22 kkt; duality

Wednesday, November 25, 2020 5:28 PM

Prop. 14.1 Let U be any nonempty subset of a normed space V.

(1) For any $u \in U$, the cone $C(u)$ of feasible directions at u is closed.

(2) Let $J : \Omega \to \mathbb{R}$ be a function defined on an open subset Ω containing U. If J has a local minimum w.r.t. U at $u \in U$, and if J_u' exists at u, then

$$J_u'(v-u) \geq 0 \quad \forall\, v \in u + C(u).$$

proof. (1) Let $(w_n)_{n\geq 0}$ be a sequence $w_n \in C(u)$ converging to $w \in V$.
May assume $w \neq 0$, since $0 \in C(u)$ by definition,
so WLOG may also assume $w_n \neq 0$.

Then $\forall n \geq 0$, $\exists (u_k^n)_{k\geq 0}$ in V and some $w_n \neq 0$ s.t.

(1) $u_k^n \in U$ and $u_k^n \neq u$ $\forall k \geq 0$, and $\lim_{k\to\infty} u_k^n = u$.

(2) $\exists (\delta_k^n)_{k\geq 0}$ in V s.t.

$$u_k^n = u + \|u_k^n - u\| \frac{w_n}{\|w_n\|} + \|u_k^n - u\| \delta_k^n, \quad \lim_{k\to\infty} \delta_k^n = 0, \quad w_n \neq 0.$$

Let $(\varepsilon_n)_{n\geq 0}$ be a sequence of real numbers $\varepsilon_n > 0$ s.t. $\lim_{n\to\infty} \varepsilon_n = 0$ (e.g. $\frac{1}{n+1}$)

Then for every fixed n, $\exists\, k(n) \in \mathbb{Z}$ s.t.

$$\|u_{k(n)}^n - u\| \leq \varepsilon_n, \quad \|\delta_{k(n)}^n\| \leq \varepsilon_n.$$

Then
$$u_{k(n)}^n = u + \|u_{k(n)}^n - u\| \frac{w}{\|w\|} + \|u_{k(n)}^n - u\| \left[\delta_{k(n)}^n + \left(\frac{w_n}{\|w_n\|} - \frac{w}{\|w\|}\right)\right]$$

Since $\lim_{n\to\infty} \frac{w_n}{\|w_n\|} = \frac{w}{\|w\|}$, $\lim_{n\to\infty} \frac{u_{k(n)}^n - u}{\|u_{k(n)}^n - u\|} = \frac{w}{\|w\|}$, so $w \in C(u)$.

(2) Let $w = v - u$ be any nonzero vector in the cone $C(u)$, and let $(u_k)_{k\geq 0}$ be the seq in $U - \{u\}$ s.t.

(1) $\lim_{k\to\infty} u_k = u$

(2) There is a sequence $(\delta_k)_{k\geq 0}$ of vectors $\delta_k \in V$ s.t.

$$u_k - u = \|u_k - u\| \frac{w}{\|w\|} + \|u_k - u\| \delta_k \quad , \quad \lim_{k\to\infty} \delta_k = 0 \quad , \quad w \neq 0.$$

(3) $J(u) \leq J(u_k)$ for all $k \geq 0$.

Since J is differentiable at u,

$$0 \leq J(u_k) - J(u) = J'_u(u_k - u) + \|u_k - u\| \varepsilon_k.$$

for some sequence $(\varepsilon_k)_{k\geq 0}$ s.t. $\lim_{k\to\infty} \varepsilon_k = 0$.

Since J'_u is linear and continuous,

$$0 \leq J'_u(u_k - u) + \|u_k - u\| \varepsilon_k$$

$$= \frac{\|u_k - u\|}{\|w\|} \left[J'_u(w) + \eta_k \right], \quad \text{where} \quad \eta_k = \|w\| \left(J'_u(\delta_k) + \varepsilon_k \right)$$

Since J'_u is continuous, $\lim_{k\to\infty} \eta_k = 0$.

$$\Rightarrow J'_u(w) \geq 0.$$

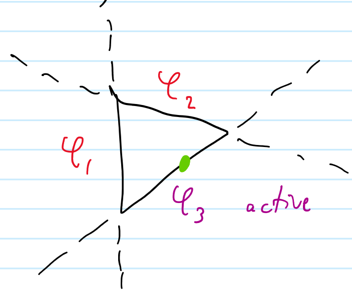
When is the cone $C(u)$ convex?

