Lecture 03
Mathematical Review in R

Oran Kittithreerapronchai[1]

[1]Department of Industrial Engineering, Chulalongkorn University
Bangkok 10330 THAILAND

last updated: August 7, 2023

Outline

1. Matrix and Linear Algebra
2. Reviews of Function Analysis & Calculus
3. Reviews of Optimization and Geometry
4. Reviews Materials for Mathematical Review
   - Extended Concept of Linear Algebra
   - Extended Concept of Convex
   - Extended Concept of Convergence

source: General references [NC20, CŻ13, Win22, Pat14]

Notation Expansion

- **Decision Variable:** single variable ⇔ matrix
- **Feasible Region:** all points ⇔ convex set, convex set
- **Visualization:** scatter plot ⇔ level set, gradient, param. function
- **Unique Opt:** $f''(\cdot)$ ⇔ convex function ≡ positive definite $(\nabla^2 f(\cdot))$
- **Efficiency:** # iterations ⇔ convergent rate
- **Quality of Solution:** reliable solution ⇔ in descent direction

Key Concepts

All iterative algorithms requires 'right' direction & 'right' step size

$$\mathbf{x_{k+1}} = \mathbf{x_k} + \gamma_k \mathbf{d_k}$$

What is Matrix?

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}$$

- **row & column:**
- **scalar product:** $\mathbf{u}^T \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$
- **norm:** $\|\mathbf{v}\| = \sqrt{\mathbf{v}^T \cdot \mathbf{v}}$
- **matrix multiplication:** $\mathbf{AB}$

source: [CŻ13]

Scalar Product/ Inner Product

Let, $$\mathbf{u} = \begin{bmatrix} 7 \\ 4 \\ 2 \end{bmatrix}$$

$$\mathbf{v} = \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}$$

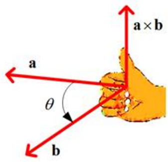
observe follows

- $\mathbf{u}^T \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}^T = \|\mathbf{u}^T\| \|\mathbf{v}\| \cos\theta$, where θ is the angle between the vectors
- $\mathbf{u}^T \cdot \mathbf{u} = \|\mathbf{u}\|^2 = \sum_i (u_i)^2$

Cross Product of Matrix

- **Meaning:** finding an orthogonal vector to other vectors
- **Use In IE:** constraint optimization

$$\vec{a} \times \vec{b} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$$



Matrix Multiplication

$$\text{Let,}\quad \mathbf{A} = \begin{bmatrix} 7 & 1 \\ 4 & -3 \\ 2 & 0 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 2 & 1 & 7 \\ 0 & -1 & 4 \end{bmatrix}$$

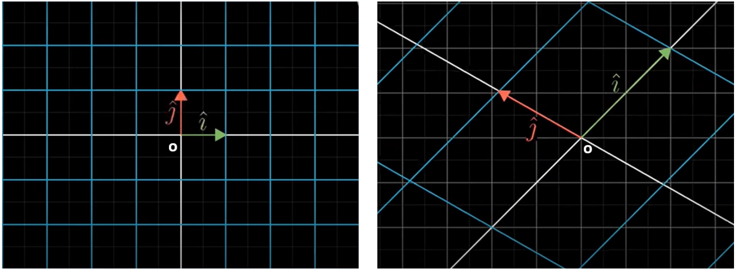
$$\mathbf{C} = \begin{bmatrix} 1 & -1 & 3 \\ 2 & 2 & 3 \\ -1 & 4 & 7 \end{bmatrix}$$

Verify these rules

- **Associative law:** $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}) = \mathbf{ABC}$
- **Distributive law:** $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$

- **No Commutative law** $\mathbf{AB} \neq \mathbf{BA}$

Multiplication = Transformation



Source. 'Matrix multiplication' by *3Blue1Brown*|youtube.com

- **What** function that take one vector and return another vector
- **Linear Transformation:** stretch, compress, and rotation at origin so the grid is evenly space and parallel

EXAMPLE

$$\begin{bmatrix} 0 & 2 \\ 1/2 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2\,y \\ \frac{1}{2}x \end{bmatrix}$$

What is System of Linear Equations?

