Quadratic Forms

CE40282-1: Linear Algebra
Hamid R. Rabiee and Maryam Ramezani
Sharif University of Technology

Symmetric Matrix

A **symmetric** matrix is a matrix A such that $A^T = A$. Such a matrix is necessarily square. Its main diagonal entries are arbitrary, but its other entries occur in pairs—on opposite sides of the main diagonal.

$$\text{Symmetric:}\quad \begin{bmatrix} 1 & 0 \\ 0 & -3 \end{bmatrix}, \quad \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & 8 \\ 0 & 8 & -7 \end{bmatrix}, \quad \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}$$

$$\text{Nonsymmetric:}\quad \begin{bmatrix} 1 & -3 \\ 3 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & -4 & 0 \\ -6 & 1 & -4 \\ 0 & -6 & 1 \end{bmatrix}, \quad \begin{bmatrix} 5 & 4 & 3 & 2 \\ 4 & 3 & 2 & 1 \\ 3 & 2 & 1 & 0 \end{bmatrix}$$

Quadratic Form

A quadratic form is any homogeneous polynomial of degree two in any number of variables. In this situation, **homogeneous** means that all the terms are of degree two. For example, the expression $7x_1x_2 + 3x_2x_4$ is homogeneous, but the expression $x_1 - 3x_1x_2$ is not. The square of the distance between two points in an inner-product space is a quadratic form. Quadratic forms were introduced by Hermite, and 70 years later they turned out to be essential in the theory of quantum mechanics! The formal definition follows.

Quadratic Form

- Given a square matrix $A \in \mathbb{R}^{n\times n}$ and a vector $x \in \mathbb{R}^n$, the scalar value x^TAx is called a ***quadratic form***.

$$x^TAx = \sum_{i=1}^{n} x_i(Ax)_i = \sum_{i=1}^{n} x_i\left(\sum_{j=1}^{n} A_{ij}x_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}x_ix_j$$

$1\times n \quad n\times n \quad n\times 1$

- A **quadratic form** on $\mathbb{R}^n$ is a function Q defined on $\mathbb{R}^n$ whose value at a vector $\mathbf{x}$ in $\mathbb{R}^n$ can be computed by an expression of the form $Q(\mathbf{x}) = \mathbf{x}^TA\mathbf{x}$, where A is an $n \times n$ symmetric matrix. The matrix A is called the **matrix of the quadratic form**.

- Simplest example of a nonzero quadratic form is ………

$\|x\|^2 = x^T I x$

$\boxed{x}$

Quadratic Form

$x^T A x$

- Example
  - Without cross-product term $A = \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix}$

$x: 2\times 1$

2×2

$\underbrace{[x_1 \ x_2]\begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\begin{bmatrix} 4x_1 \\ 3x_2 \end{bmatrix}} = 4x_1^2 + 3x_2^2$

$4x_1^2 + 3x_2^2$

  - With cross-product term $A = \begin{bmatrix} 3 & -2 \\ -2 & 7 \end{bmatrix}$

$[x_1 \ x_2]\underbrace{\begin{bmatrix} 3 & -2 \\ -2 & 7 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}\begin{bmatrix} 3x_1 - 2x_2 \\ -2x_1 + 7x_2 \end{bmatrix}$

$= 3x_1^2 - 4x_1x_2 + 7x_2^2$

$3x_1^2 + 7x_2^2 - 4x_1x_2$

Quadratic Form

- Example

For $\mathbf{x}$ in $\mathbb{R}^3$, let $Q(\mathbf{x}) = 5x_1^2 + 3x_2^2 + 2x_3^2 - x_1x_2 + 8x_2x_3$.
Write this quadratic form as $\mathbf{x}^T A \mathbf{x}$.

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad A = \begin{bmatrix} 5 & -1/2 & 0 \\ -1/2 & 3 & 4 \\ 0 & 4 & 2 \end{bmatrix}$$

- Quadratic forms are easier to use when they have no cross-product terms—that is, when the matrix of the quadratic form is a diagonal matrix

Quadratic Form

$\boxed{x^T A x}$ عدد

If **A** *and* **B** *are* $n \times n$ *real matrices connected by the relation*

$$\boxed{\cancel{\mathbf{B}} = \tfrac{1}{2}(\mathbf{A} + \mathbf{A}^T)}$$

$x^T B x = x^T A x$

then the corresponding quadratic forms of **A** *and* **B** *are identical, and* **B** *is symmetric.*

