Booth Math Camp 2018: Optimization Theory

Jianfei Cao*

September 12, 2018

1 Unconstrained Optimization

We will only talk about maximization problem. Any minimization problem can be transformed into a maximization problem by flipping the sign. An unconstrained optimization problem is: for some $S \subset \mathbb{R}^k$ and some function $f : S \to \mathbb{R}$,

$$\max_{x \in S} \; f(x),$$

which we call "the unconstrained optimization problem" through this section. Let x^* be one of the maximizers, then we write

$$x^* \in \arg\max_{x \in S} \; f(x).$$

If the optimization problem has a unique solution, we write

$$x^* = \arg\max_{x \in S} \; f(x).$$

Example 1. *(monopolist profit maximization) A monopolist firm solves the profit maximization problem*

$$\max_{p \in [0,\infty)} \; p \cdot q(p) - c(q(p)),$$

where p is price, $q(\cdot)$ is the demand function, and $c(\cdot)$ is the cost function.

Definition 1. *For an optimization problem, let S be the set of points that satisfy all constraints. Then, $x^* \in S$ is a* ***local maximum*** *if $\exists \epsilon > 0$, such that $\|x^* - x\| < \epsilon$ implies $f(x^*) \geq f(x)$. We say $x^* \in S$ is a* ***strict local maximum*** *if $\exists \epsilon > 0$, such that for $x \neq x^*$, $\|x^* - x\| < \epsilon$ implies $f(x^*) > f(x)$. Also, $x^* \in S$*

*Questions or comments: jcao0@chicagobooth.com

is a ***global maximum*** *if* $\forall x \in S$, $f(x^*) \geq f(x)$*; accordingly,* x^* *is said to be a* ***strict global maximum*** *if* $\forall x \in S$, $f(x^*) > f(x)$.

We introduce a set of necessary conditions for local maxima.

Theorem 1. *(First Order Condition/FOC)* $x \in S$ *is said to be an* ***interior point*** *of* S *if* $\exists$ *an open set* $U \subset S$ *such that* $x \in U$*. Let* f *be differentiable at some* $x^* \in S$*. For the unconstrained optimization problem, if* x^* *is an interior point of* S *and a local minimum of* f*, then we must have* $\nabla f(x^*) = 0$.

Proof. Suppose $\nabla f(x^*) \neq 0$. WLOG, let $f_1(x^*) > 0$. Pick ϵ such that $0 < \epsilon < f_1(x^*)$. Note that

$$f_1(x^*) = \lim_{h \downarrow 0} \frac{f(x^* + he_1) - f(x^*)}{h},$$

so there exists $\delta > 0$ such that $|h| < \delta$ implies

$$\left| \frac{f(x^* + he_1) - f(x^*)}{h} - f_1(x^*) \right| < \epsilon.$$

This implies for each $0 < h < \delta$,

$$f_1(x^*) - \epsilon < \frac{f(x^* + he_1) - f(x^*)}{h} < f_1(x^*) + \epsilon,$$

i.e.

$$f(x^* + he_1) > f(x^*) + (f(x^*) - \epsilon)h.$$

Thus, within any neighborhood of x^* with radius d, pick h such that $0 < h < \min\{\delta, d\}$ and we have

$$f(x^* + he_1) > f(x^*),$$

contradicting x^* being local minimum. $\square$

We call $x \in S$ a **critical point** if $\nabla f(x) = 0$. That means, when solving the unconstrained optimization problem, we only need to check all the critical points and boundary points of S.

