Contents

1

Non-linear Programming Problems

1.1 Basic Terminology and Notations

> **Definition 1.1.1 ▶ General Non-linear Programming (NLP) Problems**
>
> Define the function $f : \mathbf{R}^n \to \mathbf{R}$. Let $\boldsymbol{x} \in \mathbf{R}^n$ be a vector, then a general NLP problem aims to **optimise** (i.e. maximise or minimise) $f(\boldsymbol{x})$ subject to the constraint $\boldsymbol{x} \in S \subseteq \mathbf{R}^n$, where
> - f is known as the **objective function**;
> - S is known as the **feasible set**;
> - A solution (point) $\boldsymbol{x} \in S$ is known as a **feasible solution (point)**. Otherwise, it is known as an **infeasible solution(point)**.

> *Remark.* Note that to maximise $f(\mathbf{x})$ is equivalent to minimising $-f(\mathbf{x})$, so it suffices to only study minimisation problems.

The word "optimal", however, can be ambiguous due to its qualitative nature. Thus, we shall define what it means to be optimal quantitatively with more rigorous terms.

> **Definition 1.1.2 ▶ Optimal Solution**
>
> Consider a minimisation problem subject to constraint $\boldsymbol{x} \in S \subseteq \boldsymbol{R}^n$ whose objective function is $f(\boldsymbol{x})$. A feasible solution $\boldsymbol{x}^*$ is called an **optimal solution** if $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$ for all $\boldsymbol{x} \in S$. We can write
> $$\boldsymbol{x}^* = \operatorname*{argmin}_{\boldsymbol{x} \in S} f(\boldsymbol{x}).$$
> $f(\boldsymbol{x}^*)$ is then known as the **optimal value**.

> *Remark.* For maximisation problems, we can write
>
> $$\boldsymbol{x}^* = \operatorname*{argmax}_{\boldsymbol{x} \in S} f(\boldsymbol{x})$$

Note that not all optimisation problems have an optimal solution. We shall still expect to encounter problems for which no optimal solution nor value exists.

> **Definition 1.1.3 ▸ Unboundedness**
>
> Consider a minimisation problem subject to constraint $x \in S \subseteq R^n$ whose objective function is $f(x)$. The objective value is said to be **unbounded** if for all $K \in R$, there exists some $x \in S$ such that $f(x) < K$.

1.2   Unconstrained Non-linear Programs

To introduce the notion of an unconstrained NLP, we shall first define the openness of a set.

> **Definition 1.2.1 ▸ Open Set**
>
> Let $S \subseteq R^n$ be a set. S is called **open** if for all $x \in S$ there exists $\epsilon > 0$ such that the ball
> $$B(x, \epsilon) := \{y \in R^n : \ \|y - x\| < \epsilon\}$$
> is a subset of S.

> **Definition 1.2.2 ▸ Unconstrained NLP**
>
> An **unconstrained** NLP is an NLP whose feasible set $\mathcal{X}$ is an **open** subset of R^n.

1.3   Constrained Non-linear Programs

Similarly, to introduce the notion of a constrained NLP, we shall first define the closed-ness of a set.

> **Definition 1.3.1 ▸ Closed Set**
>
> Let $S \subseteq R^n$ be a non-empty set. S is said to be **closed** if for all convergent sequences $\{x_i\}_{i=1}^{\infty}$ with $x_i \in S$ for $i = 1, 2, \cdots$, the limit $\lim_{i \to \infty} x_i \in S$.

The empty set and Euclidean spaces $\mathbb{R}^n$ are both open and closed.

> *Remark.* Note that a set which is not open may not necessarily be closed. However, a set is open if and only if its complement is closed.

Theorem 1.3.2 ▶ Intersection of Closed Sets

If C_1 and C_2 are both closed, then $C_1 \cap C_2$ is closed.

Proof. The case where $C_1 \cap C_2 = \emptyset$ is trivial. If $C_1 \cap C_2 \neq \emptyset$, let $\{x_i\}_{i=1}^{\infty}$ be an arbitrary convergent sequence in $C_1 \cap C_2$. Since $\{x_i\}_{i=1}^{\infty} \in C_1$ which is closed, we have $\lim_{i \to \infty} x_i \in C_1$. Similarly, $\lim_{i \to \infty} x_i \in C_2$. Therefore, $\lim_{i \to \infty} x_i \in C_1 \cap C_2$.

Therefore, $C_1 \cap C_2$ is closed. $\qquad\square$

We then follow up by introducing three important closed sets.

Theorem 1.3.3

Let $g : \mathbf{R}^n \to \mathbf{R}$ be a continuous function, then the sets

$$S_1 = \{x \in \mathbf{R}^n : g(x) \leq 0\},$$
$$S_2 = \{x \in \mathbf{R}^n : g(x) \geq 0\},$$
$$S_3 = \{x \in \mathbf{R}^n : g(x) = 0\}$$

are closed.

