KKT Condition for convex program

Instructor: Jin Zhang

Department of Mathematics
Southern University of Science and Technology
Spring 2024

Geometric interpretation of KKT conditions

- **Two variables and one equality constraint:** Consider the following constrained optimization problem

$$\begin{aligned} \min\ & x_1^2 + x_2^2 \\ \text{s.t. } & x_2 = 1 - \frac{1}{4}x_1^2. \end{aligned}$$

- **Two variables and one inequality constraint:** Conside the following constrained optimization problem

$$\begin{aligned} \min\ & x_1^2 + x_2^2 \\ \text{s.t. } & x_2 \geq 1 + x_1^2. \end{aligned}$$

- **Two variables and two constraints:**

$$\begin{aligned} \min\ & f(x, y) := -x - y \\ \text{s.t. } & g_1(x, y) := x + y^2 - 5 \leq 0 \\ & g_2(x, y) := x - 2 \leq 0. \end{aligned}$$

Value functions

For an optimization problem

$$
\begin{aligned}
(P)\quad & \min\ f(x) \\
& \text{s.t. } g(x) \leq 0 \\
& \qquad x \in C \subseteq \mathbb{R}^n.
\end{aligned}
$$

where $g(x) \leq 0$ means that $g_i(x) \leq 0,\ i = 1, \cdots, m$. Consider a perturbed problem

$$
\begin{aligned}
(P_z)\quad & \min\ f(x) \\
& \text{s.t. } g(x) \leq z,\ x \in C,
\end{aligned}
$$

where $C \in \mathbb{R}^m$.

- Denote by F the set of all feasible solutions of problem (P) and $\mathcal{F}_z$ the set of all feasible solutions of (P_z), i.e., $\mathcal{F}_z := \{x \in C : g(x) \leq z,\ x \in C\}$.
- Define the value function $V(z) := \inf\{f(x) : g(x) \leq z,\ x \in C\}$.

Theorem 6.3.1 *If (P) is a convex program, then the value function $V(z)$ is a convex function and its domain is a convex subset of C. Moreover if the Slater condition holds for (P), then zero is in the interior of the domain of the value function.*

Proof. Step 1: Prove that dom$V$ is convex. Let $z_1, z_2 \in domV$ and $\lambda \in [0,1]$. Then by the definition of a domain, $\mathcal{F}_{z_1}$ and $\mathcal{F}_{z_2}$ are both nonempty. This means that there exist $x_1 \in \mathcal{F}_{z_1}$ and $x_2 \in \mathcal{F}_{z_2}$. That is, $x_1, x_2 \in C$ and $g(x_1) \leq z_1, g(x_2) \leq z_2$. By the convexity of g, we have

$$g(\lambda x_1 + (1-\lambda)x_2) \leq \lambda g(x_1) + (1-\lambda)g(x_2) \leq \lambda z_1 + (1-\lambda)z_2.$$

By the convexity of C, $\lambda x_1 + (1-\lambda)x_2 \in C$ and hence $\lambda x_1 + (1-\lambda)x_2 \in \mathcal{F}_{\lambda z_1 + (1-\lambda)z_2}$. Therefore $\mathcal{F}_{\lambda z_1 + (1-\lambda)z_2} \neq \emptyset$. Or equivalently, $\lambda z_1 + (1-\lambda)z_2 \in domV$. Hence $domV$ is a convex set.

Step 2: Prove that $V(z)$ is a convex function defined on the convex set $domV$. Let $z_1, z_2 \in domV$ and $\lambda \in [0,1]$. Then

