## Fundamentals of Linear Algebra and Optimization

The Karush-Kuhn-Tucker Conditions

Jean Gallier and Jocelyn Quaintance<br>CIS Department<br>University of Pennsylvania<br>jean@cis.upenn.edu<br>May 7, 2020

## 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 $\left(a_{1}, \ldots, a_{m}\right)$ of $m$ vectors $a_{i} \in V$ and any vector $b \in V$, for any $v \in V$,

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

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 $\varphi_{i}: \Omega \rightarrow \mathbb{R}$ be $m$ convex constraints defined on some open convex subset $\Omega$ of a finite-dimensional Euclidean vector space $V$ (more generally, a real Hilbert space $V$ ), let $J: \Omega \rightarrow \mathbb{R}$ be some function, let $U$ be given by

$$
U=\left\{x \in \Omega \mid \varphi_{i}(x) \leq 0, \quad 1 \leq i \leq m\right\},
$$

and let $u \in U$ be any point such that the functions $\varphi_{i}$ 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 $K K T$ condition hold:

$$
J_{u}^{\prime}+\sum_{i=1}^{m} \lambda_{i}(u)\left(\varphi_{i}^{\prime}\right)_{u}=0
$$

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

$$
\nabla J_{u}+\sum_{i=1}^{m} \lambda_{i}(u) \nabla\left(\varphi_{i}\right)_{u}=0,
$$

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 $\left(\lambda_{1}, \ldots, \lambda_{m}\right) \in \mathbb{R}_{+}^{m}$ 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 $\left(\lambda_{1}, \ldots, \lambda_{m}\right) \in \mathbb{R}_{+}^{m}$ 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 $\left(u_{1}, \ldots, u_{n}\right) \in \mathbb{R}^{n}$ and $\left(\lambda_{1}, \ldots, \lambda_{m}\right) \in \mathbb{R}_{+}^{m}$ :

## Minimization with Convex Constraints

$$
\begin{aligned}
\frac{\partial J}{\partial x_{1}}(u)+\lambda_{1} \frac{\partial \varphi_{1}}{\partial x_{1}}(u)+\cdots+\lambda_{m} \frac{\partial \varphi_{m}}{\partial x_{1}}(u) & =0 \\
\vdots & \vdots \\
\frac{\partial J}{\partial x_{n}}(u)+\lambda_{1} \frac{\partial \varphi_{n}}{\partial x_{1}}(u)+\cdots+\lambda_{m} \frac{\partial \varphi_{m}}{\partial x_{n}}(u) & =0 \\
\lambda_{1} \varphi_{1}(u)+\cdots+\lambda_{m} \varphi_{m}(u) & =0 \\
\varphi_{1}(u) & \leq 0 \\
\vdots & \vdots \\
\varphi_{m}(u) & \leq 0 \\
\lambda_{1}, \ldots, \lambda_{m} & \geq 0
\end{aligned}
$$

## Example of Convex Minimization

Example. Let $J, \varphi_{1}$ and $\varphi_{2}$ be the functions defined on $\mathbb{R}$ by

$$
\begin{aligned}
J(x) & =x \\
\varphi_{1}(x) & =-x \\
\varphi_{2}(x) & =x-1 .
\end{aligned}
$$

## Example of Convex Minimization

Example. Let $J, \varphi_{1}$ and $\varphi_{2}$ 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 $J, \varphi_{1}$ and $\varphi_{2}$ 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

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

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

$$
-\left(1+\lambda_{2}\right) x+\lambda_{2}(x-1)=0,
$$

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 $\Omega$, 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 $L: \Omega \times \mathbb{R}_{+}^{m} \rightarrow \mathbb{R}$ given by

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

with $\lambda=\left(\lambda_{1}, \ldots, \lambda_{m}\right)$.

## The Lagrangian

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

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

with $\lambda=\left(\lambda_{1}, \ldots, \lambda_{m}\right)$.
The function $L$ is called the Lagrangian of the Minimization Problem $(P)$ :

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

## The Lagrangian and the KKT Conditions

The KKT conditions of the preceding theorem imply that for any $u \in U$, if the vector $\lambda=\left(\lambda_{1}, \ldots, \lambda_{m}\right)$ is known and if $u$ is a minimum of $J$ on $U$, then

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

## The Lagrangian and the KKT Conditions

The KKT conditions of the preceding theorem imply that for any $u \in U$, if the vector $\lambda=\left(\lambda_{1}, \ldots, \lambda_{m}\right)$ is known and if $u$ is a minimum of $J$ on $U$, then

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

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 $\lambda=\left(\lambda_{1}, \ldots, \lambda_{m}\right)$ is known and if $u$ is a minimum of $J$ on $U$, then

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

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 $\varphi_{i}$ are affine. If so, the $m$ constraints $a_{i} x \leq b_{i}$ can be expressed in matrix form as $A x \leq b$, where $A$ is an $m \times n$ matrix whose th 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 $\varphi_{i}$ are affine. If so, the $m$ constraints $a_{i} x \leq b_{i}$ can be expressed in matrix form as $A x \leq b$, where $A$ is an $m \times n$ matrix whose th 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

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

where $\Omega$ 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

$$
\begin{aligned}
& \nabla J_{u}+A^{\top} \lambda=0 \\
& \lambda_{i} \geq 0 \text { and } \quad \text { if } a_{i} u<b_{i} \text {, then } \lambda_{i}=0, i=1, \ldots, m .
\end{aligned}
$$

## KKT Conditions with Affine Constraints

Proposition. If $U$ is given by

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

where $\Omega$ 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

$$
\begin{aligned}
& \nabla J_{u}+A^{\top} \lambda=0 \\
& \lambda_{i} \geq 0 \text { and } \quad \text { if } a_{i} u<b_{i} \text {, then } \lambda_{i}=0, i=1, \ldots, m .
\end{aligned}
$$

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