Proof. Consider S_1. Let $\{x_i\}_{i=1}^{\infty}$ be any convergent sequence with $x_i \in S_1$ for $i = 1, 2, \cdots$, then

$$g\left(\lim_{i \to \infty} x_i\right) \leq 0$$

since $x_i \leq 0$. Therefore, $\lim_{i \to \infty} x_i \in S_1$ and so S_1 is closed.
S_2 and S_3 can be proved similarly. $\qquad\square$

By Theorem 1.3.2, we know that $S_1 \cup S_2 \cup S_3$ is closed, which motivates the following definition:

Definition 1.3.4 ▶ Constrained NLP

A **constrained** NLP is an NLP whose feasible set

$$S := \left\{x \in \mathbf{R}^n : g_i(x) = 0, i = 1, 2, \cdots, p, h_j(x) \leq 0, j = 1, 2, \cdots, q\right\}$$

is **closed**, where each of the g_i's is known as an equality constraint and each of the h_j's is known as an inequality constraint.

2

Convex Functions

2.1   Convexity of Sets and Functions

Intuitively, we describe two types of shapes in natural languages: the shapes which, if you choose any of its edges, lies in the same side of that edge, and the shapes which span across both sides from some chosen edge of its.

Graphically, this means that some shapes are "convex" to all directions, where as some other shapes are "concave". We shall define this rigorously as follows:

Definition 2.1.1 ▶ Convex Set

A set $D \subseteq \mathbf{R}^n$ is said to be **convex** if for all $x, y \in D$ and for all $\lambda \in [0, 1]$,

$$\lambda x + (1 - \lambda)y \in D.$$

We can define convexity over functions as follows:

Definition 2.1.2 ▶ Convex Function

A function $f : D \to \mathbf{R}^n$ is said to be **convex** if for all $x, y \in D$ and for all $\lambda \in [0, 1]$,

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$

Definition 2.1.3 ▶ Concave Function

A function $f : D \to \mathbf{R}^n$ is said to be **concave** if for all $x, y \in D$ and for all $\lambda \in [0, 1]$,

$$f(\lambda x + (1 - \lambda)y) \geq \lambda f(x) + (1 - \lambda)f(y).$$

Remark. A function which is not convex must be concave. However, a function which is convex may not be non-concave (consider $f(x) = x$).

We may derive the following relationship between a convex set and a convex function:

Proposition 2.1.4 ▶ Relations between Convex Sets and Convex Functions

Let $D \subseteq \mathbf{R}^n$ be a convex set and let $f : D \to \mathbf{R}$ be a convex function, then for all $\alpha \in \mathbf{R}$, the set

$$S_\alpha := \{ \boldsymbol{x} \in D : f(\boldsymbol{x}) \le \alpha \}$$

is convex.

Proof. Take $\boldsymbol{x}, \boldsymbol{y} \in S_\alpha$, then $f(\boldsymbol{x}) \le \alpha$ and $f(\boldsymbol{y}) \le \alpha$. Note that for any $\lambda \in [0,1]$, we have

$$\begin{aligned} f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) &\le \lambda f(\boldsymbol{x}) + (1-\lambda)f(\boldsymbol{y}) \\ &\le \lambda \alpha + (1-\lambda)\alpha \\ &= \alpha. \end{aligned}$$

Therefore, $\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y} \in S_\alpha$, and so S_α is convex. $\qquad\square$

Next, we introduce the notion of an *epigraph*.

Definition 2.1.5 ▶ Epigraph

Let $f : D \to \mathbf{R}$ be a function over a convex set $D \subseteq \mathbf{R}^n$. The **epigraph** of f is the set $E_f \subseteq \mathbf{R}^{n+1}$ defined by

$$E_f := \{ (\boldsymbol{x}, \alpha) : \boldsymbol{x} \in D, \alpha \in \mathbf{R}, f(\boldsymbol{x}) \le \alpha \}.$$

Remark. A trivial result: $D \times \text{range}(f) \subseteq E_f$.

Note that graphically, the epigraph of a function is just the region above the graph of the function.

Proposition 2.1.6 ▶ Convexity of Epigraph

Let $f : D \to \mathbf{R}$ be a function over a convex set $D \subseteq \mathbf{R}^n$. The epigraph E_f is convex if and only if f is convex.

Proof. Suppose E_f is convex. Take any $\boldsymbol{x}, \boldsymbol{y} \in D$, then $(\boldsymbol{x}, f(\boldsymbol{x})), (\boldsymbol{y}, f(\boldsymbol{y})) \in E_f$. Let $\lambda \in [0,1]$, we have

$$\lambda(\boldsymbol{x}, f(\boldsymbol{x})) + (1-\lambda)(\boldsymbol{y}, f(\boldsymbol{y})) = (\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}, \lambda f(\boldsymbol{x}) + (1-\lambda)f(\boldsymbol{y})) \in E_f.$$

Therefore,

$$f\left(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}\right) \le \lambda f(\boldsymbol{x}) + (1-\lambda)f(\boldsymbol{y}),$$

and so f is convex.

