Chapter 4

Optimality Conditions

The objective of this chapter is to give the necessary or sufficient conditions for a point $x^\star$ to be a local minimum, for the following unconstrained optimization problem:

$$\min_{x\in\mathbb{R}^n} f(x). \tag{4.1}$$

Definition 4.0.1 (Critical Point)**.** Let $f \in C^1$. A **critical point** (or stationary point) is a point that satisfies

$$\nabla f(x^\star) = 0_{\mathbb{R}^n}.$$

Example 4.0.2. Let the function f be defined on $\mathbb{R}^2$ by

$$f(x,y) = 2x^3 + xy^2 + 5x^2 + y^2.$$

The critical points of f are the solutions of the following equation:

$$\nabla f(x,y) = 0_{\mathbb{R}^2},$$

which implies

$$\begin{pmatrix} 6x^2 + y^2 + 10x \\ 2xy + 2y \end{pmatrix} = 0_{\mathbb{R}^2}.$$

Thus,

$$S = \left\{(0,0), \left(-\tfrac{5}{3}, 0\right), (-1,2), (-1,-2)\right\}.$$

Definition 4.0.3 (Saddle Point)**.** A critical point $x^\star$ is a **saddle point** if for all $r > 0$, there exist $a, b \in B(x^\star, r)$ such that

$$f(a) \leq f(x^\star) \leq f(b).$$

Example 4.0.4. Consider the function $f(x) = x^3$. We have

$$f'(x) = 3x^2,$$

so $f'(x)$ vanishes at $x = 0$. For any $r > 0$, take $B(0,r) =]-r, r[$. For $a = -\frac{r}{2}$ and $b = \frac{r}{2}$, we obtain

$$f(a) = -\frac{r^3}{8} \leq f(0) = 0 \leq f(b) = \frac{r^3}{8}.$$

Thus, $x = 0$ is a saddle point.

4.1 First-Order Optimality Conditions

Our first result states that the first derivative must vanish whenever we have an unconstrained optimization problem.

Theorem 4.1.1 (First-Order Necessary Condition)**.** *If $x^\star$ is a local minimum of the function f on $\mathbb{R}^n$, then*

$$\nabla f(x^\star) = 0_{\mathbb{R}^n}.$$

Proof. Let $x^\star$ be a local minimum of f on $\mathbb{R}^n$. By definition of a local minimum, there exists an open ball such that

$$f(x^\star) \leq f(x), \quad \forall x \in B(x^\star, r).$$

For any vector $d \in \mathbb{R}^n$ and for a scalar $t \geq 0$ sufficiently small, the point

$$x = x^\star + td$$

belongs to $B(x^\star, r)$, and therefore

$$f(x^\star) \leq f(x^\star + td).$$

This implies that

$$\lim_{t \to 0^+} \frac{f(x^\star + td) - f(x^\star)}{t} = \nabla f(x^\star)^T d \geq 0.$$

This result holds for any $d \in \mathbb{R}^n$, hence it also holds for $-d$, i.e.

$$\nabla f(x^\star)^T d \leq 0.$$

Thus,

$$\nabla f(x^\star)^T d = 0, \quad \forall d \in \mathbb{R}^n,$$

which implies

$$\nabla f(x^\star) = 0.$$

□

Example 4.1.2. Let

$$f(x) = x_1^2 + \tfrac{1}{2}x_2^2 + 3x_2 + 92, \quad \Omega = \mathbb{R}^2.$$

We have

$$\nabla f(x) = \begin{pmatrix} 2x_1 \\ x_2 + 3 \end{pmatrix}.$$

1. For $x^\star = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$, we get

   $$\nabla f(x^\star) = \begin{pmatrix} 2 \\ 6 \end{pmatrix} \neq 0_{\mathbb{R}^2}.$$

   Therefore, $x^\star$ does not satisfy the first-order necessary condition.

2. For $x^\star = \begin{pmatrix} 0 \\ -3 \end{pmatrix}$, we obtain

$$\nabla f(x^\star) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Hence, $x^\star$ satisfies the first-order necessary condition.

Remark 4.1.3. When the function f is not convex, we can only provide a necessary condition for local optimality, but not a sufficient one. Indeed, it is possible that the differential vanishes at some point $x^\star$, while this point is not a local minimum.

Example 4.1.4. For $f(x) = x^3$, the derivative vanishes at $x^\star = 0$, but this point is a saddle point (hence neither a minimum nor a maximum).

The above condition involves the gradient vector, and is therefore called a **first-order condition**.

4.2 Second-Order Optimality Conditions

If the objective function is neither convex nor concave, the first-order conditions do not allow us to distinguish between a minimum and a maximum. To obtain such a distinction, one must study the behavior of the second derivative at $x^\star$. We then have the following optimality condition:

Theorem 4.2.1 (Second-Order Necessary Condition)**.** *If $x^\star$ is a local minimum (resp. local maximum) of the function f on $\mathbb{R}^n$, then*

$$\nabla f(x^\star) = 0 \quad \textit{and} \quad d^T \nabla^2 f(x^\star) d \geq 0, \ \forall d \in \mathbb{R}^n$$

(resp. $d^T \nabla^2 f(x^\star) d \leq 0, \ \forall d \in \mathbb{R}^n$).

