DTU

Karush-Kuhn-Tucker Conditions

Richard Lusby

Department of Management Engineering
Technical University of Denmark

Today's Topics

- Unconstrained Optimization
- Equality Constrained Optimization
- Equality/Inequality Constrained Optimization

Unconstrained Optimization

Unconstrained Optimization

Problem

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to:} \quad & \mathbf{x} \in \mathbb{R}^n \end{aligned}$$

First Order Necessary Conditions

If $\mathbf{x}^*$ is a local minimizer of $f(\mathbf{x})$ and $f(\mathbf{x})$ is continuously differentiable in an open neighbourhood of $\mathbf{x}^*$, then

$$\nabla f(\mathbf{x}^*) = \mathbf{0}$$

That is, $f(\mathbf{x})$ is stationary at $\mathbf{x}^*$

Unconstrained Optimization

Second Order Necessary Conditions

If $\mathbf{x}^*$ is a local minimizer of $f(\mathbf{x})$ and $\nabla^2 f(\mathbf{x})$ is continuously differentiable in an open neighbourhood of $\mathbf{x}^*$, then

$$\nabla f(\mathbf{x}^*) = \mathbf{0}$$

$\nabla^2 f(\mathbf{x}^*)$ is positive semi definite

Second Order Sufficient Conditions

Suppose that $\nabla^2 f(\mathbf{x})$ is continuously differentiable in an open neighbourhood of $\mathbf{x}^*$. If the following two conditions are satisfied, then $\mathbf{x}^*$ is a local minimum of $f(\mathbf{x})$.

$$\nabla f(\mathbf{x}^*) = \mathbf{0}$$

$\nabla^2 f(\mathbf{x}^*)$ is positive definite

Equality Constrained Optimization

Equality Constrained Optimization

Problem

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to:} \quad & h_i(\mathbf{x}) = 0 \; \forall i = 1, 2, \ldots m \\ & \mathbf{x} \in \mathbb{R}^n \end{aligned}$$

Equality Constrained Optimization

Consider the following example

Example

$$\begin{aligned} \text{minimize} \quad & 2x_1^2 + x_2^2 \\ \text{subject to:} \quad & x_1 + x_2 \quad = 1 \end{aligned}$$

- Let us first consider the unconstrained case
- Differentiate with respect to x_1 and x_2

$$\frac{\partial f(x_1, x_2)}{\partial x_1} = 4x_1$$

$$\frac{\partial f(x_1, x_2)}{\partial x_2} = 2x_2$$

- These yield the solution $x_1 = x_2 = 0$
- Does not satisfy the constraint

Equality Constrained Optimization
Example Continued

- Let us penalize ourselves for not satisfying the constraint
- This gives

$$L(x_1, x_2, \lambda_1) = 2x_1^2 + x_2^2 + \lambda_1(1 - x_1 - x_2)$$

- This is known as the Lagrangian of the problem
- Try to adjust the value λ_1 so we use just the right amount of resource

  $\lambda_1 = 0 \rightarrow$ get solution $x_1 = x_2 = 0, 1 - x_1 - x_2 = 1$

  $\lambda_1 = 1 \rightarrow$ get solution $x_1 = \frac{1}{4}, x_2 = \frac{1}{2}, 1 - x_1 - x_2 = \frac{1}{4}$

  $\lambda_1 = 2 \rightarrow$ get solution $x_1 = \frac{1}{2}, x_2 = 1, 1 - x_1 - x_2 = -\frac{1}{2}$

  $\lambda_1 = \frac{4}{3} \rightarrow$ get solution $x_1 = \frac{1}{3}, x_2 = \frac{2}{3}, 1 - x_1 - x_2 = 0$

Equality Constrained Optimization
Generally Speaking

Given the following Non-Linear Program

Problem

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) & \\ \text{subject to:} \quad & h_i(\mathbf{x}) & = 0 \ \forall i = 1, 2, \ldots m \\ & \mathbf{x} & \in \mathbb{R}^n \end{aligned}$$

A solution can be found using the Lagrangian

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \left(0 - h_i(\mathbf{x})\right)$$

Equality Constrained Optimization
Why is $L(x, \lambda)$ interesting?

