Matrices in Quadratic Optimisation

Pan Jing Bin

Final Year Project Introductory Talk

Outline

Mathematical Optimization

A general mathematical optimization problem is of the form

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & g_i(\boldsymbol{x}) \leq a_i, \quad i = 1, \cdots, p, \\ & h_j(\boldsymbol{x}) = b_i, \quad j = 1, \cdots, q. \end{aligned}$$

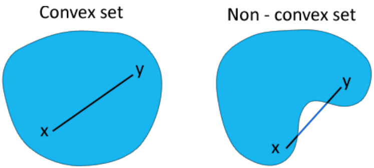
where f is the **objective function** and $g_1, \cdots, g_p, h_1, \cdots, h_q$ are the **constraint functions**. The set of points that satisfy the problem's constraints is called the **feasible set**.

Convex Optimization

A **convex optimization** is of the form

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & g_i(\boldsymbol{x}) \leq a_i, \quad i = 1, \cdots, p, \\ & h_j(\boldsymbol{x}) = b_i, \quad j = 1, \cdots, q. \end{aligned}$$

both the objective function f and the feasible set are **convex**.



Source: https://www.easycalculation.com/maths-dictionary/convex_set.html

Almost a mature technology

Quadratically Constrained Quadratic Program (QCQP)

In a **Quadratically Constrained Quadratic Program (QCQP)**, the objective and constraint functions are **quadratic functions**. It has the general form

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} \quad & \boldsymbol{x}^T \boldsymbol{C}\boldsymbol{x} + \boldsymbol{c}^T\boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{x}^T \boldsymbol{P}_i\boldsymbol{x} + \boldsymbol{p}_i^T\boldsymbol{x} \le a_i, \quad i = 1,\cdots,p, \\ & \boldsymbol{x}^T \boldsymbol{Q}_j\boldsymbol{x} + \boldsymbol{q}_j^T\boldsymbol{x} = b_i, \quad j = 1,\cdots,q. \end{aligned}$$

where $\boldsymbol{C}, \boldsymbol{P}_1, \cdots, \boldsymbol{P}_p, \boldsymbol{Q}_1, \cdots, \boldsymbol{Q}_q \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ and $\boldsymbol{c}, \boldsymbol{p}_1, \cdots, \boldsymbol{p}_p, \boldsymbol{q}_1, \cdots \boldsymbol{q}_q \in \mathbb{R}^n$.

It includes the following **integer programming** problem:

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} \quad & \boldsymbol{x}^T \boldsymbol{C}\boldsymbol{x} \\ \text{subject to} \quad & x_j^2 = 1, \quad j = 1\cdots, n \end{aligned}$$

where $C \in \mathbb{S}^n$ (symmetric $n \times n$ matrices over $\mathbb{R}$).

In general, QCQP problems are **hard**!

Semidefinite Relaxation

We study the problem

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} \quad & \boldsymbol{x}^T\boldsymbol{C}\boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{x}^T\boldsymbol{A}_i\boldsymbol{x} \leq a_i, \quad i = 1\cdots, k \end{aligned}$$

where $\boldsymbol{C}, \boldsymbol{A}_1, \cdots, \boldsymbol{A}_k \in \mathbb{S}^n$ (symmetric $n \times n$ symmetric matrices over $\mathbb{R}$).

Observe that

$$\boldsymbol{x}^T\boldsymbol{C}\boldsymbol{x} = \mathrm{Tr}(\boldsymbol{x}^T\boldsymbol{C}\boldsymbol{x}) = \mathrm{Tr}(\boldsymbol{C}\boldsymbol{x}\boldsymbol{x}^T)$$

where $\boldsymbol{X} = \boldsymbol{x}\boldsymbol{x}^T$ is a **rank one symmetric positive semidefinite (PSD)** matrix. In fact, all rank one symmetric PSD matrices is of the form $\boldsymbol{x}\boldsymbol{x}^T$ for some $\boldsymbol{x} \in \mathbb{R}^n \setminus \{0\}$. Similarly,

$$\boldsymbol{x}^T\boldsymbol{A}_i\boldsymbol{x} = \mathrm{Tr}(\boldsymbol{x}^T\boldsymbol{A}_i\boldsymbol{x}) = \mathrm{Tr}(\boldsymbol{A}_i\boldsymbol{x}\boldsymbol{x}^T)$$