$$
\begin{aligned}
& V(\lambda z_1 + (1-\lambda)z_2) \\
:= \quad & \inf\{f(x) : x \in C, g(x) \leq \lambda z_1 + (1-\lambda)z_2\} \\
\leq \quad & \inf\{f(\lambda x_1 + (1-\lambda)x_2) : x_1, x_2 \in C, g(\lambda x_1 + (1-\lambda)x_2) \leq \lambda z_1 + (1-\lambda)z_2\} \\
\leq \quad & \inf\{f(\lambda x_1 + (1-\lambda)x_2) : x_1, x_2 \in C, \lambda g(x_1) + (1-\lambda)g(x_2) \leq \lambda z_1 + (1-\lambda)z_2\} \\
\leq \quad & \inf\{f(\lambda x_1 + (1-\lambda)x_2) : x_1, x_2 \in C, g(x_1) \leq z_1, g(x_2) \leq z_2\} \\
\leq \quad & \inf\{\lambda f(x_1) + (1-\lambda)f(x_2) : x_1, x_2 \in C, g(x_1) \leq z_1, g(x_2) \leq z_2\} \\
= \quad & \lambda V(z_1) + (1-\lambda)V(z_2).
\end{aligned}
$$

So $V(z)$ is convex.

Step 3: Prove that $0 \in int(domV)$ under the Slater condition. By definition of an interior point. We want to prove that there exists a $r > 0$ such that $B(0,r) \subseteq domV$. Eqivalently, we want to find an $r > 0$ such that for all $z \in B(0,r)$, $\mathcal{F}_z \neq \emptyset$. The Slater condition means that there exists $x^0 \in C$ such that $g_i(x^0) < 0, i = 1, \dots, m$. Since the strict inequality holds, it is easy to see that for small enough the strict inequality $g_i(x^0) < z_i, i = 1, \dots, m$ still holds. Indeed, let $r = \min\{-g_i(x^0) : 1 \leq i \leq m\}$. Then $r > 0$ and for all $z \in B(0,r)$, $\sqrt{\sum_{i=1}^m z_i^2} < r$ would imply that $|z_i| < r \; i = 1, \dots, m$. So

$$g_i(x^0) \leq -r < z_i \quad i = 1, \dots, m$$

which implies that $x^0 \in \mathcal{F}_z$.

Since for any $z_1 \leq z_2$, $\mathcal{F}_{z_1} \subset \mathcal{F}_{z_2}$, the value function is a monotone decreasing extended-valued function of z. So the supporting hyperplane (if there is any) must have nonpositive slope as shown in the following theorem.

Theorem 6.3.2 *If (P) is a convex program, the Slater condition holds and $V(0)$ is finite, then $V(z)$ is finite on $domV$ and there exists $\lambda \in R^m$ such that $\lambda \geq 0$ and*

$$V(z) \geq V(0) - \lambda \cdot z \qquad \forall z \in \ domV.$$

Note that from the definition of a subgradient, $-\lambda$ is a subgradient of $V(z)$ at 0, i.e., $-\lambda \in \partial V(0)$. Hence we sometimes call λ a sensitivity vector since loosely speaking $-\lambda$ is the rate of changes of the value function with respect to the change in the right-hand side of the inequality $g(x) \leq 0$.

Proof. Since

$$domV = \{z \in R^m : \mathcal{F}_z \neq \emptyset\} = \{z \in R^m : V(z) \neq +\infty\},$$

to prove that $V(z)$ is finite on $domV$, it suffices to prove that $V(z) \neq -\infty$ for all $z \in domV$. By Theorem 6.3.1, $domV$ is a convex set and 0 is an interior point of $domV$. Let $z \neq 0$ and $z \in domV$, then there is $\lambda \in [0, 1]$ and $z_1 \in domV$ such that $0 = \lambda z_1 + (1 - \lambda)z$. This implies that

$$\begin{aligned} V(0) &= V(\lambda z_1 + (1-\lambda)z) \\ &\leq \lambda V(z_1) + (1-\lambda)V(z). \end{aligned}$$

Since $V(0) \neq -\infty$, the above inequality implies that $V(z) \neq -\infty$ as well.