A system of linear equation:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\ \vdots \qquad\qquad \vdots \qquad\qquad & \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m \end{aligned}$$

where $x_1, x_2, \ldots, x_n$ are referred to as *variables* and the a_{ij}'s and b_i's are *constants*. There are m equations in n variables.

- Above can be represent as $\mathbf{Ax} = \mathbf{b}$ or simply $\mathbf{A}|\mathbf{b}$
- A solution to SLE m equations in n variables is a set of values for the variables that satisfies all m equations.

Matrix Representation

Find the matrix representation of:

$$2x_1 + 5x_2 = 4$$
$$3x_1 + 7x_2 = 2$$

$$\begin{bmatrix} 2 & 5 \\ 3 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

Examples of Three Possible Cases

Case 1: no solution

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 1 & 2 & 1 & 3 \\ 2 & 4 & 2 & 4 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 2 & 1 & 2 \\ 0 & 0 & 0 & -2 \end{array}\right]$$

Case 2: unique solution

$$\left[\begin{array}{ccc|c} 2 & 2 & 1 & 9 \\ 2 & -1 & 2 & 6 \\ 1 & -1 & 2 & 5 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

Case 3: infinite number of solutions

$$\left[\begin{array}{ccc|c} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 3 \\ 1 & 2 & 1 & 4 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & -1 & -2 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Determinant of Matrix

Determinant

- **What:** unique property of square matrix, implying scaling of transformation
- **Use:** find linear dependent and eigenvalue
- **Useful properties:**
  - row operation + or − leaves the determinant unchanged (no scaling)
  - $\det \mathbf{AB} = (\det \mathbf{A})\,(\det \mathbf{B})$
  - $\det \mathbf{A} = \det \mathbf{A}^T$

Example Find determinant of

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

$$\det \mathbf{A} = \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -6 & -12 \end{vmatrix}$$

$$= \begin{vmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & 0 & 0 \end{vmatrix}$$

Inverse of Matrix

Inverse

- **What:** unique property of square matrix, i.e. $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$
- **Compute:** Using row operation to convert $[\mathbf{A}|\mathbf{I}]$ into $[\mathbf{I}^{-1}|\mathbf{A}]$

$$\left[\begin{array}{ccc|ccc} 2 & 0 & -1 & 1 & 0 & 0 \\ 3 & 1 & 2 & 0 & 1 & 0 \\ -1 & 0 & 1 & 0 & 0 & 1 \end{array}\right] \rightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & -5 & 1 & -7 \\ 0 & 0 & 1 & 1 & 0 & 2 \end{array}\right]$$

Positive Definite

- **What:** general condition for $f''(x) \leq 0$
- **Motivation:** If $(x, y) \neq (0,0)$, when does $f(\cdot) > 0 \; \forall (x, y) \in \mathbb{R}^2$, where $f(x, y) = a\,x^2 + 2b\,x\,y + c\,y^2$?

$$f(x, y) = \frac{1}{2} \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 2a & 2b \\ 2b & 2c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

- **Observation:**
  - Even a, b, and $c > 0$, then $f(\cdot) \not> 0$
  - At least a and $c > 0$
- **Indefinite:** $a\,c - b^2 < 0$
- **Positive definite:** $a\,c - b^2 > 0$ and $a > 0$
- **Negative definite:** $a\,c - b^2 > 0$ and $a < 0$ ($imply$ $c < 0$)

Definite Theorem

Definition

realMatrix A real symmetric matrix $\mathbf{A}$ is **positive definite** if and only if it satisfies one of following conditions:

- $\mathbf{x}^{\mathbf{T}}\mathbf{A}\mathbf{x} > 0$ for all nonzero vector $\mathbf{x}$
- All the eigenvalues of $\mathbf{A}$ satisfy $\lambda_i > 0$
- All the upper left submatrices $\mathbf{A}_{\mathbf{k}}$ has positive determinants
- All the pivots (without row exchanges) satisfy $d_i > 0$

