3.5 Quadratic Programming

Since the work in Frank and Wolfe (1956), Quadratic Programming aims to effectively solve for the numerical solution of a board class of quadratic optimization problems. Conventionally, the corresponding convex optimization problem, usually coined as a Quadratic Program (QP), is formulated as follows:

$$\underset{\boldsymbol{x}\in\mathbb{R}^D}{\text{minimize}} \quad \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x} + \boldsymbol{q}^\top \boldsymbol{x} + r\,, \tag{3.5.1}$$

$$\text{subject to} \quad \text{i)} \quad \boldsymbol{G}\boldsymbol{x} \le \boldsymbol{h}\,; \tag{3.5.2}$$

$$\text{ii)} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\,;$$

where $\boldsymbol{P} \in \mathbb{R}^{D\times D}$ is positive definite, $\boldsymbol{G} \in \mathbb{R}^{K\times D}$, and $\boldsymbol{A} \in \mathbb{R}^{M\times D}$, and the symbol $\le$ here means that every entry of vector $\boldsymbol{G}\boldsymbol{x}$ is less than or equal to the corresponding entry of the right-hand side vector $\boldsymbol{h} \in \mathbb{R}^K$. Moreover, $\boldsymbol{P} := (p_{ij})$ is usually assumed to be symmetric, since

$$\boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x} = (\boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x})^\top = \boldsymbol{x}^\top \boldsymbol{P}^\top \boldsymbol{x}\,,$$

is a scalar. Hence, we could always replace the asymmetric matrix $\boldsymbol{P}$ with $\tilde{\boldsymbol{P}} := (\boldsymbol{P} + \boldsymbol{P}^\top)/2$ as:

$$\boldsymbol{x}^\top \tilde{\boldsymbol{P}}\boldsymbol{x} := \frac{1}{2}\boldsymbol{x}^\top (\boldsymbol{P} + \boldsymbol{P}^\top)\boldsymbol{x} = \frac{1}{2}(\boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x} + \boldsymbol{x}^\top \boldsymbol{P}^\top \boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x}\,.$$

Particularly in a quadratic program, we minimize a convex quadratic function on a high dimensional polyhedron, *i.e.* the feasible set $\mathcal{P}$ as specified by the constrains (3.5.2), as depicted in Figure 3.5.1.

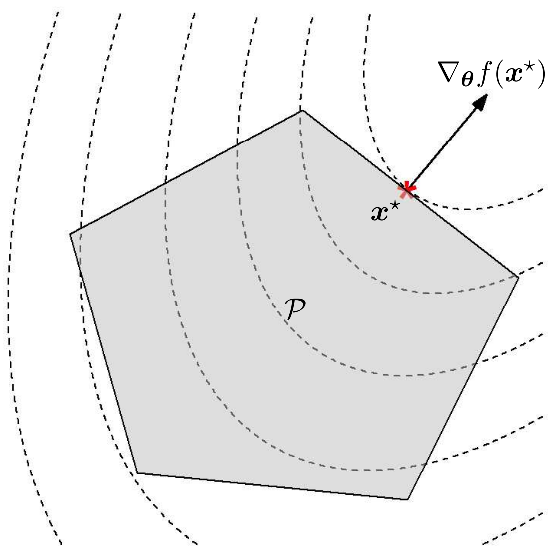

Figure 3.5.1: Geometric illustration of Quadratic Program

3.5.1 Inequality Constrainted Lagrangian Multipliers

We first state the celebrated **von Neumann's Minimax Theorem** as below:

Theorem 3.5.1. (**von Neumann's Minimax Theorem**) [4]: Let $\mathcal{X} \subset \mathbb{R}^N$ and $\mathcal{Y} \subset \mathbb{R}^M$ be compact convex sets: If $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is a continuous function that is concave-convex, *i.e.*

1. for each $y \in \mathcal{Y}$, $f(\cdot, y) : \mathcal{X} \to \mathbb{R}$ is concave in $x \in \mathcal{X}$; and
2. for each $x \in \mathcal{X}$, $f(x, \cdot) : \mathcal{Y} \to \mathbb{R}$ is convex in $y \in \mathcal{Y}$.

