Fundamentals of Linear Algebra and Optimization

The Karush–Kuhn–Tucker Conditions

Jean Gallier and Jocelyn Quaintance

CIS Department
University of Pennsylvania

jean@cis.upenn.edu

April 16, 2024

Optimization with Convex Constraints

If the domain U is defined by *convex* inequality constraints satisfying mild differentiability conditions and if the constraints at u are qualified, then there is a *necessary* condition for the function J to have a local minimum at $u \in U$ involving *generalized Lagrange multipliers*. The proof uses a version of Farkas lemma.

Farkas Lemma

We will be using the following version of Farkas lemma.

Farkas Lemma

We will be using the following version of Farkas lemma.

Proposition (*Farkas lemma*). Let V be a Euclidean space of finite dimension with inner product $\langle -, - \rangle$ (more generally, a Hilbert space). For any finite family $(a_1, \ldots, a_m)$ of m vectors $a_i \in V$ and any vector $b \in V$, for any $v \in V$,

$$\text{if } \langle a_i, v \rangle \geq 0 \text{ for } i = 1, \ldots, m \text{ implies that } \langle b, v \rangle \geq 0,$$

then there exist $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ such that

$$\lambda_i \geq 0 \text{ for } i = 1, \ldots, m, \text{ and } b = \sum_{i=1}^{m} \lambda_i a_i.$$

Optimization with Convex Constraints

We can now prove the following theorem.

Optimization with Convex Constraints

We can now prove the following theorem.

Theorem. Let $\varphi_i\colon \Omega \to \mathbb{R}$ be m *convex constraints* defined on some *open convex* subset Ω of a finite-dimensional Euclidean vector space V (more generally, a real Hilbert space V), let $J\colon \Omega \to \mathbb{R}$ be some function, let U be given by

$$U = \{x \in \Omega \mid \varphi_i(x) \leq 0,\ \ 1 \leq i \leq m\},$$

and let $u \in U$ be any point such that the functions φ_i and J are differentiable at u.

Necessary Condition for Minimization with Convex Constraints

(1) If J has a local minimum at u with respect to U, and if the constraints are *qualified*, then there exist some scalars $\lambda_i(u) \in \mathbb{R}$, such that the *KKT condition* hold:

$$J_u{}' + \sum_{i=1}^{m} \lambda_i(u)(\varphi_i')_u = 0$$

and

$$\sum_{i=1}^{m} \lambda_i(u)\varphi_i(u) = 0, \quad \lambda_i(u) \geq 0, \quad i = 1, \ldots, m.$$

Necessary Condition for Minimization with Convex Constraints

Equivalently, in terms of gradients, the above conditions are expressed as

$$\nabla J_u + \sum_{i=1}^{m} \lambda_i(u) \nabla(\varphi_i)_u = 0,$$

and

$$\sum_{i=1}^{m} \lambda_i(u)\varphi_i(u) = 0, \quad \lambda_i(u) \geq 0, \quad i = 1, \ldots, m.$$

Sufficient Condition for Minimization with Convex Constraints

(2) Conversely, if the restriction of J to U is *convex* and if there exist scalars $(\lambda_1, \ldots, \lambda_m) \in \mathbb{R}_+^m$ such that the KKT conditions hold, then the function J has a (global) minimum at u with respect to U.

Sufficient Condition for Minimization with Convex Constraints

(2) Conversely, if the restriction of J to U is *convex* and if there exist scalars $(\lambda_1, \ldots, \lambda_m) \in \mathbb{R}_+^m$ such that the KKT conditions hold, then the function J has a (global) minimum at u with respect to U.

The scalars $\lambda_i(u)$ are often called *generalized Lagrange multipliers*.

