Optimization Problem with Simple Constraints

Instructor: Jin Zhang

Department of Mathematics
Southern University of Science and Technology
Spring 2024

Optimality conditions

To motivate the derivation of a necessary optimality condition, consider a function $f : C \to R$ where C is an interval on the real line. We all know that if f has a local minimum at an interior point $\bar{x} \in C$ then $f'(\bar{x}) = 0$ which is equivalent to saying that $f'(\bar{x})(x - \bar{x}) \geq 0$ for all $x \in C$. However when a minimum is a boundary point of the interval C, one can only claim that $f'(\bar{x}) \leq 0$ and $f'(\bar{x}) \geq 0$ respectively if $\bar{x}$ is the left end point and the right end point respectively. In both cases, we have again

$$f'(\bar{x})(x - \bar{x}) \geq 0 \quad \forall x \in C.$$

In fact, it is easy to show that following result.

Theorem 5.1.1 (First order condition) *Let $f : R^n \to R$ be a differentiable function and C is a convex closed set of R^n. Then*

(a) If $\bar{x}$ is a local minimizer of f over $C \subset R^n$, then

$$\nabla f(\bar{x})^T (x - \bar{x}) \geq 0 \quad \forall x \in C.$$

(b) Moreover suppose that f is a convex function on C. Then $\bar{x}$ is a global minimum of f over C if and only if

$$\nabla f(\bar{x})^T (x - \bar{x}) \geq 0 \quad \forall x \in C.$$

Proof. (a) Let $x \in C$. Since C is a convex set and $x, \bar{x}$ are both in C, for any $t \in [0,1]$, $\bar{x} + t(x - \bar{x}) \in C$. Therefore by the local optimality of $\bar{x}$, one has

$$f(\bar{x} + t(x - \bar{x})) - f(\bar{x}) \geq 0 \qquad \forall t \text{ sufficiently small.}$$

Dividing the above inequality by t and taking $t \downarrow 0$, one has

$$f'(\bar{x}; x - \bar{x}) \geq 0 \qquad \forall x \in C.$$

The desired result follows from the relationship between the gradient and the directional derivative in Proposition 3.0.1.

(b) The necessity follows from (a). It suffices to show the sufficiency. Let f be convex on C and

$$\nabla f(\bar{x})^T (x - \bar{x}) \geq 0 \quad \forall x \in C.$$

Then by the characterization of convex functions Theorem 4.3.3 (b),

$$f(x) - f(\bar{x}) \geq \nabla f(\bar{x})^T (x - \bar{x}) \geq 0 \quad \forall x \in C$$

which means that $\bar{x}$ is a minimum of f on set C.

Definition 5.1.1 (Normal Cone) *Let $C \subset R^n$ be a convex set and $\bar{x} \in C$. η is a normal vector to C at $\bar{x}$ if and only if*

$$\eta \cdot (x - \bar{x}) \leq 0 \quad \textit{for all } x \in C.$$

The set of all normal vectors to C at $\bar{x}$ is called the normal cone of C at $\bar{x}$ and is denoted by

$$N_C(\bar{x}) = \{\eta \in R^n | \eta \cdot (x - \bar{x}) \leq 0 \quad \textit{for all } x \in C\}.$$

Theorem 5.1.2 *Let $f : R^n \to R$ be a differentiable function and C is a convex closed set of R^n. Then*

(a) If $\bar{x}$ is a local minimizer of f over $C \subset R^n$, then

$$0 \in \nabla f(\bar{x}) + N_C(\bar{x}).$$

(b) Moreover suppose that f is a convex function on C. Then $\bar{x}$ is a global minimum of f over C if and only if

$$0 \in \nabla f(\bar{x}) + N_C(\bar{x}).$$

Note that from the definition of a normal cone, when $\bar{x}$ is an interior point of C, then $N_C(\bar{x}) = \{0\}$. In this case the first order condition is reduced to the familiar form $\nabla f(\bar{x}) = 0$.