Then, we have

$$\max_{x \in \mathcal{X}} \min_{y \in \mathcal{Y}} f(x, y) = \min_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}} f(x, y) . \tag{3.5.3}$$

Remark 3.5.1. Certainly, the existance of a saddle point implies the validity of (3.5.3), as a saddle point (x_0, y_0) of f is such that

$$\min_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}} f(x, y) = \min_{y \in \mathcal{Y}} f(x_0, y) = f(x_0, y_0) = \max_{x \in \mathcal{X}} f(x, y_0) = \max_{x \in \mathcal{X}} \min_{y \in \mathcal{Y}} f(x, y) , \tag{3.5.4}$$

where we leftmost expression follows the red path in Figure 3.5.2; the rightmost term follows the blue path in the same figure. They both arrive at the same point (x_0, y_0). Nevertheless, in general, even if the saddle point of f does not exist, the Theorem remains valid.

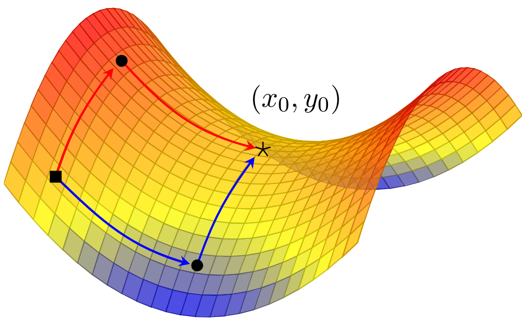

Figure 3.5.2: The two paths of the von Neumann's Minimax Theorem over a function f.

With this theorem in mind, we then consider a more general minimization problem:

$$\underset{\boldsymbol{x} \in \mathbb{R}^D}{\text{minimize}} \qquad f(\boldsymbol{x}) , \tag{3.5.5}$$

$$\text{subject to} \qquad g_k(\boldsymbol{x}) \leq 0 , \qquad k = 1, \cdots, K; \tag{3.5.6}$$

where $f, g_k : \mathbb{R}^D \to \mathbb{R}$, for $k = 1, \cdots, K$, and we denote $\boldsymbol{g}(\boldsymbol{x}) := (g_1(\boldsymbol{x}), \cdots, g_K(\boldsymbol{x}))^\top$.

By Lagrangian multiplier approach, the lagrangian with a multiplier $\boldsymbol{\lambda}$ is given by:

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) := f(\boldsymbol{x}) + \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{x}) . \tag{3.5.7}$$

Define $\mathcal{X} := \{\boldsymbol{x} \in \mathbb{R}^D : g(\boldsymbol{x}) \leq \mathbf{0}\}$. $\boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0}$, for every feasible $\boldsymbol{x} \in \mathcal{X}$, the lagrangian $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ is a decreasing function in λ_k's. Almost by definition, it is easy to see the equality that:

$$\min_{\boldsymbol{x} \in \mathcal{X}} f(\boldsymbol{x}) = \min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{\lambda} \geq \mathbf{0}} \left(f(\boldsymbol{x}) + \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{x}) \right) = \min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{\lambda} \geq \mathbf{0}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) , \tag{3.5.8}$$

which is usually called the **primal problem**. We first note that $\boldsymbol{\lambda}$ lies in the unbounded domain, hence not a compact convex set, we therefore cannot directly apply Theorem 3.5.1 to (3.5.8) immediately. Under some mild regularity condition on f, we can regard $\boldsymbol{\lambda}_{\boldsymbol{x}} : \mathcal{X} \to [0, \infty)^K$ as a continuous function in $\boldsymbol{x} \in \mathcal{X}$ such that the image $\boldsymbol{\lambda}_{\boldsymbol{x}}(\mathcal{X})$ can then be a closed and bounded, and so compact, set, let say now $\sup_{\boldsymbol{x}\in\mathcal{X}}|\boldsymbol{\lambda}_{\boldsymbol{x}}| \leq \boldsymbol{M} < \infty$, with this in mind, we can therefore apply Theorem 3.5.1 such that:

$$\min_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x}) = \max_{\boldsymbol{0}\leq\boldsymbol{\lambda}\leq\boldsymbol{M}} \min_{\boldsymbol{x}\in\mathcal{X}} \mathcal{L}(\boldsymbol{x},\boldsymbol{\lambda})\,.$$

Furthermore, if one can find a function $\ell(\boldsymbol{\lambda})$ such that $\min_{\boldsymbol{x}\in\mathcal{X}} \mathcal{L}(\boldsymbol{x},\boldsymbol{\lambda}) \geq \ell(\boldsymbol{\lambda})$ and $\ell(\boldsymbol{\lambda})$ is coercive in $\boldsymbol{\lambda}$, *i.e.* $\ell(\boldsymbol{\lambda}) \uparrow \infty$ as $\boldsymbol{\lambda} \uparrow \infty$, then we finally deduce that

$$\min_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x}) = \max_{\boldsymbol{\lambda}\geq\boldsymbol{0}} \min_{\boldsymbol{x}\in\mathcal{X}} \mathcal{L}(\boldsymbol{x},\boldsymbol{\lambda})\,.$$

