# Karush–Kuhn–Tucker conditions

## Mathematical Optimization

Georgiy Kormakov

CMC MSU

22 March 2024

# Change of variables

$$\min_x \sum_{i=1}^{N} \|A_i x + b_i\|_2 + \frac{1}{2}\|x - x_0\|_2^2.$$

# Change of variables

$$\min_x \sum_{i=1}^{N} \|A_i x + b_i\|_2 + \frac{1}{2}\|x - x_0\|_2^2.$$

$$\min_x \sum_{i=1}^{N} \|y_i\|_2 + \frac{1}{2}\|x - x_0\|_2^2$$

$$\text{s.t. } A_j x + b_j = y_j \ \forall j$$

# Change of variables

$$\min_x \sum_{i=1}^{N} \|A_i x + b_i\|_2 + \frac{1}{2}\|x - x_0\|_2^2.$$

$$\min_x \sum_{i=1}^{N} \|y_i\|_2 + \frac{1}{2}\|x - x_0\|_2^2$$
$$\text{s.t. } A_j x + b_j = y_j \ \forall j$$

$$g(\nu_1, \dots, \nu_N) = \begin{cases} \sum_{i=1}^{N} (A_i x_0 + b_i)^T \nu_j - \frac{1}{2}\|\sum_{j=1}^{N} A_j^T \nu_i\|_2^2, & \|\nu_j\|_2 \leqslant 1 \ \forall i \\ -\infty, & \text{else} \end{cases}$$

# Standart form

$$
\begin{aligned}
\min_x \;& f_0(x) \\
\text{s.t. } & f_i(x) \leqslant 0,\ i = 1, \ldots, m \\
& h_j(x) = 0,\ j = 1, \ldots, n,
\end{aligned} \tag{1}
$$

$x \in \mathbb{R}^d$, but the whole domain is $\mathrm{D} = \cap_{i=0}^m \textbf{dom } f_i \cap_{j=1}^n \textbf{dom } h_j$.

## Feasible set

$$F = \{x \in D | f_i(x) \leqslant 0, i = \overline{1, m},\ h_j(x) = 0, \forall j\}$$

## Lagrangian

For the optimisation problem 1, the *Lagrangian* is equal to

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{j=1}^n \nu_j h_j(x).$$

# Lagrangian Dual Function

## Definition

Let's define the *Lagrangian dual function* (or just the dual function) $g : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$ as follows:

$$g(\lambda, \nu) = \inf_{x \in D} \left( f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{n} \nu_j h_j(x) \right) = \inf_{x \in D} L(x, \lambda, \nu) \quad (2)$$

## Property 1

$g(\lambda, \nu)$ is always concave function of $(\lambda, \nu)$

# Lagrangian Dual Function. Properties

## Property 2

$g(\lambda, \nu)$ is smaller then optimum $p^*$ of (1) for $\lambda \succeq 0$ and any $\nu$:

$$\forall \lambda \succeq 0, \nu \in \mathbb{R}^n \quad g(\lambda, \nu) \leqslant p^*$$

# Lagrangian Dual Function. Properties

## Property 2

$g(\lambda, \nu)$ is smaller then optimum $p^*$ of (1) for $\lambda \succeq 0$ and any $\nu$:

$$\forall \lambda \succeq 0, \nu \in \mathbb{R}^n \quad g(\lambda, \nu) \leqslant p^*$$

## Duality

With such conditions, always holds a **weak duality** $\max_{\lambda,\nu} g(\lambda, \nu) = d^* \leqslant p^*$. But we want somehow the **strong** one: $d^* = p^*$

# Complementary slackness

Consider, that we have strong duality!

## Complementary slackness

Consider, that we have strong duality!

$$f_0(x^*) \quad = g(\lambda^*, \nu^*) =$$

# Complementary slackness

Consider, that we have strong duality!

$$
\begin{aligned}
f_0(x^*) &= g(\lambda^*, \nu^*) = \inf_x \left( f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{j=1}^{n} \nu_j^* h_j(x) \right) \leqslant \\
&\leqslant f_0(x^*) + \sum_{i=1}^{m} \underbrace{\lambda_i^*}_{\geqslant 0} \underbrace{f_i(x^*)}_{\leqslant 0} + \sum_{j=1}^{n} \nu_j^* \underbrace{h_j(x^*)}_{=0} \leqslant f_0(x^*),
\end{aligned}
$$

# Complementary slackness

Consider, that we have strong duality!