Theorem 5.1.3 *Let $f : R^n \to R$ be a differentiable function R^n. Then*

(a) If $\bar{x}$ is a local minimizer of f over a box $C = \{x \in R^n : a_i \leq x_i \leq b_i\}$, then the first order condition becomes

$$\begin{aligned}
&\text{if } \bar{x}_i \in (a_i, b_i), \text{ then } \frac{\partial f}{\partial x_i}(\bar{x}) = 0 \\
&\text{if } \bar{x}_i = a_i \text{ then } \frac{\partial f}{\partial x_i}(\bar{x}) \geq 0 \\
&\text{if } \bar{x}_i = b_i \text{ then } \frac{\partial f}{\partial x_i}(\bar{x}) \leq 0
\end{aligned}$$

(b) Moreover suppose that f is a convex function on C. Then $\bar{x}$ is a global minimum of f over C if and only if the first order condition in (a) holds.

Theorem 5.1.4

Suppose that f is a convex function on a bounded closed convex set C. Furthermore suppose that f has a global maximum on C. Then one can find a global maximum which lies at the boundary point of C.

Proof. If a point x is in the interior of C, a line can be drawn through x which intersects the boundary at two points, say x_1 and x_2 since C is bounded and closed. Since $f(x)$ is convex, some λ exists in the range $0 < \lambda < 1$ such that $x = \lambda x_1 + (1-\lambda)x_2$ and $f(x) \le \lambda f(x_1) + (1-\lambda)f(x_2)$. If $f(x_1) > f(x_2)$, we have

$$f(x) < \lambda f(x_1) + (1-\lambda)f(x_1) = f(x_1).$$

If $f(x_1) < f(x_2)$, we have

$$f(x) < \lambda f(x_2) + (1-\lambda)f(x_2) = f(x_2).$$

Now if $f(x_1) = f(x_2)$, then $f(x) \le f(x_1) = f(x_2)$. Evidently, in all possibilities the maximization occurs on the boundary of C.

Theorem 5.1.5

Let f be a convex function defined on a convex set C. Then x^* is a local minimizer of f over C if and only if it is a global minimizer over C.

Proof. A global minimizer is obvious a local minimizer. So it suffices to prove that if x^* is a local minimizer of f on C then it must be a global minimizer as well. By definition of a local minimizer. There exists $r > 0$ such that

$$f(x) \geq f(x^*) \qquad \forall x \in C \text{ and } x \in B(x^*, r). \tag{5.1}$$

Now pick any $y \in C$, we wish to prove that $f(y) \geq f(x^*)$. Let $\lambda \in (0,1)$ small enough such that $x^* + \lambda(y - x^*) \in B(x^*, r)$. Since C is convex and x^*, y are both in C, we have $x^* + \lambda(y - x^*) \in C$. Then by (5.1) we have that

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

By convexity of f, we have

$$\lambda f(y) + (1 - \lambda) f(x^*) \quad \geq \quad f(x^* + \lambda(y - x^*)) \geq f(x^*)$$

which implies that $\lambda f(y) - \lambda f(x^*) \geq 0$ or equivalently $f(y) \geq f(x^*)$. The proof of the theorem is complete. ■

Separation and support theorem for convex sets

Theorem 5.2.1 (The closed point theorem)

Let $C \subset \mathbf{R}^n$ be a closed convex set and $y \notin C$. Then $x^* \in C$ is the closed point in C to y if and only if

$$(y - x^*) \cdot (x - x^*) \leq 0, \frac{\forall}{x} \in C.$$

Proof. $x^* \in C$ is the closest point in C to y if and only if x^* is a global minimum of the following optimization problem with simple constraints:

$$\begin{array}{ll} \min & f(x) := \|y - x\|^2 \\ s.t. & x \in C. \end{array}$$

Since $\nabla f(x^*) = -2(y - x^*)$, the conclusion follows from Theorem 5.1.1 (b). ■