Now since 0 is an interior point of the convex set $domV$ and $V(z)$ is convex, Theorem 5.2.4 implies that there exists a vector $-\lambda \in R^m$ such that

$$V(z) \geq V(0) - \lambda \cdot z \qquad \forall z \in \text{ dom}V.$$

We remain to prove that $\lambda \geq 0$. Notice that if $z \geq 0$, then

$$\mathcal{F}_z := \{x \in R^n : g(x) \leq z, x \in C\} \supseteq \{x \in R^n : g(x) \leq 0, x \in C\}.$$

So $V(z) \leq V(0)$ if $z \geq 0$. We now show that one must have $\lambda \geq 0$ by contradiction. Otherwise if $\lambda_i < 0$ for some i. Let $z^i = (0, \ldots, r_i, 0, \ldots, 0)$ with $r_i > 0$ small enough such that $z^i \in domV$. Then

$$V(z^i) \geq V(0) - \lambda \cdot z^i > V(0)$$

which is impossible. ■

Theorem 6.3.3 *Suppose that the convex program (P) has a sensitivity vector $\lambda \geq 0$. Then*

$$V(0) := \inf_{g(x) \leq 0, x \in C} f(x) = \inf_{x \in C} \{f(x) + \lambda \cdot g(x)\}.$$

Proof. Since $\lambda \geq 0$ we have

$$f(x) + \lambda \cdot g(x) \leq f(x) \quad \forall x \text{ such that } g(x) \leq 0.$$

Hence

$$\inf_{x \in C} \{f(x) + \lambda \cdot g(x)\} \leq \inf_{x \in C, g(x) \leq 0} f(x) = V(0). \tag{6.2}$$

On the other hand, pick any $x \in C$ and let $z := g(x)$. Then

$$\begin{aligned} V(g(x)) &= \inf_{x' \in C} \{f(x') : x' \in C, g(x') \leq g(x)\} \\ &\leq f(x) \qquad \text{since } x \in C \text{ and } g(x) \leq g(x) \end{aligned}$$

which implies that

$$f(x) + \lambda \cdot g(x) \quad \geq \quad V(g(x)) + \lambda \cdot g(x).$$

But since $-\lambda$ is a sensitivity vector we have

$$V(g(x)) \geq V(0) - \lambda \cdot g(x).$$

Therefore

$$\inf_{x \in C} \{f(x) + \lambda \cdot g(x)\} \geq V(0). \tag{6.3}$$

By (6.2)-(6.3), we have the desired conclusion. ∎

Corollary 6.3.1 *If (P) is a convex program, the Slater condition holds and $V(0)$ is finite, then there exists $\lambda \geq 0$ such that*

$$V(0) := \inf_{x \in C, g(x) \leq 0} f(x) = \inf_{x \in C} \{f(x) + \lambda \cdot g(x)\}.$$

Theorem 6.3.4 (KKT condition for convex program) *Suppose that (P) is a convex program and the Slater condition holds. Let* f, g_i *be* C^1. *If* x^* *is feasible, then* x^* *is a solution to (P) if and only if there exists* $\lambda^* \in R^m$ *such that the KKT condition holds:*

(1) $\lambda_i^* \geq 0 \; i = 1, \ldots, m;$

(2) $\lambda_i^* g_i(x^*) = 0 \; i = 1, \ldots, m;$

(3) $\langle \nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*), x - x^* \rangle \geq 0 \quad \forall x \in C.$

Proof. By Corollary 6.3.1, there exists $\lambda^* \geq 0$ such that

$$V(0) = \inf_{x \in C} \{f(x) + \sum_{i=1}^{m} \lambda_i^* g_i(x)\}.$$

Let x^* be a solution of (P), then $V(0) = f(x^*)$ and hence the above equality implies that

$$\begin{aligned} f(x^*) &= \inf_{x \in C} \{f(x) + \sum_{i=1}^{m} \lambda_i^* g_i(x)\} \\ &\leq f(x^*) + \sum_{i=1}^{m} \lambda_i^* g_i(x^*) \quad \text{since } x^* \in C \\ &\leq f(x^*) \qquad \text{since } \lambda_i^* \geq 0, g_i(x^*) \leq 0. \end{aligned}$$

Therefore $f(x^*) = f(x^*) + \sum_{i=1}^{m} \lambda_i^* g_i(x^*)$ which implies (2).