Semidefinite Relaxation

We have reduced the original problem

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} \quad & \boldsymbol{x}^T\boldsymbol{C}\boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{x}^T\boldsymbol{A}_i\boldsymbol{x} \le a_i, \quad i = 1\cdots, k \end{aligned}$$

to

$$\begin{aligned} \min_{\boldsymbol{X}\in\mathbb{S}^n} \quad & \mathrm{Tr}(\boldsymbol{C}\boldsymbol{X}) \\ \text{subject to} \quad & \mathrm{Tr}(\boldsymbol{A}_i\boldsymbol{X}) \le a_i, \quad i = 1\cdots, k \\ & \boldsymbol{X} \succeq \boldsymbol{0}, \qquad \mathrm{rank}(\boldsymbol{X}) = 1 \end{aligned}$$

where $\boldsymbol{X} \succeq \boldsymbol{0}$ indicates that $\boldsymbol{X}$ is positive semidefinite.

Semidefinite Relaxation

We have reduced the original problem

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathbb{R}^n} \quad & \boldsymbol{x}^T \boldsymbol{C} \boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{x}^T \boldsymbol{A}_i \boldsymbol{x} \leq a_i, \quad i = 1 \cdots, k \end{aligned}$$

to

$$\begin{aligned} \min_{\boldsymbol{X} \in \mathbb{S}^n} \quad & \mathrm{Tr}(\boldsymbol{C}\boldsymbol{X}) \\ \text{subject to} \quad & \mathrm{Tr}(\boldsymbol{A}_i \boldsymbol{X}) \leq a_i, \quad i = 1 \cdots, k \\ & \boldsymbol{X} \succeq \boldsymbol{0}, \quad \cancel{\mathrm{rank}(\boldsymbol{X}) = 1} \end{aligned}$$

where $\boldsymbol{X} \succeq \boldsymbol{0}$ indicates that $\boldsymbol{X}$ is positive semidefinite.

This is now a **convex optimization** problem! Furthermore, there exist polynomial-time algorithms for problems of the above form.

Semidefinite Relaxation

Question: Given a solution $\boldsymbol{X}_0$ (with $\operatorname{rank}(\boldsymbol{X}_0) \neq 1$) to

$$\begin{array}{lll} \displaystyle\min_{\boldsymbol{X} \in \mathbb{S}^n} & \operatorname{Tr}(\boldsymbol{C}\boldsymbol{X}) & \\ \text{subject to} & \operatorname{Tr}(\boldsymbol{A}_i \boldsymbol{X}) \leq a_i, & i = 1 \cdots, k \\ & \boldsymbol{X} \succeq \boldsymbol{0}, & \cancel{\operatorname{rank}(\boldsymbol{X}) = 1}, \end{array}$$

how do we recover back the vector $\boldsymbol{x}$?

One heuristic: Use the **best rank one approximation** of $\boldsymbol{X}$. Write

$$\boldsymbol{X} = \sum_{i=1}^{r} \lambda_i \boldsymbol{q}_i \boldsymbol{q}_i^T$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$ and choose $\boldsymbol{x} = \sqrt{\lambda_1}\boldsymbol{q}_1$. However, this is in general not the optimal solution to the original problem.

Matrix Functions: Motivation

To understand semidefinite relaxation better, we first need a deeper understanding of **positive-semidefinite matrices** and **matrix functions**.

Given a function $f : \mathbb{R} \to \mathbb{R}$, how do we lift it to a matrix function $\widetilde{f} : \mathrm{Mat}_{n\times n}(\mathbb{R}) \to \mathrm{Mat}_{n\times n}(\mathbb{R})$ such that **interesting properties are retained**?

