

# Handout

## Review of Linear Algebra, Matrix Computations, Derivatives and Convexity

MIE 1624H

October 9, 2018

### Review of derivatives, gradients and Hessians:

- Given a function  $f$  of  $n$  variables  $x_1, x_2, \dots, x_n$ , we use the following notations to represent the vector of variables and the function:  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ ,  $f(\mathbf{x}) = f(x_1, x_2, \dots, x_n)$
- The gradient extends the notion of derivative, the Hessian matrix – that of second derivative.
- We define the *partial derivative* relative to variable  $x_i$ , written as  $\frac{\partial f}{\partial x_i}$ , to be the derivative of  $f$  with respect to  $x_i$  treating all variables except  $x_i$  as constant.
- The gradient of  $f$  at  $\mathbf{x}$ , written as  $\nabla f(\mathbf{x})$ , is

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

The gradient of  $f$  is a multivariate function of  $\mathbf{x}$ ,  $f(\mathbf{x}) \in \mathbb{R}$ ,  $\nabla f(\mathbf{x}) \in \mathbb{R}^n$ .

- The gradient vector  $\nabla f(\mathbf{x})$  gives the direction of steepest ascent of the function  $f$  at point  $\mathbf{x}$ . The gradient acts like the derivative in that small changes around a given point  $\mathbf{x}^*$  can be estimated using the gradient (see first-order Taylor series expansion and finite difference method).
- Second partial derivatives  $\frac{\partial^2 f}{\partial x_i \partial x_j}$  are obtained from  $f(\mathbf{x})$  by taking the derivative relative to  $x_i$  (this yields the first partial derivative  $\frac{\partial f}{\partial x_i}$ ) and then by taking the derivative of  $\frac{\partial f}{\partial x_i}$  relative to  $x_j$ . So, we can compute  $\frac{\partial^2 f}{\partial x_1 \partial x_1} = \frac{\partial^2 f}{\partial x_1^2}$ ,  $\frac{\partial^2 f}{\partial x_1 \partial x_2}$  and so on. These values are arranged into the *Hessian* matrix:

$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

The Hessian matrix is a symmetric matrix, that is  $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$ .

## Computing gradients and Hessians:

---

### Example

Compute the gradient and the Hessian of the function  $f(x_1, x_2) = x_1^2 - 3x_1x_2 + x_2^2$  at the point  $\mathbf{x} = (x_1, x_2)^T = (1, 1)^T$ .

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 2x_1 - 3x_2 \\ -3x_1 + 2x_2 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$$

$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{pmatrix} = \begin{pmatrix} 2 & -3 \\ -3 & 2 \end{pmatrix}$$

---

### Taylor series expansion:

Second-order Taylor series expansion:

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^T (\mathbf{x} - \mathbf{x}_0) + \frac{1}{2} (\mathbf{x} - \mathbf{x}_0)^T \nabla^2 f(\mathbf{x}_0) (\mathbf{x} - \mathbf{x}_0)$$

First-order Taylor series expansion:

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^T (\mathbf{x} - \mathbf{x}_0)$$

---

### Example

$f(x_1, x_2) = x_1^2 - 3x_1x_2 + x_2^2$ , compute  $f(1.01, 1.01)$  using first- and second-order Taylor series expansion at the point  $\mathbf{x}_0 = (1, 1)^T$ .

First-order Taylor series expansion:

$$f(1.01, 1.01) = f(1, 1) + \nabla f(1, 1)^T \begin{pmatrix} 1.01 - 1 \\ 1.01 - 1 \end{pmatrix} = -1 + (-1, -1) \begin{pmatrix} 0.01 \\ 0.01 \end{pmatrix} = -1.02$$

Second-order Taylor series expansion:

$$\begin{aligned} f(1.01, 1.01) &= f(1, 1) + \nabla f(1, 1)^T \begin{pmatrix} 0.01 \\ 0.01 \end{pmatrix} + \frac{1}{2} (0.01, 0.01) \nabla^2 f(1, 1) \begin{pmatrix} 0.01 \\ 0.01 \end{pmatrix} = \\ &= -1 + (-1, -1) \begin{pmatrix} 0.01 \\ 0.01 \end{pmatrix} + \frac{1}{2} (0.01, 0.01) \begin{pmatrix} 2 & -3 \\ -3 & 2 \end{pmatrix} \begin{pmatrix} 0.01 \\ 0.01 \end{pmatrix} = -1.0201 \end{aligned}$$

---

### Convex functions:

**Definition** A function  $f$  is convex if for any  $\mathbf{x}^1, \mathbf{x}^2 \in C$  and  $0 \leq \lambda \leq 1$

$$f(\lambda \mathbf{x}^1 + (1 - \lambda) \mathbf{x}^2) \leq \lambda f(\mathbf{x}^1) + (1 - \lambda) f(\mathbf{x}^2).$$

A square matrix  $\mathbf{A}$  said to be positive definite (PD) if  $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$  for all  $\mathbf{x} \neq 0$ .