Also by Theorem 6.3.3, a feasible point x^* is an optimal solution of (P) if and only if it is a solution to

$$\begin{array}{ll} \min & f(x) + \sum_{i=1}^{m} \lambda_i^* g_i(x) \\ s.t. & x \in C. \end{array}$$

Since C is convex and the objective function of the above problem is convex, by Thoerem 5.1.1(ii) x^* is an optimal solution of the above convex optimization problem with simple constraints if and only if

$$\langle \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*), x - x^* \rangle \geq 0 \quad \forall x \in C.$$

which is (3).

Corollary 6.3.2 *Suppose that (P) is a convex program and the Slater condition holds. Let f, g_i be C^1. If x^* is feasible and an interior point of C, then x^* is a solution to (P) if and only if there exists $\lambda^* \in R^m$ such that the KKT condition holds:*

(1) $\lambda_i^* \geq 0 \; i = 1, \dots, m;$

(2) $\lambda_i^* g_i(x^*) = 0 \; i = 1, \dots, m;$

(3) $\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) = 0$

Theorem 6.3.5 *Suppose that (P) is a convex program and that* $x^* \in C$ *is a solution of (P). If* λ^* *is a Lagrange multiplier. Then* $-\lambda^* \in \partial V(0)$.

Proof. We want to prove that if Theorem 6.3.4 (1)-(3) holds, then

$$V(z) \geq V(0) - \lambda^* z \qquad \forall z \in domV.$$

Let z be a point in $domV$. Then there exists $x \in C$ such that $g(x) \leq z$. Since $x^* \in C$ is a solution of (P) and $f(x) + \lambda^* \cdot g(x)$ is a convex function, Theorem 6.3.4 (3) implies that x^* is a global minimizer of $L(x, \lambda^*) := f(x) + \lambda^* \cdot g(x)$ on C. So

$$\begin{aligned}
V(0) &= f(x^*) + \sum_{i=1}^{m} \lambda_i^* g_i(x^*) && \text{by Theorem 6.3.4 (2)} \\
&\leq f(x) + \sum_{i=1}^{m} \lambda_i^* g_i(x) && \text{since } x^* \text{ is a global minimizer of } L(x, \lambda^*) \\
&\leq f(x) + \lambda^* \cdot z && \forall x \in C \quad \text{such that } g(x) \leq z.
\end{aligned}$$

Taking infimum on both sides of the above inequality, we have

$$V(0) \leq \inf_{x \in C, g(x) \leq z} f(x) + \lambda^* \cdot z = V(z) + \lambda^* \cdot z.$$

Dual convex program

Theorem 6.4.1 (KKT condition in saddle point form) *Suppose that (P) is a convex program and the Slater condition holds. Then a feasible solution x^* is a solution of (P) if and only if there exists $\lambda^* \in R^m$ such that*

(1) $\lambda_i^* \geq 0 \ i = 1, \ldots, m;$

(2) $\lambda_i^* g_i(x^*) = 0 \ i = 1, \ldots, m;$

(3) $L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*) \quad \forall x \in C, \lambda \geq 0.$

Proof. By Corollarly 6.3.1, one can find λ^* such that (1) holds and

$$V(0) = \inf_{x \in C} L(x, \lambda^*).$$

Since $x^* \in C$ is a solution of (P), we have

$$x^* \in C, g(x^*) \leq 0 \quad \text{and } f(x^*) = V(0).$$

Hence

$$\begin{aligned} f(x^*) &= \inf_{x \in C} \{f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)\} \\ &\leq f(x^*) + \sum_{i=1}^{m} \lambda_i^* g_i(x^*) \\ &\leq f(x^*) \end{aligned}$$

which implies (2). Let us prove (3).

$$
\begin{aligned}
L(x^*, \lambda) &:= f(x^*) + \sum_{i=1}^{m} \lambda_i g_i(x^*) \\
&\leq f(x^*) \quad \text{since } \lambda \geq 0, g(x^*) \leq 0 \\
&= f(x^*) + \sum_{i=1}^{m} \lambda_i^* g_i(x^*) \quad \text{by (2)} \\
&= L(x^*, \lambda^*).
\end{aligned}
$$

On the other hand,

$$
L(x^*, \lambda^*) = f(x^*) = \inf_{x \in C} L(x, \lambda^*) \leq L(x, \lambda^*) \qquad \forall x \in C.
$$

So (3) holds.