Theorem 5.2.2 (The basic separation theorem) *Suppose that C is a closed convex set in R^n and $y \notin C$. Then there exists $0 \neq a \in R^n$ and $\alpha \in R$ such that*

$$a \cdot x \leq \alpha < a \cdot y \qquad \forall x \in C.$$

Remark: In R^n, any hyperplane can be written as

$$H = \{R^n : a \cdot x = \alpha\}$$

for some $a \neq 0$ in R^n and $\alpha \in R$.

$$H^- \quad = \quad \{R^n : a \cdot x \leq \alpha\}$$

is a closed half space and

$$H^+ \quad = \quad \{R^n : a \cdot x > \alpha\}$$

is an open half space. So Theorem 5.2.2 says that if $y \notin C$, a closed and convex set, the there exists a hyperplane H such that $C \subset H^-$ and $y \in H^+$. The direction of the inequalities are not important. The essence is that they have to be opposite, i.e., the conclusion can be also stated as

$$a \cdot x \geq \alpha > a \cdot y \qquad \forall x \in C.$$

Proof of Theorem 5.2.2. Let x^* be the closest vector in C to y, then by the closest point theorem,

$$(y - x^*) \cdot (x - x^*) \leq 0 \quad \forall x \in C.$$

Let $a = y - x^*$, then $a \neq 0$ and

$$a \cdot (x - x^*) \leq 0 \quad \forall x \in C,$$

i.e.,

$$a \cdot x \leq a \cdot x^* \quad \forall x \in C.$$

Let $\alpha = a \cdot x^*$. Since $a \cdot y - \alpha = a \cdot (y - x^*) = \|a\|^2 > 0$, we have

$$a \cdot y > \alpha \geq a \cdot x \qquad \forall x \in C.$$

Theorem 5.2.3 (Support Theorem)

Suppose $\varnothing \neq C \subset \mathbf{R}^n$ and z is a boundary point of C. Then there exists $0 \neq a \in \mathbf{R}^n$ such that

$$a \cdot x \leq a \cdot z, \frac{\forall}{x} \in C.$$

Proof. Since z is in the boundary of C, there exists a seqence $\{z_k\}$ not in clC such that $z_k \to z$. By the basic separation theorem, corresponding to each z_k there exists a a_k with norm 1 such that $a_k \cdot z_k > a_k \cdot x$ for each $x \in clC$ (In the basic separation theorem, the normal vector can be normalized by dividing it by its norm, so that $\|a_k\| = 1$.) Since $\{a_k\}$ is bounded, it has a convergent subsequence which we lable it the same with limit p whose norm is also equal to 1. Fixing $x \in clC$ and taking limits as k approaches ∞, we get, $p \cdot (x - \bar{x}) \leq 0$. Since this is true for each $x \in clC$, the result follows.

■

Theorem 5.2.4

Suppose that $\varnothing \neq C \subset \mathbf{R}^n$ and $f\colon C \to \mathbf{R}$ is convex. If $\bar{x} \in \operatorname{int} C$. Then there is a vector $d \in \mathbf{R}^n$ such that

$$f(x) \geq f(\bar{x}) + d \cdot (x - \bar{x}), \frac{\forall}{x} \in C.$$

Proof. Since f is convex, $epif := \{(x,r) \in R^{n+1} : x \in C, r \in R, r \geq f(x)\}$ is a convex set. It is obvious that $(\bar{x}, f(\bar{x})) \in epif$ and hence $epif$ is a nonempty convex set and $(\bar{x}, f(\bar{x}))$ is a boundary point of $epif$. By the support theorem, there exists $a = (b, c) \in R^n \times R$ such that

$$b \cdot x + cr = a \cdot (x, r) \leq a \cdot (\bar{x}, f(\bar{x})) = b\bar{x} + cf(\bar{x}) \quad \forall (x, r) \in epif$$

which implies that

$$b(x - \bar{x}) \leq c(f(\bar{x}) - r) \qquad \text{if } r \geq f(x). \tag{5.2}$$

