# The Karush-Kuhn-Tucker (KKT) conditions

In this section, we will give a set of sufficient (and at most times necessary) conditions for a $\boldsymbol{x}^\star$ to be the solution of a given convex optimization problem. These are called the Karush-Kuhn-Tucker (KKT) conditions, and they play a fundamental role in both the theory and practice of convex optimization. We have derived these conditions (and have shown that they we both necessary and sufficient) in some special cases in the previous notes

We will start here by considering a general convex program with **inequality** constraints only. This is just to make the exposition easier — after we have this established, we will show how to include equality constraints (which must always be affine in convex programming). A great source for the material in this section is [Lau13, Chap. 10].

Everywhere in this section, the functionals $f_0(\boldsymbol{x}), f_1(\boldsymbol{x}), \ldots, f_M(\boldsymbol{x})$, $f_m : \mathbb{R}^N \to \mathbb{R}$, are convex and differentiable.

**KKT (inequality only)**
The KKT conditions for the convex program

$$\underset{\boldsymbol{x}}{\text{minimize}} \;\; f_0(\boldsymbol{x}) \quad \text{subject to} \quad \begin{aligned} f_1(\boldsymbol{x}) &\le 0 \\ f_2(\boldsymbol{x}) &\le 0 \\ &\vdots \\ f_M(\boldsymbol{x}) &\le 0 \end{aligned} \tag{1}$$

in $\boldsymbol{x} \in \mathbb{R}^N$ and $\boldsymbol{\lambda} \in \mathbb{R}^M$ are

$$f_m(\boldsymbol{x}) \le 0, \quad m = 1, \ldots, M, \tag{K1}$$

$$\boldsymbol{\lambda} \ge \boldsymbol{0}, \tag{K2}$$

$$\lambda_m f_m(\boldsymbol{x}) = 0, \quad m = 1, \ldots, M, \tag{K3}$$

$$\nabla f_0(\boldsymbol{x}) + \sum_{m=1}^{M} \lambda_m \nabla f_m(\boldsymbol{x}) = \boldsymbol{0}, \tag{K4}$$

We start by establishing that these are sufficient conditions for a minimizer.

If the KKT conditions hold for $\boldsymbol{x}^\star$ and some $\lambda^\star \in \mathbb{R}^M$, then $\boldsymbol{x}^\star$ is a solution to the program (1).

Below, we denote the feasible set as

$$\mathcal{C} = \{\boldsymbol{x} \in \mathbb{R}^N \;:\; f_m(\boldsymbol{x}) \le 0,\; m = 1, \ldots, M\}.$$

It should be clear that the convexity of the $f_m$ implies the convexity[1]

[1]The $f_m$ are convex functions, so their sublevel sets are convex sets, and $\mathcal{C}$

of $\mathcal{C}$. The sufficiency proof simply relies on the convexity of $\mathcal{C}$, the convexity of $f_0$, and the concept of a descent/ascent direction (see the previous notes).

Suppose $\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star$ obey the KKT conditions. The first thing to note is that if

$$\lambda_1 = \lambda_2 = \cdots = \lambda_M = 0,$$

then (K4) implies that

$$\nabla f(\boldsymbol{x}^\star) = \mathbf{0},$$

and hence $\boldsymbol{x}^\star$ is a global min, as by the convexity of $f_0$,

$$f_0(\boldsymbol{x}) \geq f_0(\boldsymbol{x}^\star) + \langle \boldsymbol{x} - \boldsymbol{x}^\star, \nabla f_0(\boldsymbol{x}^\star) \rangle = f(\boldsymbol{x}^\star),$$

for all $\boldsymbol{x} \in \mathcal{C}$.

Now suppose that $R > 0$ entries of $\boldsymbol{\lambda}^\star$ are positive — without loss of generality, we will take these to be the first $R$,

$$\lambda_1^\star > 0, \quad \lambda_2^\star > 0, \quad \cdots, \quad \lambda_R^\star > 0, \quad \lambda_{R+1}^\star = 0, \quad \cdots \quad, \lambda_M^\star = 0.$$

We can rewrite (K4) as

$$\nabla f_0(\boldsymbol{x}^\star) + \lambda_1^\star \nabla f_1(\boldsymbol{x}^\star) + \cdots + \lambda_R^\star \nabla f_R(\boldsymbol{x}^\star) = \mathbf{0}, \tag{2}$$

and note that by (K3),

$$f_1(\boldsymbol{x}^\star) = 0, \ldots, f_R(\boldsymbol{x}^\star) = 0.$$

Consider any $\boldsymbol{x} \in \mathcal{C}$, $\boldsymbol{x} \neq \boldsymbol{x}^\star$. As $\mathcal{C}$ is convex, every point in between $\boldsymbol{x}^\star$ and $\boldsymbol{x}$ must also be in $\mathcal{C}$, meaning

$$f_m(\boldsymbol{x}^\star + \theta(\boldsymbol{x} - \boldsymbol{x}^\star)) \leq 0 = f_m(\boldsymbol{x}^\star), \quad m = 1, \ldots, R,$$

is an intersection of sublevel sets.

for all $0 \leq \theta \leq 1$. This means that $\boldsymbol{x} - \boldsymbol{x}^\star$ cannot be an ascent direction, and so

$$\langle \boldsymbol{x} - \boldsymbol{x}^\star, \nabla f_m(\boldsymbol{x}^\star) \rangle \leq 0, \quad m = 1, \ldots, R.$$

It is clear, then, that

$$\langle \boldsymbol{x} - \boldsymbol{x}^\star, \nabla f_0(\boldsymbol{x}^\star) \rangle \geq 0,$$

as otherwise there is no way (2) can hold with positive $\lambda_m$. Along with the convexity of $f_0$, this means that

$$f_0(\boldsymbol{x}) \geq f_0(\boldsymbol{x}^\star) + \langle \boldsymbol{x} - \boldsymbol{x}^\star, \nabla f(\boldsymbol{x}^\star) \rangle \geq f(\boldsymbol{x}^\star).$$

Since this holds for all $\boldsymbol{x} \in \mathcal{C}$, $\boldsymbol{x}^\star$ is a minimizer.

## Necessity

To establish the necessity of the KKT conditions, we need one piece of mathematical technology that we have not been exposed to yet. The *Farkas lemma* is a fundamental result in convex analysis; we will prove it in the Technical Details section.