Proposition 6.4.1 (Application of the saddle point form) *If* $x^* \in C$ *and* $\lambda^* \geq 0$ *satisfy the saddle point condition*

$$L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*) \quad \forall x \in C, \lambda \geq 0 \tag{6.4}$$

then

$$L(x^*, \lambda^*) = \sup_{\lambda \geq 0} \inf_{x \in C} L(x, \lambda).$$

Proof. Since $x^* \in C$, for any $\lambda \geq 0$ we have $\inf\limits_{x \in C} L(x, \lambda) \leq L(x^*, \lambda)$.
By (6.4), for any $\lambda \geq 0$ we have $\inf\limits_{x \in C} L(x, \lambda) \leq L(x^*, \lambda^*)$.
Taking suprema on both sides of the above inequality we have

$$\sup_{\lambda \geq 0} \inf_{x \in C} L(x, \lambda) \leq L(x^*, \lambda^*).$$

On the other hand, from (6.4) we have for any $x \in C$,

$$L(x^*, \lambda^*) \leq L(x, \lambda^*),$$

which implies that

$$L(x^*, \lambda^*) \leq \inf_{x \in C} L(x, \lambda^*) \leq \sup_{\lambda \geq 0} \inf_{x \in C} L(x, \lambda).$$

We call our original convex program the primal program:

$$\begin{array}{lll} (P) & \min & f(x) \\ & s.t. & g(x) \leq 0 \\ & & x \in C \subseteq R^n. \end{array}$$

We associated with the primal program a new problem called the dual program of (P):

$$\begin{array}{lll} (D) & \max & h(\lambda) := \inf_{x \in C} L(x, \lambda) \\ & s.t. & \lambda \geq 0 \end{array}$$

A vector $\lambda \geq 0$ is said to be feasible for (D) if $h(\lambda) := \inf_{x \in C} L(x, \lambda) > -\infty$. We denote by VP and VD the infima for the primal program (P) and the suprima for the dual program (D) respectively. By convention if there is no feasible vector for (D), then we say $VD = \sup_{\lambda \geq 0} h(\lambda) = -\infty$.

Theorem 6.4.2 (Weak Duality Theorem) *For any convex program (P), one has* $VP \geq VD$.

Proof. If the feasible region of (P) is empty, then $VP = \inf_{x \in \mathcal{F}} f(x) = +\infty$ which means that $VP \geq Vd$. Otherwise assume that $\mathcal{F} \neq \varnothing$. Let y be a feasible solution of (P) and $\lambda \geq 0$. Then

$$f(y) \geq f(y) + \sum_{i=1}^{m} \lambda_i g_i(y) = L(y, \lambda) \geq \inf_{x \in C} L(x, \lambda).$$

Consequently,

$$f(y) \geq \sup_{\lambda \geq 0} \inf_{x \in C} L(x, \lambda), \ \forall y \in \mathcal{F},$$

which implies that

$$VP = \inf_{y \in \mathcal{F}} f(y) \geq \sup_{\lambda \geq 0} \inf_{x \in C} L(x, \lambda) = VD.$$

Definition 6.4.1 If $VP > VD$, we say (P) has a duality gap. If $VP = VD$, we say (P) has no duality gap.

Theorem 6.4.3 (Strong Duality Theorem) Suppose that (P) is a convex program and the Slater condition holds. Then (P) has no duality gap.

Proof. By Corollary 6.3.1, there exists $\lambda^* \geq 0$ such that

$$VP = \inf_{x \in C} L(x, \lambda^*).$$

Therefore

$$VP \leq \sup_{\lambda \geq 0} \inf_{x \in C} L(x, \lambda) = VD$$

which combining with the weak duality theorem yeilds $VP = DP$.

