1 Unconstrained Optimization

An unconstrained optimization problem is of the form

$$\begin{align*}
\text{minimize} \quad & f(x) \\
\text{subject to} \quad & x \in \mathbb{R}^n
\end{align*}$$

(1)

where $x \in \mathbb{R}^n$ is the optimization variable and $f : \mathbb{R}^n \to \mathbb{R}$ is the objective function. We call $x^*$, a solution to Problem (1) if it satisfies

$$f(x) \geq f(x^*), \quad \text{for all } x \in \mathbb{R}^n.$$

(2)

We also call $x^*$ a **global minimum** of $f$.

Perhaps, the simplest approach to solving Problem (1) is to do a grid search over $\mathbb{R}^n$. However, since the complexity of this approach scales exponentially in $n$, grid search is not tractable for large $n$. In fact, for many optimization problems, it might not be easy to find a global minimum at all. However, even finding a point $x^*$ such that

$$\exists \epsilon > 0 \quad \text{s.t.} \quad f(x) \geq f(x^*), \quad \text{for all } ||x - x^*|| \leq \epsilon,$$

(3)

may still be valuable for us. Any $x^*$ that satisfies (3) is called a **local minimum** of $f$. A point $x^*$ being a local minimum means that if we modify Problem (1) by restricting the feasible set to $\{x | ||x - x^*|| \leq \epsilon\}$, then $x^*$ is a solution to the modified problem.
**Definition 1** Consider Optimization Problem (1). A point \( x^* \) is called a **strict global minimum** if

\[
f(x) > f(x^*), \text{ for all } x \in \mathbb{R}^n, x \neq x^*,
\]

and a **strict local minimum** if

\[
\exists \epsilon > 0 \text{ s.t. } f(x) > f(x^*), \text{ for all } x \text{ that satisfies } ||x - x^*|| \leq \epsilon, x \neq x^*.
\]

**example:** \( f_1(x) = x^3 \) and \( f_2(x) = -|x|^3 \) neither have a global minimum nor a local minimum. \( f_3(x) = |x|^3 \) has one local minimum which is also a global minimum.

In this course, we will be studying some of the algorithms that either solve Problem (1) or at least find a local minimum for this problem.

### 1.1 Optimality Conditions

In solving Problem (1), given a point \( x \), the fundamental question is to determine whether \( x \) is a (strict) local/global minimum. We will show that if the objective function \( f \) in Problem (1) is twice differentiable, then the optimality of a point \( x^* \), necessitates the so called **optimality conditions**

\[
\nabla f(x^*) = 0, \tag{4}
\]

\[
\nabla^2 f(x^*) \succeq 0. \tag{5}
\]

Moreover, if for some \( x^* \) the **Optimality Conditions**

\[
\nabla f(x^*) = 0, \tag{6}
\]

\[
\nabla^2 f(x^*) > 0 \tag{7}
\]

hold, then \( x^* \) is a local minimum. Therefore, Optimality Conditions (6), (7) are **sufficient** for local optimality of \( x^* \).

**Theorem 1** Consider Problem (1) and assume \( f \) is twice differentiable. If \( x^* \) is a local minimum of \( f \), then \( x^* \) satisfies Optimality Conditions (4) and (5). Furthermore, if \( x^* \) satisfies Optimality Conditions (6) and (7), then \( x^* \) is a local minimum of \( f \).

**Proof:**
Suppose $x^*$ is a local minimum. Let $d \in \mathbb{R}^n$ be a unit vector and let $x(\alpha) = x^* + \alpha d$, $\forall \alpha \in \mathbb{R}$.

Note that $x(0) = x^*$. Using the first order Taylor approximation of $f$ around $x^*$, one can write

$$0 \leq \lim_{\alpha \to 0} \frac{f(x(\alpha)) - f(x^*)}{||x(\alpha) - x^*||}$$

$$= \lim_{\alpha \to 0} \frac{f(x^*) + \nabla f(x^*)^\top (x(\alpha) - x^*) + o(||x(\alpha) - x^*||) - f(x^*)}{||x(\alpha) - x^*||}$$

$$= \lim_{\alpha \to 0} \frac{\nabla f(x^*)^\top (x(\alpha) - x^*) + o(||x(\alpha) - x^*||)}{||x(\alpha) - x^*||}$$

$$= \lim_{\alpha \to 0} \frac{\nabla f(x^*)^\top (x(\alpha) - x^*)}{||x(\alpha) - x^*||}$$

$$= \frac{\nabla f(x^*)^\top d}{||d||},$$

where the first inequality follows from the local minimality of $x^*$. Now, since this holds for any direction $d$, it follows that $\nabla f(x^*) = 0$.