Assume $\mathbf{x}^*$ minimizes the following

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to:} \quad & h_i(\mathbf{x}) = 0 \; \forall i = 1, 2, \ldots m \\ & \mathbf{x} \in \mathbb{R}^n \end{aligned}$$

The following two cases are possible:

1. The vectors $\nabla h_1(\mathbf{x}^*), \nabla h_2(\mathbf{x}^*), \ldots, \nabla h_m(\mathbf{x}^*)$ are linearly dependent
2. There exists a vector $\boldsymbol{\lambda}^*$ such that

$$\frac{\partial L(\mathbf{x}^*, \boldsymbol{\lambda}^*)}{\partial x_1} = \frac{\partial L(\mathbf{x}^*, \boldsymbol{\lambda}^*)}{\partial x_2} = \frac{\partial L(\mathbf{x}^*, \boldsymbol{\lambda}^*)}{\partial x_3} =, \ldots, = \frac{\partial L(\mathbf{x}^*, \boldsymbol{\lambda}^*)}{\partial x_n} = 0$$

$$\frac{\partial L(\mathbf{x}^*, \boldsymbol{\lambda}^*)}{\partial \lambda_1} = \frac{\partial L(\mathbf{x}^*, \boldsymbol{\lambda}^*)}{\partial \lambda_2} = \frac{\partial L(\mathbf{x}^*, \boldsymbol{\lambda}^*)}{\partial \lambda_3} =, \ldots, = \frac{\partial L(\mathbf{x}^*, \boldsymbol{\lambda}^*)}{\partial \lambda_m} = 0$$

Case 1: Example

Example

$$\begin{aligned} \text{minimize} \quad & x_1 + x_2 + x_3^2 \\ \text{subject to:} \quad & x_1 = 1 \\ & x_1^2 + x_2^2 = 1 \end{aligned}$$

- The minimum is achieved at $x_1 = 1, x_2 = 0, x_3 = 0$
- The Lagrangian is:

$$L(x_1, x_2, x_3, \lambda_1, \lambda_2) = x_1 + x_2 + x_3^2 + \lambda_1(1 - x_1) + \lambda_2(1 - x_1^2 - x_2^2)$$

- Observe that:

$$\frac{\partial L(1, 0, 0, \lambda_1, \lambda_2)}{\partial x_2} = 1 \quad \forall \lambda_1, \lambda_2$$

- Observe $\nabla h_1(1, 0, 0) = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$ and $\nabla h_2(1, 0, 0) = \begin{bmatrix} 2 & 0 & 0 \end{bmatrix}$

Case 2: Example

Example

$$\begin{aligned} \text{minimize} \quad & 2x_1^2 + x_2^2 \\ \text{subject to:} \quad & x_1 + x_2 = 1 \end{aligned}$$

- The Lagrangian is:

$$L(x_1, x_2, \lambda_1) = 2x_1^2 + x_2^2 + \lambda_1(1 - x_1 - x_2)$$

- Solve for the following:

$$\frac{\partial L(x_1^*, x_2^*, \lambda_1^*}{\partial x_1}) = 4x_1^* - \lambda_1^* = 0$$

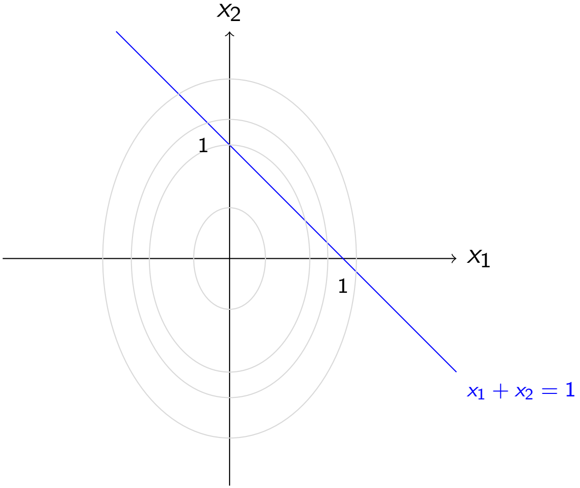
$$\frac{\partial L(x_1^*, x_2^*, \lambda_1^*}{\partial x_2}) = 2x_2^* - \lambda_1^* = 0$$

$$\frac{\partial L(x_1^*, x_2^*, \lambda_1^*)}{\partial \lambda} = 1 - x_1^* - x_2^* = 0$$

