbb{R}_{m\times n} \to \mathbb{R}$ is a function that takes as input a matrix A of size $m \times n$ and returns a real value. Then the gradient of f (with respect to $A \in \mathbb{R}_{m\times n}$) is the matrix of partial derivatives, defined as:

$$\nabla_A f(A) \in \mathbb{R}_{m\times n} = \begin{bmatrix} \frac{\partial f(A)}{\partial A_{11}} & \frac{\partial f(A)}{\partial A_{12}} & \cdots & \frac{\partial f(A)}{\partial A_{1n}} \\ \frac{\partial f(A)}{\partial A_{21}} & \frac{\partial f(A)}{\partial A_{22}} & \cdots & \frac{\partial f(A)}{\partial A_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(A)}{\partial A_{m1}} & \frac{\partial f(A)}{\partial A_{m2}} & \cdots & \frac{\partial f(A)}{\partial A_{mn}} \end{bmatrix}$$

Note that the size of $\nabla_A f(A)$ is always the same as the size of A. We have

$$\nabla_x(f(x) + g(x)) = \nabla_x f(x) + \nabla_x g(x)$$

$$\nabla_x(tf(x)) = t\nabla_x f(x) \ (t \in \mathbb{R})$$

Gradient

Let $f : \mathbb{R}_m \to \mathbb{R}$ be the function defined by $f(z) = z^T z$, such that $\nabla_z f(z) = 2z$, now consider the interpretation of $\nabla f(Ax)$.
1. In the first interpretation, recall that $\nabla_z f(z) = 2z$. Here, we

interpret $\nabla f(Ax)$ as evaluating the gradient at the point Ax, hence,

$$\nabla f(Ax) = 2(Ax) = 2Ax \in \mathbb{R}^m$$

2. In the second interpretation, we consider the quantity $f(Ax)$ as a function of the input variables x. More formally, let $g(x) = f(Ax)$. Then in this interpretation,

$$\nabla f(Ax) = \nabla_x g(x) \in \mathbb{R}^n$$

Keep the notation clear is extremely important!

Hessian

Suppose that $f : \mathbb{R}_m \to \mathbb{R}$ is a function that takes a vector in $\mathbb{R} \in n$ and returns a real number. Then the Hessian matrix with respect to x, written $\nabla_x^2 f(x)$ or simply as H is the $n \times n$ matrix of partial derivatives,

$$\nabla_x^2 f(x) \in \mathbb{R}_{n\times n} = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n x_1} & \frac{\partial^2 f(x)}{\partial x_n x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

Note that the Hessian is always symmetric, since

$$\frac{\partial^2 f(x)}{\partial x_i x_j} = \frac{\partial^2 f(x)}{\partial x_j x_i}$$

Gradient and Hessian of Quadratic and Linear Functions

For $x \in \mathbb{R}^n$, let $f(x) = b^T x$ for some known vector $b \in \mathbb{R}^n$. Then

$$f(x) = \sum_{i=1}^{n} b_i x_i$$

so

$$\frac{\partial f(x)}{\partial x_k} = \frac{\partial}{\partial x_k} \sum_{i=1}^{n} b_i x_i = b_k$$

From this we can easily see that $\nabla_x b^T x = b$.

Gradient and Hessian of Quadratic and Linear Functions

For $x \in \mathbb{R}^n$, now consider the quadratic function $f(x) = x^T Ax$ for $A \in \mathbb{S}^n$. Remember that

$$f(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j$$

To take the partial derivative, we consider the terms including x_k and x_k^2 factors separately,

$$\frac{\partial f(x)}{\partial x_k} = \frac{\partial}{\partial x_k} \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j$$

$$= \frac{\partial}{\partial x_k} [\sum_{i \neq k} \sum_{j \neq k} A_{ij} x_i x_j + \sum_{i \neq k} A_{ik} x_i x_k + \sum_{j \neq k} A_{kj} x_k x_j + A_{kk} x_k^2]$$

$$= \sum_{i \neq k} A_{ik} x_i + \sum_{j \neq k} A_{kj} x_j + 2A_{kk} x_k = 2 \sum_{i=1}^{n} A_{ki} x_i$$

Gradient and Hessian of Quadratic and Linear Functions

Note that the kth entry of $\nabla_x f(x)$ is just the inner product of the kth row of A and x. Thus $\nabla_x x^T Ax = 2Ax$.
And the Hessian of the quadratic function $f(x) = x^T Ax$ is

$$\frac{\partial^2 f(x)}{\partial x_k \partial x_l} = \frac{\partial}{\partial x_k}[\frac{\partial f(x)}{\partial x_l}] = \frac{\partial}{\partial x_k}[2\sum_{i=1}^{n} A_{ki}x_i] = 2A_{lk} = 2A_{kl}$$

Thus $\nabla_x^2 x^T Ax = 2A$ (A is symmetric)