**Farkas Lemma:**
Let $\boldsymbol{A}$ be an $M \times N$ matrix and $\boldsymbol{b} \in \mathbb{R}^M$. The exactly one of the following two things is true:

1. there exists $\boldsymbol{x} \geq \boldsymbol{0}$ such that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$;
2. there exists $\boldsymbol{\lambda} \in \mathbb{R}^M$ such that

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} \leq \boldsymbol{0}, \quad \text{and} \quad \langle \boldsymbol{b}, \boldsymbol{\lambda} \rangle > 0.$$

With this in place, we can give two different situations under which KKT is necessary. These are by no means the only situations for which this is true, but these two cover a high percentage of the cases encountered in practice.

> Suppose $\boldsymbol{x}^\star$ is a solution to a convex program with affine inequality constraints:
>
> $$\underset{\boldsymbol{x}\in\mathbb{R}^N}{\text{minimize}}\ f_0(\boldsymbol{x}) \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}.$$
>
> Then there exists a $\boldsymbol{\lambda}^\star$ such that $\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star$ obey the KKT conditions.

In this case, the constraint functions have the form

$$f_m(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{a}_m \rangle - b_m, \quad \text{and so} \quad \nabla f_m(\boldsymbol{x}) = \boldsymbol{a}_m,$$

where $\boldsymbol{a}_m^{\mathrm{T}}$ is the $m$th row of $\boldsymbol{A}$. Since $\boldsymbol{x}^\star$ is feasible, K1 must hold. If none of the constraints are "active", meaning $f_m(\boldsymbol{x}^\star)$ for $m = 1, \ldots, M$ and $\boldsymbol{x}^\star$ lies in the interior of $\mathcal{C}$, then it must be that $\nabla f(\boldsymbol{x}^\star) = \boldsymbol{0}$, and K2–K4 hold with $\boldsymbol{\lambda} = \boldsymbol{0}$.

Suppose that there are $R$ active constraints at $\boldsymbol{x}^\star$; without loss of generality, we will take these to be the first $R$:

$$\begin{aligned} f_1(\boldsymbol{x}^\star) = 0\ ,\ f_2(\boldsymbol{x}^\star) = 0\ ,\ \ldots\ ,\ f_R(\boldsymbol{x}^\star) = 0, \\ f_{R+1}(\boldsymbol{x}^\star) < 0\ ,\ \ldots,\ f_M(\boldsymbol{x}^\star) < 0. \end{aligned}$$

We start by taking $\lambda_{R+1} = \lambda_{R+2} = \cdots = \lambda_M = 0$, which means K3 will hold. Suppose that there were no $\boldsymbol{\lambda} \ge \boldsymbol{0}$ such that

$$\nabla f_0(\boldsymbol{x}^\star) + \lambda_1 \nabla f_1(\boldsymbol{x}^\star) + \cdots + \lambda_R \nabla f_R(\boldsymbol{x}^\star) = \boldsymbol{0}. \qquad (3)$$

With $\boldsymbol{A}' : R \times N$ consisting of the first $R$ rows of $\boldsymbol{A}$, and $\boldsymbol{b}' \in \mathbb{R}^R$ as the first $R$ entries in $\boldsymbol{b}$, this means that there is no $\boldsymbol{\lambda}' \in \mathbb{R}^R$ such that

$$\boldsymbol{A}'^{\mathrm{T}}\boldsymbol{\lambda}' = -\nabla f_0(\boldsymbol{x}^\star), \quad \boldsymbol{\lambda}' \geq \boldsymbol{0}.$$

By the Farkas lemma, this means that there is a $\boldsymbol{d} \in \mathbb{R}^N$ such that

$$\boldsymbol{A}'\boldsymbol{d} \leq \boldsymbol{0}, \quad \langle \boldsymbol{d}, -\nabla f_0(\boldsymbol{x}^\star)\rangle > 0,$$

which means, since $\nabla f_m(\boldsymbol{x}) = \boldsymbol{a}_m$,

$$\begin{aligned} \langle \boldsymbol{d}, \nabla f_0(\boldsymbol{x}^\star)\rangle &< 0 \\ \langle \boldsymbol{d}, \nabla f_1(\boldsymbol{x}^\star)\rangle &\leq 0 \\ &\vdots \\ \langle \boldsymbol{d}, \nabla f_R(\boldsymbol{x}^\star)\rangle &\leq 0. \end{aligned}$$

This means that $\boldsymbol{d}$ is a descent direction for $f_0$, and is not an ascent direction for $f_1, \ldots, f_R$. Because the constraint functionals are affine, if $\langle \boldsymbol{d}, \nabla f_m(\boldsymbol{x}^\star)\rangle = 0$ above, then $f_m(\boldsymbol{x}^\star + t\boldsymbol{d}) = f_m(\boldsymbol{x}^\star)$ — this means that moving in the direction $\boldsymbol{d}$ will not increase $f_1, \ldots, f_m$. Since the last $M - R$ constraints are not active, we can move at least a small amount in any direction so that they stay that way. This means that there exists a $t > 0$ such that

$$f_0(\boldsymbol{x}^\star + t\boldsymbol{d}) < f(\boldsymbol{x}^\star),$$

but also maintains feasibility:

$$f_m(\boldsymbol{x}^\star + t\boldsymbol{d}) \leq 0, \quad m = 1, \ldots, M.$$

This directly contradicts the assertion that $\boldsymbol{x}^\star$ is optimal, and so $\lambda_1, \ldots, \lambda_R \geq 0$ must exist such that (3) holds.

For general convex inequality constraints, there are various other scenarios under which the KKT conditions are necessary; these are called **constraint qualifications.** We have already seen that polygonal (affine) constraints qualify. Another set of constraint qualifications are *Slater's condition*:

**Slater's condition**: There exists at least one strictly feasible point; a $\boldsymbol{x}$ such that none of the constraints are active:

$$f_1(\boldsymbol{x}) < 0\ ,\ f_2(\boldsymbol{x}) < 0\ ,\ \cdots\ ,\ f_M(\boldsymbol{x}) < 0.$$

Suppose that Slater's condition holds for $f_1, \ldots, f_M$, and let $\boldsymbol{x}^\star$ be a solution to

$$\underset{\boldsymbol{x}\in\mathbb{R}^N}{\text{minimize}}\ f_0(\boldsymbol{x}) \quad \text{subject to} \quad f_m \leq 0, \quad m = 1, \ldots, M.$$

Then there exists a $\boldsymbol{\lambda}^\star$ such that $\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star$ obey the KKT conditions.

