PRO 5970 Métodos de Otimização Não Linear

Celma de Oliveira Ribeiro

Aula 8 - Karush Kuhn Tucker - 2023

Departmento de Engenharia de Produção
Universidade de São Paulo

Karush-Kuhn–Tucker Conditions

Consider problem PrGen

$$\begin{array}{llr} & \text{minimize } f(x) & \\ \text{s.t} & g_i(x) \leq 0 & i \in \mathcal{I} = \{1, 2, \dots m\} \\ & h_i(x) = 0 & i \in \mathcal{E} = \{1, 2, \dots l\} \end{array}$$

Necessary conditions for a constrained local optimum

Definition

x^* is KKT point if there are lagrange multipliers vectors $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^l$, such that $\begin{bmatrix} x^* & \lambda^* & \mu^* \end{bmatrix}^t$ satisfies:

$$\begin{array}{lll} \nabla_x \mathcal{L}(x^*, \lambda^* \mu^*) & = & 0 \\ g(x^*) & \leq 0 & \\ h(x^*) & = 0 & \\ \mu^* & \geq 0 & \\ \mu_i g_i(x^*) & = 0 & \forall i \in \mathcal{I} \text{ complementary slackness condition} \end{array}$$

- The complementary slackness condition applies only to inequality constraints.
- For equality constrained problems KKT become:

$$\begin{array}{rcl} \nabla_x \mathcal{L}\left(x^*, \lambda^* \mu^*\right) & = & 0 \\ h\left(x^*\right) & = & 0 \end{array}$$

- Inequality constraints introduce some complexity to the problem as complementary slackness conditions are non-linear

Exercice Write the KKT conditions and solve the resulting system

$$\min f(x) = x_1^2 + -4x_1 + x_2^2 - 6x_2$$

$$x_1 + x_2 = 3$$

Example Write the KKT conditions and solve the resulting system

$$\min f(x) = x_1^2 + -4x_1 + x_2^2 - 6x_2$$

$$\begin{gathered} x_1 + x_2 \leq 3 \\ -2x_1 + x_2 \leq 2 \end{gathered}$$

KKT

$$\min f(x) = x_1^2 + -4x_1 + x_2^2 - 6x_2$$

$$x_1 + x_2 \leq 3$$
$$-2x_1 + x_2 \leq 2$$

The Lagrangian is

KKT

$$\min f(x) = x_1^2 + -4x_1 + x_2^2 - 6x_2$$

$$x_1 + x_2 \leq 3$$
$$-2x_1 + x_2 \leq 2$$

The Lagrangian is
$\mathcal{L}(x, \lambda, \mu) = x_1^2 - 4x_1 + x_2^2 - 6x_2 + \mu_1(x_1 + x_2 - 3) + \mu_2(-2x_1 + x_2 - 2)$

Kuhn–Tucker conditions :

$$\min f(x) = x_1^2 + -4x_1 + x_2^2 - 6x_2$$

$$x_1 + x_2 \leq 3$$
$$-2x_1 + x_2 \leq 2$$

The Lagrangian is
$\mathcal{L}(x, \lambda, \mu) = x_1^2 - 4x_1 + x_2^2 - 6x_2 + \mu_1(x_1 + x_2 - 3) + \mu_2(-2x_1 + x_2 - 2)$

Kuhn–Tucker conditions :

$$\nabla_x \mathcal{L}(x^*, \lambda^* \mu^*) = 0 \quad \Rightarrow \quad \begin{cases} \frac{\partial \mathcal{L}}{\partial x_1} = 2x_1 - 4 + \mu_1 - 2\mu_2 = 0 \\ \frac{\partial \mathcal{L}}{\partial x_2} = 2x_2 - 6 + \mu_1 + \mu_2 = 0 \end{cases}$$

$$g(x^*) \leq 0 \Rightarrow \begin{cases} x_1 + x_2 \leq 3 \\ -2x_1 + x_2 \leq 2 \end{cases}$$

$\mu \geq 0$

$$\mu_i g_i(x^*) = 0 \quad \Rightarrow \quad \begin{cases} \mu_1(x_1 + x_2 - 3) = 0 \\ \mu_2(-2x_1 + x_2 - 2) = 0 \end{cases}$$

There is no simple computational procedure for the solution

KKT

- $\mu_1 = 0$ and $\mu_2 = 0$
  $\begin{cases} 2x_1 - 4 = 0 \Rightarrow x_1 = 2 \\ 2x_2 - 6 = 0 \Rightarrow x_2 = 3 \end{cases}$ This solution violates the first constraint $x_1 + x_2 \leq 3$.