ترانهاده عدد = خودش ①

عدد $x^T A x$ ②

$(x^T A x)^T = x^T A^T x = \text{Scalar}$

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

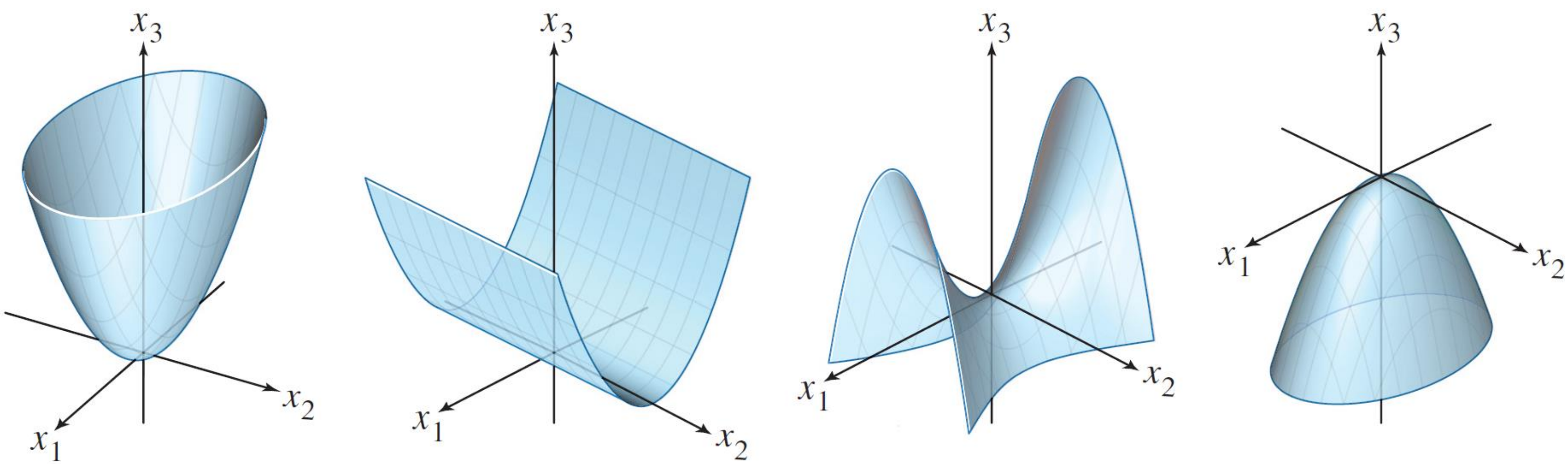
$$= x^T A x .$$

Classifying Quadratic Forms

When A is an $n \times n$ matrix, the quadratic form $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is a real-valued function with domain $\mathbb{R}^n$.

$$\text{point } (x_1, x_2, z) \text{ where } z = Q(\mathbf{x})$$

$$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$



(a) $z = 3x_1^2 + 7x_2^2$ (b) $z = 3x_1^2$ (c) $z = 3x_1^2 - 7x_2^2$ (d) $z = -3x_1^2 - 7x_2^2$

$Q(x) > 0$ $Q(x) < 0$

Classifying Quadratic Forms

$Q(x) > 0$ $x \neq 0$

- A symmetric matrix $A \in \mathbb{S}^n$ is ***positive definite*** (PD) if for all non-zero vectors $x \in \mathbb{R}^n$, $x^T A x > 0$. This is usually denoted $A \succ 0$, and often times the set of all positive definite matrices is denoted $\mathbb{S}^n_{++}$.

$Q(x) \geq 0$ x

- A symmetric matrix $A \in \mathbb{S}^n$ is ***positive semidefinite*** (PSD) if for all vectors $x^T A x \geq 0$. This is written $A \succeq 0$, and the set of all positive semidefinite matrices is often denoted $\mathbb{S}^n_+$.

- Likewise, a symmetric matrix $A \in \mathbb{S}^n$ is ***negative definite*** (ND), denoted $A \prec 0$ if for all non-zero $x \in \mathbb{R}^n$, $x^T A x < 0$.

- Similarly, a symmetric matrix $A \in \mathbb{S}^n$ is ***negative semidefinite*** (NSD), denoted $A \preceq 0$ if for all $x \in \mathbb{R}^n$, $x^T A x \leq 0$.

- Finally, a symmetric matrix $A \in \mathbb{S}^n$ is ***indefinite***, if it is neither positive semidefinite nor negative semidefinite — i.e., if there exists $x_1, x_2 \in \mathbb{R}^n$ such that $x_1^T A x_1 > 0$ and $x_2^T A x_2 < 0$.

Properties

PD ND

- Positive definite and negative definite matrices are always full rank, and hence, invertible.

- For $A \in \mathbb{R}^{m \times n}$ gram matrix is always positive semidefinite. Further, if $m \geq n$ (and we assume for convenience that A is full rank), then gram matrix is positive definite.