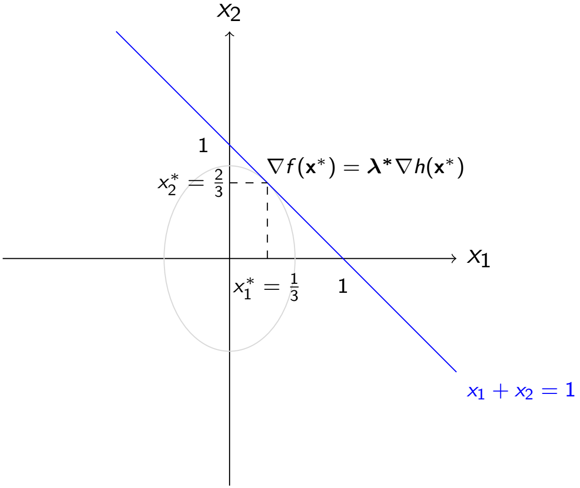
Case 2: Example continued

- Solving this system of equations yields $x_1^* = \frac{1}{3}, x_2^* = \frac{2}{3}, \lambda_1^* = \frac{4}{3}$
- Is this a minimum or a maximum?

Graphically



Graphically



Geometric Interpretation

- Consider the gradients of f and h at the optimal point
- They must point in the same direction, though they may have different lengths

$$\nabla f(\mathbf{x}^*) = \boldsymbol{\lambda}^* \nabla h(\mathbf{x}^*)$$

- Along with feasibility of $\mathbf{x^*}$, is the condition $\nabla L(\mathbf{x}^*, \boldsymbol{\lambda}^*) = 0$
- From the example, at $x_1^* = \frac{1}{3}, x_2^* = \frac{2}{3}, \lambda_1^* = \frac{4}{3}$

$$\nabla f(x_1^*, x_2^*) = \begin{bmatrix} 4x_1^* & 2x_2^* \end{bmatrix} = \begin{bmatrix} \frac{4}{3} & \frac{4}{3} \end{bmatrix}$$

$$\nabla h_1(x_1^*, x_2^*) = \begin{bmatrix} 1 & 1 \end{bmatrix}$$

Geometric Interpretation

- $\nabla f(\mathbf{x})$ points in the direction of steepest ascent
- $-\nabla f(\mathbf{x})$ points in the direction of steepest descent
- In two dimensions:
  - $\nabla f(\mathbf{x^o})$ is perpendicular to a level curve of f
  - $\nabla h_i(\mathbf{x^o})$ is perpendicular to the level curve $h_i(\mathbf{x}^o) = 0$

Equality, Inequality Constrained Optimization

Inequality Constraints

What happens if we now include inequality constraints?

General Problem

$$\begin{array}{rlll} \text{maximize} & f(\mathbf{x}) & & \\ \text{subject to:} & g_i(\mathbf{x}) \leq 0 & (\mu_i) & \forall i \in I \\ & h_j(\mathbf{x}) = 0 & (\lambda_j) & \forall i \in J \end{array}$$

- Given a feasible solution $\mathbf{x^o}$, the set of binding constraints is:

$$\mathcal{I} = \{i : g_i(\mathbf{x^o}) = 0\}$$

The Lagrangian

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(x) + \sum_{i=1}^{m} \mu_i(0 - g_i(\mathbf{x})) + \sum_{j=1}^{k} \lambda_j(0 - h_j(\mathbf{x}))$$

Inequality Constrained Optimization

Assume $\mathbf{x}^*$ maximizes the following

$$\begin{array}{lll} \text{maximize} & f(\mathbf{x}) & \\ \text{subject to:} & g_i(\mathbf{x}) \leq 0 & (\mu_i) \quad \forall i \in I \\ & h_j(\mathbf{x}) = 0 & (\lambda_j) \quad \forall i \in J \end{array}$$

