# Theory of Optimization

Jayadev Naram

## Contents

**1 Unconstrained Minimization** **2**

# 1 Unconstrained Minimization

Let $\mathbb{R}^n$ denote the $n$-dimensional Euclidean real vector space with the inner product defined for any $x, y \in \mathbb{R}^n$ as $\langle x, y\rangle = x^T y = \sum_{i=1}^n x_i y_i$, where $x = [x_1, \ldots, x_n]^T$ and $y = [y_1, \ldots, y_n]^T$ are the coordinates of $x$ and $y$ respectively. Let the norm and the metric on $\mathbb{R}^n$ be defined as $\|x\| = \sqrt{\langle x, x\rangle}$ and $d(x, y) = \|x - y\|$, respectively.

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a real-valued function on $\mathbb{R}^n$. We consider the following optimization problem

$$\min_{x\in\mathbb{R}^n} f(x). \tag{1}$$

The problem (1) is called as the unconstrained minimization problem as there are no constraints on $x$.

**Definition 1.1.** A ***local minimum*** of $f$ in the problem (1) is a vector $x^* \in \mathbb{R}^n$ for which there exists $\varepsilon > 0$ such that for all $x \in \mathbb{R}^n$ we have

$$f(x^*) \le f(x), \tag{2}$$

when $\|x - x^*\| \le \varepsilon$. A ***global minimum*** of $f$ in the problem (1) is a vector $x^* \in \mathbb{R}^n$ such that for all $x \in \mathbb{R}^n$ we have

$$f(x^*) \le f(x). \tag{3}$$

The global or local minimum $x^*$ is said to be ***strict*** if the corresponding inequality given above is strict for $x \ne x^*$. The vector $x^*$ with $\nabla f(x^*) = 0$ is referred to as a ***stationary point***.

**Proposition 1.2** (Necessary Optimality Conditions)**.** *Let $x^*$ be an unconstrained local minimum of $f : \mathbb{R}^n \to \mathbb{R}$, and assume that $f$ is continuously differentiable in an open set $U$ containing $x^*$. Then*

$$\nabla f(x^*) = 0. \qquad \text{(First Order Necessary Condition)}$$

*If in addition $f$ is twice continuously differentiable within $U$, then*

$$\nabla^2 f(x^*) \succeq 0. \qquad \text{(Second Order Necessary Condition)}$$

*Proof.* Fix some $d \in \mathbb{R}^n$. Then, using chain rule to differentiate the function $g(\alpha) = f(x^* + \alpha d)$, we have

$$0 \le \lim_{\alpha\downarrow 0} \frac{f(x^* + \alpha d) - f(x^*)}{\alpha} = \frac{dg(0)}{d\alpha} = \nabla f(x^*)^T d,$$

where the inequality follows from the assumption that $x^*$ is a local minimum and $\alpha \downarrow 0$ indicates the right-hand limit, i.e., $\alpha > 0$ and $\alpha \to 0$. Since $d$ is arbitrary, the same inequality holds with $d$ replaced by $-d$. Therefore, $\nabla f(x^*)^T d = 0$ for all $d \in \mathbb{R}^n$, which shows that $\nabla f(x^*) = 0$.

Assume that $f$ is twice continuously differentiable, and let $d$ be any vector in $\mathbb{R}^n$. For all $\alpha \in \mathbb{R}$, the second order Taylor expansion yields

$$f(x^* + \alpha d) - f(x^*) = \alpha \nabla f(x^*)^T d + \frac{\alpha^2}{2} d^T \nabla^2 f(x^*) d + o(\alpha^2).$$

Using the condition $\nabla f(x^*) = 0$ and the local optimality of $x^*$, we see that there is a sufficiently small $\varepsilon > 0$ such that for all $\alpha \in (0, \varepsilon)$,

$$0 \le \frac{f(x^* + \alpha d) - f(x^*)}{\alpha^2} = \frac{1}{2} d^T \nabla^2 f(x^*) d + \frac{o(\alpha^2)}{\alpha^2}.$$

Taking the limit $\alpha \to 0$ and using the fact that $\lim_{\alpha\to 0} o(\alpha^2)/\alpha^2 = 0$, we obtain $d^T \nabla^2 f(x^*) d \ge 0$, showing that $\nabla^2 f(x^*)$ is positive semidefinite. $\square$

**Remark 1.3.** Suppose for a function $f : \mathbb{R}^n \to \mathbb{R}$, the point $x^*$ is a local minimum of $f$ along every line that passes through $x^*$, i.e., the function

$$g(\alpha) = f(x^* + \alpha d)$$

is minimized at $\alpha = 0$ for all $d \in \mathbb{R}^n$. Then

$$0 = \left.\frac{dg}{d\alpha}\right|_{\alpha=0} = \nabla f(x^*)^T d = 0, \quad \forall\, d \in \mathbb{R}^n.$$

This shows that $\nabla f(x^*) = 0$, i.e., first order necessary condition is satisfied at $x^*$. This only shows that $x^*$ is a stationary point and it need not be a local minimum of $f$. For example, consider $f(y, z) = (z - py^2)(z - qy^2)$, where $0 < p < q$. Here $(0, 0)$ is one such stationary point that minimizes $f$ along every line passing through it but $(0, 0)$ is not a local minimum of $f$.

**Proposition 1.4** (Second Order Sufficient Optimality Conditions)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable over an open set $U$. Suppose that a vector $x^* \in U$ satisfies the conditions*

$$\nabla f(x^*) = 0, \qquad \nabla^2 f(x^*) \succeq 0.$$

*Then, $x^*$ is a strict unconstrained local minimum of $f$. In particular, there exists a scalar $\gamma > 0$ and $\varepsilon > 0$ such that for all $x \in \mathbb{R}^n$ with $\|x - x^*\| < \varepsilon$, we have*

$$f(x) \geq f(x^*) + \frac{\gamma}{2}\|x^* - x\|^2.$$

*Proof.* Denote by $\lambda > 0$ the smallest eigenvalue of $\nabla^2 f(x^*)$. Then we have

$$d^T \nabla^2 f(x^*) d \geq \lambda \|d\|^2, \qquad \forall\, d \in \mathbb{R}^n.$$

Using this relation, the hypothesis $\nabla f(x^*) = 0$, and the second order Taylor expansion, we have for all $d$

$$\begin{aligned} f(x^* + d) - f(x^*) &= \nabla f(x^*)^T d + \frac{1}{2} d^T \nabla^2 f(x^*) d + o(\|d\|^2) \\ &\geq \frac{\lambda}{2}\|d\|^2 + o(\|d\|^2) \\ &= \left(\frac{\lambda}{2} + \frac{o(\|d\|^2)}{\|d\|^2}\right)\|d\|^2. \end{aligned}$$

Choose any $\varepsilon > 0$ and $\gamma > 0$ such that for all $d \in \mathbb{R}^n$ with $\|d\| < \varepsilon$,

$$\frac{\lambda}{2} + \frac{o(\|d\|^2)}{\|d\|^2} \geq \frac{\gamma}{2}.$$

$\square$