Corollary 6.4.1 *Suppose that x^* is a feasible solution for a convex program (P) and that λ^* is a feasible solution for the dual program (D). If $f(x^*) = h(\lambda^*)$, then x^* is a solution of (P) and λ^* is a solution of (D).*

Proof. By the weak duality,

$$f(x^*) \geq VP \geq VD \geq h(\lambda^*).$$

Now the assumption implies that $f(x^*) = VP = VD = h(\lambda^*)$. So x^* is a solution of (P) and λ^* is a solution of (D). ∎

Remark 6.4.1 All results in this chapter hold for a convex program with linear constraints and polyhedral C without the Slater condition.

Consider the linear programming in the following form:

$$\begin{aligned}(LP)\ \min\ & b^T x \\ \text{s.t. } & Ax \geq c \\ & x \geq 0.\end{aligned}$$

Then dual problem of the linear program (LP) can be formulated as

$$\begin{aligned}(DLP)\ \min\ & \lambda^T c \\ \text{s.t. } & A^T \lambda \leq b \\ & \lambda \geq 0.\end{aligned}$$

Actually, (LP) can be reformulated as a convex program (P) by setting

$$
\begin{aligned}
f(x) &= b^T x, \;\; C = \{x \in \mathbb{R}^n : x \geq 0\} \\
g_i(x) &= c_i - \left(a^{(i)}\right)^T x,
\end{aligned}
$$

where $a^{(i)}$ is the ith row of A for $i = 1, 2, \cdots, m$. Then the Lagrangian of (P) is given by

$$
\begin{aligned}
L(x, \lambda) &= b^T x + \sum_{i=1}^{m} \lambda_i \left[c_i - \left(a^{(i)}\right)^T x \right] \\
&= \left[b - \sum_{i=1}^{m} \lambda_i a^{(i)} \right]^T x + \sum_{i=1}^{m} \lambda_i c_i \\
&= \left(b - A^T \lambda\right)^T x + \lambda^T c.
\end{aligned}
$$

Given $\lambda \geq 0$, if $\left(b - A^T\lambda\right)_{i_0} x_{i_0} < 0$ for some index i_0, then

$L(x, \lambda) = \left(b - A^T\lambda\right)_{i_0} x_{i_0} + \sum_{i \neq i_0} \left(b - A^T\lambda\right)_i x_i + \lambda^T c \to -\infty$ as $x_{i_0} \to +\infty$.

Therefore

$$h(\lambda) := \inf_{x \geq 0} \left\{ \left(b - A^T\lambda\right)^T x + \lambda^T c \right\} = \begin{cases} \lambda^T c & \text{if } b - A^T c \geq 0 \\ -\infty & \text{if } b - A^T c \ngeq 0. \end{cases}$$

Theorem 6.4.4 (Linear Programming Duality) If either (LP) or (DLP) has a solution, then the other has a solution and the corresponding values of the objective functions are equal.

Connection of the duality theory with the minmax theorem: By the definition of $L(x,\lambda)$,

$$\sup_{\lambda \geq 0} L(x,\lambda) = \begin{cases} f(x) & \text{if } g(x) \leq 0 \\ +\infty & \text{if } g(x) > 0 \end{cases}$$

which implies that

$$VP = \inf_{x \in x \in C, g(x) \leq 0} f(x) = \inf_{x \in C} \sup_{\lambda \geq 0} L(x,\lambda)$$

By definition, $VD = \sup_{\lambda \geq 0} \inf_{x \in C} L(x,\lambda)$. So $VP = VD$ if and only if the minmax theorem holds, i.e., if and only if inf and sup can be exchanged:

$$\inf_{x \in C} \sup_{\lambda \geq 0} L(x,\lambda) = \sup_{\lambda \geq 0} \inf_{x \in C} L(x,\lambda).$$

Example [Duffin's Duality Gap]:

$$\begin{array}{llll} (P) & \min & e^{-y} \\ & s.t. & \sqrt{x^2+y^2}-x \le 0. \end{array}$$

The dual problem is

$$\begin{array}{lll} (D) & \max & h(\lambda) := \inf\{e^{-y} + \lambda(\sqrt{x^2+y^2}-x)\} \\ & s.t. & \lambda \ge 0. \end{array}$$

Since $VP = 1 > 0 = VD$, there is a duality gap.