This is proved in much the same way as in the affine inequality case. Suppose that $\boldsymbol{x}^\star$ is a solution, and that

$$\begin{aligned} f_1(\boldsymbol{x}^\star) = 0\ ,\ f_2(\boldsymbol{x}^\star) = 0\ ,\ \ldots\ ,\ f_R(\boldsymbol{x}^\star) = 0, \\ f_{R+1}(\boldsymbol{x}^\star) < 0\ ,\ \ldots,\ f_M(\boldsymbol{x}^\star) < 0. \end{aligned}$$

We take $\lambda_{R+1} = \cdots = \lambda_M = 0$, and show that if there is not $\lambda_1, \ldots, \lambda_R \geq 0$ such that

$$\nabla f_0(\boldsymbol{x}^\star) + \sum_{m=1}^{R} \lambda_m \nabla f_m(\boldsymbol{x}^\star) = \mathbf{0}, \tag{4}$$

then there is a another feasible point with a smaller value of $f_0$.

By the Farkas lemma, if there does not exist a $\lambda_1, \ldots, \lambda_R \geq 0$ such that (4) holds, then there must be a $\boldsymbol{u} \in \mathbb{R}^N$ such that

$$\begin{aligned} \langle \boldsymbol{u}, \nabla f_0(\boldsymbol{x}^\star)\rangle &< 0 \\ \langle \boldsymbol{u}, \nabla f_1(\boldsymbol{x}^\star)\rangle &\leq 0 \\ &\vdots \\ \langle \boldsymbol{u}, \nabla f_R(\boldsymbol{x}^\star)\rangle &\leq 0. \end{aligned}$$

Now let $\boldsymbol{z}$ be a strictly feasible point, $f_m(\boldsymbol{z}) < 0$ for all $m$. We know that

$$0 > f_m(\boldsymbol{z}) \geq f_m(\boldsymbol{x}^\star) + \langle \boldsymbol{z} - \boldsymbol{x}^\star, \nabla f_m(\boldsymbol{x}^\star)\rangle \quad \Rightarrow \quad \langle \boldsymbol{z} - \boldsymbol{x}^\star, \nabla f_m(\boldsymbol{x}^\star)\rangle < 0,$$

for $m = 1, \ldots, R$, since then $f_m(\boldsymbol{x}^\star) = 0$. So $\boldsymbol{u}$ is a descent direction for $f_0$, and $\boldsymbol{z} - \boldsymbol{x}^\star$ is a descent direction for all all of the constraint functionals $f_m,\ m = 1, \ldots, R$ that are active.

We consider a convex combination of these two vectors

$$\boldsymbol{d}_\theta = (1-\theta)\boldsymbol{u} + \theta(\boldsymbol{z} - \boldsymbol{x}^\star).$$

We know that $\langle \boldsymbol{d}_\theta, \nabla f_m(\boldsymbol{x}^\star)\rangle < 0$ for all $0 < \theta \leq 1,\ m = 1, \ldots, R$. We also know that there is a $\theta$ small enough so that $\boldsymbol{d}_\theta$ is a descent direction for $f_0$; there exists $0 < \epsilon_0 < 1$ such that

$$\langle \boldsymbol{d}_{\epsilon_0}, \nabla f_0(\boldsymbol{x}^\star)\rangle < 0.$$

Finally, we also know that we can move a small enough amount in any direction and keep constraints $f_{R+1}, \ldots, f_M$ inactive. Thus there is a $t > 0$ such that

$$f_0(\boldsymbol{x}^\star + t\boldsymbol{d}_{\epsilon_0}) < f_0(\boldsymbol{x}^\star), \quad f_m(\boldsymbol{x}^\star + t\boldsymbol{d}_{\epsilon_0}) \leq 0, \quad m = 1, \ldots, M,$$

which directly contradicts the assertion that $\boldsymbol{x}^\star$ is optimal.

It should be clear from the two arguments above that Slater's condition can be refined — we only need a point which obeys $f_m(\boldsymbol{z}) < 0$ for the $f_m$ which are not affine. We now state this formally:

> Suppose that $f_1, \ldots, f_{M'}$ are affine functionals, and $f_{M'+1}, \ldots, f_M$ are convex functional which are not affine. Suppose that Slater's condition holds for $f_{M'+1}, \ldots, f_M$, and let $\boldsymbol{x}^\star$ be a solution to
>
> $$\underset{\boldsymbol{x}\in\mathbb{R}^N}{\text{minimize}} \;\; f_0(\boldsymbol{x}) \quad \text{subject to} \quad f_m(\boldsymbol{x}) \leq 0, \quad m = 1, \ldots, M.$$
>
> Then there exists a $\boldsymbol{\lambda}^\star$ such that $\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star$ obey the KKT conditions.

The above statement lets us extend the KKT conditions to optimization problems with linear equality constraints, which we now state.

**KKT (with equality constraints)**
The KKT conditions for the optimization program

$$\min_{\boldsymbol{x}} \ f_0(\boldsymbol{x}) \quad \text{subject to} \quad \begin{aligned} f_m(\boldsymbol{x}) &\leq 0, \quad m = 1, \ldots, M \\ h_p(\boldsymbol{x}) &= 0, \quad p = 1, \ldots, P \end{aligned} \tag{5}$$

in $\boldsymbol{x} \in \mathbb{R}^N$, $\boldsymbol{\lambda} \in \mathbb{R}^M$, and $\boldsymbol{\nu} \in \mathbb{R}^P$ are

$$\begin{aligned} f_m(\boldsymbol{x}) &\leq 0, \quad m = 1, \ldots, M, \\ h_p(\boldsymbol{x}) &= 0, \quad p = 1, \ldots, P \end{aligned} \tag{K1}$$

$$\boldsymbol{\lambda} \geq \boldsymbol{0}, \tag{K2}$$

$$\lambda_m f_m(\boldsymbol{x}) = 0, \quad m = 1, \ldots, M, \tag{K3}$$

$$\nabla f_0(\boldsymbol{x}) + \sum_{m=1}^{M} \lambda_m \nabla f_m(\boldsymbol{x}) + \sum_{p=1}^{P} \nu_p \nabla h_p(\boldsymbol{x}) = \boldsymbol{0}, \tag{K4}$$

We call the $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$ above **Lagrange multipliers**. Notice that $\boldsymbol{\lambda}$ is constrained to be positive, while $\boldsymbol{\nu}$ can be arbitrary. Also, if the $h_p$ are affine, which they have to be for the program above to be convex, then we can write the equality constraints

$$h_p(\boldsymbol{x}) = 0, \quad p = 1, \ldots, P \quad \text{as} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},$$

for some $\boldsymbol{A} : P \times N$ and $\boldsymbol{b} \in \mathbb{R}^P$. Also, we can rewrite (K4) as

$$\nabla f_0(\boldsymbol{x}) + \sum_{m=1}^{M} \lambda_m \nabla f_m(\boldsymbol{x}) + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\nu} = \boldsymbol{0}.$$

If the $f_m$ are convex and the $h_p$ affine, then the KKT conditions are sufficient for $\boldsymbol{x}^\star$ to be the solution to the convex program (5).