$$
\begin{aligned}
f_0(x^*) &= g(\lambda^*, \nu^*) = \inf_x \left( f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + \sum_{j=1}^n \nu_j^* h_j(x) \right) \leqslant \\
&\leqslant f_0(x^*) + \sum_{i=1}^m \underbrace{\lambda_i^*}_{\geqslant 0} \underbrace{f_i(x^*)}_{\leqslant 0} + \sum_{j=1}^n \nu_j^* \underbrace{h_j(x^*)}_{=0} \leqslant f_0(x^*),
\end{aligned}
$$

## Complementary slackness

The *complementary slackness* property is

$$
\lambda_i^* f_i(x^*) = 0, \;\; i = 1, \ldots, m. \Leftrightarrow \left[ \begin{array}{ll} \lambda_i^* > 0 & \Rightarrow f_i(x^*) = 0, \\ f_i(x^*) < 0 & \Rightarrow \lambda_i^* = 0. \end{array} \right.
$$

# Collect them ALL!

1) Logic:

$$\boxed{x^* \in F;} \qquad \text{(Primal Feasibility)}$$

## Collect them ALL!

1) Logic:

$$\boxed{x^* \in F;} \qquad \text{(Primal Feasibility)}$$

2) Desire of the strong duality:

**Definition of the set of active constraints.**

$$\text{Active}(x) = \{i \in \{1, \ldots, m\} | f_i(x) = 0\} \cup \{1, \ldots, n\}$$

Consider $(\lambda^*, \nu^*)$ is the solution for the dual problem.

$$\boxed{\lambda_i^* \geqslant 0, \; i = 1, \ldots, m;} \qquad \text{(Dual Feasibility)}$$

$$\boxed{\lambda_i^* f_i(x^*) = 0, \; i = 1, \ldots, m.} \qquad \text{(Complementary Slackness)}$$

# Always have been...

3) Optimal condition...

# Always have been...

3) Optimal condition...

**Stationarity of Lagrange function**

Consider $f_i, h_j \in C^1$, $i = \overline{0, m}$, $j = \overline{1, p}$ and $x^*$ is the local optimum, then

$$\boxed{\nabla_x L(x^*, \lambda^*, \nu^*) = 0}$$

Problems?

# Always have been...

3) Optimal condition...

**Stationarity of Lagrange function**

Consider $f_i, h_j \in C^1$, $i = \overline{0, m}$, $j = \overline{1, p}$ and $x^*$ is the local optimum, then

$$\boxed{\nabla_x L(x^*, \lambda^*, \nu^*) = 0}$$

Problems?

It's only necessary condition!

Local property.

# Karush-Kuhn-Tucker necessary conditions

## Theorem

Consider $f_i, h_j \in C^1$, $i = \overline{0, m}$, $j = \overline{1, p}$ and $x^*$ is the local optimum for (1). With the satisfied regularity conditions for optimisation problems for $\{f_i(x^*), h_j(x^*) | \quad i, j \in \text{Active}(x^*)\}$,
exists a pair $(\lambda^*, \nu^*)$, for which the following conditions (KKT) hold.

$$\boxed{\nabla_x L(x^*, \lambda^*, \nu^*) = 0} \qquad \text{(Stationarity of Lagrange function)}$$

$$\boxed{x^* \in F} \qquad \text{(Primal Feasibility)}$$

$$\boxed{\lambda_i^* \geqslant 0,\ i = 1, \ldots, m} \qquad \text{(Dual Feasibility)}$$

$$\boxed{\lambda_i^* f_i(x^*) = 0,\ i = 1, \ldots, m} \qquad \text{(Complementary Slackness)}$$

# Karush-Kuhn-Tucker necessary conditions

## Theorem

Consider $f_i, h_j \in C^1$, $i = \overline{0, m}$, $j = \overline{1, p}$ and $x^*$ is the local optimum for (1). With the satisfied regularity conditions for optimisation problems for $\{f_i(x^*), h_j(x^*) | \quad i, j \in \text{Active}(x^*)\}$,
exists a pair $(\lambda^*, \nu^*)$, for which the following conditions (KKT) hold.

$$\boxed{\nabla_x L(x^*, \lambda^*, \nu^*) = 0} \qquad \text{(Stationarity of Lagrange function)}$$

$$\boxed{x^* \in F} \qquad \text{(Primal Feasibility)}$$

$$\boxed{\lambda_i^* \geqslant 0,\ i = 1, \ldots, m} \qquad \text{(Dual Feasibility)}$$

$$\boxed{\lambda_i^* f_i(x^*) = 0,\ i = 1, \ldots, m} \qquad \text{(Complementary Slackness)}$$

Strong Duality? Nash?

