CE 272
Traffic Network Equilibrium

Lecture 3

Review of Convex Optimization - Part II

Previously on Traffic Network Equilibrium...

Definition (Convex Set)

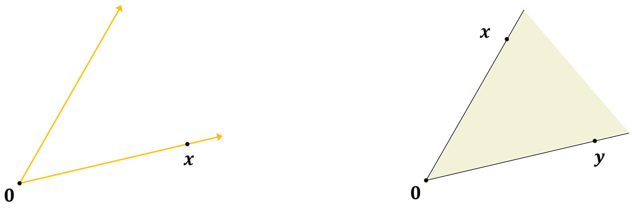
A set X is convex iff the convex combination of any two points in the set also belongs to the set. Mathematically,

$$X \subseteq \mathbb{R}^n \text{ is convex} \Leftrightarrow \forall \mathbf{x}, \mathbf{y} \in X \text{ and } \forall \lambda \in [0,1], \lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in X$$

Previously on Traffic Network Equilibrium...

Definition (Cone)

A set C is called a cone if for every $\mathbf{x} \in C$ and $\lambda \geq 0$, $\lambda\mathbf{x} \in C$.



Definition (Convex Cone)

A set C is called a convex cone if it is convex and a cone, i.e., $\forall\, \mathbf{x}, \mathbf{y} \in C$ and $\lambda_1, \lambda_2 \geq 0$, $\lambda_1\mathbf{x} + \lambda_2\mathbf{y} \in C$.

Previously on Traffic Network Equilibrium...

Definition (Convexity of General Functions)

A function $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ is convex if $\forall \mathbf{x}, \mathbf{y} \in X, \lambda \in [0, 1]$,

$$f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y})$$

Definition (Convexity of Differentiable Functions)

A **differentiable** function $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ is convex iff

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) \,\forall\, \mathbf{x}, \mathbf{y} \in X$$

Definition (Convexity of Twice-Differentiable Functions)

A **twice-differentiable** function $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ is convex iff $\nabla^2 f(\mathbf{x}) \succeq 0 \,\forall\, \mathbf{x} \in X$.

Previously on Traffic Network Equilibrium...

For unconstrained problems,

Proposition (Necessary Conditions)

$\mathbf{x}^*$ *is a local minimum of a differentiable function* $f : X \subset \mathbb{R}^n \to \mathbb{R}$ $\Rightarrow \nabla f(\mathbf{x}^*) = \mathbf{0}$

Proposition (Necessary and Sufficient Conditions)

$\mathbf{x}^*$ *is a global minimum of a differentiable convex function* $f : X \subset \mathbb{R}^n \to \mathbb{R} \Leftrightarrow \nabla f(\mathbf{x}^*) = \mathbf{0}$

Lecture Outline

1. Duality
2. KKT Conditions
3. Exercises

Lecture Outline

Duality

Duality

Primal Problem

Let's call the optimization problem in the standard form the **primal**. Suppose that f^* is an optimal solution to the primal.

Definition (Primal Problem)

$$\begin{aligned} \min_{\mathbf{x}} \; & f(\mathbf{x}) & \\ \text{s.t.} \; & g_i(\mathbf{x}) \leq 0 & \forall\, i = 1, 2, \ldots, l \\ & h_i(\mathbf{x}) = 0 & \forall\, i = 1, 2, \ldots, m \end{aligned}$$

- For now, let's not make any assumptions on convextiy.
- Also, recall that X is the set of feasible points that satisfy the implicit constraints.

Duality

Primal Problem

Note down the following example. We will use it to illustrate the concepts defined in this lecture.

$$\begin{aligned} \min_{x_1, x_2} \; & x_1^2 + x_2^2 \\ \text{s.t. } & x_1^2 + x_2^2 \leq 5 \\ & x_1 + 2x_2 = 4 \\ & x_1, x_2 \geq 0 \end{aligned}$$

Duality

Lagrangian

Definition (Lagrangian)

The Lagrangian function $\mathcal{L} : X \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}$ of the primal is defined as

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \sum_{i=1}^{l} \lambda_i g_i(\mathbf{x}) + \sum_{i=1}^{m} \mu_i h_i(\mathbf{x})$$

The $\boldsymbol{\lambda}$s and $\boldsymbol{\mu}$s are referred to as **Lagrange multipliers**.

- If the Lagrange multipliers are zeros, we recover the primal objective.
- Otherwise, we may interpret the Lagrangian as the objective plus a penalty (reward) for violating (satisfying) a constraint.