Def. 14.3 Given m functions $\varphi_i : \Omega \to \mathbb{R}$ defined on an open subset Ω of a vector space V, let U be defined by

$$U = \{ x \in \Omega \mid \varphi_i(x) \leq 0, \quad 1 \leq i \leq m \}$$

For any $u \in U$, a constraint φ_i is said to be active at u if $\varphi_i(u) = 0$, else inactive at u if $\varphi_i(u) < 0$.

e.g.



Def. 14.4 Given $U = \{ x \in \Omega \mid \varphi_i(x) \leq 0, \ 1 \leq i \leq n \}$, define

$$I(u) = \{ i \in \{1, \dots, m\} \mid \varphi_i(u) = 0 \}$$

as the set of indices where the constraints are active.

Define $C^*(u) = \{ v \in V \mid (\varphi_i')_u (v) \leq 0, \ i \in I(u) \}$

Note: Each $(\varphi_i')_u$ is a linear form, so $C^*(u)$ is an intersection of half-spaces passing through the origin, so it is both convex and a cone.

half-spaces passing through the origin, so it is both convex and a polyhedral cone.
If $I(u) = \emptyset$, then $C^*(u) = V$.

There is an entire theory of polyhedra which we do not cover here.

Def. 14.5 For any $u \in U$ as defined above, if φ_i are differentiable at u for all $i \in I(u)$, we say that the constraints are *qualified* at u if

(a) Either φ_i is affine $\forall i \in I(u)$, or

(b) $\exists w \in V$, $w \neq 0$ s.t. $\forall i \in I(u)$

(i) $(\varphi_i')_u(w) \leq 0$

(ii) If φ_i is not affine, then $(\varphi_i')_u(w) < 0$.

implies that u is not a crit. pt. for all φ_i's, so no singularity.

Intuitively, constraints are qualified if the boundary of U near u behaves nicely.

Prop. 14.2 Let $u \in U = \{x \in \Omega \mid \varphi_i(x) \leq 0,\ 1 \leq i \leq m\}$, where Ω is an open subset of a vector space V, and assume that φ_i are differentiable at u (for $i \in I(u)$). Then

(1) The cone $C(u)$ of feasible directions at u is contained in the convex cone $C^*(u)$.

(2) If the constraints are qualified at u (and φ_i are continuous at u for $i \notin I(u)$), then $C(u) = C^*(u)$.

proof. (1) For every $i \in I(u)$, since $\varphi_i(v) \leq 0$ $\forall v \in U$ and $\varphi_i(u) = 0$, the function $-\varphi_i$ has a local minimum at u w.r.t. U, so by Prop. 14.1(2), we have

$$(-\varphi_i')_u(v) \geq 0 \quad \forall v \in C(u),$$

so $(\varphi_i')_u(v) \leq 0$ $\forall v \in C(u)$ and for all $i \in I(u)$.

Thus, $u \in C^*(u) \Rightarrow C(u) \subseteq C^*(u)$.

(2) We will prove this only in the affine case; see the book for the general case.

If φ_i is affine, then $\varphi_i(v) = h_i(v) + c_i$, where h_i is a linear form, $c_i \in \mathbb{R}$.

Then $(\varphi_i')_u(v) = h_i(v)$ $\forall v \in V$.

Pick $w \in C^*(u)$, $w \neq 0$, so $(\varphi_i')_u(w) \leq 0 \quad \forall i \in I(u)$.

For any sequence $(\varepsilon_k)_{k\geq 0}$ of reals $\varepsilon_k > 0$ s.t. $\lim_{k\to\infty} \varepsilon_k = 0$,
let $(u_k)_{k\geq 0}$ be the seq given by $u_k = u + \varepsilon_k w$.

Then $u_k - u = \varepsilon_k w \neq 0$ and $\lim_{k\to\infty} u_k = u$.

By continuity of the φ_i's,

$$0 > \varphi_i(u) = \lim_{k\to\infty} \varphi_i(u_k).$$

Since φ_i is affine and $\varphi_i(u) = 0 \ \forall i \in I$, $\varphi_i(u) = h_i(u) + c_i = 0$, so

$$\varphi_i(u_k) = h_i(u_k) + c_i = h_i(u_k) - h_i(u) = h_i(u_k - u) = (\varphi_i')_u(u_k - u) = \varepsilon_k (\varphi_i')_u(w) \leq 0$$

$\Rightarrow u_k \in U$ for all k large enough so that the inactive constraints are met.

