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 functions $f(\boldsymbol{x}), g_1(\boldsymbol{x}), \ldots, g_M(\boldsymbol{x})$, $g_m : \mathbb{R}^N \to \mathbb{R}$, are convex and differentiable.

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

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}}\;\; f(\boldsymbol{x}) \quad \text{subject to} \quad g_1(\boldsymbol{x}) &\le 0 \\ g_2(\boldsymbol{x}) &\le 0 \\ &\vdots \\ g_M(\boldsymbol{x}) &\le 0 \end{aligned} \tag{1}$$

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

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

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

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

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

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

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

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

[1] The g_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, and the concept of a descent/ascent direction.

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

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

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,

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

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,

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

We can rewrite (K4) as

$$\nabla f(\boldsymbol{x}^\star) + \lambda_1^\star \nabla g_1(\boldsymbol{x}^\star) + \cdots + \lambda_R^\star \nabla g_R(\boldsymbol{x}^\star) = \mathbf{0}, \qquad (2)$$

and note that by (K3),

$$g_1(\boldsymbol{x}^\star) = 0, \ldots, g_R(\boldsymbol{x}^\star) = 0.$$

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

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

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

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

It is now clear that

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

as otherwise there is no way (2) can hold with positive λ_m. Along with the convexity of f, this means that

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

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:

$$\underset{\boldsymbol{x}\in\mathbb{R}^N}{\text{minimize}}\ f(\boldsymbol{x}) \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}.$$

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

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

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

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

$$\begin{aligned} g_1(\boldsymbol{x}^\star) = 0\ ,\ g_2(\boldsymbol{x}^\star) = 0\ ,\ \ldots\ ,\ g_R(\boldsymbol{x}^\star) = 0,\\ g_{R+1}(\boldsymbol{x}^\star) < 0\ ,\ \ldots,\ g_M(\boldsymbol{x}^\star) < 0.\end{aligned}$$

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

$$\nabla f(\boldsymbol{x}^\star) + \lambda_1 \nabla g_1(\boldsymbol{x}^\star) + \cdots + \lambda_R \nabla g_R(\boldsymbol{x}^\star) = \mathbf{0}. \tag{3}$$

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

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

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

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

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

$$\begin{aligned} \langle \boldsymbol{d}, \nabla f(\boldsymbol{x}^\star)\rangle &< 0 \\ \langle \boldsymbol{d}, \nabla g_1(\boldsymbol{x}^\star)\rangle &\leq 0 \\ &\vdots \\ \langle \boldsymbol{d}, \nabla g_R(\boldsymbol{x}^\star)\rangle &\leq 0. \end{aligned}$$

This means that $\boldsymbol{d}$ is a descent direction for f, and is not an ascent direction for $g_1, \ldots, g_R$. Because the constraint functionals are affine, if $\langle \boldsymbol{d}, \nabla g_m(\boldsymbol{x}^\star)\rangle = 0$ above, then $g_m(\boldsymbol{x}^\star + t\boldsymbol{d}) = g_m(\boldsymbol{x}^\star)$ — this means that moving in the direction $\boldsymbol{d}$ will not increase $g_1, \ldots, g_m$. 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

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

but also maintains feasibility:

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

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:

$$g_1(\boldsymbol{x}) < 0 \ , \ g_2(\boldsymbol{x}) < 0 \ , \ \cdots \ , \ g_M(\boldsymbol{x}) < 0.$$

Suppose that Slater's condition holds for $g_1, \ldots, g_M$, and let $\boldsymbol{x}^\star$ be a solution to

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

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

$$\begin{aligned} g_1(\boldsymbol{x}^\star) = 0 \ , \ g_2(\boldsymbol{x}^\star) = 0 \ , \ \ldots \ , \ g_R(\boldsymbol{x}^\star) = 0, \\ g_{R+1}(\boldsymbol{x}^\star) < 0 \ , \ \ldots, \ g_M(\boldsymbol{x}^\star) < 0. \end{aligned}$$

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

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

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

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

$$\begin{aligned} \langle \boldsymbol{u}, \nabla f(\boldsymbol{x}^\star)\rangle &< 0 \\ \langle \boldsymbol{u}, \nabla g_1(\boldsymbol{x}^\star)\rangle &\leq 0 \\ &\vdots \\ \langle \boldsymbol{u}, \nabla g_R(\boldsymbol{x}^\star)\rangle &\leq 0. \end{aligned}$$

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

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

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

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

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

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

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 $g_m(\boldsymbol{z}) < 0$ for the g_m which are not affine. We now state this formally:

Suppose that $g_1, \ldots, g_{M'}$ are affine functionals, and $g_{M'+1}, \ldots, g_M$ are convex functional which are not affine. Suppose that Slater's condition holds for $g_{M'+1}, \ldots, g_M$, and let $\boldsymbol{x}^\star$ be a solution to

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

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

$$\underset{\boldsymbol{x}}{\text{minimize}}\ f(\boldsymbol{x}) \quad \text{subject to} \quad \begin{aligned} g_m(\boldsymbol{x}) &\le 0, \quad m = 1, \ldots, M \\ h_p(\boldsymbol{x}) &= 0, \quad p = 1, \ldots, P \end{aligned} \tag{5}$$

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

$$\begin{aligned} g_m(\boldsymbol{x}) &\le 0, \quad m = 1, \ldots, M, && \text{(K1)} \\ h_p(\boldsymbol{x}) &= 0, \quad p = 1, \ldots, P && \\ \boldsymbol{\lambda} &\ge \mathbf{0}, && \text{(K2)} \\ \lambda_m g_m(\boldsymbol{x}) &= 0, \quad m = 1, \ldots, M, && \text{(K3)} \\ \nabla f(\boldsymbol{x}) + \sum_{m=1}^{M} \lambda_m \nabla g_m(\boldsymbol{x}) + \sum_{p=1}^{P} \nu_p \nabla h_p(\boldsymbol{x}) &= \mathbf{0}, && \text{(K4)} \end{aligned}$$

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

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

If the g_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 g_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 λ_1 is the Lagrange multiplier for $\langle \boldsymbol{x}, \boldsymbol{a} \rangle - b$ and λ_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.