CO 367

272284444

September 2023

Contents

1 Introduction

1.1 Lecture 1-Preliminaries

Definition 1 (Quadratic Form)**.** Let A be a symmetric matrix and $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$. The **quadratic form** Q of the matrix A is defined as

$$Q = x^T A x$$

Example 1. Consider the matrix $A = \begin{pmatrix} 5 & -5 \\ -5 & 1 \end{pmatrix}$. The quadratic form of A is

$$Q(x) = 5x_1^2 - 10x_1x_2 + x_2^2$$

Definition 2 (Classification of Quadratic Forms)**.** Let Q be a quadratic form of a matrix A. Then Q is

1. positive definite if $Q(x) > 0$ for all non-zero vectors x, and $Q(x) = 0$ if and only if $x = 0$. Or all eigenvalues of A are positive.
2. positive semidefinite if $Q(x) \geqslant 0$ for all vectors x, with $Q(x) = 0$ occurring for some non-zero vectors x. Or all eigenvalues of A are non-negative.
3. negative definite if $Q(x) < 0$ for all non-zero vectors x, and $Q(x) = 0$ if and only if $x = 0$. Or all eigenvalues of A are negative.
4. negative semidefinite if $Q(x) \leqslant 0$ for all vectors x, with $Q(x) = 0$ occuring for some non-zero vectors x. Or all eigenvalues of A are non-negative.
5. indefinite if $Q(x)$ can be positive or negative. Or there are positive and negative eigenvalues for A.

Definition 3 (Big O and little o)**.** Big O is basically the rate of growth of that function. A function $f(n)$ is of order 1, or $O(1)$ if there exists some non zero constant c such that

$$\frac{f(n)}{c} \to 1$$

as $n \to \infty$.

Little o is the upper bound of the rate of growth of that function. Therefore, a function $f(n)$ is of order 1, or $o(1)$ if for all constants $c > 0$,

$$\frac{f(n)}{c} \to 0$$

as $n \to \infty$.

Definition 4 (Differentiability Based on Big o and Little o)**.** If f is differentiable at $x = a$, then

$$f(a+h) = f(a) + f'(a)h + o(h)$$

Conversely, if there exists constants A and B such that

$$f(a+h) = A + Bh + o(h)$$

then f is differentiable at $x = a$. Moreover, $A = f(a)$ and $B = f'(a)$.

Definition 5 (Product Rule)**.** If f, g are differentiable at $x = a$, then

$$f(a+h) = f(a) + f'(a)h + o(h), \quad g(a+h) = g(a) + g'(a)h + o(h)$$

Then

$$\begin{aligned} p(a+h) &= f(a+h)g(a+h) \\ &= f(a)g(a) + [f(a)g'(a) + g(a)f(a)]h + o(h) \end{aligned}$$

Then by above theorem, $p = fg$ is differentiable at $x = a$, and $p'(a) = f(a)g'(a) + g(a)f'(a)$.

Definition 6 (Chain Rule)**.** WIP

Definition 7 (Inner Product Space)**.** Let $x \in \mathbb{R}^n$, represented as:

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

The inner product space is defined as:

$$\langle x, y \rangle = x^T y = \sum_{i=1}^{n} x_i y_i \quad \text{(dot product)}$$

The angle between vectors x and y is given by $\cos(\theta) = \frac{\langle x,y \rangle}{\|x\|}$.

With corresponding norm to be the Euclidean Norm

Definition 8 (Open ball)**.** Given $\delta > 0$, $\bar{x} \in \mathbb{R}^n$, the open ball $B_\delta(\bar{x}) = \{x \in \mathbb{R}^n \; | \|x - \bar{x}\| < \delta\}$

Definition 9 (map)**.** Suppose the map $f : \mathbb{R}^n - > \mathbb{R}$.