Since r can go to infinity, c must be nonpositive. We now show that $c < 0$. If $c = 0$ then by (5.2) we have

$$b(x - \bar{x}) \leq 0 \qquad \forall x \in C.$$

But since $\bar{x}$ is an interior point of C, the above implies that $b = 0$ but this is not possible since $a = (b, c) \neq 0$. Hence $c < 0$. Since $c < 0$, (5.2) is equivalent to

$$\frac{b}{c}(x - \bar{x}) \geq (f(\bar{x}) - r) \qquad \text{if } r \geq f(x).$$

In particular let $r = f(x)$ in the above. We have

$$f(x) - f(\bar{x}) \geq -\frac{b}{c}(x - \bar{x}).$$

The proof is done if we let $d = -\frac{b}{c}$. ■

Subgradients of convex functions

Definition 5.3.1 (Subgradient) *Let $C \subset R^n$ be a convex set and f be a convex function defined on C. A vector $d \in R^n$ satisfying*

$$f(x) \geq f(\bar{x}) + d \cdot (x - \bar{x}) \quad \forall x \in C$$

is called a subgradient of f at $\bar{x}$.

Theorem 5.2.4 guarantees that if $\bar{x}$ is an interior point of a convex set C and f is a convex function then there exists at least one subgradient of f at $\bar{x}$, i.e., $\partial f(\bar{x}) \neq \emptyset$.

If f is not differentiable, then the subgradient may not be unique. We denote the set of all subgradients by

$$\partial f(\bar{x}) := \{d \in R^n : f(x) \geq f(\bar{x}) + d \cdot (x - \bar{x}) \quad \forall x \in C\}.$$

In the case when f is convex and differentiable at $\bar{x}$,

$$\partial f(\bar{x}) = \{\nabla f(\bar{x})\}.$$

Theorem 5.3.1 *Let $C \subset R^n$ be an open convex set and $f : R^n \to R$ be a convex function. Then x^* is a global minimizer of f on C if and only if $0 \in \partial f(x^*)$.*

Proof. x^* is a global minimizer of f on C if and only if

$$f(x) \geq f(x^*) \quad \forall x \in C,$$

or equivalently

$$f(x) \geq f(x^*) + 0 \cdot (x - x^*) \quad \forall x \in C$$

which by definition of the subgradient is equivalent to $0 \in \partial f(x^*)$.

■

Value functions and envelope theorem

Envelope Theorems for Unconstrained Problems

The objective function in economic optimization problems usually involves parameters like prices in addition to choice variables like quantities. Consider an objective function wih a parameter vector α of the form $f(x, \alpha) = f(x_1, \ldots, x_n, \alpha_1, \ldots, \alpha_k)$, where $x \in S \subseteq R^n$ and $\alpha \in R^k$. For each fixed α suppose that we have found the minimum of $f(x, \alpha)$ when x varies in S. The minimum value of $f(x, \alpha)$ usually depends on α. We denote this value by $V(\alpha)$ and call it the value function. Thus

$$V(\alpha) = \min_{x \in S} f(x, \alpha).$$

The vector x that minimizes $f(x, \alpha)$ depends on α and is denoted by $x^*(\alpha)$. Then $V(\alpha) = f(x^*(\alpha), \alpha)$. Note that there may be several choices of x that minimize $f(x,\alpha)$ for a given parameter vector α. Then we let $x^*(\alpha)$ denote one of these choices, and we may try to select x for different values of α so that $x^*(\alpha)$ is a differentiable function of α.