If Slater's condition holds for the non-affine $f_m$, then they are also necessary. Almost nothing changes in the proofs above — we could simply separate an equality constraint of the form $\langle \boldsymbol{x}, \boldsymbol{a} \rangle = b$ into $\langle \boldsymbol{x}, \boldsymbol{a} \rangle - b \leq 0$ and $\langle \boldsymbol{x}, -\boldsymbol{a} \rangle + b \leq 0$. Then we can recombine the result, taking $nu = \lambda_1 - \lambda_2$, where $\lambda - 1$ is the Lagrange multiplier for $\langle \boldsymbol{x}, \boldsymbol{a} \rangle - b$ and $\lambda_2$ is the same for $\langle \boldsymbol{x}, -\boldsymbol{a} \rangle + b$.

## Technical Details: Proof of the Farkas Lemma

We prove the Farkas Lemma: if $\boldsymbol{A}$ is an $M \times N$ matrix and $\boldsymbol{b} \in \mathbb{R}^M$ is a given vector, then exactly one of the following two things is true:

1. there exists $\boldsymbol{x} \geq \boldsymbol{0}$ such that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$;
2. there exists $\boldsymbol{v} \in \mathbb{R}^M$ such that

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{v} \leq \boldsymbol{0}, \quad \text{and} \quad \langle \boldsymbol{b}, \boldsymbol{v} \rangle > 0.$$

It is clear that if the first condition holds, the second cannot, as $\langle \boldsymbol{b}, \boldsymbol{v} \rangle = \langle \boldsymbol{x}, \boldsymbol{A}^{\mathrm{T}}\boldsymbol{v} \rangle$ for any $\boldsymbol{x}$ such that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, and $\langle \boldsymbol{x}, \boldsymbol{A}^{\mathrm{T}}\boldsymbol{v} \rangle \leq 0$ for any $\boldsymbol{x} \geq \boldsymbol{0}$ and $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{v} \leq \boldsymbol{0}$.

It is more difficult to argue that if the first condition does not hold, the second must. This ends up being a direct result of the separating hyperplane theorem. Let $\mathcal{C}(\boldsymbol{A})$ be the (convex) cone generated by the columns $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$ of $\boldsymbol{A}$:

$$\mathcal{C}(\boldsymbol{A}) = \left\{ \boldsymbol{v} \in \mathbb{R}^M \;:\; \boldsymbol{v} = \sum_{n=1}^{N} \theta_n \boldsymbol{a}_n, \quad \theta_n \geq 0, \; n = 1, \ldots, N \right\}.$$

Then 1 above is clearly equivalent to $\boldsymbol{b} \in \mathcal{C}(\boldsymbol{A})$. Since $\mathcal{C}(\boldsymbol{A})$ is closed and convex, and $\boldsymbol{b}$ is a single point, we know that if $\boldsymbol{b} \notin \mathcal{C}(\boldsymbol{A})$, then $\mathcal{C}(\boldsymbol{A})$ and $\boldsymbol{b}$ are strongly separated by a hyperplane. That is, if $\boldsymbol{b} \notin \mathcal{C}(\boldsymbol{A})$ implies that there exists a $\boldsymbol{v} \in \mathbb{R}^M$ such that

$$\boldsymbol{v}^{\mathrm{T}}\boldsymbol{b} > \boldsymbol{v}^{\mathrm{T}}\boldsymbol{\lambda} \quad \text{for all} \quad \boldsymbol{\lambda} \in \mathcal{C}(\boldsymbol{A}),$$

which is the same as saying

$$\boldsymbol{v}^{\mathrm{T}}\boldsymbol{b} > \sup_{\boldsymbol{\lambda} \in \mathcal{C}(\boldsymbol{A})} \boldsymbol{v}^{\mathrm{T}}\boldsymbol{\lambda} = \sup_{\boldsymbol{x} \geq \boldsymbol{0}} \boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}.$$

We know that $\mathbf{0} \in \mathcal{C}(\boldsymbol{A})$, so we must have $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{b} > 0$. The above equation also gives a finite upper bound (namely whatever the actual value of $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{b}$ is) on the function $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$ for all $\boldsymbol{x} \geq \mathbf{0}$. But this means that $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{v} \leq \mathbf{0}$, as otherwise we would have the following contradiction. If there were some index $n$ such that $(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{v})[n] = \epsilon > 0$, then with $\boldsymbol{e}_m \geq \mathbf{0}$ as the unit vector

$$\boldsymbol{e}_n[k] = \begin{cases} 1, & k = n, \\ 0, & k \neq n \end{cases},$$

we have

$$\sup_{\boldsymbol{x} \geq \mathbf{0}} \boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} \geq \sup_{\alpha \geq 0} \boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}(\alpha\boldsymbol{e}_n) = \sup_{\alpha \geq 0} \alpha\epsilon = \infty,$$

which contradicts the existence of this upper bound.

# References


[Lau13] N. Lauritzen. *Undergraduate Convexity.* World Scientific, 2013.

# MATH4230 - Optimization Theory - 2019/20

Tieyong Zeng
zeng@math.cuhk.edu.hk
March, 2020

## Plan (March 10-11, 2020)

1. Review of subgradient

2. Duality

3. Kuhn-Tucker theorem

**Theorem 3.1 (Kuhn-Tucker – no equality constraints)** *Let* $f, g_1, \cdots, g_m : \mathbb{R}^n \to (-\infty, +\infty]$ *be convex functions and let* $S \subset \mathbb{R}^n$ *be a convex set. Consider the* convex programming problem

$$(P)\quad \inf_{x\in S}\{f(x) : g_1(x) \le 0, \cdots, g_m(x) \le 0\}.$$

*Let* $\bar{x}$ *be a feasible point of* $(P)$*; denote by* $I(\bar{x})$ *the set of all* $i \in \{1, \cdots, m\}$ *for which* $g_i(\bar{x}) = 0$.

