Karush-Kuhn-Tucker optimality conditions and Lagrangian duality

G. Mastroeni

Department of Computer Science, University of Pisa

Optimization Methods and Game Theory
Master of Science in Artificial Intelligence and Data Engineering
University of Pisa – A.Y. 2024/25

Contents of the lessons

- The Abadie constraints qualification (ACQ);
- Karush-Kuhn-Tucker optimality conditions;
- Lagrangian duality.

First-order optimality conditions for constrained optimization problems

Consider the constrained optimization problem

$$\begin{cases} \min \ f(x) \\ x \in X := \{x \in \mathbb{R}^n : g_j(x) \le 0, \quad j = 1, \dots, m, h_k(x) = 0, \ k = 1, \dots, p\} \end{cases} \qquad \text{(P)}$$

where f, g_j and h_k are continuously differentiable for any j, k.

Definition

- The *Tangent cone* at $x^* \in X$, is defined by

$$T_X(x^*) = \left\{ d \in \mathbb{R}^n : \exists \ \{z_k\} \subset X, \ \exists \ \{t_k\} > 0, \ z_k \to x^*, \ t_k \to 0, \ \lim_{k \to \infty} \frac{z_k - x^*}{t_k} = d \right\}$$

- $\mathcal{A}(x^*) = \{j : \ g_j(x^*) = 0\}$ is the set of inequality constraints which are active at $x^* \in X$.
- The set

$$D(x^*) = \left\{ d \in \mathbb{R}^n : \begin{array}{ll} d^\mathsf{T} \nabla g_j(x^*) \le 0 & \forall \ j \in \mathcal{A}(x^*), \\ d^\mathsf{T} \nabla h_k(x^*) = 0 & \forall \ k = 1, \dots, p \end{array} \right\}$$

is the *first-order feasible direction cone* at $x^* \in X$.

Definition – Abadie constraint qualification (ACQ)

We say that the Abadie constraint qualification (ACQ) holds at a point $x^* \in X$, if $T_X(x^*) = D(x^*)$.

Theorem 1 (Sufficient conditions for ACQ)

a) (*Affine constraints*)
If g_j and h_k are affine for all $j = 1, \dots, m$ and $k = 1, \dots, p$, then ACQ holds at any $x \in X$.

b) (*Slater condition for convex problems*)
If g_j are convex for all $j = 1, \dots, m$, h_k are affine for all $k = 1, \dots, p$ and there exists $\bar{x} \in X$ s.t. $g(\bar{x}) < 0$ and $h(\bar{x}) = 0$, then ACQ holds at any $x \in X$.

c) (*Linear independence of the gradients of active constraints*)
If $x^* \in X$ and the vectors

$$\begin{cases} \nabla g_j(x^*) & \text{for } j \in \mathcal{A}(x^*), \\ \nabla h_k(x^*) & \text{for } k = 1, \dots, p \end{cases}$$

are linearly independent, then ACQ holds at x^*.

Theorem 2 (Karush-Kuhn-Tucker)

If x^* is a local minimum and ACQ holds at x^*, then there exist $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ s.t. (x^*, λ^*, μ^*) satisfies the KKT system:

$$\begin{cases} \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^{p} \mu_j^* \nabla h_j(x^*) = 0 \\ \lambda_i^* g_i(x^*) = 0 \qquad \forall\, i = 1, \dots, m \\ \lambda^* \geq 0 \\ g(x^*) \leq 0 \\ h(x^*) = 0 \end{cases}$$

Remark

The previous theorem can be proved considering the first order optimality condition for (P) given by

$$\nabla f(x^*)^T d \geq 0, \quad \forall d \in T_X(x^*)$$

and taking into account the Abadie constraint qualification at x^*.