KKT

- $\mu_1 = 0$ and $\mu_2 = 0$
  $\begin{cases} 2x_1 - 4 = 0 \Rightarrow x_1 = 2 \\ 2x_2 - 6 = 0 \Rightarrow x_2 = 3 \end{cases}$ This solution violates the first constraint $x_1 + x_2 \leq 3$.

- $\mu_1 \neq 0$ and $\mu_2 = 0$
  $\begin{cases} 2x_1 + \mu_1 = 4 \\ 2x_2 + \mu_1 = 6 \end{cases}$
  Due to the complementary slackness condition $\mu_1 \neq 0 \Rightarrow x_1 + x_2 = 3$
  Solution of the system: $x_1 = 1, x_2 = 2, \mu_1 = 2$. It satisfies the remaining Kuhn Tucker conditions.

KKT

- $\mu_1 = 0$ and $\mu_2 \neq 0$
  The resulting system is:
  $$\begin{cases} 2x_1 - 2\mu_2 = 4 \\ 2x_2 + \mu_2 = 6 \\ -2x_1 + x_2 = 2 \end{cases}$$
  Solution: $x_1 = 4/5, x_2 = 18/5 \mu_2 = -6/5$ Violates $\mu_2 \geq 0$ and $x_1 + x_2 \leq 3$

KKT

- $\mu_1 = 0$ and $\mu_2 \neq 0$
  The resulting system is:
  $$\begin{cases} 2x_1 - 2\mu_2 = 4 \\ 2x_2 + \mu_2 = 6 \\ -2x_1 + x_2 = 2 \end{cases}$$
  Solution: $x_1 = 4/5, x_2 = 18/5 \mu_2 = -6/5$ Violates $\mu_2 \geq 0$ and $x_1 + x_2 \leq 3$

- $\mu_1 \neq 0$ and $\mu_2 \neq 0$
  Due to the complementary slackness condition
  $$\begin{cases} x_1 + x_2 = 3 \\ -2x_1 + x_2 = 2 \end{cases}$$
  With solution $x_1 = 1/3 x_2 = 8/3$. Substituting in the equations $\mathcal{L}(x, \lambda, \mu) = 0$, gives $u_2 = 89 \leq 0$

The inspection of the graph of the feasible solutions illustrates that $x_1 = 1, x_2 = 2, u_1 = 2, u_2 = 0$ is indeed the optimal solution .

Is KKT a necessary condition?

*Usually it is necessary that the feasible set of the original problem satisfies some regularity assumption (***Constraint qualification***) in order to derive optimality conditions.*

- KKT conditions can be useful in assessing potential local minima
-

Active set

The active set is the equality constraints indices, together with the indices of the inequality constraints for which $g_i(x) = 0$, that is $\mathcal{A}(x) = \mathcal{E} \cap \{i \in \mathcal{I} | g_i(x) = 0\}$

In the KKT conditions, if $g_j(x^*) < 0$ (an inactive constraint) at x^* then $\mu_j = 0$

Thus

$$\sum_{j \in \mathcal{I}} \mu_j^* \nabla g_j(x^*) = \sum_{j \in \mathcal{I} \cap \mathcal{A}} \mu_j^* \nabla g_j(x^*)$$

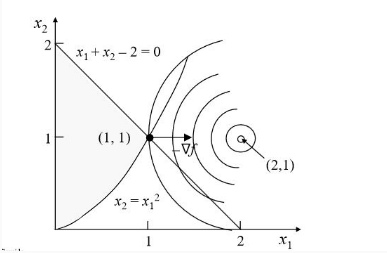
Geometric interpretation

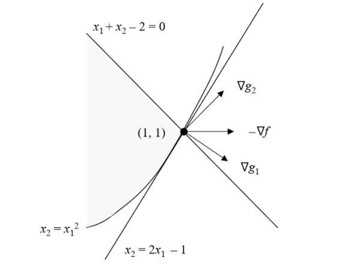
$\min_{x \in \mathbb{R}^2} (x_1 - 2)^2 + (x_2 - 1)^2$