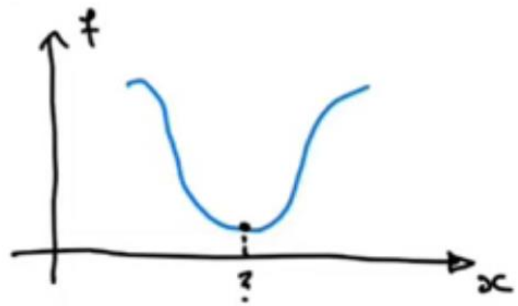
$x^T A^T A x$ $\boxed{A^T A}$

Vector and matrix derivatives

CE40282-1: Linear Algebra
Hamid R. Rabiee and Maryam Ramezani
Sharif University of Technology

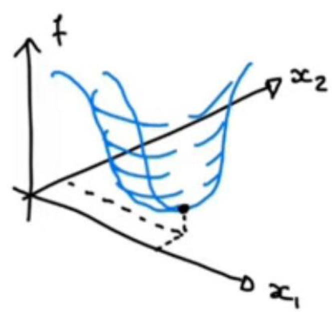
Motivation

- Machine Learning training requires one to evaluate how one vector changes with respect to another
  - How output changes with respect to parameters
- How do we find minimum of a scalar function?



$\frac{\partial f(x)}{\partial x}$

- How do we find minimum of two variables?



$+ \; \frac{\partial f(x)}{\partial x_1} \quad \frac{\partial f(x)}{\partial x_2}$ $\quad *+$

$\frac{\partial f(x)}{\partial x} = \begin{bmatrix} * \\ *+ \end{bmatrix}$

Good Resource

- http://en.wikipedia.org/wiki/Matrix_calculus
- https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf
- http://www.kamperh.com/notes/kamper_matrixcalculus13.pdf

Definitions

- Derivative of a scalar function $f : \mathbb{R}^N \to \mathbb{R}$ with respect to vector $\mathbf{x} \in \mathbb{R}^N$:

[illegible]

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \triangleq \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_N} \end{bmatrix}$$

- Derivative of a vector function $\boldsymbol{f} : \mathbb{R}^N \to \mathbb{R}^M$ with respect to vector $\mathbf{x} \in \mathbb{R}^N$:

M

$$\frac{\partial \boldsymbol{f}(\mathbf{x})}{\partial \mathbf{x}} \triangleq \begin{bmatrix} \frac{\partial \boldsymbol{f}(\mathbf{x})}{\partial x_1} \\ \frac{\partial \boldsymbol{f}(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial \boldsymbol{f}(\mathbf{x})}{\partial x_N} \end{bmatrix} \underset{N}{=} \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \frac{\partial f_2(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_M(\mathbf{x})}{\partial x_1} \\ \frac{\partial f_1(\mathbf{x})}{\partial x_2} & \frac{\partial f_2(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f_M(\mathbf{x})}{\partial x_2} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_1(\mathbf{x})}{\partial x_N} & \frac{\partial f_2(\mathbf{x})}{\partial x_N} & \cdots & \frac{\partial f_M(\mathbf{x})}{\partial x_N} \end{bmatrix}$$

Definitions

- Derivative of a scalar function $f : \mathbb{R}^{M\times N} \to \mathbb{R}$ with respect to matrix $\mathbf{X} \in \mathbb{R}^{M\times N}$:

$$\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} \triangleq \begin{bmatrix} \frac{\partial f(\mathbf{X})}{\partial X_{1,1}} & \frac{\partial f(\mathbf{X})}{\partial X_{1,2}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial X_{1,N}} \\ \frac{\partial f(\mathbf{X})}{\partial X_{2,1}} & \frac{\partial f(\mathbf{X})}{\partial X_{2,2}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial X_{2,N}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f(\mathbf{X})}{\partial X_{M,1}} & \frac{\partial f(\mathbf{X})}{\partial X_{M,2}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial X_{M,N}} \end{bmatrix}$$

- Using the above definitions, we can generalise the chain rule. Given $\mathbf{u} = \boldsymbol{h}(\mathbf{x})$ (i.e. $\mathbf{u}$ is a function of $\mathbf{x}$) and $\boldsymbol{g}$ is a vector function of $\mathbf{u}$, the vector-by-vector chain rule states:

$$\boxed{\frac{\partial \boldsymbol{g}(\mathbf{u})}{\partial \mathbf{x}} = \frac{\partial \mathbf{u}}{\partial \mathbf{x}} \frac{\partial \boldsymbol{g}(\mathbf{u})}{\partial \mathbf{u}}}$$

chain rule

Scalar and vectors

$$\left(\frac{\partial a}{\partial x}\right)_i = \frac{\partial a_i}{\partial x} \longrightarrow$$