Suppose conversely that f is convex. For any $\boldsymbol{x}, \boldsymbol{y} \in D$ and any $\alpha, \beta \in \mathbf{R}$ such that $f(\boldsymbol{x}) \le \alpha$ and $f(\boldsymbol{y}) \le \beta$, we have $(\boldsymbol{x}, \alpha), (\boldsymbol{y}, \beta) \in E_f$. For all $\lambda \in [0, 1]$, consider

$$\lambda(\boldsymbol{x}, \alpha) + (1-\lambda)(\boldsymbol{y}, \beta) = (\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}, \lambda \alpha + (1-\lambda)\beta).$$

Since f is convex, we have

$$f\left(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}\right) \le \lambda f(\boldsymbol{x}) + (1-\lambda)f(\boldsymbol{y})$$
$$\le \lambda \alpha + (1-\lambda)\beta$$

for all $\lambda \in [0, 1]$. Therefore, $(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}, \lambda \alpha + (1-\lambda)\beta) \in E_f$, and so E_f is convex. $\square$

Lastly, we generalise the notion of a *convex combination*.

Proposition 2.1.7 ▶ Generalised Convex Combination

Let $k \in \mathbb{N}^+$ and let $f : S \to \mathbf{R}$ be a convex function on the convex set $S \subseteq \mathbf{R}^n$ and let $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_k \in S$, then

$$f\left(\sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i\right) \le \sum_{i=1}^{k} \lambda_i f(\boldsymbol{x}_i),$$

where $\sum_{i=1}^{k} \lambda_i = 1$ and $\lambda_i \ge 0$ for $i = 1, 2, \cdots, k$.

Proof. The case where $k = 1$ is trivial.

Suppose that there exists some $n \in \mathbb{N}^+$ such that

$$f\left(\sum_{i=1}^{n} \lambda_i \boldsymbol{x}_i\right) \le \sum_{i=1}^{n} \lambda_i f(\boldsymbol{x}_i),$$

$\square$

2.2   Gradient Vectors

Proposition 2.2.1 ▸ An Alternative Expression for Directional Derivatives

Let f be a function over $D \subseteq \mathbf{R}^n$ and let $\boldsymbol{d} \in \mathbf{R}^n$ be non-zero, then

$$\nabla f(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{d} = \lim_{\lambda \to 0} \frac{f(\boldsymbol{x} + \lambda\boldsymbol{d}) - f(\boldsymbol{x})}{\lambda}.$$

Proposition 2.2.2 ▸ Tangent Plane Characterisation of Convex Functions

Let f be a function over an open convex set $S \subseteq \mathbf{R}^n$ with continuous first partial derivatives, then f is convex if and only if

$$f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{x}) \leq f(\boldsymbol{y})$$

for all $\boldsymbol{x}, \boldsymbol{y} \in S$. In particular, f is strictly convex if and only if the above inequality is strict.

Proposition 2.2.3 ▸ Global Minimiser of Convex Functions

Let $f \colon C \to \mathbf{R}$ be a convex and continuously differentiable function over a convex set $C \subseteq \mathbf{R}^n$. Then $\boldsymbol{x}^* \in C$ is a global minimiser for the minimisation problem

$$\min\{f(\boldsymbol{x}) \colon \boldsymbol{x} \in C\}$$

if and only if

$$\nabla f(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^*) \geq 0$$

for all $\boldsymbol{x} \in C$.

3

Unconstrained NLPs

3.1   Coercive Functions

Definition 3.1.1 ▶ Coercive Function

A **continuous** function $f : \mathbf{R}^n \to \mathbf{R}$ is said to be **coercive** if

$$\lim_{\|x\|\to\infty} f(x) = +\infty.$$

More formally, f is coercive if and only if for all $M > 0$, there is some $r > 0$ such that $f(x) > M$ whenever $\|x\| > r$.

Theorem 3.1.2 ▶ Global Minimiser of Coercive Functions

If a function $f : \mathbf{R}^n \to \mathbf{R}$ is coercive, then f has at least one global minimiser.

Proof. Take $M = |f(0)| + 1 > 0$. Since f is coercive, there exists some $r > 0$ such that $f(x) > M > f(0)$ whenever $\|x\| > r$. Consider

$$B(0, r) = \{x \in \mathbf{R}^n : \|x\| \leq r\}$$

which is a compact set. So by Weiestrass' Theorem, there exists some $x^* \in B(0, r)$ such that $f(x^*) \leq f(x) \leq f(0)$ for all $x \in B(0, r)$. Now, this means that for all $x \in \mathbf{R}^n$, we have $(x^*) \leq f(x)$, which means that x^* is a global minimiser for f.   □

Definition 3.1.3 ▶ Stationary (Critical) Point

Let $X \subseteq \mathbf{R}^n$ be an open set and let $f : X \to \mathbf{R}$ be a function. An interior point x^* is called a **stationary point** of f if $\nabla f(x^*) = 0$.

Definition 3.1.4 ▶ Saddle Point

A stationary point x^* of a function f which is neither a local minimisr nor a local maximiser is called a **saddle point**.

Corollary 3.1.5

Let $\boldsymbol{x}^*$ be a stationary point of f. If $H_f(\boldsymbol{x}^*)$ is indefinite, then $\boldsymbol{x}^*$ is a saddle point.