Definition 10 (open set)**.** Let D $\subset \mathbb{R}^n$, D open set. $\forall x \in D$, $\exists \delta > 0$, s.t $B_\delta(x) \subset D$

Definition 11 (differ)**.** We define f to be in C^1, C^2 on an open set $D \subseteq \mathbb{R}^n$, denoted $f \in C^1(D), C^2(D)$, respectively, if the partial first $\frac{\partial f(x)}{\partial x_i}$ and second $\frac{\partial^2 f(x)}{\partial x_i \partial x_j}$ derivatives exist and are continuous for all i, j, respectively. We then get the gradient vector in $\mathbb{R}^n$ and the $n \times n$ symmetric Hessian matrix, respectively denoted as:

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_i} \right) \in \mathbb{R}^n, \quad \nabla^2 f(x) = \left[\frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right] \in \mathbb{S}^n.$$

Here, $\mathbb{S}^n$ is the vector space of $n \times n$ symmetric matrices.

Definition 12 (General Nonlinear opt. function NLO)**.** The general problem of nonlinear optimization, denoted NLO, is defined as follows: Given C^2-smooth functions $f, g_i, h_j : D \subseteq \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \ldots, m$ and $j = 1, \ldots, p$, where D is an open subset of $\mathbb{R}^n$, the objective is to find the optimal value p^* and an optimum x^* of NLO, represented as:

$$p^* := \min f(x) \text{s.t.} \quad g_i(x) \leqslant 0, \quad \forall i = 1, \ldots, m h_j(x) = 0, \quad \forall j = 1, \ldots, p x \in D$$

If f, g_i, h_i are all **affine** function and D=R^2, then we have an LP

Definition 13 (affine).

$$f(x) = Ax + b \tag{1}$$

where b≠0

Definition 14 (Types of Minimality). Consider $f : \mathbb{R}^n \to \mathbb{R}$ and $D \subset \mathbb{R}^n$. Then $\bar{x} \in D$ is:

- a *global minimizer* for f on D if $f(\bar{x}) \leqslant f(x)$ for all $x \in D$.
- a *strict global minimizer* for f on D if $f(\bar{x}) < f(x)$ for all $x \in D$ where $x \neq \bar{x}$.
- a *local minimizer* for f on D if there exists $\delta > 0$ such that $f(\bar{x}) \leqslant f(x)$ for all $x \in D \cap B_\delta(\bar{x})$.
- a *strict local minimizer* for f on D if there exists $\delta > 0$ such that $f(\bar{x}) < f(x)$ for all $x \in D \cap B_\delta(\bar{x})$ where $x \neq \bar{x}$.

Definition 15 (Linear Approximation). Suppose f is a function that is differentiable on an interval I containing the point a. The **linear approximation** to f at a is the linear function

$$L(x) = f(a) + f'(a)(x - a)$$

for $x \in I$.

Definition 16 (Quadratic Approximation). Similar as above, the **quadratic approximation** to f at a is the quadratic function

$$Q(x) = f(a) + f'(a)(x - a) + \frac{1}{2}f''(a)(x - a)^2$$

for $x \in I$.

Definition 17 (Formal Definition of Derivative). The **derivative** of f at a is defined as

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

if the limit exists.

An alternate definition is

$$f'(a) = \lim_{h \to 0} \frac{f(a + h) - f(a)}{h}$$

1.2 Lecture 2

Definition 18 (General NLO/NLP). A **Non-linear Optimization Problem** (NLP) is of the following form:

$$\underbrace{p^*}_{\text{Optimal Value}} = \min \underbrace{f(x)}_{\text{Objective function}}$$

s.t.

$$g(x) = (g_i(x)) \leqslant 0 \in \mathbb{R}^m$$
$$h(x) = (h_j(x)) = 0 \in \mathbb{R}^p$$

Example 2.

