Mathematical Background 2

Source: Ahmadi, ORF 363 slides.

Eigenvalues and eigenvectors

Positive definite and positive semidefinite matrices

Differential calculus

Little o and big O notation

Taylor expansion

Eigenvalues and Eigenvectors

Let A be an $n \times n$ square matrix. A scalar λ and a nonzero vector v satisfying the equation:

$$Av = \lambda v$$

are respectively called an **eigenvalue** and an **eigenvector** of A.

In general, both λ and v may be complex.

Example

$$A = \begin{pmatrix} 4 & 1 \\ -1 & 2 \end{pmatrix} \text{ has } \lambda = 3 \text{ and } v = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Eigenvalues of a symmetric matrix

A is a **symmetric matrix** if $A^T = A$.

All eigenvalues of a real symmetric matrix are real.

Eigenvectors of a symmetric matrix

Any real symmetric $n \times n$ matrix A has a set of n real eigenvectors that are mutually orthogonal.

Eigenvectors of a symmetric matrix

Any real symmetric $n \times n$ matrix A has a set of n real eigenvectors that are mutually orthogonal.

Consequence: Any vector in $\mathbb{R}^n$ can we written as a linear combination of eigenvectors of A.

Eigenvalues and eigenvectors

Positive definite and positive semidefinite matrices

Differential calculus

Little o and big O notation

Taylor expansion

Positive definite and Positive semidefinite matrices

A symmetric $n \times n$ matrix Q is said to be

- Positive semidefinite (psd) if $x^T Q x \geq 0$ for all $x \in \mathbb{R}^n$ ($Q \succeq 0$)

Positive definite and Positive semidefinite matrices

A symmetric $n \times n$ matrix Q is said to be

- Positive semidefinite (psd) if $x^T Q x \geq 0$ for all $x \in \mathbb{R}^n$ ($Q \succeq 0$)

- Positive definite (pd) if $x^T Q x > 0$ for all $x \in \mathbb{R}^n, x \neq 0$ ($Q \succ 0$)

Positive definite and Positive semidefinite matrices

A symmetric $n \times n$ matrix Q is said to be

- Positive semidefinite (psd) if $x^T Qx \geq 0$ for all $x \in \mathbb{R}^n$ ($Q \succeq 0$)
- Positive definite (pd) if $x^T Qx > 0$ for all $x \in \mathbb{R}^n, x \neq 0$ ($Q \succ 0$)
- Negative semidefinite if $x^T Qx \leq 0$ for all $x \in \mathbb{R}^n$ ($Q \preceq 0$)
- Negative definite if $x^T Qx < 0$ for all $x \in \mathbb{R}^n, x \neq 0$ ($Q \prec 0$)
- Indefinite if it is neither positive semidefinite nor negative semidefinite.

Link with the eigenvalues

A symmetric matrix Q is positive semidefinite (resp. positive definite) iff all eigenvalues of are nonnegative (resp. positive).

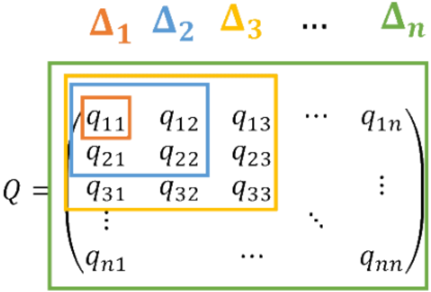
Link with the eigenvalues

A symmetric matrix Q is positive semidefinite (resp. positive definite) iff all eigenvalues of are nonnegative (resp. positive).

A symmetric matrix Q is negative semidefinite (resp. negative definite) iff the eigenvalues of are nonpositive (resp. negative).

Sylvester's criterion

A symmetric matrix Q is **positive definite** iff $\det(\Delta_1), \det(\Delta_2), \ldots, \det(\Delta_n)$ are positive, where $\Delta_1, \Delta_2, \ldots, \Delta_n$ are submatrices defined as in the drawing below. These determinants are called the leading principal minors of the matrix.



Sylvester's criterion

A symmetric matrix Q is **positive semidefinite** iff $\det(\Delta_1), \det(\Delta_2), \ldots, \det(\Delta_{2^n-1})$ are positive, where $\Delta_1, \Delta_2, \ldots, \Delta_{2^n-1}$ are submatrices obtained by choosing the same row and column indices of Q. These determinants are called the **leading principal minors** of the matrix.

Eigenvalues and eigenvectors

Positive definite and positive semidefinite matrices

Differential calculus

Little o and big O notation

Taylor expansion

Partial derivative

The **partial derivative** of a function $f : \mathbb{R}^n \to \mathbb{R}$ **with respect to a variable** x_i is denoted by $\dfrac{\partial f}{\partial x_i}$.