Define the Lagrangian function $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ by

$$L(x, \lambda, \mu) := f(x) + \sum_{i=1}^{m} \lambda_i \, g_i(x) + \sum_{j=1}^{p} \mu_j \, h_j(x)$$

Then the KKT system can be formulated as:

$$\begin{cases} \nabla_x L(x, \lambda, \mu) \; = \; 0 \\ \lambda_i g_i(x) \; = \; 0, \quad i = 1, .., m \\ \lambda \geq 0, \;\; h(x) = 0, \quad g(x) \leq 0 \end{cases} \tag{1}$$

Note that condition $\lambda_i g_i(x) \; = \; 0$, per $i = 1, .., m$, is equivalent to $\lambda^T g(x) = 0$ or also, $\langle \lambda, g(x) \rangle = 0$.

Example

Find the the minimum points of the function $f(x_1, x_2) = 2x_1 + x_2$ on the set $X = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 \leq 4\}$

Note that X is compact so that from Weierstrass Theorem it follows that f admits global maximum and global minimum on X.
The Lagrangian function is:

$$L(x_1, x_2, \lambda) = 2x_1 + x_2 + \lambda(x_1^2 + x_2^2 - 4)$$

The KKT system is given by:

$$\begin{cases} 2 + 2\lambda x_1 = 0 \\ 1 + 2\lambda x_2 = 0 \\ \lambda(x_1^2 + x_2^2 - 4) = 0 \\ x_1^2 + x_2^2 \leq 4 \\ \lambda \geq 0 \end{cases}$$

Note that for $\lambda = 0$ the system is impossible.

We are led to solve the system:

$$\begin{cases} x_1 = -\frac{1}{\lambda} \\ x_2 = -\frac{1}{2\lambda} \\ x_1^2 + x_2^2 = 4 \\ \lambda \geq 0 \end{cases}$$

Then:

$$\frac{1}{\lambda^2} + \frac{1}{4\lambda^2} = 4$$

from which

$$16\lambda^2 = 5, \quad \lambda = \pm\frac{\sqrt{5}}{4}.$$

- $\lambda = \dfrac{\sqrt{5}}{4} \quad \Rightarrow \quad x_1 = -\dfrac{4}{\sqrt{5}} = -\dfrac{4\sqrt{5}}{5}, \quad x_2 = -\dfrac{2}{\sqrt{5}} = -\dfrac{2\sqrt{5}}{5}$

It follows that $\bar{x} = \left(-\frac{4\sqrt{5}}{5}, -\frac{2\sqrt{5}}{5}\right)$ is a global minimum point.

Remark

Note that in the previous example, ACQ is fulfilled for every $x \in X$. Indeed, there is only the constraint $g(x) = x_1^2 + x_2^2 - 4 \leq 0$, with $\nabla g(x_1, x_2) \neq (0,0)$, for every $x \in X, s.t.\ x_1^2 + x_2^2 - 4 = 0$.

What about the maximum points of f on X ?

Notice that it is enough to set $\lambda \leq 0$ in the KKT system. In fact, in order to find the maxima of f we have to look for the minima of $-f$. The KKT system for the problem

$$\min(-f(x)) \quad x \in X$$

is:

$$\begin{cases} -2 + 2\lambda x_1 = 0 \\ -1 + 2\lambda x_2 = 0 \\ \lambda(x_1^2 + x_2^2 - 4) = 0 \\ x_1^2 + x_2^2 \leq 4 \\ \lambda \geq 0 \end{cases}$$

which is equivalent to set $\lambda \leq 0$ in the original one.
Choosing in the original system: $\lambda = -\frac{\sqrt{5}}{4}$ we obtain
$x_1 = \frac{4}{\sqrt{5}} = \frac{4\sqrt{5}}{5}, \quad x_2 = \frac{2}{\sqrt{5}} = \frac{2\sqrt{5}}{5}$
so that $\hat{x} = \left(\frac{4\sqrt{5}}{5}, \frac{2\sqrt{5}}{5}\right)$ is a global maximum point for f on the set X.

Remark

ACQ assumption is crucial in the KKT Theorem.

Example

$$\begin{cases} \min x_1 + x_2 \\ (x_1 - 1)^2 + (x_2 - 1)^2 - 1 \leq 0 \\ x_2 \leq 0 \end{cases}$$

$x^* = (1, 0)$ is the global optimum.