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

s.t.

$$x_1^2 - x_2 \leqslant 0 \qquad (g_1(x) \leqslant 0)$$
$$x_1 + x_2 - 2 \leqslant 0 \qquad (g_2(x) \leqslant 0)$$

Definition 19 (Contour)**.** For $\alpha \in \mathbb{R}$, the **contour** of a function f is

$$C_\alpha = \{x \in \mathbb{R}^n : f(x) = \alpha\}$$

Definition 20 (Feasible Set)**.** The **feasible set** is

$$F = \{x \in \mathbb{R}^n : g(x) \leqslant 0, h(x) = 0, x \in D\}$$

(Is D the domain??)

Definition 21 (Gradient)**.** The **gradient** of f is

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

For the optimal solution x^*, we have

$$\alpha \nabla f(x^*) = \lambda_1 \nabla g_1(x^*) + \lambda_2 \nabla g_2(x^*)$$

for some $\alpha, \lambda_1, \lambda_2 \in \mathbb{R}$.

We will see later that we can choose $\alpha = 1$ and we need $\lambda_1 \geqslant 0, \lambda_2 \geqslant 0$.

Example 3 (Max-cut Problem)**.** Given a weighted graph $G = (\underbrace{V}_{\text{vertices}}, \underbrace{E}_{\text{edges}}, \underbrace{w}_{\text{weight}})$, a **cut** is $U \subseteq V, U \neq \varnothing$.

The objective function

$$\max \quad \frac{1}{2} \sum_{\substack{i \in U, j \notin U \\ (i,j) \in E}} w_{i,j}$$

maximizes the sum of edges in a cut.

Formulating as an NLP, we introduce variables $x_i \in \{\pm 1\}, i = 1, \ldots, n$. Then the Max-cut problem (MC) is as follows:

$$\max \quad \frac{1}{2} \sum_{ij \in E} w_{ij}(1 - x_i x_j)$$

Why 1/2 s.t.

$$x_i \in \{\pm 1\} \quad (\text{equivalent to } x_i^2 = 1) \quad \forall i = 1, \ldots, n$$

This works because

$$1 - x_i x_j = \begin{cases} 0 & \text{if } x_i = x_j \quad (i, j \text{ in the same set, } U \text{ or } U^c) \\ 2 & \text{otherwise} \end{cases}$$

MC is a **quadratically constrained quadratic program** (QOP) since each constraint $x_i \in \{-1, 1\}$ is equivalent to the quadratic constraint $x_i^2 = 1$. Note that MC is an NP-hard problem.

2 Unconstrained Optimization

2.1 Lecture 2

Example 4 (Simplest Case - No Constraints)**.** Let $\Omega \subseteq \mathbb{R}^n$ be an open set. Assume f is sufficiently smooth (differentiable) then the NLP with no constraints is

$$\min_{x \in \Omega} \quad f(x)$$

Theorem 1 (Taylor's Theorem on the real line)**.** Let $f : (a, b) \to \mathbb{R}$, and $\bar{x}, x \in (a, b)$, then there exists z strictly between $x, \bar{x}$ such that

$$f(x) = f(\bar{x}) + f'(\bar{x})(x - \bar{x}) + \frac{f''(z)}{2}(x - \bar{x})^2$$

or equivalently

$$f(\bar{x} + \delta x) = \underbrace{f(x) + f'(x)\delta x}_{\text{Linear approximation}} + o(|\delta x|)(\text{little O})$$

Lemma 1 (Directional Derivative)**.** Let $f : \mathbb{R}^n \to \mathbb{R}, \bar{x}, d \in \mathbb{R}^n$ where d is the direction. We define

$$\phi(\epsilon) = f(\bar{x} + \epsilon d) : \mathbb{R} \to \mathbb{R}$$

Then the **directional derivative**, denoted $f'(x; d)$ of f at x at the direction d is

$$f'(x; d) = \phi'(0) = \lim_{\epsilon \to 0} \frac{f(x + \epsilon d) - f(x)}{\epsilon} = \nabla f(x)^T d$$