$x_1^2 - x_2 \leq 0$

$x_1 + x_2 - 2 \leq 0.$

Analyse at $\begin{bmatrix} 1 & 1 \end{bmatrix}^t$





Geometric interpretation

$$\begin{array}{rl} & \text{minimize } f(x) \\ \text{s.t} & f_1(x) \leq 0 \\ & f_2(x) \leq 0 \\ & \vdots \\ & f_m(x) \leq 0 \end{array}$$

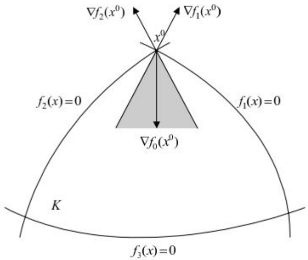


Fig. 2.1. Kuhn–Tucker conditions.

Optimality conditions

Example 1

$$\min_{x\in\mathbb{X}} (x_1 - 2)^2 + \left(x_2 - 0.5\left(3 - \sqrt{5}\right)\right)^2$$

$$\mathbb{X} = \left\{x \in \mathbb{R}^2 | x_1 + x_2 \leq 1 \quad x_1^2 \leq x_2\right\}$$

Consider the optimal solution $x^* = \frac{1}{2}\left[\begin{array}{cc} -1+\sqrt{5} & 3-\sqrt{5} \end{array}\right]^t$

- Which are the active constraints at x*?
- Write the gradients of $f(x)$ and $g_i(x)$ and obtain $\nabla f(x^*)$ and $\nabla g_i(x^*)$
- Can $\nabla f(x^*)$ be written as a linear combination of $\nabla g_i(x^*)$?

Example 1

$$\nabla f(x^*) = \left[\begin{array}{c} -5+\sqrt{5} \\ 0 \end{array} \right]$$

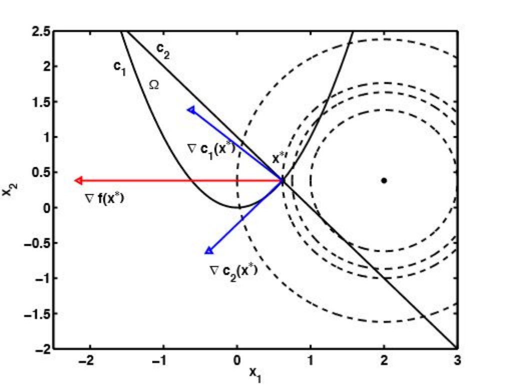
$$\nabla g_1(x^*) = \left[\begin{array}{c} 1-\sqrt{5} \\ 1 \end{array} \right] \quad \nabla g_2(x^*) = \left[\begin{array}{c} -1 \\ -1 \end{array} \right]$$

$$g_1(x^*) = g_2(x^*) = 0$$

$\nabla f(x^*) = \mu_1^* \mu g_1(x^*) + \mu_2^* \mu g_2(x^*)$ with $\mu_1^* = \mu_2^* = \sqrt{5} - 1 > 0$

$\boxed{\text{It is possible to find a KKT point}}$

Optimality conditions



Example 2 - The same problem

$$\min_{x \in \mathbb{X}} (x_1 - 2)^2 + \left(x_2 - 0.5\left(3 - \sqrt{5}\right)\right)^2$$

$$\mathbb{X} = \left\{x \in \mathbb{R}^2 | -x_1 - x_2 + 1 \geq 0 \quad x_2 - x_1^2 \geq 0\right\}$$