Example: $f(x_1, x_2) = x_1^2 \cos(x_2)$

Jacobian matrix

For a function $f : \mathbb{R}^n \to \mathbb{R}^m$ given as $f(x) = (f_1(x), \ldots, f_m(x))^T$,the **Jabocian matrix** is the $m \times n$ matrix of first partial derivatives:

$$J_f = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

Jacobian matrix

The **Jabocian matrix, evaluated at a point** x is:

$$J_f(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \cdots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \cdots & \frac{\partial f_m}{\partial x_n}(x) \end{pmatrix}$$

First order approximation

The first order approximation of f near a point x_0 is obtained using the Jacobian matrix:

$$A(x) = f(x_0) + J_f(x_0)^T(x - x_0)$$

Note that this is an affine function of x.

The gradient vector

The gradient of a real-valued function $f : \mathbb{R}^n \to \mathbb{R}$ is denoted by ∇f and is given by

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x1}(x) \\ \vdots \\ \frac{\partial f}{\partial xn}(x) \end{pmatrix} = J_f(x)^T$$

Level sets

For a scalar $\alpha \in \mathbb{R}$, the **α-level set** of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

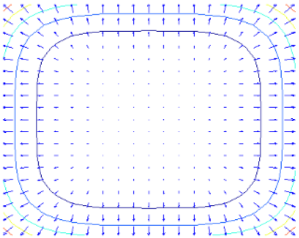
$$S_\alpha = \{x \in \mathbb{R}^n \mid f(x) = \alpha\}$$

and the **α-sublevel set** of f is given by

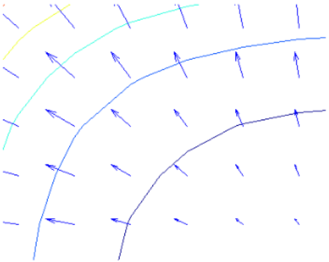
$$S_\alpha = \{x \in \mathbb{R}^n \mid f(x) \leq \alpha\}$$

Connection to gradient vector

At any point $x_0 \in \mathbb{R}^n$, the gradient vector $\nabla f(x_0)$ is orthogonal to the tangent to the level set going through x_0.



Level sets and gradient vectors of a function.



Zooming in on the same picture to see orthogonality.

The Hessian matrix

For a function $f : \mathbb{R}^n \to \mathbb{R}$, the **Hessian matrix** is the $n \times n$ matrix of second derivatives:

$$\nabla_f^2(x) = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1}(x) & \cdots & \dfrac{\partial f}{\partial x_n^2}(x) \end{pmatrix}$$

Practical rules for differentiation

The sum rule

If $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ then $J_{f+g}(x) = J_f(x) + J_g(x)$

Practical rules for differentiation

The sum rule

If $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ then $J_{f+g}(x) = J_f(x) + J_g(x)$

The product rule

If $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^n \to \mathbb{R}^m$. Define the function $h : \mathbb{R}^n \to \mathbb{R}$ by $h(x) = f(x)^T g(x)$. Then

$$J_h(x) = f(x)^T J_g(x) + g(x)^T J_f(X)$$

Practical rules for differentiation

The chain rule

If $f : \mathbb{R} \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}$. Define the function $h : \mathbb{R} \to \mathbb{R}$ by $h(x) = g(f(x))$. Then

$$h'(x) = \nabla g(f(x))^T \begin{pmatrix} f_1'(x) \\ \vdots \\ f_n'(x) \end{pmatrix}$$

Practical rules for differentiation

The chain rule

If $f : \mathbb{R} \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}$. Define the function $h : \mathbb{R} \to \mathbb{R}$ by $h(x) = g(f(x))$. Then

$$h'(x) = \nabla g(f(x))^T \begin{pmatrix} f_1'(x) \\ \vdots \\ f_n'(x) \end{pmatrix}$$

A special case that comes up a lot

Let x and y be two **fixed** vectors in $\mathbb{R}^n$ and let $g : \mathbb{R}^n \to \mathbb{R}$. Define a univarite function $h(t) = g(x + ty)$. Then

$$h'(t) = \nabla g(x + ty)^T y$$

Examples

Gradients and Hessians of affine function

$f(x) = c^T x + b$, where $x \in \mathbb{R}^n, c \in \mathbb{R}^n, b \in \mathbb{R}$.

$$\nabla f(x) = c \text{ and } \nabla^2 f(x) = 0_{n \times n}$$

Examples

Gradients and Hessians of affine function

$f(x) = c^T x + b$, where $x \in \mathbb{R}^n, c \in \mathbb{R}^n, b \in \mathbb{R}$.

$$\nabla f(x) = c \text{ and } \nabla^2 f(x) = 0_{n \times n}$$

Gradients and Hessians of quadratic form

$f(x) = x^T Q x$, where $x \in \mathbb{R}^n, Q \in \mathbb{R}^{n \times n}$ is symmetric.

$$\nabla f(x) = 2Qx \text{ and } \nabla^2 f(x) = 2Q$$

Eigenvalues and eigenvectors

Positive definite and positive semidefinite matrices

Differential calculus

Little o and big O notation

Taylor expansion

Little o and big O notation

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^n \to \mathbb{R}$ be functions