A square matrix  $\mathbf{A}$  said to be positive semidefinite (PSD) if  $\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0$  for all  $\mathbf{x}$ .

Hessian  $\nabla^2 f(\mathbf{x})$  is PD  $\implies$  strictly convex function.

Hessian  $\nabla^2 f(\mathbf{x})$  is PSD  $\implies$  convex function.

Gradient  $\nabla f(\bar{\mathbf{x}}) = 0$  and Hessian  $\nabla^2 f(\bar{\mathbf{x}})$  is PSD  $\implies \bar{\mathbf{x}}$  is a minimum of the function  $f$ .

Gradient  $\nabla f(\bar{\mathbf{x}}) = 0$  and Hessian  $\nabla^2 f(\bar{\mathbf{x}})$  is PD  $\implies \bar{\mathbf{x}}$  is a strict minimum of the function  $f$ .

### Checking a matrix for PD and PSD by computing principal minors:

Leading principal minors  $D_k, k = 1, 2, \dots, n$  of a matrix  $\mathbf{A} = (a_{ij})_{[n \times n]}$  are defined as

$$D_k = \det \begin{pmatrix} a_{11} & \dots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \dots & a_{kk} \end{pmatrix}$$

A square matrix  $\mathbf{A}$  is PD  $\Leftrightarrow D_k > 0$  for all  $k = 1, 2, \dots, n$ .

---

### Example

Consider the function  $f(\mathbf{x}) = 3x_1^2 + 3x_2^2 + 5x_3^2 - 2x_1x_2$ . The corresponding Hessian matrix is

$$\nabla^2 f(\mathbf{x}) = 2 \begin{pmatrix} 3 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix}$$

Leading principal minors of  $\nabla^2 f(\mathbf{x})$  are

$$D_1 = 2 \cdot 3 = 6 > 0, \quad D_2 = 2 \cdot \det \begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix} = 2[3 \cdot 3 - (-1)(-1)] = 2 \cdot 8 = 16 > 0,$$

$$\begin{aligned} D_3 &= 2 \cdot \det \begin{pmatrix} 3 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix} \\ &= 2([3 \cdot 3 \cdot 5 + 0 \cdot 0 \cdot (-1) + 0 \cdot 0 \cdot (-1)] - [0 \cdot 0 \cdot 3 + 0 \cdot 0 \cdot 3 + (-1) \cdot (-1) \cdot 5]) \\ &= 2 \cdot 40 = 80 > 0 \end{aligned}$$

So, the Hessian is positive definite (PD) and the function is strictly convex.

---

A square matrix  $\mathbf{A}$  is PSD  $\Leftrightarrow$  all the principal minors of  $\mathbf{A}$  are  $\geq 0$ .

The *principal minor* is

$$\det \begin{pmatrix} a_{i_1 i_1} & \dots & a_{i_1 i_p} \\ \vdots & & \vdots \\ a_{i_p i_1} & \dots & a_{i_p i_p} \end{pmatrix}, \text{ where } 1 \leq i_1 < i_2 < \dots < i_p \leq n, p \leq n.$$

### Checking if symmetric matrix is PD or PSD by computing its eigenvalues:

**Definition** Any number  $\lambda$  such that the equation  $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$  has a non-zero vector-solution  $\mathbf{x}$  is called an eigenvalue (or a characteristic root) of the equation.

A symmetric matrix is PD if its eigenvalues  $\lambda_i > 0$  for all  $i = 1, 2, \dots, n$  and PSD if  $\lambda_i \geq 0$ .

How to calculate eigenvalues:  $\mathbf{A}\mathbf{x} - \lambda\mathbf{x} = 0 \Rightarrow (\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = 0$ . Since  $\mathbf{x}$  is non-zero, the determinant of  $(\mathbf{A} - \lambda\mathbf{I})$  should vanish. Therefore all eigenvalues can be calculated as roots of the equation (which is often called the characteristic equation of  $\mathbf{A}$ ):

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0.$$

---

### Example

Consider the Hessian matrix

$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} 3 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix}$$

Computing eigenvalues

$$\det(\nabla^2 f(\mathbf{x}) - \lambda\mathbf{I}) = \begin{vmatrix} 3-\lambda & -1 & 0 \\ -1 & 3-\lambda & 0 \\ 0 & 0 & 5-\lambda \end{vmatrix} = (5-\lambda)(\lambda^2 - 6\lambda + 8) = (5-\lambda)(\lambda-2)(\lambda-4) = 0.$$

Therefore, the eigenvalues are  $\lambda = 2$ ,  $\lambda = 4$  and  $\lambda = 5$ . As all of them are strictly positive, the Hessian is positive definite (PD).