For some differentiable function $f : S \to \mathbb{R}$ with $S \subset \mathbb{R}^k$, the gradient of f is

$$\nabla f(x_1, \ldots, x_k) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x_1, \ldots, x_k) \\ \vdots \\ \frac{\partial}{\partial x_k} f(x_1, \ldots, x_k) \end{bmatrix}.$$

If f is twice differentiable, we write the second partial derivative for some $i, j \in \{1, 2, \ldots, k\}$ as

$$f_{ij}(x_1, \ldots, x_k) = \frac{\partial^2}{\partial x_i \partial x_j} f(x_1, \ldots, x_k) = \frac{\partial}{\partial x_j} \left(\frac{\partial}{\partial x_i} f(x_1, \ldots, x_k) \right).$$

We define the Hessian of f at $x = (x_1, \ldots, x_k)$ as

$$\nabla^2 f(x_1, \ldots, x_k) = \begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1k} \\ f_{21} & f_{22} & \cdots & f_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ f_{k1} & f_{k2} & \cdots & f_{kk} \end{bmatrix}.$$

Definition 2. *A function $f : S \to \mathbb{R}$ is **continuously differentiable** if the partial derivative $\partial f / \partial x_i$ is continuous for each i. We say f is **twice continuously differentiable** if for each $i, j \in \{1, 2, \ldots, k\}$, $f_{i,j} : S \to \mathbb{R}$ is a continuous function.*

Theorem 2. *(Second Order Condition/SOC) Suppose f is twice continuously differentiable in the unconstrained optimization problem. If x^* is an interior point of S and a local maximum, then $\nabla^2 f(x^*)$ is negative semi-definite.*

Unless the problem is univariate, we rarely use the SOC because it is painful to check whether a matrix is negative semi-definite.

Theorem 3. *(Envelope Theorem) Let $f(x; a)$ be a function of $x \in S \subset \mathbb{R}^k$ and $a \in A \subset \mathbb{R}$. Assume $f(x; a)$ is continuously differentiable in $x \in S$ for each $a \in A$ and S is open. For each $a \in A$, let $x^*(a) = \arg\max_{x \in S} f(x; a)$ and assume $x^*(a)$ is continuously differentiable in a. Then,*

$$\frac{d}{da} f(x^*(a); a) = \frac{\partial}{\partial a} f(x^*(a); a) = \left. \frac{\partial}{\partial a} f(x; a) \right|_{x = x^*(a)}.$$

Proof. Write out the total derivative and apply the first order condition. □

Example 2. *Suppose the profit of a firm is determined by*

$$\pi(x; p) = px - c(x),$$

where x is quantity, p is the price of the product, and $c(x)$ is differentiable. We assume the firm is a price-taker such that we treat p as given. Assume c is smooth. The FOC of the profit maximization problem *requires $p = c'(x^*)$, where x^* is the maximizer. Then, the Envelope Theorem says*

$$\frac{d}{dp}\pi(x^*(p); p) = x^*(p).$$

2 Constrained optimization

Example 3. *(consumer utility maximization) A consumer facing budget constraint solves the optimization problem*

$$\begin{aligned}
\max_{x_1,\ldots,x_n} \quad & u(x_1, \ldots, x_n) \\
s.t. \quad & p_1x_1 + \cdots + p_nx_n \leq m \\
& x_i \geq 0 \textit{ for } i = 1, \ldots, n,
\end{aligned}$$

where $u(\cdot)$ is the utility function, p_i the price for x_i, and m is the budget.

A constrained optimization is following: for some open set $S \subset \mathbb{R}^k$,

$$\begin{aligned}
\max_{x \in S} \quad & f(x) \\
s.t. \quad & g_i(x) \geq 0, \ \forall i = 1, \ldots, n \\
& h_j(x) = 0, \ \forall j = 1, \ldots, m.
\end{aligned}$$

The Lagrangian of this optimization problem is

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \sum_{i=1}^{n} \lambda_i g_i(x) + \sum_{j=1}^{m} \mu_j h_j(x) = f(x) + \lambda^T g(x) + \mu^T h(x),$$

for some $\lambda = (\lambda_1, \ldots, \lambda_n)'$, $\mu = (\mu_1, \ldots, \mu_m)'$, $g(x) = (g_1(x), \ldots, g_n(x))'$, and $h(x) = (h_1(x), \ldots, h_m(x))'$. The Karush-Kuhn-Tucker conditions (KKT) are given by

1. (primal constraints) $g_i(x) \geq 0$, $\forall i = 1, \ldots, n$; $h_j(x) = 0$, $\forall j = 1, \ldots, m$.