# Tangent space

# Tangent space

$$S(x_0) = \{x(t) | x(t) \in C^1,\ x(0) = x_0,\\ x(t) \in F\ \forall t \in [0, t_{max}]\}$$

$x_0$ — local minimum, $\boxed{\nabla f_0(x_0)^T d \geqslant 0}$ is true for $d$ from the tangent space:
It exists if we can reparametrise our function

$$\frac{d}{dt} f_0(x(t)) = \sum_{i=1}^{n} \frac{\partial f_0}{\partial x_i} \frac{\partial x_i}{\partial t} = \nabla f_0(x(t))^T \frac{dx}{dt}$$

## Tangent space

$$T_F(x) = \{d \in \mathbb{R}^d | \exists x(t) \in S(x_0) :\ d = \frac{dx}{dt}\Big|_{t=0}\}$$

# Case 1

$$\begin{cases} f(x) \to \min\limits_x \\ h(x) = 0 \end{cases}$$

# Regularity conditions

| | | |
|---|---|---|
| Linearity constraint qualification | LCQ | If $f_i$ and $h_j$ are affine functions |
| Linear independence constraint qualification | LICQ | $\{\nabla f_i(x^*), \nabla h_j(x^*) \mid \quad i, j \in \text{Active}(x^*)\}$ are linearly independent |
| Slater's condition | SC | For a convex problem ($f_0$, $f_i$ —convex, $h_j$ — affine), $\exists \tilde{x} : h_j(\tilde{x}) = 0$ and $f_i(\tilde{x}) < 0$. |
| Weak Slater's condition | wSC | $\exists \tilde{x} : h_j(\tilde{x}) = 0$, $f_i(\tilde{x}) \leqslant 0$ for affine $f_i$<br>$f_i(\tilde{x}) < 0$ for other $f_i$ |

# Case 2

$$
\begin{cases}
f(x) \to \min\limits_{x} \\
f_1(x) \leqslant 0
\end{cases}
$$

# Case 3

$$\begin{cases} f(x) \to \min\limits_{x} \\ f_1(x) \leqslant 0 \\ f_2(x) \leqslant 0 \end{cases}$$

# General case

$$\begin{cases} f(x) \to \min\limits_x \\ f_i(x) \leqslant 0 \\ h_j(x) = 0 \end{cases}$$

$$T_F(x_0) = \{d | \nabla f_i(x_0)^T d \leqslant 0, \nabla h_j(x_0)^T d = 0, \; \forall i, j \in \text{Active}(x^*)\}$$

Then, Farkas' lemma for this cone

$$N(x_0) = \left\{ -\sum_{i=1}^{m} \lambda_i \nabla f_i(x_0) - \sum_{j=1}^{n} \nu_j \nabla h_j(x_0) \Big| \lambda_i \geqslant 0, \; \forall \nu_j \right\}$$

# Sufficient conditions

## Slater case

For Slater regularity conditions, the KKT conditions becomes criteria.

# Sufficient conditions

## Slater case

For Slater regularity conditions, the KKT conditions becomes criteria.

## Critical cone definition

$$C(x_0, \lambda_0) = \{d \in T_F(x_0) | \text{ if } \lambda_{0i} > 0 \text{ then} \nabla f_i(x_0)^T d = 0\}$$

## Second order conditions

With $f_i, h_j \in C^2$, $i = \overline{0, m}$, $j = \overline{1, p}$ and in $x_0$ holds regularity conditions, then $\exists (\lambda_0, \nu_0)$ : holds 4 conditions from KKT AND

$$\boxed{d^T \nabla_{xx} L(x_0, \lambda_0, \nu_0) d \geqslant 0, \ \forall d \in C(x_0, \lambda_0)}$$

# Sufficient conditions

## Slater case

For Slater regularity conditions, the KKT conditions becomes criteria.

## Critical cone definition

$$C(x_0, \lambda_0) = \{d \in T_F(x_0) | \text{ if } \lambda_{0i} > 0 \text{ then} \nabla f_i(x_0)^T d = 0\}$$

## Second order conditions

With $f_i, h_j \in C^2$, $i = \overline{0, m}$, $j = \overline{1, p}$ and in $x_0$ holds regularity conditions, then $\exists(\lambda_0, \nu_0)$ : holds 4 conditions from KKT AND

$$\boxed{d^T \nabla_{xx} L(x_0, \lambda_0, \nu_0) d \geqslant 0, \ \forall d \in C(x_0, \lambda_0)}$$

The last means, actually, this: $[\nabla_x f_i(x_0), \nabla_x h_j(x_0)]^T d = 0$

# The simplest one

$$\begin{cases} x^2 \to \min\limits_x \\ x > 0 \end{cases}$$

# A bit harder

$$\begin{cases} x^2 \to \min\limits_x \\ x \geqslant 0 \end{cases}$$

## Trickiest one

$$\begin{cases} x \to \min\limits_{x} \\ x^2 \leqslant 0 \end{cases}$$

# Hopefully normal

$$
\begin{aligned}
\min_x \ & \frac{1}{2}x^T P x + q^T x + r \\
\text{s.t.} \ & Ax = b,
\end{aligned}
$$

where $P \in \mathcal{S}^d_+$, $A \in \mathbb{R}^{m \times d}$.