Definition (Dual Function)

We define the dual function $\mathcal{F} : \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}$ as

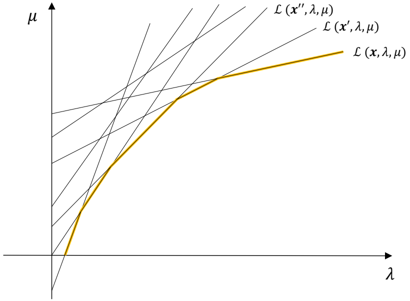
$$\begin{aligned}\mathcal{F}(\boldsymbol{\lambda}, \boldsymbol{\mu}) &= \inf_{\mathbf{x} \in X} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \\ &= \inf_{\mathbf{x} \in X} \left(f(\mathbf{x}) + \sum_{i=1}^{l} \lambda_i g_i(\mathbf{x}) + \sum_{i=1}^{m} \mu_i h_i(\mathbf{x}) \right)\end{aligned}$$

Duality

Dual Function

Proposition (Concavity)

The dual function $\mathcal{F}$ is concave in $(\boldsymbol{\lambda}, \boldsymbol{\mu})$.



For each $\mathbf{x}$, the Lagrangian is an affine function in $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ and the infimum of affine functions is concave.

Proposition (Lower Bound)

If $\boldsymbol{\lambda} \geq \mathbf{0}$, then $\mathcal{F}(\boldsymbol{\lambda}, \boldsymbol{\mu})$ is a lower bound on f^* for any $(\boldsymbol{\lambda}, \boldsymbol{\mu})$.

Proof.

Consider a primal feasible solution $\bar{\mathbf{x}}$. Since, it is feasible, $h_i(\bar{\mathbf{x}}) = 0$ for all $i = 1, \ldots, m$. Hence, the Lagrangian at $\bar{\mathbf{x}}$ can be written as

$$\begin{aligned} \mathcal{L}(\bar{\mathbf{x}}, \boldsymbol{\lambda}, \boldsymbol{\mu}) &= f(\bar{\mathbf{x}}) + \sum_{i=1}^{l} \lambda_i g_i(\bar{\mathbf{x}}) \\ &\leq f(\bar{\mathbf{x}}) \end{aligned}$$

The last inequality is true since $\boldsymbol{\lambda} \geq \mathbf{0}$, $g_i(\bar{\mathbf{x}}) \leq 0 \, \forall \, i = 1, \ldots, l$. Now consider the dual function

$$\begin{aligned} \mathcal{F}(\boldsymbol{\lambda}, \boldsymbol{\mu}) &= \inf_{\mathbf{x} \in X} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \\ &\leq \mathcal{L}(\bar{\mathbf{x}}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \leq f(\bar{\mathbf{x}}) \end{aligned}$$

■

Duality

Dual Problem

Given a $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ such that $\boldsymbol{\lambda} \geq \mathbf{0}$, you can use dual function and generate a lower bound to the primal. Can we find the best possible lower bound?

Definition (Dual Problem)

$$\max_{\boldsymbol{\lambda}, \boldsymbol{\mu}} \; \mathcal{F}(\boldsymbol{\lambda}, \boldsymbol{\mu})$$
$$\text{s.t. } \boldsymbol{\lambda} \geq \mathbf{0}$$

The Lagrange multipliers are also thus called **dual variables**.

This is a very powerful result! Using a **convex** program (Why?) we can generate a lower bound to the primal problem (even if the primal is not convex)!!

And if we know an upper bound, we can bound the optimal value. (Do we know any upper bound?)

Duality

Recap

- Primal Problem: $\min f(\mathbf{x})$ s.t. $g_i(\mathbf{x}) \leq 0, h_i(\mathbf{x}) = 0$
- Lagrangian: $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \sum \lambda_i g_i(\mathbf{x}) + \sum \mu_i h_i(\mathbf{x})$
- Dual Function: $\mathcal{F}(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$
- Dual Problem: $\max \mathcal{F}(\boldsymbol{\lambda}, \boldsymbol{\mu})$ s.t. $\boldsymbol{\lambda} \geq \mathbf{0}$

Duality

Weak and Strong Duality

Suppose f^* and $\mathcal{F}^*$ denote the optimal values of the primal and dual problems. The term $f^* - \mathcal{F}^*$ is referred to as **duality gap**.

Definition (Weak Duality)

Weak duality holds if $\mathcal{F}^* \leq f^*$. (Always true)

Definition (Strong Duality)

Strong duality is said to hold if $\mathcal{F}^* = f^*$.

When does strong duality hold? In many cases, some of which do not even require convexity! The conditions (called constraint qualifications) however are usually complicated and we do not need to know much about it for this course.