How does $V(\alpha)$ vary as the ith parameter α_i changes? Provided that $V(\alpha)$ is differentiable we have the following so-called envelope theorem:

$$\frac{\partial V}{\partial \alpha_i}(\alpha) = \frac{\partial f}{\partial \alpha_i}(x^*(\alpha),\alpha) \qquad i = 1,\ldots,k. \tag{5.3}$$

To see why, assume an interior solution and that $V(\alpha)$ is differentiable. Assume that there is only one parameter α. Then because $x = x^*(\alpha)$ minimizes $f(x,\alpha)$ with respect to x, all the partial derivatives $\partial f(x^*(\alpha),\alpha)/\partial x_i$ must be zero. Hence by the chain rule,

$$\begin{aligned} V'(\alpha) = \frac{\partial}{\partial \alpha} f(x^*(\alpha),\alpha) &= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x^*(\alpha),\alpha)\frac{dx_i^*}{d\alpha}(\alpha) + \frac{\partial f}{\partial \alpha}(x^*(\alpha),\alpha) \\ &= \frac{\partial f}{\partial \alpha}(x^*(\alpha),\alpha) \end{aligned}$$

The general case where $k > 1$ can be shown similarly and is left as an exercise.

Theorem 5.4.1 *(Envelope Theorem A) Let* $f(x,\alpha) : R^n \times R^k \to R$*. Let* $S \subseteq R^n$ *and consider the problem* $\min_{x\in S} f(x,\alpha)$*. Suppose that* $x^*(\alpha)$ *is a solution of this problem for every* α *in some open ball* $B(\bar{\alpha},\delta)$*, with* $\delta > 0$*. Furthermore assume that the mapping* $\alpha \to f(x^*(\bar{\alpha}),\alpha)$ *(with* $\bar{\alpha}$ *fixed) and the value function* $V(\alpha)$ *are both differentiable at* $\bar{\alpha}$*. Then*

$$\nabla V(\bar{\alpha}) = \nabla_\alpha f(x^*(\bar{\alpha}),\bar{\alpha}).$$

Proof. Define the function $\varphi(\alpha) := f(x^*(\bar{\alpha}), \alpha) - V(\alpha)$. Because $x^*(\bar{\alpha})$ is a minimum of $f(x, \alpha)$ when $\alpha = \bar{\alpha}$, one has $\varphi(\bar{\alpha}) = 0$. Also the definition of the value function implies that $\varphi(\alpha) \geq 0$ for all $\alpha \in B(\bar{\alpha}, \delta)$. Hence φ has an interior minimum at $\alpha = \bar{\alpha}$. The envelop result follows from the fact that $\nabla\varphi(\bar{\alpha}) = 0$. ■

A Geometric Illustration of the Envelope Theorem

Figure x illustrates the envelop result in the case where there is only one parameter α. For each fixed value of x, there is a curve K_x in the αy-plane, given by the equation $y = f(x, \alpha)$. The figure shows some of these curves together with the graph of $V(\alpha)$—that is, the curve $y = V(\alpha)$.

For all x and all α we have $f(x, \alpha) \geq \min_x f(x, \alpha) = V(\alpha)$. It follows that none of the K_x-curves can ever lie below the cure $y = V(\alpha)$. On the other hand, for each value of α there is at least one value x^* of x such that $f(x^*, \alpha) = V(\alpha)$, namely the choice of x^* that solves the minimization problem for the given value of α. The curve K_{x^*} will then just tourch the curve $y = V(\alpha)$ at point $(x^*, V(\alpha)) = (x^*, f(x^*, \alpha))$, and so much have exactly the same tangent as the graph of $V(\alpha)$ at this point. Moreover the slope of this common tangent must be both $V'(\alpha)$, the slope of the tangent to the graph of V, and $\partial f(x^*, \alpha)/\partial\alpha$, the slope of the tangent to the curve K_{x^*}, which is the graph of $f(x^*, \alpha)$ when x^* is fixed.

As the figure suggests, the graph of $y = V(\alpha)$ is the highest curve with the property that it lies on or below all the curves K_x, so its graph is like an envelope that is used to "wrap" all these curves; that is why we call the graph of $V(\alpha)$ the envelope of the family of K_x-curves.

