MATHEMATICS OF DATA SCIENCE

October 31, 2025

Today: Duality Pt 2 (KKT conditions)

November 11: Correction KKT conditions (valid under constraint qualifications)

Previously: . For any optimization problem (primal problem), there exists a convex optimization problem called the dual

. When strong duality holds, both problems are equivalent

Today: what kind of optimality conditions can we get from the dual?

Recall: For $\underset{x \in \mathbb{R}^n}{\text{minimize}}\ f(x)$ s.t. $x \in F$ convex set

$f : \mathbb{R}^n \to \mathbb{R}$ C^1 convex

$x^* \in \operatorname{argmin}\{f(x) \text{ s.t. } x \in F\}$ $\iff$ $x^* \in F$ and $\nabla f(x^*)^T (z - x^*) \geq 0 \quad \forall z \in F$

1 Constraint qualifications and derivatives

Our problem today:

(P) $\underset{x \in \mathbb{R}^n}{\text{minimize}}\ f(x)$ s.t. $g(x) \leq 0_{\mathbb{R}^m}$ $\iff$ $g_i(x) \leq 0$ $i = 1..m$

$h(x) = 0_{\mathbb{R}^l}$ $\iff$ $h_i(x) = 0$ $i = 1..l$

Assumption: $f, g_1, \ldots, g_m, h_1, \ldots, h_l$ are C^1 on $\mathbb{R}^n$

($\Rightarrow$ $\operatorname{dom}(f) = \operatorname{dom}(g_i) = \operatorname{dom}(h_i) = \mathbb{R}^n$, problem domain is $\mathbb{R}^n$)

$\hookrightarrow$ To fully exploit dual information, we need strong duality, but optimality conditions only require (some form of) constraint qualification

Four examples of constraint qualifications (CQ)

1. g_i, h_i linear $\Rightarrow$ linear CQ holds

2. h_i linear, g_i convex and $\exists\, x \in \mathrm{ri}\left(\{x \mid g(x) \le 0,\ h(x) = 0\}\right)$ such that $h(x) = Ax - b = 0$

$h(x) = Ax - b$

$g_i(x) < 0 \quad \forall i = 1..m$

↑ x is "strictly feasible"

$\Rightarrow$ slater CQ holds

⊕ Guarantees strong duality for convex problems!

NB: 1 and 2 do not involve derivatives of g_i and h_i

3. LICQ (Linear Independence Constraint Qualification)

LICQ holds at a feasible point $\bar{x}$ if

$$\{\nabla g_i(\bar{x})\}_{i \in A(\bar{x})} \cup \{\nabla h_i(\bar{x})\} \text{ are linearly}$$

↑ Gradients of active inequality constraints

↑ Gradients of all equality constraints

independent, where

$$A(\bar{x}) = \{ i \mid g_i(\bar{x}) = 0 \} \quad (\text{Active set at } \bar{x})$$

($v_1, \ldots, v_q$ linearly independent if $\sum_{i=1}^{q} \alpha_i v_i = 0 \Rightarrow \alpha_i = 0 \ \forall i$)

Remark: $i \notin A(\bar{x}) \Rightarrow g_i(\bar{x}) < 0$

$\bar{x}$ is strictly feasible w.r.t. constraint i

4. MFCQ (Mangasarian – Fromovitz CQ)

MFCQ holds at $\bar{x}$ feasible if

i) $\{\nabla h_i(\bar{x})\}$ are linearly independent

ii) $\exists\, d \in \mathbb{R}^n$ such that

$$\nabla h_i(\bar{x})^T d = 0 \quad i = 1..l$$

$$\nabla g_i(\bar{x})^T d < 0 \quad i \in A(\bar{x})$$

$\{\nabla g_i(\bar{x})\}$ need not be linearly independent

One can show that LICQ $\Rightarrow$ MFCQ

The converse is not true

$x_1 \leq 0 \qquad x_1 - 1 \leq 0$

Toy problem ($\underset{x \in \mathbb{R}^2}{\text{minimize}} \; \|x\|^2$ s.t. $x_1 \leq 0,\; x_1 \leq 0,\; x_2 \leq 0$

$g_1(x) \leq 0 \qquad g_2(x) \leq 0 \qquad g_3(x) \leq 0$

At $\bar{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, LICQ does not hold (because of the "duplicated" constraint)

$$A(\bar{x}) = \{1, 2, 3\}$$

$$\{\nabla g_i(\bar{x})\} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$$

MFCQ holds with $d = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$ $\qquad \{\nabla g_i(\bar{x})^T d\} = \{-1, -1, -1\}$

NB: Relationships between CQ are not trivial.

$\Rightarrow$ Need one CQ to use duality!