**Theorem 3.1 (Kuhn-Tucker – no equality constraints)** *Let $f, g_1, \cdots, g_m : \mathbb{R}^n \to (-\infty, +\infty]$ be convex functions and let $S \subset \mathbb{R}^n$ be a convex set. Consider the* convex programming problem

$$(P) \quad \inf_{x \in S} \{f(x) : g_1(x) \leq 0, \cdots, g_m(x) \leq 0\}.$$

*Let $\bar{x}$ be a feasible point of $(P)$; denote by $I(\bar{x})$ the set of all $i \in \{1, \cdots, m\}$ for which $g_i(\bar{x}) = 0$.*

*(i) $\bar{x}$ is an optimal solution of $(P)$ if there exist vectors of* multipliers *$\bar{u} := (\bar{u}_1, \cdots, \bar{u}_m) \in \mathbb{R}^m_+$ and $\bar{\eta} \in \mathbb{R}^n$ such that the following three relationships hold:*

$$\bar{u}_i g_i(\bar{x}) = 0 \text{ for } i = 1, \cdots, m \quad \textbf{(complementary slackness)},$$

$$0 \in \partial f(\bar{x}) + \sum_{i \in I(\bar{x})} \bar{u}_i \partial g_i(\bar{x}) + \bar{\eta} \quad \textbf{(normal Lagrange inclusion)},$$

$$\bar{\eta}^t (x - \bar{x}) \leq 0 \text{ for all } x \in S \quad \textbf{(obtuse angle property)}.$$

**Theorem 3.1 (Kuhn-Tucker – no equality constraints)** *Let* $f, g_1, \cdots, g_m : \mathbb{R}^n \to (-\infty, +\infty]$ *be convex functions and let* $S \subset \mathbb{R}^n$ *be a convex set. Consider the* convex programming problem

$$(P) \quad \inf_{x \in S} \{f(x) : g_1(x) \leq 0, \cdots, g_m(x) \leq 0\}.$$

*Let* $\bar{x}$ *be a feasible point of* $(P)$*; denote by* $I(\bar{x})$ *the set of all* $i \in \{1, \cdots, m\}$ *for which* $g_i(\bar{x}) = 0$.

*(ii) Conversely, if* $\bar{x}$ *is an optimal solution of* $(P)$ *and if* $\bar{x} \in \text{int dom } f \cap \bigcap_{i \in I(\bar{x})} \text{int dom } g_i$*, then there exist multipliers* $\bar{u}_0 \in \{0, 1\}$*,* $\bar{u} \in \mathbb{R}^m_+$*,* $(\bar{u}_0, \bar{u}) \neq (0, 0)$*, and* $\bar{\eta} \in \mathbb{R}^n$ *such that the complementary slackness relationship and obtuse angle property of part (i) hold, as well as the following:*

$$0 \in \bar{u}_0 \partial f(\bar{x}) + \sum_{i \in I(\bar{x})} \bar{u}_i \partial g_i(\bar{x}) + \bar{\eta} \ \textbf{(Lagrange inclusion)}.$$

**Theorem 2.9 (Moreau-Rockafellar)** *Let* $f, g : \mathbb{R}^n \to (-\infty, +\infty]$ *be convex functions. Then for every* $x_0 \in \mathbb{R}^n$

$$\partial f(x_0) + \partial g(x_0) \subset \partial(f + g)(x_0).$$

*Moreover, suppose that* $\operatorname{int} \operatorname{dom} f \cap \operatorname{dom} g \neq \emptyset$. *Then for every* $x_0 \in \mathbb{R}^n$ *also*

$$\partial(f + g)(x_0) \subset \partial f(x_0) + \partial g(x_0).$$

**Theorem 2.10** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function and let $S \subset \mathbb{R}^n$ be a nonempty convex set. Consider the optimization problem*

$$(P) \quad \inf_{x \in S} f(x).$$

*Then $\bar{x} \in S$ is an optimal solution of $(P)$ if and only if there exists a subgradient $\bar{\xi} \in \partial f(\bar{x})$ such that*

$$\bar{\xi}^t (x - \bar{x}) \geq 0 \text{ for all } x \in S. \tag{1}$$

*Here the* normal *case is said to occur when* $\bar{u}_0 = 1$ *and the* abnormal *case when* $\bar{u}_0 = 0$.

**Remark 3.2 (minimum principle)** *By Theorem 2.9, the normal Lagrange inclusion in Theorem 3.1 implies*

$$-\bar{\eta} \in \partial(f + \sum_{i \in I(\bar{x})} \bar{u}_i g_i)(\bar{x}).$$

*So by Theorem 2.10 and Remark 2.11 it follows that*

$$\bar{x} \in \operatorname{argmin}_{x \in S}[f(x) + \sum_{i \in I(\bar{x})} \bar{u}_i g_i(x)] (\textbf{minimum principle}).$$

*Likewise, under the additional condition* $\operatorname{dom} f \cap \cap_{i \in I(\bar{x})} \operatorname{int} \operatorname{dom} g_i \neq \emptyset$, *this minimum principle implies the normal Lagrange inclusion by the converse parts of Theorem 2.10/Remark 2.11 and Theorem 2.9.*

**Remark 3.3 (Slater's constraint qualification)** *The following* Slater constraint qualification *guarantees normality: Suppose that there exists* $\tilde{x} \in S$ *such that* $g_i(\tilde{x}) < 0$ *for* $i = 1, \cdots, m$. *Then in part (ii) of Theorem 3.1 we have the normal case* $\bar{u}_0 = 1$.