Try computing eigenvalues and determinants in Python:

```
import numpy as np
H = np.matrix( ((3,-1,0), (-1,3,0), (0,0,5)) )
eigenvalues, eigenvectors = np.linalg.eig(H)
determinant = np.linalg.det(H)
```

---

### Properties of convex functions:

- if  $f$  is convex function, its sublevel set  $f(\mathbf{x}) \leq \alpha$  is convex;
- positive multiple of convex function is convex:  
 $f$  convex,  $\alpha \geq 0 \implies \alpha f$  convex
- sum of convex functions is convex:  
 $f_1, f_2$  convex  $\implies f_1 + f_2$  convex
- pointwise maximum of convex functions is convex:  
 $f_1, f_2$  convex  $\implies \max\{f_1(\mathbf{x}), f_2(\mathbf{x})\}$  convex  
(corresponds to intersections of epigraphs)
- affine transformation of domain:  
 $f$  convex  $\implies f(\mathbf{A}\mathbf{x} + \mathbf{b})$  convex

### Composition rules:

Composite function

$$f(x) = h(g(x))$$

is convex if:

- $g$  convex;  $h$  convex nondecreasing
- $g$  concave;  $h$  convex nonincreasing

*Proof* (differentiable functions,  $x \in \Re$ ):

$$f'' = h''(g')^2 + g''h'$$

Examples:

- $f(x) = e^{g(x)}$  is convex if  $g$  is convex
- $f(x) = 1/g(x)$  is convex if  $g$  is concave, positive
- $f(x) = g(x)^p$ ,  $p \geq 1$  is convex if  $g(x)$  is convex, positive

---

## Examples

**Show that the function  $e^x + \frac{1}{2}x^2$  is convex and solve  $\min e^x + \frac{1}{2}x^2$ .**

First derivative: A function is increasing if  $f' > 0$ , decreasing if  $f' < 0$  and neither if  $f' = 0$ .

Second derivative: A function is convex if  $f'' > 0$  and concave if  $f'' < 0$ .

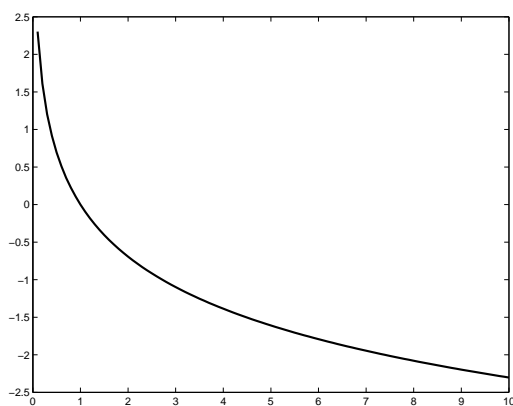
Answer:  $f'(x) = e^x + x$  and  $f''(x) = e^x + 1 > 0$ . So,  $f$  is convex.

Thus, we can find a solution to an optimization problem by solving  $f'(x) = 0$ , given  $f$  is convex.

**Find the local/global minimum of the functions if exists:**

- $-\ln x$

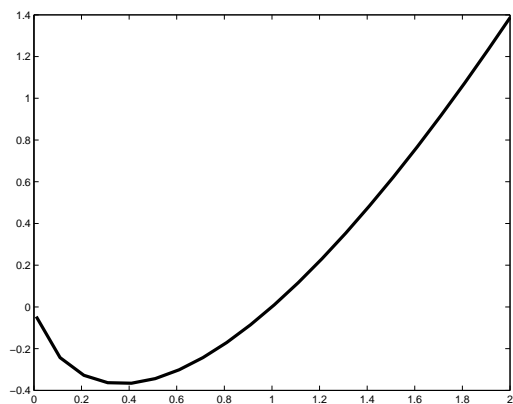
$f'(x) = -1/x$ ,  $f''(x) = 1/x^2 > 0$  - strictly convex function.  $f'(x) = -1/x = 0 \implies x \rightarrow \infty$



- $x \ln x$

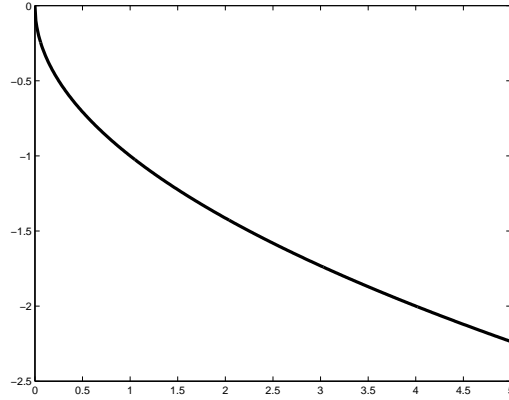
$f'(x) = 1 + \ln x$ ,  $f''(x) = 1/x > 0$  on the domain of  $\ln x \implies$  strictly convex function.