The following two cases are possible:

1. $\nabla h_1(\mathbf{x}^*), \ldots, \nabla h_k(\mathbf{x}^*), \nabla g_1(\mathbf{x}^*), \ldots, \nabla g_m(\mathbf{x}^*)$ are linearly dependent
2. There exist vectors $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}*$ such that

$$\nabla f(\mathbf{x}^*) - \sum_{j=1}^{k} \lambda_j \nabla h_j(\mathbf{x}^*) - \sum_{i=1}^{m} \mu_i \nabla g_i(\mathbf{x}^*) = 0$$

$$\mu_i^* g_i(\mathbf{x}^*) = 0$$

$$\boldsymbol{\mu}^* \geq 0$$

Inequality Constrained Optimization

- These conditions are known as the Karush-Kuhn-Tucker Conditions
- We look for candidate solutions $\mathbf{x}^*$ for which we can find $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^*$
- Solve these equations using complementary slackness
- At optimality some constraints will be binding and some will be slack
- Slack constraints will have a corresponding μ_i of zero
- Binding constraints can be treated using the Lagrangian

Constraint qualifications

KKT constraint qualification

$$\nabla g_i(\mathbf{x^o}) \text{ for } i \in I \text{ are linearly independent}$$

Slater constraint qualification

- $g_i(\mathbf{x})$ for $i \in I$ are convex functions
- A non boundary point exists: $g_i(\mathbf{x}) < 0$ for $i \in I$

Case 1 Example

The Problem

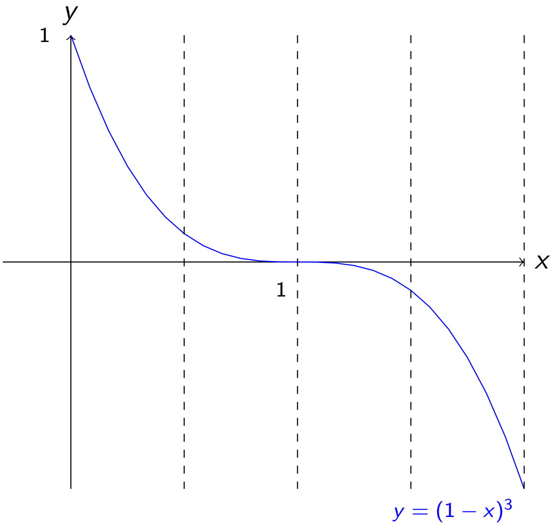
$$\begin{aligned} \text{maximize} \quad & x \\ \text{subject to:} \quad & y \le (1-x)^3 \\ & y \ge 0 \end{aligned}$$

- Consider the global max: $(x, y) = (1, 0)$
- After reformulation, the gradients are

$$\begin{aligned} \nabla f(x,y) &= (1,0) \\ \nabla g_1 &= (3(x-1)^2, 1) \\ \nabla g_2 &= (0, -1) \end{aligned}$$

- Consider $\nabla f(x,y) - \sum_{i=1}^{2} \mu_i \nabla g_i(x,y)$

Graphically



Case 1 Example

We get:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \mu_1 \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \mu_2 \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

- No μ_1 and μ_2 exist such that:

$$\nabla f(x, y) - \sum_{i=1}^{2} \mu_i \nabla g_i(x, y) = \mathbf{0}$$

Case 2 Example

The Problem

$$\begin{aligned} \text{maximize} \quad & -(x-2)^2 - 2(y-1)^2 \\ \text{subject to:} \quad & x + 4y \le 3 \\ & x \ge y \end{aligned}$$

The Problem (Rearranged)

$$\begin{aligned} \text{maximize} \quad & -(x-2)^2 - 2(y-1)^2 \\ \text{subject to:} \quad & x + 4y \le 3 \\ & -x + y \le 0 \end{aligned}$$