$T_X(x^*) = \{0\}$, $D(x^*) = \{d \in \mathbb{R}^2 : \ d_2 = 0\}$, hence ACQ does not hold at x^*.

$\nabla g_1(x^*) = (0, -2)$, $\nabla g_2(x^*) = (0, 1)$, $\nabla f(x^*) = (1, 1)$, hence there is no λ^* s.t. (x^*, λ^*) solves KKT system.

KKT Theorem gives necessary optimality conditions, but not sufficient ones.

Example

$$\begin{cases} \min x_1 + x_2 \\ -x_1^2 - x_2^2 + 2 \leq 0 \end{cases}$$

$x^* = (1, 1)$, $\lambda^* = \dfrac{1}{2}$ solves KKT system, but x^* is not a local optimum.

Theorem 3

If the optimization problem is convex and (x^*, λ^*, μ^*) solves KKT system, then x^* is a global optimum.

Recall that (P) is convex if f and g are convex and h is affine.

Exercise 4.1. Prove the previous theorem.

Lagrangian relaxation and dual function

Denote by $v(P)$ denotes the optimal value of (P).

Definition

Given $\lambda \geq 0$ and $\mu \in \mathbb{R}^p$, the problem

$$\begin{cases} \inf \ L(x, \lambda, \mu) \\ x \in \mathbb{R}^n \end{cases}$$

is called Lagrangian relaxation of (P) and
$\varphi(\lambda, \mu) = \inf_{x \in \mathbb{R}^n} L(x, \lambda, \mu)$ is the Lagrangian dual function.

Lagrangian relaxation provides a lower bound to $v(P)$.

Theorem 4

For any $\lambda \geq 0$ and $\mu \in \mathbb{R}^p$, we have $\varphi(\lambda, \mu) \leq v(P)$.

Proof. If $x \in X$, i.e. $g(x) \leq 0, h(x) = 0$, then

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i \, g_i(x) \leq f(x),$$

hence

$$\varphi(\lambda, \mu) = \inf_{x \in \mathbb{R}^n} \; L(x, \lambda, \mu) \leq \inf_{x \in X} \; L(x, \lambda, \mu) \leq \inf_{x \in X} \; f(x) = v(P).$$

□

Properties of the dual function

The dual function φ

- is concave because inf of affine functions w.r.t (λ, μ)
- may be equal to $-\infty$ at some point
- may be not differentiable at some point

Definition

The problem

$$\begin{cases} \max \ \varphi(\lambda, \mu) \\ \lambda \geq 0 \end{cases} \tag{D}$$

is called Lagrangian dual problem of (P) [and (P) is called primal problem].

- The dual problem (D) consists in finding the best lower bound of $v(P)$.
- (D) is always equivalent to a convex problem, even if (P) is a non-convex problem, indeed, it is a maximization of a concave function on a convex set.

Theorem 4, can be equivalently stated as:

Theorem 4 (weak duality)

For any optimization problem (P), we have $v(D) \leq v(P)$.
The previous inequality is called "weak duality property".

Example - Linear Programming.

Primal problem:

$$\begin{cases} \min \ c^{\mathsf{T}} x \\ Ax \geq b \end{cases} \tag{P}$$

Lagrangian function: $L(x, \lambda) = c^{\mathsf{T}} x + \lambda^{\mathsf{T}}(b - Ax) = \lambda^{\mathsf{T}} b + (c^{\mathsf{T}} - \lambda^{\mathsf{T}} A)x$
Dual function:

$$\varphi(\lambda) = \inf_{x \in \mathbb{R}^n} L(x, \lambda) = \begin{cases} -\infty & \text{if } c^{\mathsf{T}} - \lambda^{\mathsf{T}} A \neq 0 \\ \lambda^{\mathsf{T}} b & \text{if } c^{\mathsf{T}} - \lambda^{\mathsf{T}} A = 0 \end{cases}$$

Dual problem:

$$\begin{cases} \max \ \varphi(\lambda) \\ \lambda \geq 0 \end{cases} \longrightarrow \begin{cases} \max \ \lambda^{\mathsf{T}} b \\ \lambda^{\mathsf{T}} A = c^{\mathsf{T}} \\ \lambda \geq 0 \end{cases} \tag{D}$$

is a linear programming problem.