Minimization with Convex Constraints

If $V = \mathbb{R}^n$, the necessary conditions of the preceding theorem are expressed as the following system of equations and inequalities in the unknowns $(u_1, \ldots, u_n) \in \mathbb{R}^n$ and $(\lambda_1, \ldots, \lambda_m) \in \mathbb{R}^m_+$:

Minimization with Convex Constraints

$$
\begin{aligned}
\frac{\partial J}{\partial x_1}(u) + \lambda_1 \frac{\partial \varphi_1}{\partial x_1}(u) + \cdots + \lambda_m \frac{\partial \varphi_m}{\partial x_1}(u) &= 0 \\
&\vdots \quad\quad \vdots \\
\frac{\partial J}{\partial x_n}(u) + \lambda_1 \frac{\partial \varphi_1}{\partial x_n}(u) + \cdots + \lambda_m \frac{\partial \varphi_m}{\partial x_n}(u) &= 0 \\
\lambda_1 \varphi_1(u) + \cdots + \lambda_m \varphi_m(u) &= 0 \\
\varphi_1(u) &\leq 0 \\
&\vdots \quad \vdots \\
\varphi_m(u) &\leq 0 \\
\lambda_1, \ldots, \lambda_m &\geq 0.
\end{aligned}
$$

Example of Convex Minimization

Example. Let J, φ_1 and φ_2 be the functions defined on $\mathbb{R}$ by

$$J(x) = x$$
$$\varphi_1(x) = -x$$
$$\varphi_2(x) = x - 1.$$

Example of Convex Minimization

Example. Let J, φ_1 and φ_2 be the functions defined on $\mathbb{R}$ by

$$\begin{aligned} J(x) &= x \\ \varphi_1(x) &= -x \\ \varphi_2(x) &= x - 1. \end{aligned}$$

In this case

$$U = \{x \in \mathbb{R} \mid -x \leq 0,\ x - 1 \leq 0\} = [0, 1].$$

Example of Convex Minimization

Example. Let J, φ_1 and φ_2 be the functions defined on $\mathbb{R}$ by

$$J(x) = x$$
$$\varphi_1(x) = -x$$
$$\varphi_2(x) = x - 1.$$

In this case

$$U = \{x \in \mathbb{R} \mid -x \leq 0,\ x - 1 \leq 0\} = [0, 1].$$

Since the constraints are *affine*, they are automatically *qualified* for any $u \in [0, 1]$.

Example of Convex Minimization

The system of equations and inequalities shown above becomes

$$\begin{aligned} 1 - \lambda_1 + \lambda_2 &= 0 \\ -\lambda_1 x + \lambda_2 (x - 1) &= 0 \\ -x &\leq 0 \\ x - 1 &\leq 0 \\ \lambda_1, \lambda_2 &\geq 0. \end{aligned}$$

Example of Convex Minimization

The system of equations and inequalities shown above becomes

$$\begin{aligned} 1 - \lambda_1 + \lambda_2 &= 0 \\ -\lambda_1 x + \lambda_2 (x - 1) &= 0 \\ -x &\leq 0 \\ x - 1 &\leq 0 \\ \lambda_1, \lambda_2 &\geq 0. \end{aligned}$$

The first equality implies that $\lambda_1 = 1 + \lambda_2$.

The second equality then becomes

$$-(1 + \lambda_2)x + \lambda_2 (x - 1) = 0,$$

which implies that $\lambda_2 = -x$.

Example of Convex Minimization

Since $0 \leq x \leq 1$, or equivalently $-1 \leq -x \leq 0$, and $\lambda_2 \geq 0$, we conclude that $\lambda_2 = 0$ and $\lambda_1 = 1$ is the solution associated with $x = 0$, the minimum of $J(x) = x$ over $[0, 1]$.

Example of Convex Minimization

Since $0 \leq x \leq 1$, or equivalently $-1 \leq -x \leq 0$, and $\lambda_2 \geq 0$, we conclude that $\lambda_2 = 0$ and $\lambda_1 = 1$ is the solution associated with $x = 0$, the minimum of $J(x) = x$ over $[0, 1]$.

Observe that the case $x = 1$ corresponds to the maximum and not a minimum of $J(x) = x$ over $[0, 1]$.

The Karush–Kuhn–Tucker Conditions

It is important to note that when *both* the constraints, the domain of definition Ω, *and* the objective function J are *convex*, if the KKT conditions hold for some $u \in U$ and some $\lambda \in \mathbb{R}^m_+$, the preceding theorem implies that J has a (global) minimum at u with respect to U, *independently* of any assumption on the qualification of the constraints.

The Lagrangian

The above theorem suggests introducing the function $L\colon \Omega \times \mathbb{R}_+^m \to \mathbb{R}$ given by

$$L(v, \lambda) = J(v) + \sum_{i=1}^{m} \lambda_i \varphi_i(v),$$

with $\lambda = (\lambda_1, \ldots, \lambda_m)$.