Example 1: Is $\mathbf{A}$ positive definite

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$$

Examples of positive definite

EXAMPLE 2: Show that this matrix is $\mathbf{A}$ is not positive semi-definite

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

HINT: check determine of $\mathbf{A_2}$

EXAMPLE 3: For what range of numbers b and c are matrices $\mathbf{B}$ and $\mathbf{C}$ are positive definite

$$\mathbf{B} = \begin{bmatrix} b & 1 & 1 \\ 1 & b & 1 \\ 1 & 1 & b \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} 2 & 2 & 4 \\ 2 & c & 8 \\ 4 & 8 & 7 \end{bmatrix}$$

$$b \in (-2, 1) \qquad\qquad c = \emptyset$$

Eigenvalue and Eigenvector

- **Eigenvector:** vector that preserve its direction after linear transformation
- **Definition:** Given matrix $\mathbf{A}$, a scalar λ and vector $\mathbf{v}$ are called eigenvalue and eigenvector of matrix $\mathbf{A}$ if and only if $\mathbf{A}\,\mathbf{x} = \lambda\mathbf{x}$
- **Properties:** $\det(\lambda\mathbf{I} - \mathbf{A}) = 0$ (why?)
- **Theorem:** all eigenvalue of a symmetric matrix are real

Find all eigenvalues of matrix $\mathbf{A}$

$$\mathbf{A} = \begin{bmatrix} 1 & 4 & 5 \\ 0 & \frac{3}{4} & 6 \\ 0 & 0 & \frac{1}{2} \end{bmatrix} \quad \begin{bmatrix} \lambda - 1 & 4 & 5 \\ 0 & \lambda - \frac{3}{4} & 6 \\ 0 & 0 & \lambda - \frac{1}{2} \end{bmatrix}$$

Summary: Geometry of Matrix

- **Vector:** at origin
- **Addition:** result of connecting vectors head to tail
- **Negative:** reverse direction (rotate 180^o)
- **Transpose:** reflection of vector (across diagonal)
- **Multiplicative:** linear projection/transformation
- **Inversion:** reversing linear projection/transformation
- **Determinant:** scaling of transformation/ signed 'area' of parallelogram
- **Rank:** minimal necessary dimension to represent collection of vectors

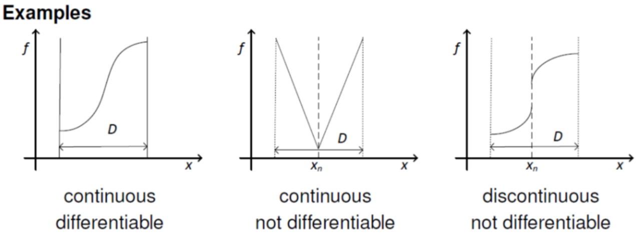
- **More Reading:** https://www.gastonsanchez.com/matrix4sl/

Differentiability

- **What:** a function $f(\cdot)$ that maps $\mathbb{D} \subseteq \mathbb{R}^n$ to $\mathbb{R}$

$$f'(\mathbf{x_0} \in \mathbb{D}) \equiv \lim_{\|\boldsymbol{\epsilon}\| \to 0} \frac{f(\mathbf{x_0} + \boldsymbol{\epsilon}) - f(\mathbf{x_0})}{\|\boldsymbol{\epsilon}\|}$$

- **Meaning:** change of function within $\mathbf{x_0}$
- **Note:** derivative may not exist!



Taylor Series

If a function $f\colon \mathbb{R}^n \rightarrow \mathbb{R}$ is m times continuously differentiable (i.e., $f \in \mathcal{C}^m$) on $[\mathbf{a}, \mathbf{b}]$, then

$$f(\mathbf{b}) = f(\mathbf{a}) + \frac{\mathbf{b}-\mathbf{a}}{1!} f^{(1)}(\mathbf{a}) + \frac{(\mathbf{b}-\mathbf{a})^2}{2!} f^{(2)}(\mathbf{a}) + \ldots + \frac{(\mathbf{b}-\mathbf{a})^{m-1}}{(m-1)!} f^{(m-1)}(\mathbf{a}) + R_m$$

where,

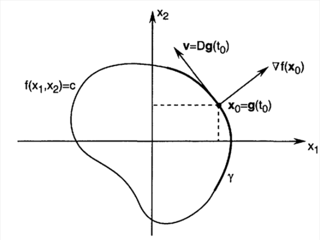
$$f^{(k)}(\cdot) = \nabla^k f(\mathbf{x})$$

$$R_m = \frac{(b-a)^m}{m!} f^{(m)}\left(a + \lambda(b-a)\right) \text{ and } \lambda \in [0,1]$$