Case 2 Example

- The Lagrangian is:

$$L(x_1, y, \mu_1, \mu_2) = -(x-2)^2 - 2(y-1)^2 + \mu_1(3-x-4y) + \mu_2(0+x-y)$$

- This gives the following KKT conditions

$$\frac{\partial L}{\partial x} = -2(x-2) - \mu_1 + \mu_2 = 0$$

$$\frac{\partial L}{\partial y} = -4(y-1) - 4\mu_1 - \mu_2 = 0$$

$$\mu_1(3-x-4y) = 0$$

$$\mu_2(x-y) = 0$$

$$\mu_1, \mu_2 \geq 0$$

Case 2 Example
Continued

We have two complementarity conditions $\rightarrow$ check 4 cases

1. $\mu_1 = \mu_2 = 0 \rightarrow x = 2, y = 1$
2. $\mu_1 = 0, x - y = 0 \rightarrow x = \frac{4}{3}, \mu_2 = -\frac{4}{3}$
3. $3 - x - 4y = 0, \mu_2 = 0 \rightarrow x = \frac{5}{3}, y = \frac{1}{3}, \mu_1 = \frac{2}{3}$
4. $3 - x - 4y = 0, x - y = 0 \rightarrow x = \frac{3}{5}, y = \frac{3}{5}, \mu_1 = \frac{22}{25}, \mu_2 = -\frac{48}{25}$

Optimal solution is therefore $x^* = \frac{5}{3}, y^* = \frac{1}{3}, f(x^*, y^*) = -\frac{4}{9}$

Case 2 Example
Continued

We have two complementarity conditions $\rightarrow$ check 4 cases

1. $\mu_1 = \mu_2 = 0 \rightarrow x = 2, y = 1$
2. $\mu_1 = 0, x - y = 0 \rightarrow x = \frac{4}{3}, \mu_2 = -\frac{4}{3}$
3. $3 - x - 4y = 0, \mu_2 = 0 \rightarrow x = \frac{5}{3}, y = \frac{1}{3}, \mu_1 = \frac{2}{3}$
4. $3 - x - 4y = 0, x - y = 0 \rightarrow x = \frac{3}{5}, y = \frac{3}{5}, \mu_1 = \frac{22}{25}, \mu_2 = -\frac{48}{25}$

Optimal solution is therefore $x^* = \frac{5}{3}, y^* = \frac{1}{3}, f(x^*, y^*) = -\frac{4}{9}$

Continued

The Problem

$$\begin{aligned} \text{minimize} \quad & (x-3)^2 + (y-2)^2 \\ \text{subject to:} \quad & x^2 + y^2 \leq 5 \\ & x + 2y \leq 4 \\ & x, y \geq 0 \end{aligned}$$

The Problem (Rearranged)

$$\begin{aligned} \text{maximize} \quad & -(x-3)^2 - (y-2)^2 \\ \text{subject to:} \quad & x^2 + y^2 \leq 5 \\ & x + 2y \leq 4 \\ & -x, -y \leq 0 \end{aligned}$$

Inequality Example

- The gradients are:

$$
\begin{aligned}
\nabla f(x,y) &= (6-2x, 4-2y) \\
\nabla g_1(x,y) &= (2x, 2y) \\
\nabla g_2(x,y) &= (1, 2) \\
\nabla g_3(x,y) &= (-1, 0) \\
\nabla g_4(x,y) &= (0, -1)
\end{aligned}
$$

Inequality Example
Continued

- Consider the point $(x, y) = (2, 1)$

$$\text{It is feasible } \mathcal{I} = \{1, 2\}$$

- This gives

$$\begin{bmatrix} 2 \\ 2 \end{bmatrix} - \mu_1 \begin{bmatrix} 4 \\ 2 \end{bmatrix} - \mu_2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

- $\mu_1 = \frac{1}{3}, \mu_2 = \frac{2}{3}$ satisfy this

Sufficient condition

General Problem

$$\begin{aligned} \text{maximize} \quad & f(\mathbf{x}) \\ \text{subject to:} \quad & g_i(\mathbf{x}) \leq 0 \quad \forall i \in I \end{aligned}$$

Theorem

If $f(\mathbf{x})$ is concave and $g_i(\mathbf{x})$ for $i \in I$ are convex functions then a feasible KKT point is optimal

- An equality constraint is equivalent to two inequality constraints:

$$h_j(\mathbf{x}) = 0 \Leftrightarrow h_j(\mathbf{x}) \leq 0 \text{ and } -h_j(\mathbf{x}) \leq 0$$