If f is **real-analytic**, then we may use its Taylor expansion:

$$\exp(\boldsymbol{A}) = \boldsymbol{I}_n + \boldsymbol{A} + \frac{1}{2!}\boldsymbol{A}^2 + \frac{1}{3!}\boldsymbol{A}^3 + \cdots$$

Other functions that arise frequently: $\log(x)$, $\dfrac{1}{x}$.

Matrix Function

Alternative way for **Hermitian matrices**:

For a diagonal matrix $\boldsymbol{D} = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$, define

$$f(\boldsymbol{D}) = \mathrm{diag}\big(f(\lambda_1), f(\lambda_2), \cdots, f(\lambda_n)\big).$$

This definition can be extended to all Hermitian matrices $\boldsymbol{A}$ by

$$f(\boldsymbol{A}) = \boldsymbol{U} f(\boldsymbol{D}) \boldsymbol{U}^*$$

where $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^*$ with $\boldsymbol{U}$ unitary and $\boldsymbol{D}$ diagonal.

This definition agrees with the power series definition! (when their domain of definition coincide)

Löewner Order

Define a partial order on the set of $n \times n$ matrices as follows:

$$\boldsymbol{A} \succeq \boldsymbol{B} \iff \boldsymbol{A} - \boldsymbol{B} \text{ is positive semi-definite.}$$

This allows us to generalise the notion of monotonicity and convexity to matrix functions:

1. A function $f : (a, b) \to \mathbb{R}$ is **matrix monotone** if for all $n \times n$ Hermitian matrices $\boldsymbol{A}$ and $\boldsymbol{B}$,
$$\boldsymbol{A} \succeq \boldsymbol{B} \implies f(\boldsymbol{A}) \succeq f(\boldsymbol{B}).$$
2. A function $f : (a, b) \to \mathbb{R}$ is **matrix convex** if for all $\lambda \in [0, 1]$ and for all $n \times n$ Hermitian matrices $\boldsymbol{A}$ and $\boldsymbol{B}$,
$$f(\lambda\boldsymbol{A} + (1 - \lambda)\boldsymbol{B}) \preceq \lambda f(\boldsymbol{A}) + (1 - \lambda)f(\boldsymbol{B}).$$

Relationship With Complex Analysis

Theorem (Löwener)

Let $f : (-1,1) \to \mathbb{R}$ be a non-constant matrix monotone function. Then there exists a unique probability measure μ on $[-1,1]$ such that

$$f(t) = f(0) + f'(0) \int_{-1}^{1} \frac{t}{1-\lambda t}\, d\mu(\lambda).$$

If t is replaced by a complex variable z, we have a holomorphic function

$$f(z) = f(0) + f'(0) \int_{-1}^{1} \frac{z}{1-\lambda z}\, d\mu(\lambda)$$

defined on $\mathbb{C} \setminus (-\infty,-1] \cup [1,\infty)$.

Relationship With Complex Analysis

$$f(z) = f(0) + f'(0) \int_{-1}^{1} \frac{z}{1 - \lambda z} \, d\mu(\lambda) \qquad (\star)$$

As a holomorphic function, f maps the upper-half plane $\mathbb{H} = \{z \in \mathbb{C} \ : \ \mathrm{Im}(z) > 0\}$ to itself since

$$\mathrm{Im}\left(\frac{z}{1 - \lambda z}\right) = \frac{z - \overline{z}}{2|1 - \lambda z|^2} = \frac{\mathrm{Im}(z)}{|1 - \lambda z|^2}.$$

Holomorphic functions with the above property are known as **Pick functions**. Let $P(a, b)$ be the class of functions that take on **purely real** values on the interval (a, b). The analytic continuation $(\star)$ induces a bijection

$$\left\{\begin{matrix}\text{Matrix monotone} \\ \text{functions on } (a, b)\end{matrix}\right\} \longleftrightarrow \left\{\begin{matrix}\text{Functions of} \\ \text{class } P(a, b)\end{matrix}\right\}.$$

The End

Thank you for your attention!