# Entropy maximisation

$$
\begin{aligned}
\min_x \quad & \sum_{i=1}^{d} x_i \log x_i \\
\text{s.t.} \quad & Ax \preceq b, \\
& 1^T x = 1,
\end{aligned}
$$

with the domain $\mathbb{R}^d_{++}$

# Entropy maximisation

$$\begin{aligned} \min_x \; & \sum_{i=1}^d x_i \log x_i \\ \text{s.t. } & Ax \preceq b, \\ & 1^T x = 1, \end{aligned}$$

$$\begin{aligned} \max_{\lambda,\nu} \; & \left[ -b^T\lambda - \nu - e^{-\nu-1} \sum_{i=1}^d e^{-a_i^T\lambda} \right] \\ \text{s.t. } & \lambda \succeq 0, \end{aligned}$$

with the domain $\mathbb{R}^d_{++}$

# Separable functions

$$\min_x \sum_{i=1}^{d} f_i(x_i)$$
$$\text{s.t. } a^T x = b,$$

$f_i$ are strictly convex and differentiable functions.

# Separable functions

$$\min_x \sum_{i=1}^{d} f_i(x_i)$$
$$\text{s.t. } a^T x = b,$$

$f_i$ are strictly convex and differentiable functions.

$$L(x,\nu) = \sum_{i=1}^{d} f_i(x_i) + \nu(a^T x - b) = -b\nu + \sum_{i=1}^{d}(f_i(x_i) + \nu a_i x_i),$$

# Separable functions

$$\min_x \sum_{i=1}^{d} f_i(x_i)$$
$$\text{s.t. } a^T x = b,$$

$f_i$ are strictly convex and differentiable functions.

$$L(x,\nu) = \sum_{i=1}^{d} f_i(x_i) + \nu(a^T x - b) = -b\nu + \sum_{i=1}^{d} (f_i(x_i) + \nu a_i x_i),$$

$$g(\nu) = -b\nu + \inf_x \left( \sum_{i=1}^{d} (f_i(x_i) + \nu a_i x_i) \right) = -b\nu + \sum_{i=1}^{d} \inf_{x_i} (f_i(x_i) + \nu a_i x_i)$$
$$= -b\nu - \sum_{i=1}^{d} f_i^*(-\nu a_i),$$

# Separable functions. Final actions

Dual problem:

$$\max_{\nu} \left[ -b\nu - \sum_{i=1}^{d} f_i^*(-\nu a_i) \right],$$

with scalar $\nu$

# Piece-wise linear

$$\min_{x} \max_{i=1,\dots,m} (a_i^T x + b_i)$$

# Piece-wise linear

$$\min_{x} \max_{i=1,\dots,m} (a_i^T x + b_i)$$

$$\begin{aligned} \min_{x} \max_{i=1,\dots,m} & \; y_i \\ \text{s.t.} & \; a_j^T x + b_j = y_j \; \forall j. \end{aligned}$$

# Piece-wise linear

$$\min_x \max_{i=1,\dots,m} (a_i^T x + b_i)$$

$$\min_x \max_{i=1,\dots,m} y_i$$
$$\text{s.t. } a_j^T x + b_j = y_j \ \forall j.$$

$$\inf_y (\max_i y_i - \nu^T y) = \begin{cases} 0, \ \nu \succeq 0, \ 1^T \nu = 1 \\ -\infty, \quad \text{else} \end{cases}$$

# Piece-wise linear

$$\min_x \max_{i=1,\dots,m} (a_i^T x + b_i)$$

$$\min_x \max_{i=1,\dots,m} y_i$$
$$\text{s.t. } a_j^T x + b_j = y_j \ \forall j.$$

$$\inf_y (\max_i y_i - \nu^T y) = \begin{cases} 0, \ \nu \succeq 0, \ 1^T \nu = 1 \\ -\infty, \quad \text{else} \end{cases}$$

$$g(\nu) = \begin{cases} b^T \nu, \ \sum_{i=1}^m \nu_i a_i = 0, \ \nu \succeq 0, \ 1^T \nu = 1 \\ -\infty, \quad \text{else} \end{cases}$$