Proof. Using the second-order Taylor expansion of f around the local minimum $x^\star$, we obtain for all $d \in \mathbb{R}^n$ and for $t \geq 0$ sufficiently small:

$$f(x^\star) \leq f(x^\star + td) = f(x^\star) + t\nabla f(x^\star)^T d + \tfrac{1}{2} t^2 d^T \nabla^2 f(x^\star) d + t^2 \|d\|^2 \varepsilon(td).$$

Since $\nabla f(x^\star) = 0$, we have

$$f(x^\star + td) = f(x^\star) + \tfrac{1}{2} t^2 d^T \nabla^2 f(x^\star) d + t^2 \|d\|^2 \varepsilon(td).$$

Dividing by t^2 and taking the limit, we find

$$\lim_{t \to 0^+} \frac{f(x^\star + td) - f(x^\star)}{t^2} = \tfrac{1}{2} d^T \nabla^2 f(x^\star) d \geq 0.$$

□

Example 4.2.2. Let $f(x) = x_1^2 - x_2^2$ with $\Omega = \mathbb{R}^2$. For $x^\star = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ we have

$$\nabla f(x^\star) = 0_{\mathbb{R}^2}, \quad \nabla^2 f(x^\star) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}.$$

We observe that $\nabla^2 f(x^\star)$ is not positive semidefinite. Thus $x^\star$ is neither a local minimum nor a local maximum.

Moreover, if the second-order condition is strictly satisfied, we obtain the following sufficient condition:

Theorem 4.2.3 (Second-Order Sufficient Condition)**.** *Let $x^\star \in \mathbb{R}^n$. If $x^\star$ satisfies*

$$\nabla f(x^\star) = 0_{\mathbb{R}^n} \quad \textit{and} \quad d^T \nabla^2 f(x^\star) d > 0, \quad \forall d \neq 0 \in \mathbb{R}^n,$$

then $x^\star$ is a strict local minimum of the function f on $\mathbb{R}^n$.

The proof of this theorem relies on the following lemma:

Lemma 4.2.4. *Let f be a twice continuously differentiable function on $\mathbb{R}^n$, and let $x^\star \in \mathbb{R}^n$ be such that $\nabla^2 f(x^\star)$ is positive definite. Then there exists $\delta > 0$ such that*

$$\nabla^2 f(x) \succ 0, \quad \forall x \in B(x^\star, \delta).$$

Proof. From Lemma 4.1, there exists $\delta > 0$ such that $\nabla^2 f(x) \succeq 0$ for all $x \in B(x^\star, \delta)$. Let $x \in B(x^\star, \delta)$ with $x \neq x^\star$. Define

$$\Phi(t) = f\big(x^\star + t(x - x^\star)\big), \quad t \in [0, 1].$$

It is clear that Φ is twice continuously differentiable on $[0, 1]$. Moreover, we have

$$\Phi(0) = f(x^\star), \tag{4.2}$$

and

$$\Phi(1) = f(x). \tag{4.3}$$

By Taylors theorem, there exists $\hat{t} \in [0, 1]$ such that

$$\Phi(t) = \Phi(0) + \Phi'(0)t + \tfrac{1}{2}\Phi''(\hat{t})t^2, \tag{4.4}$$

with

$$\Phi'(0) = \nabla f(x^\star)^T (x - x^\star) = 0,$$

and

$$\Phi''(\hat{t}) = (x - x^\star)^T \nabla^2 f\big(x^\star + \hat{t}(x - x^\star)\big)(x - x^\star) \geq 0.$$

For $t = 1$, we obtain

$$\Phi(1) - \Phi(0) = f(x) - f(x^\star) = \tfrac{1}{2}\Phi''(\hat{t}) \geq 0.$$

Thus, $x^\star$ is a strict local minimum. $\square$

Example 4.2.5. Let $f(x) = x_1^2 + x_2^2$ with $\Omega = \mathbb{R}^2$. For $x^\star = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, we have

$$\nabla f(x^\star) = 0_{\mathbb{R}^2}, \quad \nabla^2 f(x^\star) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \succeq 0.$$

Thus, $x^\star$ is a strict local minimum of f on $\mathbb{R}^2$.

Example 4.2.6. Consider the function f from Example 4.2. Its Hessian is

$$\nabla^2 f(x, y) = \begin{pmatrix} 12x + 10 & 2y \\ 2y & 2x + 2 \end{pmatrix}.$$

At $(0, 0)$, we have

$$\nabla^2 f(0, 0) = \begin{pmatrix} 10 & 0 \\ 0 & 2 \end{pmatrix} \succeq 0,$$

so $(0, 0)$ is a strict local minimum.

At $\left(-\frac{5}{3}, 0\right)$, we have

$$\nabla^2 f\left(-\tfrac{5}{3}, 0\right) = \begin{pmatrix} -10 & 0 \\ 0 & -\frac{4}{3} \end{pmatrix} \preceq 0,$$

hence this point is not a local minimum. Thus, the point

$$\left(-\tfrac{5}{3}, 0\right)$$

is a strict local maximum.

For the point $(-1, 2)$, we have

$$\nabla^2 f(-1, 2) = \begin{bmatrix} -2 & 4 \\ 4 & 0 \end{bmatrix}.$$

The sign of the Hessian matrix $\nabla^2 f(-1, 2)$ is indefinite, therefore $(-1, 2)$ is a saddle point.

$$\nabla^2 f(-1, -2) = \begin{bmatrix} -2 & -4 \\ -4 & 0 \end{bmatrix}.$$

The sign of the Hessian matrix $\nabla^2 f(-1, -2)$ is indefinite, therefore $(-1, -2)$ is a saddle point.

Exercises

Exercise 4.1 We consider the optimization problem

$$\min_{(x_1,x_2)\in\mathbb{R}^2} f(x) = x_1^2 + x_2^2 - 2x_2 + 5.$$

Is the first-order necessary condition for a local minimum satisfied at the points

$$[1,1]^\top, \quad [-1,-1]^\top, \quad [0,1]^\top, \quad [0,-1]^\top\,?$$