The Lagrangian

The above theorem suggests introducing the function $L\colon \Omega \times \mathbb{R}_+^m \to \mathbb{R}$ given by

$$L(v, \lambda) = J(v) + \sum_{i=1}^{m} \lambda_i \varphi_i(v),$$

with $\lambda = (\lambda_1, \dots, \lambda_m)$.

The function L is called the *Lagrangian* of the *Minimization Problem* (P):

$$\begin{aligned} &\text{minimize} && J(v) \\ &\text{subject to} && \varphi_i(v) \leq 0, \quad i = 1, \dots, m. \end{aligned}$$

The Lagrangian and the KKT Conditions

The KKT conditions of the preceding theorem imply that for any $u \in U$, if the vector $\lambda = (\lambda_1, \ldots, \lambda_m)$ is known and if u is a minimum of J on U, then

$$\frac{\partial L}{\partial u}(u) = 0$$
$$J(u) = L(u, \lambda).$$

The Lagrangian and the KKT Conditions

The KKT conditions of the preceding theorem imply that for any $u \in U$, if the vector $\lambda = (\lambda_1, \ldots, \lambda_m)$ is known and if u is a minimum of J on U, then

$$\frac{\partial L}{\partial u}(u) = 0$$
$$J(u) = L(u, \lambda).$$

The Lagrangian technique "absorbs" the constraints into the new objective

function L and reduces the problem of finding a constrained minimum of the function J, to the problem of finding an *unconstrained* minimum of the function $L(v, \lambda)$.

The Lagrangian and the KKT Conditions

The KKT conditions of the preceding theorem imply that for any $u \in U$, if the vector $\lambda = (\lambda_1, \ldots, \lambda_m)$ is known and if u is a minimum of J on U, then

$$\frac{\partial L}{\partial u}(u) = 0$$
$$J(u) = L(u, \lambda).$$

The Lagrangian technique "absorbs" the constraints into the new objective

function L and reduces the problem of finding a constrained minimum of the function J, to the problem of finding an *unconstrained* minimum of the function $L(v, \lambda)$.

This is the *main point* of Lagrangian duality which will be treated in the next lesson.

KKT Conditions with Affine Constraints

A case that arises often in practice is the case where the constraints φ_i are affine. If so, the m constraints $a_i x \leq b_i$ can be expressed in matrix form as $Ax \leq b$, where A is an $m \times n$ matrix whose ith row is the row vector a_i.

KKT Conditions with Affine Constraints

A case that arises often in practice is the case where the constraints φ_i are affine. If so, the m constraints $a_i x \leq b_i$ can be expressed in matrix form as $Ax \leq b$, where A is an $m \times n$ matrix whose ith row is the row vector a_i.

The KKT conditions of the preceding theorem yield the following corollary.

KKT Conditions with Affine Constraints

Proposition. If U is given by

$$U = \{x \in \Omega \mid Ax \leq b\},$$

where Ω is an open convex subset of $\mathbb{R}^n$ and A is an $m \times n$ matrix, and if J is differentiable at u and J has a local minimum at u, then there exist some vector $\lambda \in \mathbb{R}^m$, such that

$$\begin{aligned} &\nabla J_u + A^\top \lambda = 0 \\ &\lambda_i \geq 0 \ \text{ and } \ \text{ if } a_i u < b_i, \text{ then } \lambda_i = 0, \ i = 1, \ldots, m. \end{aligned}$$

KKT Conditions with Affine Constraints

Proposition. If U is given by

$$U = \{x \in \Omega \mid Ax \leq b\},$$

where Ω is an open convex subset of $\mathbb{R}^n$ and A is an $m \times n$ matrix, and if J is differentiable at u and J has a local minimum at u, then there exist some vector $\lambda \in \mathbb{R}^m$, such that

$$\nabla J_u + A^\top \lambda = 0$$
$$\lambda_i \geq 0 \text{ and } \text{ if } a_i u < b_i, \text{ then } \lambda_i = 0, \; i = 1, \ldots, m.$$

If the function J is *convex*, then the above conditions are also *sufficient* for J to have a minimum at $u \in U$.