2. (dual constraints) $\lambda_i \geq 0$, $\forall i = 1, \ldots, n$.

3. (complementary slackness) $\lambda_i g_i(x) = 0$, $\forall i = 1, \ldots, n$.

4. (vanishing gradient) $\nabla_x \mathcal{L} = 0$.

KKT is a set of conditions that is necessary for optimality under regularity conditions (say, f, g_i, h_j are continuously differentiable).

Example 4.

$$\max_{x,y}\ x - y^2$$
$$s.t.\ x \geq 0, y \geq 0$$
$$x^2 + y^2 = 4.$$

KKT conditions specify:

$$\begin{aligned} Primal &: x \geq 0, y \geq 0, x^2 + y^2 = 4 \\ Dual &: \lambda_1 \geq 0, \lambda_2 \geq 0 \\ Complementary\ Slackness &: \lambda_1 x = 0, \lambda_2 y = 0 \\ Gradient &: 1 + \lambda_1 + 2\mu x = 0, -2y + \lambda_2 + 2\mu y = 0, \end{aligned}$$

which yields $(x^*, y^*) = (0, 4)$.

Example 5. *(simplified utility maximization) Suppose there are two goods and ignore the positivity constraint. A consumer facing budget constraint solves the optimization problem*

$$\max_{x_1,x_2}\ u(x_1, x_2)$$
$$s.t.\ p_1 x_1 + p_2 x_2 \leq m$$

The KKT conditions specify:

$$\begin{aligned} Primal &: m - p_1 x_1 - p_2 x_2 \geq 0 \\ Dual &: \lambda \geq 0 \\ Complementary\ Slackness &: \lambda(m - p_1 x_1 - p_2 x_2) = 0 \\ Gradient &: \partial u/\partial x_1 - \lambda p_1 = 0, \partial u/\partial x_2 - \lambda p_2 = 0, \end{aligned}$$

implying

$$\lambda = \frac{\partial u/\partial x_1}{p_1} = \frac{\partial u/\partial x_2}{p_2},$$

i.e. utility increment of spending one more dollar on x_1 is equal to that of x_2.

3 Convex Optimization

Definition 3. *A set $X \subset \mathbb{R}^k$ is a **convex set** if for each $\theta \in (0,1)$ and $x_1, x_2 \in X$, $\theta x_1 + (1-\theta)x_2 \in X$, where $x = \theta x_1 + (1-\theta)x_2$ is called a **convex combination** of x_1 and x_2.*

That is, a convex set is closed under convex combination. The **convex hull** of a set S is the set of all convex combinations of points in S.

Proposition 1. *Suppose $A, B \subset X$. If A and B are convex, then $A \cap B$ is convex.*

Proof. Use definition. □

A set X is called a **hyperplane** if $X = \{x \in \mathbb{R}^k | a^T x = b\}$ for some nonzero $a \in \mathbb{R}^k$ and $b \in \mathbb{R}$.

Theorem 4. *(Separating Hyperplane Theorem) If $C, D \subset \mathbb{R}^k$ are nonempty disjoint convex sets, there exists a nonzero vector $a \in \mathbb{R}^k$ and $b \in \mathbb{R}$ such that $a^T x \le b$ for each $x \in C$ and $a^T x \ge b$ for each $x \in D$.*

Theorem 5. *(Supporting Hyperplane Theorem) A supporting hyperplane to a set C at a boundary point x_0 is a hyperplane X such that $X = \{x \in \mathbb{R}^k | a^T x = a^T x_0\}$ for some nonzero vector $a \in \mathbb{R}^k$, and $a^T x \le a^T x_0$ for each $x \in C$. If C is a convex set, then there exists a supporting hyperplane to C at each boundary point of C.*

Definition 4. *A function $f : X \to \mathbb{R}$ is a **convex function** if X is convex and for each $\theta \in (0,1)$ and $x_1, x_2 \in X$, $f(\theta x_1 + (1-\theta)x_2)) \le \theta f(x_1) + (1-\theta) f(x_2)$. We say f is **strictly convex** if X is convex and for each $\theta \in (0,1)$ and $x_1, x_2 \in X$, $f(\theta x_1 + (1-\theta)x_2)) < \theta f(x_1) + (1-\theta) f(x_2)$.*