Let $x^* = \begin{bmatrix} 0 & 0 \end{bmatrix}^t$

$$\nabla f(x^*) = \begin{bmatrix} -4 \\ \sqrt{5} - 3 \end{bmatrix}$$

$$\nabla g_1(x^*) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

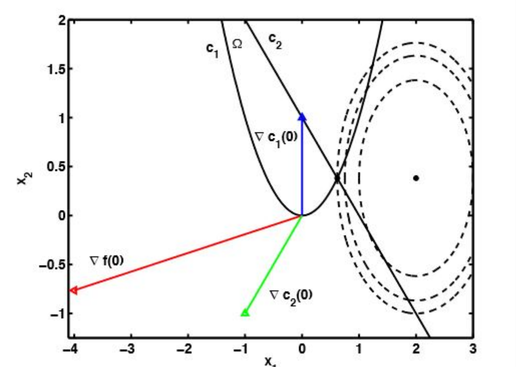
Example 2

$$g_1(x^*) = 0 \text{ active}$$

$$g_2(x^*) = 1 \;\; \text{inactive} \Rightarrow \lambda_2 = 0$$

It is not possible to write

$$\nabla f(x^*) = \lambda_1^* \nabla g_1(x^*) \quad \text{with} \quad \lambda_1^* \geq 0$$



Optimality conditions - Constraint qualification

- We would like KKT to be a necessary condition for a given optimal solution. In this case, we could verify if a candidate to optimal can be discarded.

- We would like KKT to be a necessary condition for a given optimal solution. In this case, we could verify if a candidate to optimal can be discarded. (x^* optimal $\Rightarrow$ KKT is satisfied. Then, If x^* does not satisfy KKT it cannot be optimal)

- We would like KKT to be a necessary condition for a given optimal solution. In this case, we could verify if a candidate to optimal can be discarded. (x^* optimal $\Rightarrow$ KKT is satisfied. Then, If x^* does not satisfy KKT it cannot be optimal)
- When the problem is unconstrained , the KKT conditions reduce to $\nabla f(x^*) = 0$ which is a necessary optimality condition.
- Usually it is necessary that the feasible set of the original problem satisfies some regularity assumption (**Constraint qualfication**) in order to derive optimality conditions.

- We would like KKT to be a necessary condition for a given optimal solution. In this case, we could verify if a candidate to optimal can be discarded. (x^* optimal $\Rightarrow$ KKT is satisfied. Then, If x^* does not satisfy KKT it cannot be optimal)
- When the problem is unconstrained , the KKT conditions reduce to $\nabla f(x^*) = 0$ which is a necessary optimality condition.
- Usually it is necessary that the feasible set of the original problem satisfies some regularity assumption (**Constraint qualfication**) in order to derive optimality conditions.

Main idea

Theorem (First order necessary conditions)

Under suitable constraint qualfications,

$$x^* \text{ is a local minimizer} \Rightarrow x^* \text{ KKT point .}$$

Linear independence constraint qualification

PrGen satisfies linear independence constraint qualification at a given $x \Leftrightarrow$ $\{\nabla g_i(x), i \in \mathcal{A}(x)\}$ is linearly independent.

Exercice (entrega aula) Give an example where this constraint qualification holds and show that the optimal solution satisfies KKT (may be the example below if you explain what is going on. See Nocedal ex 12.3)

Exercice (entrega aula) Give an example where this constraint qualification does not hold.

Optimality conditions - Constraint qualification

Linear independence constraint qualification

PrGen satisfies linear independence constraint qualification at a given $x \Leftrightarrow$ $\{\nabla g_i(x), i \in \mathcal{A}(x)\}$ is linearly independent.

Exercice (entrega aula) Give an example where this constraint qualification holds and show that the optimal solution satisfies KKT (may be the example below if you explain what is going on. See Nocedal ex 12.3)

Exercice (entrega aula) Give an example where this constraint qualification does not hold.

Slater constraint qualification

PrGen satisfies Slater constraint qualification $x \Leftrightarrow$ if there exists some feasible solution x for which all inequality constraints are strictly satisfied (i.e. $\exists x|$ $g_i(x) < 0 \forall i \in \mathcal{I},\ h_j(x) = 0 \forall j \in \mathcal{E}$)

Both fail for $\min_{x \in \mathbb{X}} x_1 + x_2$ $\mathbb{X} = \left\{ x| \ \ x_1^2 + (x_2 - 1)^2 \leq 1 \quad x_2 \geq 0 \ \right\}$