Note

- Accuracy of series depends on number of terms
- $\lim_{m\to\infty} R_m = 0$
- $\nabla f(\mathbf{x})$ is called Gradient vector
- $\nabla^2 f(\mathbf{x})$ or $\mathbf{F}(\mathbf{x})$ is called Hessian matrix

Gradient Vector



Source. Chong & Zak. 2001 pp 68 [CŻ13]

- **What:** first-order partial derivative at a given point
- **Gradient ($\nabla f(\mathbf{x_0})$):** the direction maximize increasing rate of $f(\cdot)$

$$f(x_1, y_1) = 2\,x_1 - x_2. \quad \nabla f(\mathbf{x}) = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

Hessian Matrix

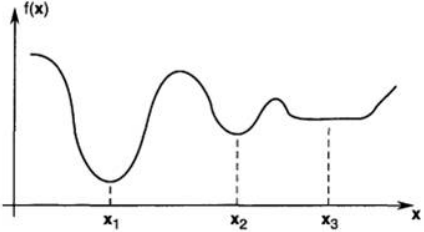
$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix} \quad \text{and} \quad \nabla^2 f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

Source. Chong & Zak. 2001 pp 68 [CŻ13]

- **What:** second-order partial derivative; always symmetric matrix
- **Hessian ($\nabla^2 f(\mathbf{x_0})$):** the local curvature of of $f(\cdot)$

$$f(x_1, y_1) = 2\,x_1^2 - \sin(x_2). \quad \nabla^2 f(\mathbf{x}) = \begin{bmatrix} 4 & 0 \\ 0 & -\sin(x_2) \end{bmatrix}$$

Local and Global Minimizer



Source. Chong & Zak. 2001 pp 72 [CŻ13]

Local Minimizer

Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a function defined on $\mathcal{F} \subset \mathbb{R}^n$. A point $\mathbf{x}^* \in \mathcal{F}$ is a *local minimizer* of $f(\cdot)$ if $\exists\ \epsilon > 0$ such that $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for all $\mathbf{x} \in \mathcal{F} \backslash \mathbf{x}^*$ and $\|\mathbf{x} - \mathbf{x}^*\| < \epsilon$

Global Minimizer

A point $\mathbf{x}^* \in \mathcal{F}$ is a *global minimizer* of $f(\cdot)$ if $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for all $\mathbf{x} \in \mathcal{F} \backslash \mathbf{x}^*$

Optimal Condition

- **What:** Is this point local optima?
- **First Order:** $\nabla f(\mathbf{x}_*) = 0$
- **Second Order:** $\nabla^2 f(\mathbf{x}_*)$ is positive definite (for minimization)

Example: Check Optimality Conditions:

- $f(x_1, x_2) = \frac{1}{3}x_1^3 + \frac{1}{2}x_1^2 + 2x_1 x_2 + \frac{1}{2}x_2^2 - x_2 + 9$

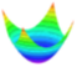
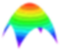
$$\nabla f(\mathbf{x}) = \begin{bmatrix} x_1^2 + x_1 + 2x_2 \\ 2x_1 + x_2 - 1 \end{bmatrix}$$

$(x_1 - 2)(x_1 - 1) = 0$ or $\mathbf{x} = \{[2, -3]^T, [1, -1]^T\}$

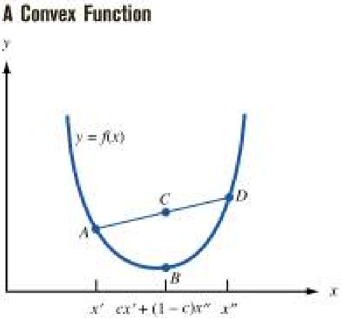
$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} 2x_1 + 1 & 2 \\ 2 & 1 \end{bmatrix}$$

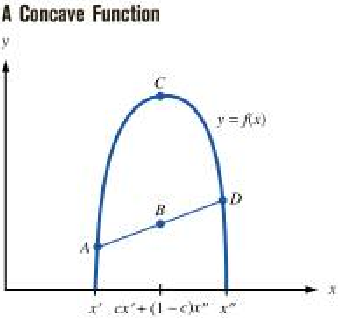
$\nabla^2 f(\mathbf{x_a}) = \begin{bmatrix} 5 & 2 \\ 2 & 1 \end{bmatrix}$ and $\nabla^2 f(\mathbf{x_b}) = \begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix}$