Exercise 4.2. Find the dual of (D).

Example - Least norm solution of linear equations

Primal problem:

$$\begin{cases} \min \ \dfrac{1}{2}x^{\mathsf{T}}x \\ Ax = b \end{cases} \tag{P}$$

Lagrangian function: $L(x,\mu) = \frac{1}{2}x^{\mathsf{T}}x + \mu^{\mathsf{T}}(b - Ax)$.
Dual function: $\varphi(\mu) = \inf_{x\in\mathbb{R}^n} L(x,\mu)$.
$L(x,\mu)$ is quadratic and strongly convex with respect to x, therefore

$$\varphi(\mu) = \inf_{x\in\mathbb{R}^n} L(x,\mu) = \min_{x\in\mathbb{R}^n} L(x,\mu),$$

thus the global optimum is the stationary point:

$$\nabla_x L = x - A^{\mathsf{T}}\mu = 0 \iff x = A^{\mathsf{T}}\mu,$$

hence $\varphi(\mu) = -\frac{1}{2}\mu^{\mathsf{T}}AA^{\mathsf{T}}\mu + b^{\mathsf{T}}\mu$.
Dual problem:

$$\begin{cases} \max \ -\dfrac{1}{2}\mu^{\mathsf{T}}AA^{\mathsf{T}}\mu + b^{\mathsf{T}}\mu \\ \mu \in \mathbb{R}^p \end{cases} \tag{D}$$

is an unconstrained convex quadratic programming problem.

Exercise 4.3. Find the dual problem of a general convex quadratic programming problem

$$\begin{cases} \min\ \frac{1}{2}x^{\mathsf{T}}Qx + c^{\mathsf{T}}x \\ Ax \leq b \end{cases} \tag{P}$$

where Q is a symmetric positive definite matrix.

Definition (Strong duality)

We say that strong duality holds for (P) if $v(D) = v(P)$ and (D) admits an optimal solution.

Strong duality does not hold in general.

Example. Consider the following (non-convex) problem:

$$\begin{cases} \min \ -x^2 \\ x - 1 \leq 0 \\ -x \leq 0 \end{cases} \tag{P}$$

It is easy to check that $v(P) = -1$.
The Lagrangian function is $L(x, \lambda) = -x^2 + \lambda_1(x-1) - \lambda_2 x$, hence

$$\varphi(\lambda) = \inf_{x \in \mathbb{R}} L(x, \lambda) = -\infty \qquad \forall\ (\lambda_1, \lambda_2) \in \mathbb{R}^2,$$

hence $v(D) = -\infty$.

Strong duality

Next theorem provides sufficient conditions which guarantee strong duality for (P).

Theorem 5

Suppose f, g, h are continuously differentiable, the primal problem

$$\begin{cases} \min \ f(x) \\ g(x) \leq 0 \\ h(x) = 0 \end{cases} \tag{P}$$

is convex, there exists a global optimum x^* and ACQ holds at x^*. Then:

- Strong duality holds ($v(D) = v(P)$ and (D) admits an optimal solution);
- (λ^*, μ^*) is optimal for (D) if and only if (λ^*, μ^*) is a KKT multipliers vector associated with x^*.

Proof. $L(x, \lambda, \mu)$ is convex with respect to x since (P) is convex.
Let (λ^*, μ^*) be any KKT vector of multipliers associated with x^*. Then,

$$\nabla_x L(x^*, \lambda^*, \mu^*) = 0 \qquad \lambda^* \geq 0, \qquad (\lambda^*)^{\mathsf{T}} g(x^*) = 0.$$

Thus,

$$\begin{aligned} v(D) \geq \varphi(\lambda^*, \mu^*) = \inf_x L(x, \lambda^*, \mu^*) &\underset{[L \text{ convex}]}{=} L(x^*, \lambda^*, \mu^*) \\ = f(x^*) + (\lambda^*)^{\mathsf{T}} g(x^*) + (\mu^*)^{\mathsf{T}} h(x^*) = f(x^*) = v(P) &\underset{[\text{weak duality}]}{\geq} v(D). \end{aligned}$$