$f'(x) = 1 + \ln x = 0 \implies x = 0.37$  (global minimum).



- $-\sqrt{x}$  when  $x \geq 0$

$f'(x) = -0.5x^{-1/2}$ ,  $f''(x) = 0.25x^{-3/2} \geq 0$  when  $x \geq 0 \Rightarrow$  convex function.  $f'(x) = -0.5x^{-1/2} = 0 \Rightarrow x \rightarrow \infty$ .



- $(x_1 - 2)^2 + (x_2 + 1)^2 - 2$

$$\nabla f(x) = \begin{pmatrix} 2(x_1 - 2) \\ 2(x_2 + 1) \end{pmatrix}$$

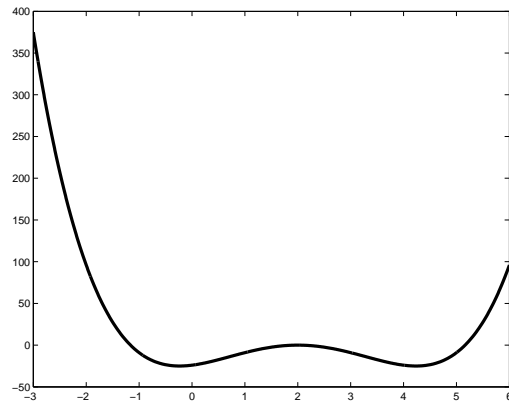
$$\nabla^2 f(x) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \succ 0$$

As  $\nabla^2 f(x)$  is PD,  $f(x)$  is strictly convex function.

$$\nabla f(x) = 0 \Rightarrow x = \begin{pmatrix} 2 \\ -1 \end{pmatrix} \text{ (global minimum).}$$

- $(x - 2)^4 - 10(x - 2)^2$

$f'(x) = 4(x - 2)^3 - 20(x - 2)$ ,  $f''(x) = 12(x - 2)^2 - 20$  - non-convex, non-concave function.



### Example of Newton Method

Consider minimizing the function  $f(x_1, x_2) = e^{x_1+x_2-2} + (x_1 - x_2)^2$ . Given  $\mathbf{x}^0 = (1, 1)^T$ , apply a full Newton step and compute  $\mathbf{x}^1$ .

$$\begin{aligned}\nabla f(\mathbf{x}) &= \begin{pmatrix} e^{x_1+x_2-2} + 2(x_1 - x_2) \\ e^{x_1+x_2-2} - 2(x_1 - x_2) \end{pmatrix} \\ \nabla^2 f(\mathbf{x}) &= \begin{pmatrix} e^{x_1+x_2-2} + 2 & e^{x_1+x_2-2} - 2 \\ e^{x_1+x_2-2} - 2 & e^{x_1+x_2-2} + 2 \end{pmatrix} \\ \mathbf{x}^1 &= \mathbf{x}^0 - (\nabla^2 f(\mathbf{x}^0))^{-1} \nabla f(\mathbf{x}^0)\end{aligned}$$

So

$$\begin{aligned}\nabla f(\mathbf{x}^0) &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ \nabla^2 f(\mathbf{x}^0) &= \begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix}\end{aligned}$$

Instead of inverting matrix  $\nabla^2 f(\mathbf{x}^0)$ , which is costly, we can solve the system of equations. Please note that if want to compute  $y = A^{-1}b$ , we can solve the system of equations  $Ay = b$  to find  $y$ . So we can solve  $\nabla^2 f(\mathbf{x}^0)y = \nabla f(\mathbf{x}^0)$  to get  $y = (\nabla^2 f(\mathbf{x}^0))^{-1} \nabla f(\mathbf{x}^0)$ .

$$\begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \Rightarrow \begin{cases} y_1 = 0.5 \\ y_2 = 0.5 \end{cases}$$

Thus,

$$\begin{aligned}\mathbf{x}^1 &= \mathbf{x}^0 - (\nabla^2 f(\mathbf{x}^0))^{-1} \nabla f(\mathbf{x}^0) \\ &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}\end{aligned}$$

$$f(\mathbf{x}^1) = 0.3679 < f(\mathbf{x}^0) = 1$$