Then, since $\dfrac{u_k - u}{\|u_k - u\|} = \dfrac{w}{\|w\|}$ for all $k \geq 0$, $w \in C(u)$. $\Rightarrow C^*(u) \subseteq C(u)$.

The nonaffine case is a bit more work. See book. $\square$

Karush-Kuhn-Tucker conditions

Several lectures ago, we proved the Farkas-Minkowski lemma saying that to check if a point is in a polyhedral cone, we just need to check for separating hyperplanes.

Thm 14.1/14.5 Let $\varphi_i : \Omega \to \mathbb{R}$ be m constraints defined on some open subset Ω of a real Hilbert space V. Let $J : \Omega \to \mathbb{R}$, and let

$$U = \{ x \in \Omega \mid \varphi_i(x) \leq 0, \ 1 \leq i \leq m \}.$$

For any $u \in U$, let

$$I(u) = \{ i \in \{1, \dots, m\} \mid \varphi_i(u) = 0 \}$$

and assume that φ_i are differentiable at u $\forall i \in I(u)$
and continuous at u $\forall i \notin I(u)$.

If J is differentiable at u, has a local minimum at u w.r.t. U, and if the constraints are qualified at u, then there exist scalars $\lambda_i(u) \in \mathbb{R}$ $\forall i \in I(u)$, s.t.

$$J_u' + \sum_{i \in I(u)} \lambda_i(u)(\varphi_i')_u = 0, \quad \text{and} \quad \lambda_i(u) \geq 0 \quad \forall\, i \in I(u).$$

The above conditions are called the Karush-Kuhn-Tucker (KKT) optimality conditions. (compare to Thm 4.1 on Lagrange multipliers)

Equivalently, $\nabla J_u + \sum_{i \in I(u)} \lambda_i(u) \nabla(\varphi_i)_u = 0$ and $\lambda_i(u) \geq 0 \quad \forall\, i \in I(u)$.

proof. By Prop. 14.1(2),

$$J_u'(w) \geq 0 \quad \forall\, w \in C(u)$$

(i.e. in every feasible direction, J_u' is pos.)

By Prop. 14.2(2), $C(u) = C^*(u)$, where

$$C^*(u) = \{ v \in V \mid (\varphi_i')_u(v) \leq 0, \; i \in I(u) \}$$

$\Rightarrow \forall\, w \in V$, if $w \in C^*(u)$, then $J_u'(w) \geq 0$.

Or, if $-(\varphi_i')_u(w) \geq 0 \;\; \forall\, i \in I(u)$, then $J_u'(w) \geq 0$.

By using the Riesz representation thm, $J_u'(w) = \langle w, \nabla J_u \rangle$

and $(\varphi_i')_u(w) = \langle w, \nabla(\varphi_i)_u \rangle$.

Thus, if $\langle w, -\nabla(\varphi_i)_u \rangle \geq 0 \;\; \forall\, i \in I(u)$, then $\langle w, \nabla J_u \rangle \geq 0$.

By Farkas-Minkowski, $\exists$ scalars $\lambda_i(u)$ $\forall\, i \in I(u)$ s.t. $\lambda_i(u) \geq 0$ and

$$\nabla J_u = \sum_{i \in I(u)} \lambda_i(u)(-\nabla(\varphi_i)_u)$$

(i.e. ∇J_u is in the polyhedral cone of $\{-\nabla(\varphi_i)_u\}_{i \in I(u)}$)

$$\Rightarrow \nabla J_u + \sum_{i \in I(u)} \lambda_i(u) \nabla(\varphi_i)_u = 0$$

$$\Rightarrow J_u' + \sum_{i \in I(u)} \lambda_i(u)(\varphi_i')_u = 0.$$

$\blacksquare$

Sometimes will see KKT conditions written without the index set notation, i.e.

$$J_u' + \sum_{i=1}^{m} \lambda_i(u)(\varphi_i')_u = 0, \quad \underbrace{\sum_{i=1}^{m} \lambda_i(u)\varphi_i(u) = 0}, \quad \lambda_i(u) \geq 0, \quad i = 1, \dots, m$$

i.e. if φ_i is active, then λ_i can be pos.
otherwise, λ_i has to be 0.

Note: Sometimes this notation of KKT are referred to as complementary slackness conditions.

The scalars $\lambda_i(u)$ are often called generalized Lagrange multipliers.

When the constraints are convex, under certain conditions, KKT is also sufficient.