Optimal VS Definite

Nature of $\mathbf{x}^*$	Definiteness of H	$\mathbf{x}^T H \mathbf{x}$	λ_i	Illustration
Minimum	positive definite	> 0	> 0	
Valley	positive semi-definite	≥ 0	≥ 0	
Saddle point	indefinite	$\neq 0$	$\neq 0$	
Ridge	negative semi-definite	≤ 0	≤ 0	
Maximum	negative definite	< 0	< 0	

Convex Function & Concave Function





Source. Winston Section 11.3 pp 42 [Win22]

Verify Convex & Concave Functions

- $f_1(x) = \ln(x)$, where $\mathcal{S} \in \ (0, \infty)$

  concave function $f_1''(x) = -x^{-2}$

- $f_2(x_1, x_2) = x_1^3 + 3x_1x_2 + x_2^2$, where $\mathcal{S} \in \ \mathbb{R}^2$

  either convex nor concave

$$f_2''(x_1, x_2) \ = \ \begin{bmatrix} 6x_1 & 3 \\ 3 & 2 \end{bmatrix}$$

- $f_3(x_1, x_2) = x_1^2 + x_2^2$, where $\mathcal{S} \in \ \mathbb{R}^2$

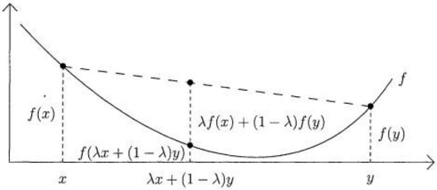
  convex function

$$f_3''(x_1, x_2) \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

- $f_4(x_1, x_2, x_3) = -x_1^2 - x_2^2 - 2x_3^2 + \frac{1}{2}x_1x_2$, where $\mathcal{S} \in \ \mathbb{R}^3$

  concave function because eigenvalue $= [-1.5, -2.5, -4]$

Convex Function



Source. Chong & Zak. 2001 pp 42 [CŻ13]

Definition (Convex Function)

A function $f(\cdot)$ is *convex* on a convex set $\mathcal{S}$ if it satisfies

$$f(\lambda\,\mathbf{u} + (1-\lambda)\mathbf{v}) \le \lambda\, f(\mathbf{u}) + (1-\lambda)\, f(\mathbf{v})$$

- **well-known:** $e^{a\,x}$ for $a \in \mathbb{R}$; x^a for $a \ge 1$; sum of convex function

Verify First & Second Order Conditions

- $f_1(x_1, x_2) = x_1^2 + e^{x_2} - 3x_1\, x_2$

$$\nabla f_1(\mathbf{x}) = [2x_1 - 3x_2, e^{x_2} - 3x_1]^T$$

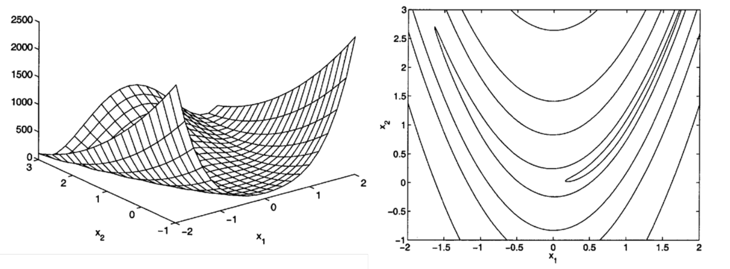
$$\nabla^2 f_1(\mathbf{x}) = \begin{bmatrix} 2 & -3 \\ -3 & e^{x_2} \end{bmatrix}$$

- $f_2(x_1, x_2) = (x_1 + x_2)\, e^{-(x_1 + x_2)}$

$$\nabla f_2(\mathbf{x}) = \begin{bmatrix} -(x_1 + x_2)\, e^{-(x_1+x_2)} + e^{-(x_1+x_2)} \\ -(x_1 + x_2)\, e^{-(x_1+x_2)} + e^{-(x_1+x_2)} \end{bmatrix}$$

$$\nabla^2 f_2(\mathbf{x}) = \begin{bmatrix} \quad \end{bmatrix}$$

Level Sets



$f(\mathrm{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$ or Rosenbrock's function

Source. Chong & Zak. 2001 pp 67 [CŻ13]

Definition (Level Set)

The level set is a function $f : \mathbb{R}^n \to \mathbb{R}$ at level c is the set of points

$$\mathcal{S} = \{\mathbf{x} : f(\mathbf{x}) = c\}$$

Reference

[CŻ13] E. Chong and S. Żak.
An introduction to optimization.
John Wiley & Sons, 2013.