*Indeed, suppose we had* $\bar{u}_0 = 0$. *For* $\bar{u}_0 = 0$ *instead of* $\bar{u}_0 = 1$ *the proof of the minimum principle in Remark 3.2 can be mimicked and gives*

$$\sum_{i=1}^{m} \bar{u}_i g_i(\bar{x}) \leq \sum_{i=1}^{m} \bar{u}_i g_i(\tilde{x}).$$

*Since* $(\bar{u}_1, \cdots, \bar{u}_m) \neq (0, \cdots, 0)$, *this gives* $\sum_{i=1}^{m} \bar{u}_i g_i(\bar{x}) < 0$, *in contradiction to complementary slackness.*

**Theorem 3.1 (Kuhn-Tucker – no equality constraints)** *Let* $f, g_1, \cdots, g_m : \mathbb{R}^n \to (-\infty, +\infty]$ *be convex functions and let* $S \subset \mathbb{R}^n$ *be a convex set. Consider the* convex programming problem

$$(P) \quad \inf_{x \in S} \{f(x) : g_1(x) \leq 0, \cdots, g_m(x) \leq 0\}.$$

*Let* $\bar{x}$ *be a feasible point of* $(P)$*; denote by* $I(\bar{x})$ *the set of all* $i \in \{1, \cdots, m\}$ *for which* $g_i(\bar{x}) = 0$.

*(i)* $\bar{x}$ *is an optimal solution of* $(P)$ *if there exist vectors of* multipliers $\bar{u} := (\bar{u}_1, \cdots, \bar{u}_m) \in \mathbb{R}^m_+$ *and* $\bar{\eta} \in \mathbb{R}^n$ *such that the following three relationships hold:*

$$\bar{u}_i g_i(\bar{x}) = 0 \text{ for } i = 1, \cdots, m \quad \textbf{(complementary slackness)},$$

$$0 \in \partial f(\bar{x}) + \sum_{i \in I(\bar{x})} \bar{u}_i \partial g_i(\bar{x}) + \bar{\eta} \quad \textbf{(normal Lagrange inclusion)},$$

$$\bar{\eta}^t (x - \bar{x}) \leq 0 \text{ for all } x \in S \quad \textbf{(obtuse angle property)}.$$

PROOF OF THEOREM 3.1. Let us write $I := I(\bar{x})$. $(i)$ By Remark 3.2 the minimum principle holds, i.e., for any $x \in S$ we have

$$f(x) + \sum_{i \in I} \bar{u}_i g_i(x) \geq f(\bar{x})$$

(observe that $\sum_{i \in I} \bar{u}_i g_i(\bar{x}) = 0$ by complementary slackness). Hence, for any *feasible* $x \in S$ we have

$$f(x) \geq f(x) + \sum_{i \in I} \bar{u}_i g_i(x) \geq f(\bar{x}),$$

by nonnegativity of the multipliers. Clearly, this proves optimality of $\bar{x}$.

**Theorem 2.17 (Dubovitskii-Milyutin)** *Let* $f_1, \cdots, f_m : \mathbb{R}^n \to (-\infty, +\infty]$ *be convex functions and let* $x_0$ *be a point in* $\cap_{i=1}^m \text{int dom } f_i$. *Let* $f : \mathbb{R}^n \to (-\infty, +\infty]$ *be given by*

$$f(x) := \max_{1 \le i \le m} f_i(x)$$

*and let* $I(x_0)$ *be the (nonempty) set of all* $i \in \{1, \cdots, m\}$ *for which* $f_i(x_0) = f(x_0)$. *Then*

$$\partial f(x_0) = \text{co } \cup_{i \in I(x_0)} \partial f_i(x_0).$$

**Theorem 3.1 (Kuhn-Tucker – no equality constraints)** *Let* $f, g_1, \cdots, g_m : \mathbb{R}^n \to (-\infty, +\infty]$ *be convex functions and let* $S \subset \mathbb{R}^n$ *be a convex set. Consider the* convex programming problem

$$(P)\quad \inf_{x\in S}\{f(x) : g_1(x) \le 0, \cdots, g_m(x) \le 0\}.$$

*Let* $\bar{x}$ *be a feasible point of* $(P)$*; denote by* $I(\bar{x})$ *the set of all* $i \in \{1, \cdots, m\}$ *for which* $g_i(\bar{x}) = 0$.

*(ii) Conversely, if* $\bar{x}$ *is an optimal solution of* $(P)$ *and if* $\bar{x} \in \text{int dom } f \cap \cap_{i\in I(\bar{x})} \text{int dom } g_i$, *then there exist multipliers* $\bar{u}_0 \in \{0, 1\}$, $\bar{u} \in \mathbb{R}^m_+$, $(\bar{u}_0, \bar{u}) \ne (0, 0)$, *and* $\bar{\eta} \in \mathbb{R}^n$ *such that the complementary slackness relationship and obtuse angle property of part (i) hold, as well as the following:*

$$0 \in \bar{u}_0 \partial f(\bar{x}) + \sum_{i\in I(\bar{x})} \bar{u}_i \partial g_i(\bar{x}) + \bar{\eta} \ \textbf{(Lagrange inclusion)}.$$

$(ii)$ Consider the auxiliary optimization problem

$$(P') \ \inf_{x \in S} \phi(x),$$

where $\phi(x) := \max[f(x) - f(\bar{x}), \max_{1 \leq i \leq m} g_i(x)]$. Since $\bar{x}$ is an optimal solution of $(P)$, it is not hard to see that $\bar{x}$ is also an optimal solution of $(P')$ (observe that $\phi(\bar{x}) = 0$ and that $x \in S$ is feasible if and only if $\max_{1 \leq i \leq m} g_i(x) \leq 0$). By Theorem 2.10 and Remark 2.11 there exists $\bar{\eta}$ in $\mathbb{R}^n$ such that $\bar{\eta}$ has the obtuse angle property and $-\bar{\eta} \in \partial\phi(\bar{x})$. By Theorem 2.17 this gives

$$-\bar{\eta} \in \partial\phi(\bar{x}) = \operatorname{co}(\partial f(\bar{x}) \cup \cup_{i \in I} \partial g_i(\bar{x})).$$

$$-\bar{\eta} \in \partial\phi(\bar{x}) = \mathrm{co}(\partial f(\bar{x}) \cup \cup_{i\in I}\partial g_i(\bar{x})).$$

Since subdifferentials are convex, we get the existence of $(u_0, \xi_0) \in \mathbb{R}_+ \times \partial f(\bar{x})$ and $(u_i, \xi_i) \in \mathbb{R}_+ \times \partial g_i(\bar{x})$, $i \in I$, such that $\sum_{i\in\{0\}\cup I} u_i = 1$ and

$$-\bar{\eta} = \sum_{i\in\{0\}\cup I} u_i \xi_i.$$

In case $u_0 = 0$, we are done by setting $\bar{u}_i := u_i$ for $i \in \{0\} \cup I$ and $\bar{u}_i := 0$ otherwise. Observe that in this case $(\bar{u}_1, \cdots, \bar{u}_m) \neq (0, \cdots, 0)$ by $\sum_{i\in I} u_i = 1$. In case $u_0 \neq 0$, we know that $u_0 > 0$, so we can set $\bar{u}_i := u_i/u_0$ for $i \in \{0\} \cup I$ and $\bar{u}_i := 0$ otherwise. QED