② KKT conditions

Recall:

Lagrangian of (P) $\quad \mathcal{L}(x, \lambda, \mu) = f(x) + \lambda^T g(x) + \mu^T h(x)$

Dual problem of (P)

(D) $\underset{(\lambda,\mu)}{\text{maximize}} \quad \inf_x \mathcal{L}(x,\lambda,\mu) \quad$ s.t. $\lambda \geq 0$

((P) $\iff \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \sup_{\lambda,\mu} \{\mathcal{L}(x,\lambda,\mu)\}$ s.t. $g(x) \leq 0$, $h(x) = 0$)

Under strong duality, if x^* solves (P), then (λ^*,μ^*) solve (D)

$\forall (x,\lambda,\mu)$,

$$\underbrace{\mathcal{L}(x^*,\lambda,\mu) \leq \mathcal{L}(x^*,\lambda^*,\mu^*)}_{\text{maximize } \mathcal{L} \text{ w.r.t. } (\lambda,\mu)} \leq \mathcal{L}(x,\lambda^*,\mu^*)$$

(the underline of $\mathcal{L}(x^*,\lambda^*,\mu^*) \leq \mathcal{L}(x,\lambda^*,\mu^*)$: minimize $\mathcal{L}$ w.r.t. x)

Theorem (Karush-Kuhn-Tucker conditions)

Suppose that constraint qualification holds between (P) and (D) and that f, g_i, h_i are C^1 on $\mathbb{R}^n$.

Then, if (x^*,λ^*,μ^*) is a primal-dual solution (i.e. x^* solves (P) and (λ^*,μ^*) solves (D)), $\boxed{\text{we have}}$

KKT conditions

$$\begin{cases} (i) & \nabla_x \mathcal{L}(x^*,\lambda^*,\mu^*) = 0 \\ (ii) & \nabla_\mu \mathcal{L}(x^*,\lambda^*,\mu^*) = h(x^*) = 0 \\ (iii) & \nabla_\lambda \mathcal{L}(x^*,\lambda^*,\mu^*) = g(x^*) \leq 0 \\ (iv) & \lambda^* \geq 0 \\ (v) & \lambda_i^* g_i(x^*) = 0 \quad \forall i = 1..m \end{cases}$$

(i): Optimality w.r.t. x

(ii), (iii): Primal feasibility (constraints of (P))

(iv): Constraint of (D)

(v): Complementarity condition

Moreover, if (P) is convex (f, g_i convex + h_i linear), then

(x^*, λ^*, μ^*) primal-dual solution $\iff$ (x^*, λ^*, μ^*) satisfies (KKT)

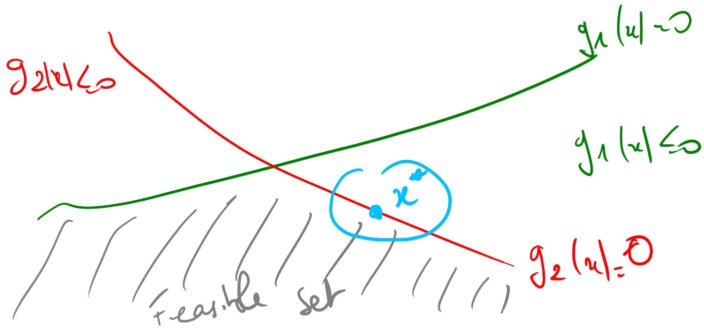
↳ (iii) – (v) show the difficulty of dealing with inequality constraints

g_i is "like an equality constraint" at the solution

$\left[\forall i \in \mathcal{A}(x^*),\ g_i(x^*) = 0 \Rightarrow \text{(v) holds}\right.$

$\left[\forall i \notin \mathcal{A}(x^*),\ g_i(x^*) < 0 \Rightarrow \text{we must have } \lambda_i^* = 0 \text{ for (v) to hold}\right.$

g_i does not matter at the solution



(v) (Complementarity condition)

characterizes which inequality constraints matter at the solution

<u>Remark:</u> • Primal version of (KKT)

If x^* solution of (P) (aka primal solution)

Then $\exists (\lambda^*, \mu^*)$ solution of (D) such that

(x^*, λ^*, μ^*) satisfies (KKT)

<u>Example:</u> Linear programming

$$\underset{x \in \mathbb{R}^n}{\text{minimize}}\ c^T x \quad \text{s.t.}\ Ax = b,\ x \geq 0$$

$A \in \mathbb{R}^{m \times n}$

($\text{rank}(A) = m$ for simplicity)

Dual

$$\underset{\substack{\mu \in \mathbb{R}^m \\ \lambda \in \mathbb{R}^n}}{\text{maximize}} \quad b^T \mu \qquad \text{s.t.} \quad A^T\mu - \lambda + c = 0 \,, \quad \lambda \geq 0$$

KKT for LP

$$\mathcal{L}(x, \lambda, \mu) = c^T x - \lambda^T x + \mu^T (Ax - b)$$

(i) $A^T\mu^* - \lambda^* + c = 0$

(ii) $Ax^* = b$

(iii) $x^* \geq 0$

(iv) $\lambda^* \geq 0$

(v) $\lambda_i^* x_i^* = 0$

(ii), (iii): Primal feasibility

(i), (iv): Dual feasibility

(v): Complementarity condition that connects the primal to the dual variables

→ Many classes of algorithms satisfy 4 of the 5 conditions for (LP) compute iterates that satisfy 4 of the 5 conditions and work towards satisfying the fifth one.