- $f(x) = O(g(x))$ (pronounced "f is big Oh of g") means there is $K > 0$ such that:

$$\frac{\|f(x)\|}{|g(x)|} \leq K \text{ when } x \text{ is small}$$

Little o and big O notation

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^n \to \mathbb{R}$ be functions

- $f(x) = O(g(x))$ (pronounced "f is big Oh of g") means there is $K > 0$ such that:

$$\frac{\|f(x)\|}{|g(x)|} \leq K \text{ when } x \text{ is small}$$

- $f(x) = o(g(x))$ (pronounced "f is little oh of g") means that:

$$\frac{\|f(x)\|}{|g(x)|} \to 0 \text{ as } x \to 0$$

Examples

$f(x) = O(g(x))$

- $x = O(x)$ as $\frac{|x|}{|x|} = 1$

Examples

$f(x) = O(g(x))$

- $x = O(x)$ as $\frac{|x|}{|x|} = 1$

- $x^2 = O(x)$

Examples

$f(x) = O(g(x))$

- $x = O(x)$ as $\frac{|x|}{|x|} = 1$
- $x^2 = O(x)$
- $\cos(x) = O(1)$

Examples

$f(x) = O(g(x))$

- $x = O(x)$ as $\frac{|x|}{|x|} = 1$
- $x^2 = O(x)$
- $\cos(x) = O(1)$
- $x \neq O(x^2)$

Examples

$f(x) = O(g(x))$

- $x = O(x)$ as $\frac{|x|}{|x|} = 1$
- $x^2 = O(x)$
- $\cos(x) = O(1)$
- $x \neq O(x^2)$
- $\sin(x) = O(x)$

Examples

$$f(x) = o(g(x))$$

- $x^2 = o(x)$

Examples

$$f(x) = o(g(x))$$

- $x^2 = o(x)$

- $\begin{pmatrix} x^3 \\ 2x^2 \end{pmatrix} = o(x)$

Examples

$$f(x) = o(g(x))$$

- $x^2 = o(x)$

- $\begin{pmatrix} x^3 \\ 2x^2 \end{pmatrix} = o(x)$

- $x^3 = o(x^2)$

Examples

$$f(x) = o(g(x))$$

- $x^2 = o(x)$
- $\begin{pmatrix} x^3 \\ 2x^2 \end{pmatrix} = o(x)$
- $x^3 = o(x^2)$
- $x = o(1)$

Eigenvalues and eigenvectors

Positive definite and positive semidefinite matrices

Differential calculus

Little o and big O notation

Taylor expansion

Taylor expansion

- The idea behind Taylor expansion is to **approximate a function around a given point** by functions that are ”simpler”; in this case **by polynomials**.

- As we increase the order of the Taylor expansion, we increase the degree of this polynomial and we reduce the error in our approximation.

- The little o and big O notation that we just introduced nicely capture how our error of approximation scales around the point we are approximating.

Taylor expansion: univariate function

Version 1: Let $f : \mathbb{R} \to \mathbb{R}$. The Taylor expansion of f at point $a \in \mathbb{R}$ is

$$f(x) = f(a) + \frac{f^{(1)}(a)}{1!}(x-a) + \frac{f^{(2)}(a)}{2!}(x-a)^2 + \ldots + \frac{f^{(m)}(a)}{m!}(x-a)^m + o(|x-a|^m)$$

Taylor expansion: univariate function

Version 1: Let $f:\mathbb{R}\to\mathbb{R}$. The Taylor expansion of f at point $a\in\mathbb{R}$ is

$$f(x)=f(a)+\frac{f^{(1)}(a)}{1!}(x-a)+\frac{f^{(2)}(a)}{2!}(x-a)^2+\ldots+\frac{f^{(m)}(a)}{m!}(x-a)^m+o(|x-a|^m)$$

Version 2: Let $f:\mathbb{R}\to\mathbb{R}$. The Taylor expansion of f at point $a\in\mathbb{R}$ is

$$f(x)=f(a)+\frac{f^{(1)}(a)}{1!}(x-a)+\frac{f^{(2)}(a)}{2!}(x-a)^2+\ldots+\frac{f^{(m)}(a)}{m!}(x-a)^m+O(|x-a|^{m+1})$$

Taylor expansion: multivariate function

Let $f : \mathbb{R}^n \to \mathbb{R}$. **First order**:

Version 1: If f is C^1

$$f(x) = f(a) + \nabla f(a)^T (x - a) + o(\|x - a\|)$$

Version 2: If f is C^2

$$f(x) = f(a) + \nabla f(a)^T (x - a) + O(\|x - a\|)^2$$

Taylor expansion: multivariate function

Let $f : \mathbb{R}^n \to \mathbb{R}$. **Second order**:

Version 1: If f is C^2

$$f(x) = f(a) + \nabla f(a)^T(x - a) + \frac{1}{2}(x - a)^T \nabla^2 f(a)^T (x - a) + o(\|x - a\|^2)$$

Version 2: If f is C^3

$$f(x) = f(a) + \nabla f(a)^T(x - a) + \frac{1}{2}(x - a)^T \nabla^2 f(a)^T (x - a) + O(\|x - a\|^3)$$