**Example 3.4** Consider the following optimization problem:

$$(P) \text{ minimize } (x_1 - \frac{9}{4})^2 + (x_2 - 2)^2$$

over all $(x_1, x_2) \in \mathbb{R}^2_+$ such that

$$\begin{aligned} x_1^2 - x_2 &\leq 0 \\ x_1 + x_2 - 6 &\leq 0 \\ -x_1 + 1 &\leq 0 \end{aligned}$$

**Example 3.4** Consider the following optimization problem:

$$(P) \text{ minimize } (x_1 - \frac{9}{4})^2 + (x_2 - 2)^2$$

over all $(x_1, x_2) \in \mathbb{R}^2_+$ such that

$$\begin{aligned} x_1^2 - x_2 &\leq 0 \\ x_1 + x_2 - 6 &\leq 0 \\ -x_1 + 1 &\leq 0 \end{aligned}$$

Since Slater's constraint qualification clearly holds, we get that a feasible point $(\bar{x}_1, \bar{x}_2)$ is optimal if and only if there exists $(\bar{u}_1, \bar{u}_2, \bar{u}_3) \in \mathbb{R}^3_+$ such that

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 2(\bar{x}_1 - \frac{9}{4}) \\ 2(\bar{x}_2 - 2) \end{pmatrix} + \bar{u}_1 \begin{pmatrix} 2\bar{x}_1 \\ -1 \end{pmatrix} + \bar{u}_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \bar{u}_3 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + \begin{pmatrix} \bar{\eta}_1 \\ \bar{\eta}_2 \end{pmatrix}$$

for some $\bar{\eta} := (\bar{\eta}_1, \bar{\eta}_2)^t$ with

$$\bar{\eta}^t (x - \bar{x}) \leq 0 \text{ for all } x \in \mathbb{R}^2_+$$

and such that

$$\begin{aligned} \bar{u}_1(\bar{x}_1^2 - \bar{x}_2) &= 0 \\ \bar{u}_2(\bar{x}_1 + \bar{x}_2 - 6) &= 0 \\ \bar{u}_3(-\bar{x}_1 + 1) &= 0 \end{aligned}$$

$$\bar{\eta}^t(x - \bar{x}) \leq 0 \text{ for all } x \in \mathbb{R}^2_+$$

and such that

$$\begin{array}{rcl} x_1^2 - x_2 & \leq & 0 \\ x_1 + x_2 - 6 & \leq & 0 \\ -x_1 + 1 & \leq & 0 \end{array}$$

$$\begin{array}{rcl} \bar{u}_1(\bar{x}_1^2 - \bar{x}_2) & = & 0 \\ \bar{u}_2(\bar{x}_1 + \bar{x}_2 - 6) & = & 0 \\ \bar{u}_3(-\bar{x}_1 + 1) & = & 0 \end{array}$$

Let us first deal with $\bar{\eta}$: observe that the above obtuse angle property forces $\bar{\eta}_1$ and $\bar{\eta}_2$ to be nonpositive, and $\bar{x}_i > 0$ even implies $\bar{\eta}_i = 0$ for $i = 1, 2$ (this can be seen as a form of complementarity). Since $\bar{x}_1 \geq 1$, this means $\bar{\eta}_1 = 0$. Also, $\bar{x}_2 = 0$ stands no chance, because it would mean $\bar{x}_1^2 \leq 0$. Hence, $\bar{\eta} = 0$.

*Let $\bar{x}$ be a feasible point of $(P)$; denote by $I(\bar{x})$ the set of all $i \in \{1, \cdots, m\}$ for which $g_i(\bar{x}) = 0$.*

no chance, because it would mean $\bar{x}_1^2 \leq 0$. Hence, $\bar{\eta} = 0$. We now distinguish the following possibilities for the set $I := I(\bar{x})$:

*Case 1 ($I = \emptyset$):* By complementary slackness, $\bar{u}_1 = \bar{u}_2 = \bar{u}_3 = 0$, so the Lagrange inclusion gives $\bar{x}_1 = 9/4$, $\bar{x}_2 = 2$, which violates the first constraint ($(9/4)^2 \not\leq 2$).

$$\begin{aligned} x_1^2 - x_2 &\le 0 \\ x_1 + x_2 - 6 &\le 0 \\ -x_1 + 1 &\le 0 \end{aligned}$$

$$\begin{aligned} \bar{u}_1(\bar{x}_1^2 - \bar{x}_2) &= 0 \\ \bar{u}_2(\bar{x}_1 + \bar{x}_2 - 6) &= 0 \\ \bar{u}_3(-\bar{x}_1 + 1) &= 0 \end{aligned}$$

*Case 2 ($I = \{1\}$):* By complementary slackness, $\bar{u}_2 = \bar{u}_3 = 0$. The Lagrange inclusion gives $\bar{x}_1 = \frac{9}{4}(1 + \bar{u}_1)^{-1}$, $\bar{x}_2 = \bar{u}_1/2 + 2$, so, since $\bar{x}_1^2 = \bar{x}_2$, by definition of $I$, we obtain the equation $\bar{u}_1^3 + 6\bar{u}_1^2 + 9\bar{u}_1 = 49/8$, which has $\bar{u}_1 = 1/2$ as its only solution. It follows then that $\bar{x} = (3/2, 9/4)^t$.

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 2(\bar{x}_1 - \frac{9}{4}) \\ 2(\bar{x}_2 - 2) \end{pmatrix} + \bar{u}_1 \begin{pmatrix} 2\bar{x}_1 \\ -1 \end{pmatrix} + \bar{u}_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \bar{u}_3 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + \begin{pmatrix} \bar{\eta}_1 \\ \bar{\eta}_2 \end{pmatrix}$$

At this stage we can already stop: Theorem 3.1($i$) guarantees that, in fact, $\bar{x} = (3/2, 9/4)^t$ is an optimal solution of $(P)$. Moreover, since the objective function $(x_1, x_2) \mapsto (x_1 - \frac{9}{4})^2 + (x_2 - 2)^2$ is *strictly* convex, it follows that any optimal solution of $(P)$ must be unique. So $\bar{x} = (3/2, 9/4)^t$ is the unique optimal solution of $(P)$.