- The corresponding two nonnegative multipliers may be combined to one free one

$$\lambda_{j+} \nabla h(\mathbf{x}) + \lambda_{j-}(-\nabla h(\mathbf{x})) = \lambda_j \nabla h(\mathbf{x})$$

Equality constraints

General Problem

$$\begin{aligned} \text{maximize} \quad & f(\mathbf{x}) & & \\ \text{subject to:} \quad & g_i(\mathbf{x}) & \leq 0 & \quad \forall i \in I \\ & h_j(\mathbf{x}) & = 0 & \quad \forall j \in J \end{aligned}$$

- Let $\mathbf{x^o}$ be a feasible solution
- As before, $\mathcal{I} = \{i : g_i(\mathbf{x^o}) = 0\}$
- Assume constraint qualification holds

Equality constraints
Continued

KKT Necessary Optimality Conditions

If $\mathbf{x^o}$ is a local maximum, there exist multipliers $\mu_i \geq 0 \; \forall i \in I$ and λ_j $\forall j \in J$ such that

$$\nabla f(\mathbf{x^o}) - \sum_{i \in I} \mu_i \nabla g_i(\mathbf{x^o}) - \sum_{j} \lambda_j \nabla h_j(\mathbf{x^o}) = \mathbf{0}$$

KKT Sufficient Optimality Conditions

If $f(\mathbf{x})$ is concave, $g_i(\mathbf{x}) \; \forall \; i \in I$ are convex functions and $h_j \; \forall j \in J$ are affine (linear) then a feasible KKT point is optimal

KKT Conditions - Summary

General Problem

$$\begin{aligned} \text{maximize} \quad & f(\mathbf{x}) & & \\ \text{subject to:} \quad & g_i(\mathbf{x}) & \leq 0 & \quad \forall i \in I \\ & h_j(\mathbf{x}) & = 0 & \quad \forall j \in J \end{aligned}$$

KKT conditions

$$\begin{aligned} \nabla f(\mathbf{x^o}) - \sum_i \mu_i \nabla g_i(\mathbf{x^o}) - \sum_j \lambda_j \nabla h_j(\mathbf{x^o}) & = \mathbf{0} & \\ \mu_i g_i(\mathbf{x^o}) & = 0 & \forall i \in I \\ \mu_i & \geq 0 & \forall i \in I \\ \mathbf{x^o} \text{ feasible} & & \end{aligned}$$

Alternative Formulation
Vector Function Form

General Problem

$$\begin{aligned} \text{maximize} \quad & f(\mathbf{x}) \\ \text{subject to:} \quad & \mathbf{g}(\mathbf{x}) \leq 0 \\ & \mathbf{h}(\mathbf{x}) = 0 \end{aligned}$$

KKT Conditions

$$\begin{aligned} \nabla f(\mathbf{x^o}) - \boldsymbol{\mu}\nabla \mathbf{g}(\mathbf{x^o}) - \boldsymbol{\lambda}\nabla \mathbf{h}(\mathbf{x^o}) &= \mathbf{0} \\ \boldsymbol{\mu}\mathbf{g}(\mathbf{x^o}) &= \mathbf{0} \\ \boldsymbol{\mu} &\geq \mathbf{0} \\ \mathbf{x^o} \text{ feasible} & \end{aligned}$$

Class Exercise 1

The Problem

$$\begin{aligned} \text{maximize} \quad & \ln(x+1) + y \\ \text{subject to:} \quad & 2x + y \leq 3 \\ & x, y \geq 0 \end{aligned}$$

Class Exercise 2

The problem

$$\begin{aligned} \text{minimize} \quad & x^2 + y^2 \\ \text{subject to:} \quad & x^2 + y^2 && \leq 5 \\ & x + 2y && = 4 \\ & x, y && \geq 0 \end{aligned}$$

Class Exercise 3

Write the KKT conditions for

$$\begin{aligned} \text{maximize} \quad & \mathbf{c}^T\mathbf{x} \\ \text{subject to:} \quad & A\mathbf{x} \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{aligned}$$