In order to prove $\nabla^2 f(x^*) \succeq 0$, we can use the second order Taylor approximation of $f$ to write

$$0 \leq \lim_{\alpha \to 0} \frac{f(x) - f(x^*)}{||x - x^*||^2}$$

$$= \lim_{\alpha \to 0} \frac{\nabla f(x^*)^\top (x - x^*) + \frac{1}{2}(x - x^*)^\top \nabla^2 f(x^*)(x - x^*) + o(||x - x^*||^2)}{||x - x^*||^2}$$

$$= \lim_{\alpha \to 0} \frac{\nabla^2 f(x^*) d}{\alpha^2 ||d||^2}$$

$$= \frac{1}{2} d^\top \nabla^2 f(x^*) d.$$

This inequality holds for any direction $d$; therefore, $\nabla^2 f(x^*)$ is positive semidefinite.

The proof of the sufficient conditions is similar. (A more rigorous proof would be based on the mean value theorem.)

Note that in a similar manner, the Taylor approximations of order higher that two can also yield other optimality conditions which might be use-
ful in solving more complex problems. Also, note that Optimality Conditions (4), (5) are not sufficient.

**Example:** \( f(x) = x^3 \) at \( x = 0 \), \( f'(0) = 0 \), \( f''(0) \geq 0 \), but \( x = 0 \) is not a local/global minimum. Sufficient optimality conditions in this case would be

\[
f'(x^*) = 0, \quad f''(x^*) > 0
\]

**Remark 1** The optimality conditions are useful because

- they provide tractable conditions for optimality,
- they help narrow down the list of potential solutions,
- they are useful in the design and analysis of algorithms.

### 1.2 Existence of optimal Solution

Optimization Problem (1) does not always have a solution. For example

1. \( f(x) = x^2 - x^4 \) has no global minimum and two local minima. Note that
   \[
f'(x) = 2x - 4x^3, \quad f''(x) = 2 - 12x^2.
\]

2. \( g(x) = e^{-|x|} \) although satisfies \( \inf_{x \in \mathbb{R}} g(x) = 0 \), it does not have any global minima.

However, there are conditions which guarantee the existence of solution for Problem (1).

**Theorem 2 (Bolzano-Weierstrass)** Every continuous function \( f \) attains its minimum over any **compact** set \( \mathcal{X} \). In other words, \( \exists x^* \in \mathcal{X} \) s.t. \( f(x^*) = \inf_{x \in \mathcal{X}} f(x) \).

Note that a set is compact if it is closed and bounded. Theorem 2 implies that if the level set \( \{ x \in \mathbb{R}^n | f(x) \leq f(x^0) \} \) is compact, then a global minimum exists.

**Definition 2** A function \( f : \mathbb{R}^n \to \mathbb{R} \) is called coercive, if it satisfies

\[
\lim_{|x| \to +\infty} f(x) = +\infty. \tag{8}
\]
It can be shown that if a function \( f \) is coercive and continuous on \( \mathbb{R}^n \), then global minimum exists.

### 1.3 Unconstrained quadratic optimization

An unconstrained quadratic optimization problem is of the form

\[
\text{minimize } x \in \mathbb{R}^n \quad F(x) := \frac{1}{2} \mathbf{x}^\top \mathbf{Q} \mathbf{x} + \mathbf{b}^\top \mathbf{x},
\]

where, without loss of generality, we assume \( \mathbf{Q} \) is a symmetric matrix. (Note that for any \( \mathbf{Q} \), one can write \( \mathbf{x}^\top \mathbf{Q} \mathbf{x} = \mathbf{x}^\top \frac{\mathbf{Q} + \mathbf{Q}^\top}{2} \mathbf{x} \) where \( \frac{\mathbf{Q} + \mathbf{Q}^\top}{2} \) is a symmetric matrix.) As an illustrative example, we will discuss Optimality Conditions (4), (5) for Problem (9).

- **Necessary conditions**

  \[
  \nabla F(x) = \mathbf{Q} \mathbf{x} + \mathbf{b} = 0, \quad \nabla^2 F(x) = \mathbf{Q} \succeq 0.
  \]

  - if \( \mathbf{Q} \mathbf{x} + \mathbf{b} = 0 \) is not feasible, then \( \inf f(x) = -\infty \).
  - if \( \mathbf{Q} \) is not positive semidefinite, then we can show that it does not have a global minimum. Let \( \mathbf{Q} = \sum_i \lambda_i \mathbf{u}_i \mathbf{u}_i^\top \) be an eigenvalue decomposition of \( \mathbf{Q} \), where \( \mathbf{u}_i \) are orthonormal and assume \( \lambda_1 < 0 \). One can write

    \[
    \frac{1}{2} \mathbf{x}^\top \mathbf{Q} \mathbf{x} + \mathbf{b}^\top \mathbf{x} = \frac{1}{2} \mathbf{x}^\top \sum_i \lambda_i \mathbf{u}_i \mathbf{u}_i^\top \mathbf{x} + \mathbf{b}^\top \mathbf{x} = \frac{1}{2} \sum_i \lambda_i (\mathbf{x}^\top \mathbf{u}_i)^2 + (\mathbf{b}^\top \mathbf{u}_1) \mathbf{x}.
    \]