[NC20] Fred Nwanganga and Mike Chapple.
Practical machine learning in R.
John Wiley & Sons, 2020.

[Pat14] Manas A Pathak.
Beginning data science with R.
Springer, 2014.

[Win22] Wayne L Winston.
Operations research: applications and algorithms.
Cengage Learning, 2022.

Special Matrix

Unit Vector A unit vector, $\mathbf{e}_j$, is a vector where the 1 appears in the jth position and 0's elsewhere.

Diagonal Matrix A square matrix whose off-diagonal elements are all equal to zero. For example:

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

Identity Matrix A diagonal matrix in which all diagonal elements are equal to 1. An identity matrix of order m is designated as either $\mathbf{I}_m$ or just $\mathbf{I}$. For example:

$$\mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Orthogonal Matrix A matrix in which all unit column vectors are perpendicular to one others

source: [CŻ13]

Special Matrix

Null or Zero Matrix A null matrix has all of its elements equal to zero. For example: $\mathbf{0} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

Symmetric Matrix A symmetric matrix is one whose transpose and the matrix itself are equal. That is $\mathbf{A} = \mathbf{A}^T$. For example:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 6 & 4 \\ 3 & 4 & 9 \end{bmatrix} = \mathbf{A}^T$$

Augmented Matrix An augmented matrix is one in which rows or columns of another matrix, of appropriate order, are appended to the original matrix. For example:

$$\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 5 & 6 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \quad \mathbf{A}|\mathbf{b} = \left[\begin{array}{cc|c} 1 & 4 & 3 \\ 5 & 6 & 1 \end{array}\right]$$

source: [CŻ13]

Special Matrix

Lower/Upper Triangular A special sparse square matrix. For example:

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 \\ 4 & -3 & 0 \\ 1 & 1 & 3 \end{bmatrix} \quad \mathbf{U} = \begin{bmatrix} 2 & 1 & 4 \\ 0 & -3 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

Null Space (matrix) The set vectors that orthogonal to rows of a given matrix. $\mathcal{N}(\mathbf{A}) = \{\mathbf{p} \in \mathbb{R}^n : \mathbf{Ap} = \mathbf{0}\}$ For example:

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \quad \mathbf{Ap} = \mathbf{0}, \qquad \mathbf{p} = \begin{bmatrix} v_1 \\ v_1 \\ v_2 \\ -v_2 \end{bmatrix}$$

source: [CŻ13]

Diagonal Form of Matrix

Definition

Diagonal Form of Matrix If matrix $\mathbf{A}$ has n linearly independent eigenvectors, then it can be diagonalize in $\mathbf{S}\boldsymbol{\Lambda}\mathbf{S}^{-1}$ where $\mathbf{S}^{-1}$ is eigenvector matrix and $\boldsymbol{\Lambda}$ is eigenvalues matrix.

Example: Each year $\frac{1}{10}$ of the people outside BKK move in, and $\frac{2}{10}$ of the people inside BKK move out. What is the population at year k.

$$\begin{bmatrix} y_{n+1} \\ z_{n+1} \end{bmatrix} = \begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix} \begin{bmatrix} y_n \\ z_n \end{bmatrix}$$

$$\begin{aligned} \begin{bmatrix} y_k \\ z_k \end{bmatrix} &= \begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix}^k \begin{bmatrix} y_0 \\ z_0 \end{bmatrix} = \left(\mathbf{S}^{-1}\boldsymbol{\Lambda}\mathbf{S}\right)^k \begin{bmatrix} y_0 \\ z_0 \end{bmatrix} \\ &= \begin{bmatrix} 0.894 & -0.707 \\ -0.707 & 0.707 \end{bmatrix} \begin{bmatrix} 1.0^k & \\ & 0.7^k \end{bmatrix} \begin{bmatrix} 0.745 & 0.745 \\ -0.471 & 0.942 \end{bmatrix} \begin{bmatrix} y_0 \\ z_0 \end{bmatrix} \end{aligned}$$