Consider $x^* = \left[\begin{array}{cc} 0 & 0 \end{array} \right]^t$

Some specific problems are easier

Linear Equality-constrained problem

Theorem First order necessary condition

Let $f : \mathbb{R}^n \to \mathbb{R}$,be partially differentiable with continuous partial derivatives, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ if $x^* \in \mathbb{R}^n$ is a local minimizer of

$$\min_{x\in\mathbb{R}^n} \{f(x)|Ax = b\}$$

then

$$\exists \lambda^* \in \mathbb{R}^m \text{ such that } \begin{cases} \nabla f(x^*) + A^t \lambda^* = 0 \\ Ax^* = b \end{cases}$$

$\boxed{\text{These are the KKT}}$

$$\text{Observe that } \begin{cases} \nabla_x \mathcal{L}(x,\lambda) = \nabla f(x^*) + A^t \lambda^* \\ \nabla_\lambda \mathcal{L}(x,\lambda) = Ax^* - b \end{cases}$$

Sufficient conditions

Some specific problems are easier

Linear Equality-constrained problem

Theorem Sufficient condition

Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is partially differentiable with continuous partial derivatives, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ Let $x^* \in \mathbb{R}^n$ and $\lambda^* \in \mathbb{R}^m$ satisfy:

$$\begin{array}{ll} \nabla f(x^*) + A^t \lambda^* & = 0 \\ Ax^* & = b \\ \text{f is } \textbf{convex} \text{ on} & \{x \in \mathbb{R}^n | Ax = b\} \end{array}$$

Then x^* is a **global** minimizer of

$$\min_{x \in \mathbb{R}^n} \{f(x) | Ax = b\}$$

Example

Consider $\min_{x \in \mathbb{R}^n} \{f(x) | Ax = b\}$ with

$f(x) = (x_1 - 1)^2 + (x_2 - 3)^2$,

$A = \begin{bmatrix} 1 & 1 \end{bmatrix}$, $b = \begin{bmatrix} 0 \end{bmatrix}$

a) Verify if $x^* = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$, and $\lambda^* = \begin{bmatrix} -2 \end{bmatrix}$ Satisfy KKT conditions

b) Is x^* is optimal?

Some specific problems are easier

Linear Equality-constrained problem

Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is twice partially differentiable with continuous second partial derivatives, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ Let $x^* \in \mathbb{R}^n$ and $\lambda^* \in \mathbb{R}^m$ satisfy:

$$\left\{ \begin{array}{lcl} \nabla f(x^*) + A^t \lambda^* & = & 0 \\ Ax^* & = & b \\ (A\Delta x = 0 \ \text{ and } \ \Delta x \neq 0) & \Rightarrow & (\Delta x^t \nabla^2 f(x^*) \Delta x > 0) \end{array} \right.$$

not necessarily positive definite in $\mathbb{R}^n$

Then x^* is a local minimizer of

$$\min_{x \in \mathbb{R}^n} \{f(x) | Ax = b\}$$

Some specific problems are easier

Convex programming

Convex programming problems

An optimization problem $\min_{x \in \mathbb{X}} f(x)$ is called a **convex programming problem** if $f(x)$ is a convex function and $\mathbb{X}$ is a convex set.

Theorem Sufficient optimality conditions for convex problems

Let PrGen be a convex programming problem. Then

x^* is a KKT point of PrGen $\Rightarrow$ x^* is a (global) minimizer of PrGen

IMPORTANT

- When the problem is not convex, the KKT conditions are not, in general, sufficient for optimality
- One needs positive definiteness of the Hessian of the Lagragian function along *feasible* directions.

Idea for algorithms

i Minimize the Lagrangian over x for a fixed λ then adjust λ

ii Find critical points of $\mathcal{L}$ and try to assure optimality

$$\boxed{\text{First order necessary conditions are not sufficient!}}$$

Example

$\min_{x \in \mathbb{R}^n} \{f(x) | Ax = b\}$ with $f(x) = -\frac{1}{2}x_1^2 - \frac{1}{2}x_2^2$, $A = \begin{bmatrix} 1 & -1 \end{bmatrix}$, $b = \begin{bmatrix} 0 \end{bmatrix}$

- Find the necessary conditions
- Analyze the point $\hat{x} = 0$, $\hat{\lambda} = 0$