بردار

عدد

$$a = \begin{bmatrix} x^2 \\ x^3 \\ x^5 \end{bmatrix} \qquad \frac{\partial a}{\partial x} = \begin{bmatrix} 2x \\ 3x^2 \\ 5x^4 \end{bmatrix}$$

عدد

$$\left(\frac{\partial x}{\partial a}\right)_i = \frac{\partial x}{\partial a_i}$$

بردار

بردار

$$f(x) = x_1 x_2 x_3^2 \qquad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

عدد

بردار

عدد

$$\frac{\partial f(x)}{\partial x} = \nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1}, & \frac{\partial f(x)}{\partial x_2}, & \frac{\partial f(x)}{\partial x_3} \end{bmatrix}^T$$

بردار

گرادیان

$$\begin{bmatrix} x_2 x_3^2 \\ x_1 x_3^2 \\ 2x_1 x_2 x_3 \end{bmatrix}$$

Vectors and vectors

بردار

$$\left(\frac{\partial a}{\partial b}\right)_{ij} = \frac{\partial a_i}{\partial b_j}$$

بردار

$$a = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \qquad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

$$\frac{\partial a}{\partial b} = \begin{bmatrix} \frac{\partial a_1}{\partial b_1} & \frac{\partial a_1}{\partial b_2} \\ \frac{\partial a_2}{\partial b_1} & \frac{\partial a_2}{\partial b_2} \end{bmatrix}$$

ژاکوبین

بردار

$$\frac{\partial (x \cdot a)}{\partial x} = \frac{\partial (x^T a)}{\partial x} = \frac{\partial (a^T x)}{\partial x} = a$$

Matrices and vectors

$$\frac{\partial AB}{\partial X} = \frac{\partial A}{\partial X} B + A \frac{\partial B}{\partial X}$$

chain rule

$$\cancel{\frac{\partial A}{\partial X} B + \frac{\partial B}{\partial X} A}$$

Imp !!

$$\frac{\partial (X^T A X)}{\partial X} = (A + A^T) X$$

اگر A متقارن باشد: $\longrightarrow A = A^T \longrightarrow 2AX$

Another View

- **Finding the Derivative**

To find $f'(x)$, we use a four-step process:

Step 1. Find $f(x+h)$

Step 2. Find $f(x+h)-f(x)$

Step 3. Find $\dfrac{f(x+h)-f(x)}{h}$

Step 4. Find $\lim\limits_{h\to 0} \dfrac{f(x+h)-f(x)}{h}$

- Example: find the derivation of quadratic form

Conclusion

$$\frac{\partial(\boldsymbol{u}(\mathbf{x})+\boldsymbol{v}(\mathbf{x}))}{\partial \mathbf{x}} = \frac{\partial \boldsymbol{u}(\mathbf{x})}{\partial \mathbf{x}} + \frac{\partial \boldsymbol{v}(\mathbf{x})}{\partial \mathbf{x}}$$

$$\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}^\top$$

$$\frac{\partial \mathbf{x}^\top \mathbf{a}}{\partial \mathbf{x}} = \mathbf{a} \quad \checkmark$$

$$\frac{\partial \mathbf{x}^\top \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = (\mathbf{A}+\mathbf{A}^\top)\mathbf{x} \quad \checkmark$$

$$\frac{\partial \mathbf{x}^\top \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x} \text{ if } \mathbf{A} \text{ is symmetric} \quad \checkmark$$

$$\frac{\partial |\mathbf{X}|}{\partial \mathbf{X}} = |\mathbf{X}|(\mathbf{X}^{-1})^\top$$

$$\frac{\partial \ln |\mathbf{X}|}{\partial \mathbf{X}} = (\mathbf{X}^{-1})^\top$$

Conclusion

1. Derivative of a linear function:

$$\frac{\partial}{\partial \vec{x}} \vec{a} \cdot \vec{x} \quad = \quad \frac{\partial}{\partial \vec{x}} \vec{a}^\top \vec{x} \quad = \quad \frac{\partial}{\partial \vec{x}} \vec{x}^\top \vec{a} \quad = \quad \vec{a}$$

   (If you think back to calculus, this is just like $\frac{d}{dx} ax = a$).

2. Derivative of a quadratic function: if A is symmetric, then

$$\frac{\partial}{\partial \vec{x}} \vec{x}^\top A \vec{x} = 2A\vec{x}$$

   (Again, thinking back to calculus this is just like $\frac{d}{dx} ax^2 = 2ax$).

   If you ever need it, the more general rule (for non-symmetric A) is:

$$\frac{\partial}{\partial \vec{x}} \vec{x}^\top A \vec{x} = (A + A^\top)\vec{x},$$

   which of course is the same thing as $2A\vec{x}$ when A is symmetric.