Let's look at one instance called **Slater's condition**. If our primal was of the form

$$\begin{aligned} &\min_{\mathbf{x}} f(\mathbf{x}) \\ \text{s.t. } &g_i(\mathbf{x}) \leq 0 \;\; \forall\, i = 1, 2, \ldots, l \\ &\mathbf{Ax} = \mathbf{b} \end{aligned}$$

where f and gs are all convex and there exist a feasible $\mathbf{x}$ such that $g_i(\mathbf{x}) < 0 \,\forall\, i = 1, \ldots l$, then strong duality holds.

Duality

Complementary Slackness

Suppose strong duality holds. Let $\mathbf{x}^*$ be optimal to the primal problem, and $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ be optimal to the dual problem.

$$
\begin{aligned}
f(\mathbf{x}^*) &= \mathcal{F}(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \\
&= \inf_{x \in X} \left(f(\mathbf{x}) + \sum_{i=1}^{l} \lambda_i^* g_i(\mathbf{x}) + \sum_{i=1}^{m} \mu_i^* h_i(\mathbf{x}) \right) \\
&\leq f(\mathbf{x}^*) + \sum_{i=1}^{l} \lambda_i^* g_i(\mathbf{x}^*) + \cancelto{0}{\sum_{i=1}^{m} \mu_i^* h_i(\mathbf{x}^*)}
\end{aligned}
$$

$\Rightarrow \sum_{i=1}^{l} \lambda_i^* g_i(\mathbf{x}^*) \geq 0$, but from primal and dual feasibility, $\sum_{i=1}^{l} \lambda_i^* g_i(\mathbf{x}^*) \leq 0$.

$$\therefore \sum_{i=1}^{l} \lambda_i^* g_i(\mathbf{x}^*) = 0$$

Duality

Complementary Slackness

Recall that $\lambda_i^* \geq 0$ and $g_i(\mathbf{x}^*) \leq 0$. Hence, $\sum_{i=1}^{l} \lambda_i^* g_i(\mathbf{x}^*) = 0$ implies that the following **complementary slackness conditions** must hold

$$\lambda_i^* g_i(\mathbf{x}^*) = 0 \, \forall \, i = 1, \ldots, l$$

Which implies

- $\lambda_i^* > 0 \Rightarrow g_i(\mathbf{x}^*) = 0$
- $g_i(\mathbf{x}^*) < 0 \Rightarrow \lambda_i^* = 0$

Duality

Complementary Slackness

Proposition

Let $\mathbf{x}^*$ and $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ be optimal to the primal and dual respectively. Suppose strong duality holds. Then $\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \mathbf{0}$

Proof.

$$\begin{aligned}
\mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) &= f(\mathbf{x}^*) + \cancelto{0}{\sum_{i=1}^{l} \lambda_i^* g_i(\mathbf{x}^*)} + \cancelto{0}{\sum_{i=1}^{m} \mu_i^* h_i(\mathbf{x}^*)} \\
&= f(\mathbf{x}^*)
\end{aligned}$$

From strong duality,

$$f(\mathbf{x}^*) = \mathcal{F}(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \inf_{x \in X} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$$

Hence, $\mathbf{x}$ minimizes $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$. Therefore, $\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \mathbf{0}$ ■

Duality

Recap

- Primal Problem: $\min f(\mathbf{x})$ s.t. $g_i(\mathbf{x}) \leq 0, h_i(\mathbf{x}) = 0$
- Lagrangian: $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \sum \lambda_i g_i(\mathbf{x}) + \sum \mu_i h_i(\mathbf{x})$
- Dual Function: $\mathcal{F}(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$
- Dual Problem: $\max \mathcal{F}(\boldsymbol{\lambda}, \boldsymbol{\mu})$ s.t. $\boldsymbol{\lambda} \geq \mathbf{0}$

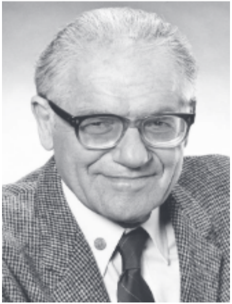
- Weak Duality: $\mathcal{F}^* \leq f^*$
- Strong Duality: $\mathcal{F}^* = f^*$
- Complementary Slackness: $\mathcal{F}^* = f^* \Rightarrow \lambda_i^* g_i(\mathbf{x}^*) = 0$

KKT Conditions

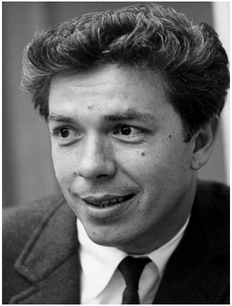
KKT Conditions

History Break

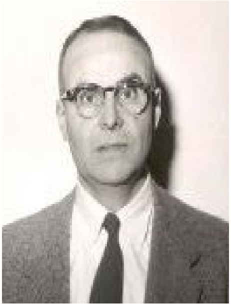
Karush-Kuhn-Tucker (KKT) conditions are named after William Karush, Harold Kuhn, and Albert Tucker.