Therefore, $v(P) = v(D)$ and (λ^*, μ^*) is optimal for (D).
Viceversa, if (λ^*, μ^*) is any optimal solution for (D), then

$$\begin{aligned} f(x^*) = v(P) = v(D) = \varphi(\lambda^*, \mu^*) = \inf_x L(x, \lambda^*, \mu^*) \leq L(x^*, \lambda^*, \mu^*) \\ = f(x^*) + (\lambda^*)^{\mathsf{T}} g(x^*) + (\mu^*)^{\mathsf{T}} h(x^*) = f(x^*) + (\lambda^*)^{\mathsf{T}} g(x^*) \leq f(x^*), \end{aligned}$$

thus $(\lambda^*)^{\mathsf{T}} g(x^*) = 0$ and $\inf_x L(x, \lambda^*, \mu^*) = L(x^*, \lambda^*, \mu^*)$, hence $\nabla_x L(x^*, \lambda^*, \mu^*) = 0$, i.e., (λ^*, μ^*) is a KKT multipliers vector associated with x^*. $\square$

Strong duality may hold also for some non-convex problems.

Example

Consider the (non-convex) problem

$$\begin{cases} \min & -x_1^2 - x_2^2 \\ x_1^2 + x_2^2 - 1 \leq 0 \end{cases}$$

We have $v(P) = -1$. The Lagrangian function is

$$L(x, \lambda) = -x_1^2 - x_2^2 + \lambda(x_1^2 + x_2^2 - 1) = (\lambda - 1)x_1^2 + (\lambda - 1)x_2^2 - \lambda.$$

The dual function is

$$\varphi(\lambda) = \inf_{x \in \mathbb{R}} L(x, \lambda) = \begin{cases} -\infty & \text{if } \lambda < 1 \\ -\lambda & \text{if } \lambda \geq 1 \end{cases}$$

The dual problem is

$$\begin{cases} \max & -\lambda \\ \lambda \geq 1 \end{cases}$$

hence its optimal solution is $\lambda^* = 1$ and $v(D) = -1$.

Theorem 6 (characterization of strong duality)

(x^*, λ^*, μ^*) is a saddle point of L, i.e.

$$L(x^*, \lambda, \mu) \leq L(x^*, \lambda^*, \mu^*) \leq L(x, \lambda^*, \mu^*) \qquad \forall\, x \in \mathbb{R}^n,\ \lambda \in \mathbb{R}^m_+,\ \mu \in \mathbb{R}^p,$$

if and only if x^* is optimum of (P), (λ^*, μ^*) is optimum of (D) and $v(P) = v(D)$.

Proof. If (x^*, λ^*, μ^*) is a saddle point of L, then we can prove that $x^* \in X$, $\langle \lambda^*, g(x^*) \rangle = 0$ which implies $\varphi(\lambda^*, \mu^*) = f(x^*)$.
Viceversa, we have that

$$f(x^*) = \varphi(\lambda^*, \mu^*) = \inf_{x \in \mathbb{R}^n} L(x, \lambda^*, \mu^*) \leq L(x^*, \lambda^*, \mu^*) = f(x^*) + \langle \lambda^*, g(x^*) \rangle,$$

hence $\langle \lambda^*, g(x^*) \rangle = 0$ and $L(x^*, \lambda^*, \mu^*) = f(x^*) = \varphi(\lambda^*, \mu^*) \leq L(x, \lambda^*, \mu^*)$ for all $x \in \mathbb{R}^n$. Moreover

$$L(x^*, \lambda, \mu) = f(x^*) + \langle \lambda, g(x^*) \rangle \leq f(x^*) = L(x^*, \lambda^*, \mu^*) \qquad \forall\, \lambda \in \mathbb{R}^m_+,\ \mu \in \mathbb{R}^p.$$

$\square$