Remark 3.5.2. However, if the regularity conditions are not met or in the absence of suitable concavity and convexity, we still have the **max–min inequality** [5]:

$$\max_{\boldsymbol{y}\in\mathcal{Y}} \min_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x},\boldsymbol{y}) \leq \min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathcal{Y}} f(\boldsymbol{x},\boldsymbol{y})\,, \tag{3.5.9}$$

where $f : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}$; indeed, to see this, for all $\boldsymbol{x}\in\mathcal{X}$ and $\boldsymbol{y}\in\mathcal{Y}$, we clearly have $\min_{\boldsymbol{x}'\in\mathcal{X}} f(\boldsymbol{x}',\boldsymbol{y}) \leq f(\boldsymbol{x},\boldsymbol{y})$. Taking maximum with respect to $\boldsymbol{y}$ on both sides yields:

$$\max_{\boldsymbol{y}\in\mathcal{Y}} \min_{\boldsymbol{x}'\in\mathcal{X}} f(\boldsymbol{x}',\boldsymbol{y}) \leq \max_{\boldsymbol{y}\in\mathcal{Y}} f(\boldsymbol{x},\boldsymbol{y})\,, \qquad \forall\,\boldsymbol{x}\in\mathcal{X}\,.$$

Since the inequality above holds for any $\boldsymbol{x}\in\mathcal{X}$, (3.5.8) can be obtained by considering only the minimum of the right-hand side. Finally, with the **max–min inequality** in (3.5.9), we can re-write the primal problem of (3.5.8) into a **dual problem**:

$$\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{\lambda}\geq\boldsymbol{0}} \mathcal{L}(\boldsymbol{x},\boldsymbol{\lambda}) \geq \max_{\boldsymbol{\lambda}\geq\boldsymbol{0}} \min_{\boldsymbol{x}\in\mathcal{X}} \mathcal{L}(\boldsymbol{x},\boldsymbol{\lambda}) =: \max_{\boldsymbol{\lambda}\geq\boldsymbol{0}} \mathcal{D}(\boldsymbol{\lambda})\,. \tag{3.5.10}$$

This result warrants the duality nature of primal and dual problems, such that the solution to the dual (maximization) problem provides a lower bound to the solution of the primal (minimization) problem, and the difference between the optimal values of primal and dual is called the duality gap; see the illustration in Figure 3.5.3, which can be referred to the Fenchel-Rockafellar duality (see Theorem 1.9 in Villani (2021)) of convex optimization problems, and this duality gap can be removed under a convex constraints as specified in the Minimax Theorem 3.5.1.

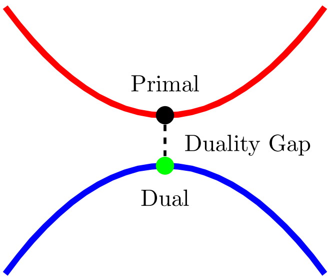

Figure 3.5.3: Duality Gap

3.5.2 Lagrange Multiplier in Quadratic Programming

Next, we shall reformulate the problem in (3.5.1) with constraint in (3.5.2) by using the Lagrangian multiplier approach, and the Lagrangian $\mathcal{L}$ with $\boldsymbol{\lambda}_1 \in \mathbb{R}^K, \boldsymbol{\lambda}_2 \in \mathbb{R}^M$ is given by:

$$\begin{aligned}\mathcal{L}(\boldsymbol{x},\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2) &= \frac{1}{2}\boldsymbol{x}^\top\boldsymbol{P}\boldsymbol{x} + \boldsymbol{q}^\top\boldsymbol{x} + r + \boldsymbol{\lambda}_1^\top(\boldsymbol{G}\boldsymbol{x}-\boldsymbol{h}) + \boldsymbol{\lambda}_2^\top(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})\\ &= \frac{1}{2}\boldsymbol{x}^\top\boldsymbol{P}\boldsymbol{x} + (\boldsymbol{c}+\boldsymbol{G}^\top\boldsymbol{\lambda}_1+\boldsymbol{A}^\top\boldsymbol{\lambda}_2)^\top\boldsymbol{x} - \boldsymbol{\lambda}_1^\top\boldsymbol{h} - \boldsymbol{\lambda}_2^\top\boldsymbol{b} + r\,,\end{aligned} \tag{3.5.11}$$