William Karush



Harold Kuhn



Albert Tucker

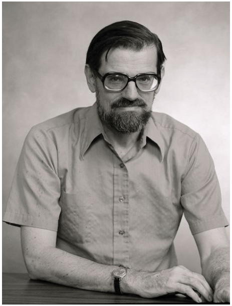
These were popularly known as Kuhn-Tucker conditions after the authors who discovered them in 1951 but Karush had derived similar results in his master's thesis in 1939. See [PDF] for a historical account of the KKT conditions.

KKT Conditions

History Break

Incidentally, Albert Tucker was the one who formalized 'Prisoner's Dilemma' and also produced these two PhDs, both of whom won the Nobel in economics.

John Nash



Lloyd Shapley

Although they never worked on traffic, we'll see some of their connections with this course later.

KKT Conditions

Necessary Conditions

The results that we have derived so far are essentially the necessary conditions for optimality.

KKT Conditions

Necessary Conditions

Proposition (Necessary KKT Conditions)

Assuming strong duality holds, any $\mathbf{x}^*$ *and* $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ *that are optimal for the primal and dual problems must satisfy*

- *Primal Feasibility*

$$g_i(\mathbf{x}^*) \leq 0 \, \forall \, i = 1, \ldots, l$$
$$h_i(\mathbf{x}^*) = 0 \, \forall \, i = 1, \ldots, m$$

- *Dual Feasibility*

$$\boldsymbol{\lambda}^* \geq \mathbf{0}$$

- *Complementary Slackness*

$$\lambda_i^* g_i(\mathbf{x}^*) = 0 \, \forall \, i = 1, \ldots, m$$

- *Gradient of the Lagrangian vanishes*

$$\nabla_{\mathbf{x}} f(\mathbf{x}^*) + \sum_{i=1}^{l} \lambda_i^* \nabla_{\mathbf{x}} g_i(\mathbf{x}^*) + \sum_{i=1}^{m} \mu_i^* \nabla_{\mathbf{x}} h_i(\mathbf{x}^*) = \mathbf{0}$$

KKT Conditions

Sufficient Conditions

As is the case with unconstrained optimization, any $\mathbf{x}$ and $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ that satisfy the KKT conditions are not optimal to the primal and dual. We need additional assumptions for them to be sufficient.

KKT Conditions

Sufficient Conditions

Proposition (Sufficient KKT Conditions)

Suppose f, g_i, and h_i are all differentiable and convex. Then, any $\bar{\mathbf{x}}$ and $(\bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}})$ that satisfy the following KKT conditions are optimal to the primal and dual and the duality gap is 0.

$$g_i(\bar{\mathbf{x}}) \leq 0 \,\forall\, i = 1, \ldots, l$$

$$h_i(\bar{\mathbf{x}}) = 0 \,\forall\, i = 1, \ldots, m$$

$$\bar{\boldsymbol{\lambda}} \geq \mathbf{0}$$

$$\bar{\lambda}_i g_i(\bar{\mathbf{x}}) = 0 \,\forall\, i = 1, \ldots, m$$

$$\nabla_{\mathbf{x}} f(\bar{\mathbf{x}}) + \sum_{i=1}^{l} \bar{\lambda}_i \nabla_{\mathbf{x}} g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{m} \bar{\mu}_i \nabla_{\mathbf{x}} h_i(\bar{\mathbf{x}}) = \mathbf{0}$$

KKT Conditions
Visualizing KKT Conditions

Suppose we wish to optimize the following problem.

$$\begin{aligned} \min_{\mathbf{x}} \ & f(\mathbf{x}) \\ \text{s.t.} \ & g_1(\mathbf{x}) \leq 0 \\ & g_2(\mathbf{x}) \leq 0 \end{aligned}$$

$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \lambda_1 g_1(\mathbf{x}) + \lambda_2 g_2(\mathbf{x})$ and one of the KKT conditions is