Theorem 5.4.2 *(Envelope Theorem B) Let $f(x,\alpha)$ be a C^2 function for all x in an open convex set $S \subset R^n$ and for each α in an open ball $B(\bar{\alpha},\delta) \subseteq R^k$. Assume that for each fixed α in $B(\bar{\alpha},\delta)$, the function $x \to f(x,\alpha)$ is convex, and that when $\alpha = \bar{\alpha}$ the Hessian matrix of the function f with respect to x is positive definite. Moreover, assume that x^* is a minimum for $x \to f(x,\bar{\alpha})$ in S. Then $V(\alpha) = \min_{x \in S} f(x,\alpha)$ is defined for all α in an open ball around $\bar{\alpha}$. Moreover the value function $V(\alpha)$ is C^1 at $\bar{\alpha}$ and*

$$\frac{\partial V}{\partial \alpha_i}(\bar{\alpha}) = \frac{\partial f}{\partial \alpha_i}(x^*,\bar{\alpha}) \qquad i = 1,\ldots,k.$$

Proof. Consider the first order condition:

$$0 = \nabla_x f(x, \alpha).$$

Since f is a C^2 function, the Hessian matrix $\nabla^2_{xx} f(x, \alpha)$ is positive definite in some open ball centered at (x^*, α) and hence nonsingular. By the implicite function theorem, it follows that the equality system $0 = \nabla_x f(x, \alpha)$ in the unknown vector x has a unique solution $x(\alpha)$ which is a C^1 function of α in some ball $B(\bar{\alpha}, \varepsilon)$, and moreover $x(\bar{\alpha}) = x^*$. Provided that α lies in $B(\bar{\alpha}, \varepsilon) \cap B(\bar{\alpha}, \delta)$, the function $x \to f(x, \alpha)$ is convex, so $x(\alpha)$ is a minimum of $x \to f(x, \alpha)$ in S. Because $x(\alpha)$ is differentiable at $\alpha = \alpha^*$, so is $V(\alpha) = f(x(\alpha), \alpha)$. In particular, Theorem 5.4.1 applies.

Theorem 5.4.3 *(Envelope Theorem C) Suppose that* $V(\alpha) = \inf_{x \in S} f(x, \alpha)$ *is finite and convex in* $\alpha \in A$*, where* A *is an open convex set in* R^k*, and* $S \subseteq R^n$*. Assume that the point* $(x^*, \bar{\alpha}) \in S \times A$ *satisfies* $f(x^*, \bar{\alpha}) = V(\bar{\alpha})$ *and that the gradient vector* $\nabla_\alpha f$ *exists at* $(x^*, \bar{\alpha})$*. Then the value function* $V(\alpha)$ *is diffentiable at* $\bar{\alpha}$ *and*

$$\nabla V(\bar{\alpha}) = \nabla_\alpha f(x^*, \bar{\alpha}).$$

Proof. Because A is open convex set and $V(\alpha)$ is a convex function, for any $\xi \in \partial V(\bar{\alpha})$ one has

$$f(x^*,\alpha) - f(x^*,\bar{\alpha}) \geq V(\alpha) - V(\bar{\alpha}) \geq \xi \cdot (\alpha - \bar{\alpha}) \qquad \forall \alpha \in A.$$

This implies that any ξ in $\partial V(\bar{\alpha})$ is a subgradient of the function $\alpha \to f(x^*,\alpha)$ at $\bar{\alpha}$. But by assumption the function $\alpha \to f(x^*,\alpha)$ at $\bar{\alpha}$ is differentiable at $(x^*,\bar{\alpha})$,

therefore the subgradient $\partial V(\bar{\alpha})$ must be a singleton equal to $\{\nabla_\alpha f(x^*,\bar{\alpha})\}$. This means that V is differentiable at $\bar{\alpha}$ and

$$\nabla V(\bar{\alpha}) = \nabla_\alpha f(x^*,\bar{\alpha}).$$