Definition 5. *A function $f : X \to \mathbb{R}$ is **concave** if $-f$ is convex. We say f is **strictly concave** if $-f$ is strictly convex.*

Lemma 1. *Let $X \subset \mathbb{R}^k$. If $f : X \to \mathbb{R}$ is convex, then $C = \{(x,y) \in \mathbb{R}^{k+1} : y \ge f(x), x \in X\}$ is convex.*

Proof. Let $z_1 = (x_1, y_1)$, $z_2 = (x_2, y_2) \in C$. Then we have $y_1 \ge f(x_1)$ and $y_2 \ge f(x_2)$. Want to show convex combination $z = \lambda z_1 + (1-\lambda) z_2 \in C$ for $\lambda \in (0,1)$. Note that by convexity $f(\lambda x_1 + (1-\lambda)x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2) \le \lambda y_1 + (1-\lambda) y_2$, so $z = (\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda) y_2) \in C$. □

Theorem 6. *(Jensen's Inequality) If $S \subset \mathbb{R}^k$ and $f : S \to \mathbb{R}$ is convex, then for some random variable X such that $\Pr(X \in S) = 1$ and $E[|X|] < \infty$, we have $f(E[X]) \le E[f(X)]$.*

Proof. Let $z_0 = (E[X]^T, f(E[X]))^T$. By the previous lemma and the supporting hyperplane theorem, at $z_0 \in \mathbb{R}^k + 1$, there exists a supporting hyperplane to the set $C = \{(x, y) \in \mathbb{R}^{k+1} : y \geq f(x), x \in S\}$ such that $Z = \{z \in \mathbb{R}^{k+1} : a^T z = a^T z_0\}$, where we write $a = (a_x^T, a_y)^T$ to denote the x part and y part, separately. Note $a_y \neq 0$, unless S is singleton, in which case Jensen's inequality trivially holds. So for each $(x^T, y)^T \in Z$, we can write $y = -a_y^{-1} a_x^T x - a_y^{-1} a^T z_0$ and note $y \leq f(x)$. Therefore,

$$E[f(X)] \geq E[-a_y^{-1} a_x^T X - a_y^{-1} a^T z_0] = -a_y^{-1} a_x^T E[X] - a_y^{-1} a^T z_0 = f(E[X]).$$

The first equality is because of linearity. The last equality is because $z_0 \in Z$. □

A convex optimization problem is a constrained optimization problem where f and $g_1, \ldots, g_n$ are all concave functions, and $h_1, \ldots, h_m$ are affine functions.

Lemma 2. *If $g : S \to \mathbb{R}$ is concave, then $\{x \in S : g(x) \geq 0\}$ is convex.*

Proof. Use definition. □

Theorem 7. *A local maximum of a convex optimization problem is a global maximum.*

Proof. Let S be the set of points that satisfy all constraints. Then S is convex by previous lemma and proposition. The optimization problem becomes $\max_{x \in S} f(x)$, for S convex and f concave. Let x_0 be a local maximization to this problem, and suppose for contradiction that there exists $x_1 \in S$ such that $f(x_1) > f(x_0)$. For each $\epsilon > 0$, pick λ such that $0 < \lambda < \epsilon / \|x_0 - x_1\|$. Let the convex combination be $z = \lambda x_1 + (1 - \lambda) x_0$, and we have $\|x_0 - z\| = \lambda \|x_0 - x_1\| \leq \epsilon$, so $z \in B_\epsilon(x_0)$. But $f(z) \geq \lambda f(x_1) + (1 - \lambda) f(x_0) > f(x_0)$, contradicting x_0 being local maximum. □

References

[1] Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*, Chapter 2-4.

[2] Mas-Colell, A., Whinston, M. D., & Green, J. R. (1995). *Microeconomic theory*, Mathematical Appendix.

[3] Simon, C. P., & Blume, L. (1994). *Mathematics for economists*, Chapter 17-19.