Matrix Decomposition

- **What:** processing original matrix $\mathbf{A}$ into multiplication of matrix with special properties
- **Popular Decomposition:**
  - LU Decomposition for compute determinant and find inverse

$$\mathbf{A} = \mathbf{L}\,\mathbf{U} \qquad \begin{bmatrix} 4 & 3 \\ 6 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1.5 & 1 \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 0 & -1.5 \end{bmatrix}$$

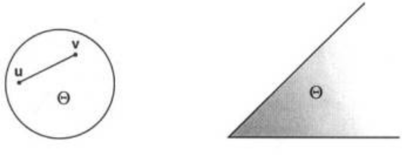
  - QR Decomposition for solve SLE and linear regression

$$\mathbf{A} = \mathbf{Q}\,\mathbf{R} \qquad \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & -2 \end{bmatrix} = \begin{bmatrix} -\sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} \\ -\sqrt{\frac{1}{2}} & +\sqrt{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} -\sqrt{2} & -\sqrt{\frac{1}{2}} & \sqrt{2} \\ 0 & \sqrt{\frac{1}{2}} & -\sqrt{2} \end{bmatrix}$$

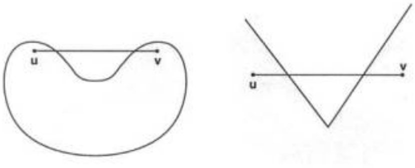
  - Single Value Decomposition for image transformataion

$$\mathbf{A} = \mathbf{U}\,\mathbf{\Lambda}\,\mathbf{V}^T \qquad \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} \sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} \\ \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} \sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} \\ \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} \end{bmatrix}$$

Convex Sets



(a) convex set



(b) non-convex set

Source. Chong & Zak. 2001 pp 42 [CŻ13]

Definition (Convex Set)

Set is *convex set* if, for any elements $\mathbf{u},\ \mathbf{v} \in \mathcal{S}\ \lambda\mathbf{u} + (1-\lambda)\mathbf{v} \in \mathcal{S}\ \lambda \in [0,1]$

Example of Convex Sets

- **Definition:** empty set, line segment, hyperplane, $\mathbb{R}^n$
- **Properties:**
  - SCALING: If Θ is convex set and β is a real number, then $\beta\Theta$ is convex set
  - INTERSECTION If Θ_1 and Θ_2 are convex sets, then $\Theta_1 \bigcap \Theta_2$ is also convex
  - ADDITIVE: If Θ_1 and Θ_2 are convex sets, then

$$\Theta_1 + \Theta_2 = \{\mathbf{x} : \mathbf{x} = \mathbf{v_1} + \mathbf{v_2}, \mathbf{v_1} \in \Theta_1, \mathbf{v_2} \in \Theta_2\}$$

is also convex

Results of Convex Function

Properties

- **Invert:** convex function = - concave function
- **Preserve:**
  - summation of convex functions is convex function
  - maximum value of convex functions is convex function
- **Trivial:** linear function both convex and concave

Determine convex function

- **In 1D:** $f^{''}(x) > 0$ for all $x \in \mathcal{S}$
- **In general:** $\nabla^2 f(\mathbf{x}) > \mathbf{0}$ for all $\mathbf{x} \in \mathcal{S}$

  Hessian matrix must be positive-definite for all $\mathbf{x} \in \mathcal{S}$
  - **Definition:** $\mathbf{x}^{\mathrm{T}}\mathbf{Q}\,\mathbf{x} > 0$ for all $\mathbf{x}$, where $\mathbf{Q}$ is symmetric matrix
  - **Sylvester'Criteria:** all leading principle minors of $\mathbf{Q}$ are positive $\rightarrow$ all eigenvalues are positive

SUFFICIENT CONDITION

- **What:** Is this point local optima?
- **First Order:** $\nabla f(\mathbf{x}_*) = 0$
- **Second Order:** $\nabla^2 f(\mathbf{x}_*)$ is positive definite (for minimization)