Thm 14.2/14.6 Let $\varphi_i : \Omega \to \mathbb{R}$ be convex, $\Omega \subseteq V$ an open subset of a real Hilbert space. Let $J : \Omega \to \mathbb{R}$ and $U = \{ x \in \Omega \mid \varphi_i(x) \leq 0,\ 1 \leq i \leq m \}$ and let $u \in U$ be s.t. φ_i and J are differentiable at u. Then

(1) If J has a local minimum at u w.r.t. U, and if the constraints are qualified, then $\exists\, \lambda_i(u) \in \mathbb{R}$ s.t. the KKT conditions hold: $J'_u + \sum_{i=1}^m \lambda_i(u)(\varphi_i')_u = 0$ and $\sum_{i=1}^m \lambda_i(u)\varphi_i(u) = 0$, $\lambda_i(u) \geq 0$, $i = 1, \dots, m$.

(2) Conversely, if the restriction of J to U is convex, and if there exist scalars $(\lambda_1, \dots, \lambda_m) \in \mathbb{R}^m_+$ s.t. the KKT conditions hold, then J has a global minimum at u w.r.t. U.

Lagrangian duality

Consider the Minimization Problem (P): (Call it the Primal)

$$\text{minimize } J(v)$$
$$\text{subject to } \varphi_i(v) \leq 0, \quad i = 1, \dots, m.$$

Def. 14.8 The Lagrangian of the Minimization Problem (P) is the function

$$L(v, \mu) = J(v) + \sum_{i=1}^m \mu_i \varphi_i(v),$$

with $\mu = (\mu_1, \dots, \mu_m)$, where μ_i are called the generalized Lagrange multipliers.

$M = \mathbb{R}^m_+$ for us.

Def. 14.7 Let $L : \Omega \times M \to \mathbb{R}$, where Ω and M are open subsets of normed vector spaces. A pt. $(u, \lambda) \in \Omega \times M$ is a saddle pt of L if u is a minimum of the function $v \mapsto L(v, \lambda)$ given fixed λ and λ is a maximum of the function $\mu \mapsto L(u, \mu)$ given fixed μ.

Equivalently, $\sup_{\mu \in M} L(u, \mu) = L(u, \lambda) = \inf_{v \in \Omega} L(v, \lambda)$.

Prop. 14.11/14.13 If (μ, λ) is a saddle pt of a function $L: \Omega \times M \to \mathbb{R}$, then $\sup_{\mu \in M} \inf_{v \in \Omega} L(v,\mu) = L(u,\lambda) = \inf_{v\in\Omega} \sup_{\mu \in M} L(v,\mu)$.

Intuition: At saddle pts, you can swap optimizing over either Ω or M.

Thm 14.4/14.14 Consider Problem (P) defined above where $J: \Omega \to \mathbb{R}$ and the constraints $\varphi_i : \Omega \to \mathbb{R}$, where $\Omega \subseteq V$ open subset of a Hilbert space. Then

(1) If $(u,\lambda) \in \Omega \times \mathbb{R}^m_+$ is a saddle pt of the Lagrangian L, then $u \in U$ is a solution of (P) and $J(u) = L(u,\lambda)$

(2) If Ω is open, if φ_i and J are convex and differentiable at $u \in U$, if the constraints are qualified, and if $u \in U$ is a minimum of (P), then $\exists \lambda \in \mathbb{R}^m_+$ s.t. (u,λ) is a saddle pt of L.

proof. (1) Since (u,λ) is a saddle pt, $\sup_{u \in \mathbb{R}^m_+} L(u,\mu) = L(u,\lambda)$

$$\Rightarrow L(u,\mu) \le L(u,\lambda) \quad \text{for all } \mu \in \mathbb{R}^m_+$$

$$\Rightarrow J(u) + \sum_{i=1}^m \mu_i \varphi_i(u) \le J(u) + \sum_{i=1}^m \lambda_i \varphi_i(u)$$

$$\Rightarrow \sum_{i=1}^m (\mu_i - \lambda_i)\varphi_i(u) \le 0 \qquad \forall \mu \in \mathbb{R}^m_+.$$

Thus, $\varphi_i(u) \le 0$ because otherwise could just choose μ_i large

$\Rightarrow u \in U$.

Letting $\mu = 0$, we also get $\sum_{i=1}^m \lambda_i \varphi_i(u) \ge 0$.

But $\lambda_i \ge 0$, so $\sum_{i=1}^m \lambda_i \varphi_i(u) = 0$.

$$\Rightarrow J(u) = L(u,\lambda).$$

We also know that $L(u,\lambda) \le L(v,\lambda)$ because it's a saddle pt.

$$\Rightarrow J(u) \le J(v) + \sum_{i=1}^m \lambda_i \varphi_i(v) \le J(v).$$

$\Rightarrow u$ is a minimum of J on U.