    Now, if we let \( \mathbf{x} = \alpha \mathbf{u}_1 \), then we have

    \[
    \frac{1}{2} \mathbf{x}^\top \mathbf{Q} \mathbf{x} + \mathbf{b}^\top \mathbf{x} = \frac{1}{2} \lambda_1 \alpha^2 + (\mathbf{b}^\top \mathbf{u}_1) \alpha.
    \]

    Clearly, \( F(\alpha \mathbf{u}_1) \) goes to \(-\infty\) as \( \alpha \to +\infty \).

- **Claim:**

  - The above necessary conditions are also sufficient. (Positive definiteness of \( \mathbf{Q} \) is not required in this particular example.)
Any local optimum is also globally optimum (True for any convex optimization).

2 Convexity

There are classes of functions for which any local minimum is also a global minimum. Perhaps the most important of such classes is the class of convex functions.

Definition 3 A function \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is convex if it satisfies

\[
f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y), \quad \forall \alpha \in (0, 1), \ x, y \in \mathbb{R}^n. \tag{10}
\]

Moreover, \( f \) is called strictly convex if the inequality in (10) is strict for any \( x \neq y \).

Convexity of a function means that any line segment connecting two points in the graph of function should lie above the graph. It can be shown that a continuous function is convex if and only if

\[
f(y) \geq f(x) + \nabla f(x)^\top(y - x), \quad \forall x, y.
\tag{11}
\]

This means that for any \( x \in \mathbb{R}^n \), the graph of linear approximation of a convex function \( f \) around \( x \) should lie below the graph of \( f \).

Definition 4 For any set \( A \subset \mathbb{R}^n \), the indicator function associated to \( A \) is defined as

\[
I_A(x) := \begin{cases} 
0 & \text{if } x \in A, \\
+\infty & \text{otherwise}.
\end{cases}
\]

Definition 5 A set \( S \subset \mathbb{R}^n \) is convex if \( I_S \) is convex.

An equivalent definition for a convex set \( S \) would be

\( S \) is convex if \( x, y \in S \Rightarrow \gamma x + (1 - \gamma)y \in S, \ \forall \gamma \in [0, 1] \).
Definition 6 The optimization problem

\[
\min_x f(x) \quad \text{subject to } x \in X
\]

is convex if \( f \) is a convex function and \( X \) is a convex set.

Remark 2 Note that

- If \( f \) is twice continuously differentiable, then
  \[
  f \text{ is convex } \iff \nabla^2 f(x) \succeq 0, \quad \forall x.
  \]

  This follows from the fact that a function \( f \) is convex if and only if the restriction of \( f \) to any direction \( d \) starting at any point \( x_0 \), defined as
  \[
  f_{d,x_0}(t) := f(x_0 + td),
  \]
  is convex.

- For convex optimization problems, local optimality implies global optimality.

  Simply, let \( x \) and \( y \) be two local minima. The line segment connecting \( x \) and \( x \) passes through the balls \( \{z \mid \|z - x\| \leq \epsilon\} \) and \( \{z \mid \|z - y\| \leq \epsilon\} \) for any \( \epsilon > 0 \). Therefore, neither \( f(x) < f(y) \) nor \( f(y) < f(x) \). This implies that any two local minima have the same objective value.

- The set of solutions of a convex optimization problem is convex.
  (The proof is straightforward using Definition 3.)

- If the objective function is strictly convex, then the minimum is unique.

Example:

\[
\text{minimize } f(x, y) = \frac{1}{2}(\alpha x^2 + \beta y^2) - x \quad \text{subject to } (x, y) \in \mathbb{R}^2.
\]

The Hessian is

\[
\nabla^2(x, y) = \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix}.
\]
• \( \alpha, \beta > 0 \implies \) strict convexity, \((x^*, y^*) = \left( \frac{1}{\alpha}, 0 \right)\).

• \( \alpha = 0, \beta > 0 \implies \) convexity, there is not local/global minima.

• \( \alpha > 0, \beta = 0 \implies \) convexity (not strictly), \((x^*, y^*) = \left( \frac{1}{\alpha}, y \right), y \in \mathbb{R} \).

• \( \alpha > 0, \beta < 0 \implies \) non-convexity.