CO 367

272284444

September 2023

Contents

1 Introduction

1.1 Lecture 1

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. https://math.stackexchange.com/questions/4061952/differentiability-using-little-oh-notation

1.2 Lecture 2

Definition 4 (General NLO/NLP)**.** A **Non-linear Optimization Problem** (NLP) is of the following form:

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

s.t.

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

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

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

Definition 6 (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 7 (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:

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

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

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

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

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

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

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