Example 5. Let $f(x, y, z) = x^2 z + y^3 z^2 - xyz$ with $d = \begin{pmatrix} -1 \\ 0 \\ 3 \end{pmatrix}$ Then the **directional derivative** in the direction d is

$$\nabla f(x, y, z)^T d = \begin{pmatrix} 2xz - yz \\ 3y^2 z^2 - xz \\ x^2 + 2y^3 z - xy \end{pmatrix}^T \begin{pmatrix} -1 \\ 0 \\ 3 \end{pmatrix} = -2xz + yz + 3x^2 + 6y^3 z - 3xy$$

Corollary 1. Let $f : (a, b) \to \mathbb{R}$

1. If $\bar{x}$ is a **local minimizer** of f on (a, b), then $f'(\bar{x}) = 0$ and $f''(\bar{x}) \geqslant 0$.
2. If $f(\bar{x}) = 0, f''(\bar{x}) > 0$ then $\bar{x}$ is a **strict local minimizer** of f.

Definition 22 (Hessian)**.** The **Hessian** of f at $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ is the matrix

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

2.2 Lecture 3

Definition 23 (Matrix Norm)**.**

$$\|Q\| = max_{\|x\|=1} \|Qx\| = \text{Largest singular value of A}$$

Definition 24. Define f,D,$\bar{x}$, D is an open set, suppose $\bar{x}$ is a local minmum for f on D, Then:

1. Nec: $\nabla f(\bar{x}) = 0, \nabla^2 f(\bar{x}) \geqslant 0$ positive semidefinite

2. Suff: $\nabla f(\bar{x}) = 0, \nabla^2 f(\bar{x}) > 0$ positive definite $\implies$ $\bar{x}$ is a strict local minmum on Define

Proof. 1. updated later after I confirmed some details with the professor ☐

Example 6 (Algorithm to find Min using Nec above)**.** Given f:$\mathbb{R} \implies \mathbb{R}$ and $f^{'}(\bar{x}) \neq 0$, then $x_{new} = \bar{x} - slop * f^{'}(\bar{x})$.

Example 7. Given f:$\mathbb{R}^n \implies \mathbb{R}, \phi(\epsilon) = f(\bar{x} + \epsilon d)$
using tylar expansion $f(\bar{x} + \epsilon d)$= $f(\bar{x})$+ϵ $\nabla f(\bar{x})^T d$ +$o(\|\epsilon\|)$ shouldnt be ϵd? or d is the unit vector

let $d = -\nabla f(\bar{x})$, f($\bar{x}$)-$\epsilon\|f(\bar{x})\|^2$+$o(\epsilon)$. $< f(x)$ (if $\nabla f(\bar{x}) \neq 0$)
i.e test nec condition
If $\nabla f(\bar{x}) \neq 0$, then $x_{new} = \bar{x} + \epsilon(-\nabla f(\bar{(x)}))$ Move to the deapest direction

Definition 25 (Cauchy's method of steepest descant)**.** https://www.math.usm.edu/lambers/mat419/lecture10.pdf
$x_0 \in \mathbb{R}^n$.

$$Is \nabla f(x_k) \approxeq 0?\text{IF yes Stop}$$

.

$$O.W, \text{ find a Stop} \alpha > 0$$

$$x_{k+1} = x_k - \alpha \nabla f(x_k)$$

$$repeat$$

Example 8. $f(x, y) = x^3 - 12xy + 8y^3$
$\nabla f(x, y) = \begin{pmatrix} 3x^2 - 12y \\ -12x + 24y^2 \end{pmatrix}$. $\nabla^2 f(x, y) = \begin{pmatrix} 6x & -12 \\ -12 & 48y \end{pmatrix}$.

$$\nabla f = 0 \implies (x, y) = (0, 0) or (2, 1)$$

$$\nabla^2 f(2, 1) = \begin{pmatrix} 12 & -2 \\ -12 & 48 \end{pmatrix}$$

All leading principal minors are ppsitive $\implies$ strict local minimum
while (0,0) are indefinite and this is called **inflection point**