**Corollary 3.5 (Kuhn-Tucker – general case)** *Let* $f, g_1, \cdots, g_m : \mathbb{R}^n \to (-\infty, +\infty]$ *be convex functions, let* $S \subset \mathbb{R}^n$ *be a convex set. Also, let* $A$ *be a* $p \times n$*-matrix and let* $b \in \mathbb{R}^p$. *Define* $L := \{x : Ax = b\}$. *Consider the* convex programming problem

$$(P) \quad \inf_{x \in S} \{f(x) : g_1(x) \leq 0, \cdots, g_m(x) \leq 0, Ax - b = 0\}.$$

*Let* $\bar{x}$ *be a feasible point of* $(P)$*; denote by* $I(\bar{x})$ *the set of all* $i \in \{1, \cdots, m\}$ *for which* $g_i(\bar{x}) = 0$.

*(i)* $\bar{x}$ *is an optimal solution of* $(P)$ *if there exist vectors of multipliers* $\bar{u} \in \mathbb{R}^m_+$, $\bar{v} \in \mathbb{R}^p$ *and* $\bar{\eta} \in \mathbb{R}^n$ *such that the complementary slackness relationship and the obtuse angle property hold just as in Theorem 3.1(i), as well as the following version of the normal Lagrange inclusion:*

$$0 \in \partial f(\bar{x}) + \sum_{i \in I(\bar{x})} \bar{u}_i \partial g_i(\bar{x}) + A^t \bar{v} + \bar{\eta}.$$

*(ii) Conversely, if* $\bar{x}$ *is an optimal solution of* $(P)$ *and if both* $\bar{x} \in \operatorname{int} \operatorname{dom} f \cap \cap_{i \in I(\bar{x})} \operatorname{int} \operatorname{dom} g_i$ *and* $\operatorname{int} S \cap L \neq \emptyset$, *then there exist multipliers* $\bar{u}_0 \in \{0, 1\}$, $\bar{u} \in \mathbb{R}^m_+$, $(\bar{u}_0, \bar{u}) \neq (0, 0)$, *and* $\bar{v} \in \mathbb{R}^p$, $\bar{\eta} \in \mathbb{R}^n$ *such that the complementary slackness relationship and obtuse angle property of part (i) hold, as well as the following Lagrange inclusion:*

$$0 \in \bar{u}_0 \partial f(\bar{x}) + \sum_{i \in I(\bar{x})} \bar{u}_i \partial g_i(\bar{x}) + A^t \bar{v} + \bar{\eta}.$$

PROOF. Observe that $\partial\chi_L(\bar{x}) = \text{im } A^t$. Indeed, $\eta \in \partial\chi_L(\bar{x})$ is equivalent to $\eta^t(x - \bar{x}) \leq 0$ for all $x \in L$, i.e., to $\eta^t(x - \bar{x}) = 0$ for all $x \in \mathbb{R}^n$ with $A(x - \bar{x}) = 0$. But the latter states that $\eta$ belongs to the bi-orthoplement of the linear subspace im $A^t$, so it belongs to im $A^t$ itself. This proves the observation. Let us note that the above problem $(P)$ is precisely the same problem as the one of Theorem 3.1, but with $S$ replaced by $S' := S \cap L$. Thus, parts $(i)$ and $(ii)$ follow directly from Theorem 3.1, but now $\bar{\eta}$ as in Theorem 3.1 has to be replaced by an element (say $\eta'$) in $\partial\chi_{S'}$. From Theorem 2.9 we know that

$$\partial\chi_{S'}(\bar{x}) = \partial\chi_S(\bar{x}) + \partial\chi_L(\bar{x}),$$

in view of the condition int $S \cap L \neq \emptyset$. Therefore, $\eta'$ can be decomposed as $\eta' = \bar{\eta} + \eta$, with $\bar{\eta} \in \partial\chi_S(\bar{x})$ (this amounts to the obtuse angle property, of course), and with $\eta \in \partial\chi_L(\bar{x})$. By the above there exists $\bar{v} \in \mathbb{R}^m$ with $\eta = A^t\bar{v}$ and this finishes the proof. QED

https://math.stackexchange.com/questions/1205388/is-the-formula-textker-a-perp-textim-at-necessarily-true

**Example 3.6** Let $c_1, \cdots, c_n$, $a_1, \cdots, a_n$ and $b$ be positive real numbers. Consider the following optimization problem:

$$(P) \ \text{minimize} \sum_{i=1}^{n} \frac{c_i}{x_i}$$

over all $x = (x_1, \cdots, x_n)^t \in \mathbb{R}^n_{++}$ (the strictly positive orthant) such that

$$\sum_{i=1}^{n} a_i x_i = b.$$

Let us try to meet the sufficient conditions of Corollary 3.5$(i)$. Thus, we must find a feasible $\bar{x} \in \mathbb{R}^n$ and multipliers $\bar{v} \in \mathbb{R}$, $\bar{\eta} \in \mathbb{R}^n$ such that

$$\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} -\frac{c_1}{\bar{x}_1^2} \\ \vdots \\ -\frac{c_n}{\bar{x}_n^2} \end{pmatrix} + \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \bar{v} + \bar{\eta}.$$

and such that the obtuse angle property holds for $\bar{\eta}$. To begin with the latter, since we seek $\bar{x}$ in the open set $S := \mathbb{R}^n_{++}$, the only $\bar{\eta}$ with the obtuse angle property is $\bar{\eta} = 0$. The above Lagrange inclusion gives $\bar{x}_i = (c_i/(\bar{v}a_i))^{1/2}$ for all $i$. To determine $\bar{v}$, which must certainly be positive, we use the constraint: $b = \sum_i a_i \bar{x}_i = \sum_i (a_i c_i/\bar{v})^{1/2}$, which gives $\bar{v} = (\sum_i (a_i c_i)^{1/2}/b)^2$. Thus, all conditions of Corollary 3.5($i$) are seen to hold: an optimal solution of ($P$) is $\bar{x}$, given by

$$\bar{x}_i = \sqrt{\frac{c_i}{a_i}} \frac{b}{\sum_{j=1}^n \sqrt{a_j c_j}},$$

and it is implicit in our derivation that this solution is unique (exercise).