EXAMPLE: Check Optimality Conditions:

- $f(x_1, x_2) = \frac{1}{3}x_1^3 + \frac{1}{2}x_1^2 + 2x_1 x_2 + \frac{1}{2}x_2^2 - x_2 + 9$

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

$(x_1 - 2)(x_1 - 1) = 0$ or $\mathbf{x} = \{[2, -3]^T, [1, -1]^T\}$

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} 2x_1 + 1 & 2 \\ 2 & 1 \end{bmatrix}$$

$\nabla^2 f(\mathbf{x_a}) = \begin{bmatrix} 5 & 2 \\ 2 & 1 \end{bmatrix}$ and $\nabla^2 f(\mathbf{x_b}) = \begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix}$

Verify First & Second Order Conditions

- $f_1(x_1, x_2) = x_1^2 + e^{x_2} - 3x_1 x_2$

$$\nabla f_1(\mathbf{x}) = [2x_1 - 3x_2, e^{x_2} - 3x_1]^T$$

$$\nabla^2 f_1(\mathbf{x}) = \begin{bmatrix} 2 & -3 \\ -3 & e^{x_2} \end{bmatrix}$$

```
expr  <- expression(x1^2+exp(x2)-3*x1*x2)
exprD1 <- expression(NA,NA)
exprD1[[1]] <- D(expr,"x1")
exprD1[[2]] <- D(expr,"x2")


exprD2 <- expression(NA,NA,NA,NA)
exprD2[[1]] <- D(D(expr,"x1"),"x1")
exprD2[[2]] <- D(D(expr,"x2"),"x1")
exprD2[[3]] <- D(D(expr,"x1"),"x2")
exprD2[[4]] <- D(D(expr,"x2"),"x2")

expr.all <- function(x1,x2){ }
body(expr.all) <- deriv3(expr,c("x1","x2"))
```

Rate of Convergence

- **What:** Does algorithm converge? If so, How fast?
- **Define:** $e_k \equiv \; x_k - x_*$
- **Converge:** $\lim_{k\to\infty} \frac{\|e_{k+1}\|}{\|e_k\|^r} = C$

EXAMPLE: Find the converge rate of these sequence:

- **Sq#1** 2, 1.1, 1.01, 1.001, 1.0001, ... , $1 + 10^{-k}$

  $x_* = 1$ and $e_k = x_k - x_* = 10^{-k}$

  $\lim_{k\to\infty} \frac{10^{-k-1}}{10^{-k}} = \frac{1}{10}$

- **Sq#2** 4, 2.5, 2.05, 2.00060975, ... , $x_{k+1} = \frac{x_k}{2} + \frac{2}{x_k}$

  if $x_0 = 4$ then $x_* = 2$ and $e_{k+1} = \frac{x_k}{2} + \frac{2}{x_k} - 2 = \frac{1}{2x_k} e_k^2$

  $\lim_{k\to\infty} \frac{\frac{1}{2x_k} e_k^2}{\|e_k\|^2} = \frac{1}{4}$

Guaranteeing Convergence

- **Line Search:** right directions & many steps

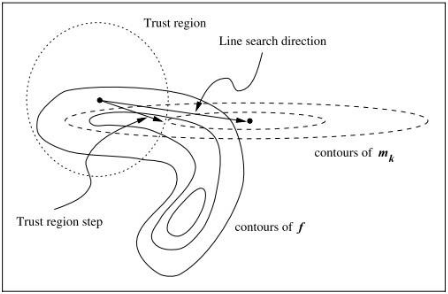
  $\mathbf{x_{k+1}} = \mathbf{x_k} + \gamma_k \, \mathbf{d_k}$ so that $f(\mathbf{x_k} + \gamma_k \, \mathbf{d_k}) < f(\mathbf{x_k})$

  In general: $f(\mathbf{x_0}) > \ldots > f(\mathbf{x_k}) > \ldots > f(\mathbf{x_*})$

  In other words: $d_k^T \nabla f(\mathbf{x}) < 0, \quad \forall \ k$

  In addition: $\gamma_k > 0, \quad \forall \ k$ and $\sum_{k=1}^{\infty} \gamma_k = \infty$

- **Trust Region:** adjustable steps depending on results



Source. Nocedal & Wright. 1999 pp 66