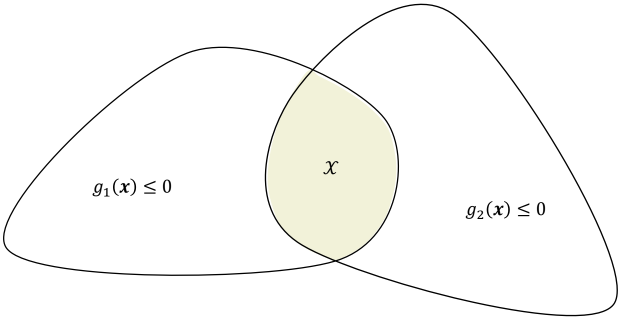
$$\begin{aligned} \nabla f(\mathbf{x}) + \lambda_1 \nabla g_1(\mathbf{x}) + \lambda_2 \nabla g_2(\mathbf{x}) = \mathbf{0} \\ \Rightarrow -\nabla f(\mathbf{x}) = \lambda_1 \nabla g_1(\mathbf{x}) + \lambda_2 \nabla g_2(\mathbf{x}) \end{aligned}$$

Let's try to relate this to our normal cone version of the optimality conditions.

KKT Conditions

Visualizing KKT Conditions

Suppose the feasible region looks as shown below

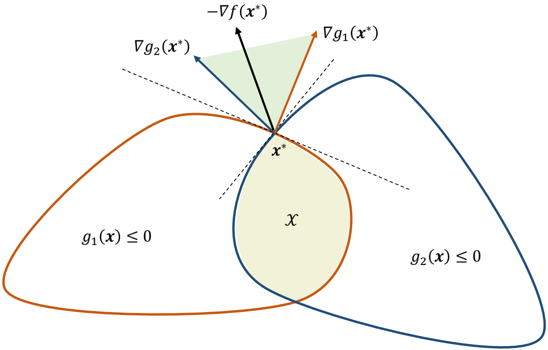


The boundaries of the constraints are $g_1(\mathbf{x}) = 0$ and $g_2(\mathbf{x}) = 0$, and the $\mathbf{x}$ values satisfying these points are the level sets of g_1 and g_2.

KKT Conditions

Visualizing KKT Conditions

Therefore, $\nabla g_1(\mathbf{x}^*)$ and $\nabla g_2(\mathbf{x}^*)$ are orthogonal to boundaries.



Since $-\nabla f(\mathbf{x}^*) = \lambda_1 \nabla g_1(\mathbf{x}^*) + \lambda_2 \nabla g_2(\mathbf{x}^*)$, it belongs to the cone formed by $\nabla g_1(\mathbf{x}^*)$ and $\nabla g_2(\mathbf{x}^*)$, which is also the normal cone at $\mathbf{x}^*$.

KKT Conditions

Summary

	Unconstrained	Constrained
Necessary Conditions	$\mathbf{x}^*$ is optimal $\Rightarrow \nabla f(\mathbf{x}^*) = \mathbf{0}$	$\mathbf{x}^*$ is primal optimal and $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is dual optimal and strong duality holds $\Rightarrow$ KKT conditions are satisfied
Sufficient Conditions	$\nabla f(\mathbf{x}^*) = \mathbf{0}$ for a convex function $\Rightarrow \mathbf{x}^*$ is optimal	Objective and constraints involve convex functions and $\bar{\mathbf{x}}$ and $(\bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}})$ satisfy KKT conditions $\Rightarrow \bar{\mathbf{x}}$ and $(\bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}})$ are optimal for the primal and dual and the duality gap is 0

Lecture Outline

Exercises

Exercises

Exercise 1

Using KKT conditions solve

$$\begin{aligned} \min_{x_1, x_2} \; & (x_1 - 1)^2 + x_2 - 2 \\ \text{s.t.} \; & x_1 + x_2 \leq 2 \\ & x_2 - x_1 = 1 \end{aligned}$$

- Are the objective and constraints convex?
- Is the solution optimal?

Write the KKT conditions for the following problem

$$\begin{aligned} \min_{\mathbf{x}} \ & f(\mathbf{x}) \\ \text{s.t.} \ & \mathbf{x} \geq \mathbf{0} \end{aligned}$$

Exercises

Exercise 3

Write the KKT conditions for the following problem

$$\min_{\mathbf{x}} \; f(\mathbf{x})$$
$$\text{s.t.} \;\; \mathbf{A}\mathbf{x} = \mathbf{b}$$
$$\mathbf{x} \geq \mathbf{0}$$

Where $\mathbf{A}\mathbf{x} = \mathbf{b}$ is

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m \end{aligned}$$

Supplementary Reading

Boyd, S., & Vandenberghe, L. (2004). Convex optimization. Cambridge university press. [PDF]

Your Moment of Zen