(2) Because we satisfy the conditions of Thm 14.2/6 (1), if $u \in U$ is a soln to (P), then $\exists \lambda \in \mathbb{R}^m_+$ such that the KKT conditions hold:

$$J'(u) + \sum_{i=1}^m \lambda_i (\varphi_i')_u = 0 \quad \text{and} \quad \sum_{i=1}^m \lambda_i \varphi_i(u) = 0.$$

Then $L(u,\mu) = J(u) + \sum_{i=1}^m \mu_i \varphi_i(u) \le J(u) = J(u) + \sum_{i=1}^m \lambda_i \varphi_i(u) = L(u,\lambda)$.

Then, because the function $v \mapsto J(v) + \sum_{i=1}^m \lambda_i \varphi_i(v) = L(v,\lambda)$ is convex,

by Thm 4.5(4)/4.11(4), $J'(u) + \sum_{i=1}^m \lambda_i (\varphi_i')_u = 0$ says that the derivative is 0, so it is sufficient to guarantee the existence of a minimum.

$\Rightarrow L(u,\lambda) \le L(v,\lambda)$ for all $v \in \Omega$.

$\Rightarrow (u,\lambda)$ is a saddle point of L. $\boxed{/\!/}$

i.e. If we know the second argument λ of a saddle point (u,λ), then we can replace the constrained Problem (P) with the unconstrained problem (P_λ): find $u_\lambda \in \Omega$ s.t.

$$L(u_\lambda,\lambda) = \inf_{v\in\Omega} L(v,\lambda).$$

How do we find such a $\lambda \in \mathbb{R}^m_+$?

Note that $L(u_\lambda,\lambda) = \inf_{v\in\Omega} L(v,\lambda) = \sup_{\mu\in\mathbb{R}^m_+} \inf_{v\in\Omega} L(v,\mu)$.

So let's introduce $G: \mathbb{R}^m_+ \to \mathbb{R}$ given by $G(\mu) = \inf_{v\in\Omega} L(v,\mu)$.

Then λ will be the solution to

find $\lambda \in \mathbb{R}^m_+$ s.t.

$$G(\lambda) = \sup_{\mu\in\mathbb{R}^m_+} G(\mu)$$

Dual problem

Which is equivalent to the **Maximation Problem** (D)

a convex optimization problem
$$\left\{ \begin{array}{l} \text{maximize } G(\mu) \leftarrow \text{(Lagrange) dual function, is concave} \\ \text{subject to } \mu \in \mathbb{R}^m_+ . \end{array}\right.$$

convex constraint

a convex optimization problem: subject to $\mu \in \mathbb{R}^m_+$. ← convex constraint

Note: $G(\mu) \le L(u,\mu) \le J(u)$ for all $u \in U$ and $\mu \in \mathbb{R}^m_+$.

If (P) has a minimum p^* at optimum u^*
(D) has a maximum d^* at optimum λ^*,
then $d^* \le p^*$ and $G(\lambda^*) \le J(u^*)$.

If $p^* = -\infty$, then (P) is unbounded below, so (D) is infeasible.
$d^* = +\infty$, then (D) is unbounded above, so (P) is infeasible.

Def. §14.10 $p^* - d^* \ge 0$ is called the *optimal duality gap*.
If $p^* = d^*$, then *strong duality holds*.

Thm 14.5/14.16 Consider the Minimization Problem (P)

$$\text{minimize } J(v)$$
$$\text{subject to } \varphi_i(v) \le 0, \quad i = 1, \dots, m.$$

where J and φ_i are defined on $\Omega \subseteq V$, an open subset of a Hilbert space.

(1) Suppose φ_i are continuous, and that $\forall \mu \in \mathbb{R}^m_+$, the problem (P_μ):

$$\text{minimize } L(v,\mu)$$
$$\text{subject to } v \in \Omega$$

has a unique solution u_μ, so that

$$L(u_\mu, \mu) = \inf_{v \in \Omega} L(v,\mu) = G(\mu)$$

and the function $\mu \mapsto u_\mu$ is continuous. Then the function G is differentiable for all $\mu \in \mathbb{R}^m_+$, and

$$G'_\mu(\xi) = \sum_{i=1}^m \xi_i \varphi_i(u_\mu) \quad \text{for all } \xi \in \mathbb{R}^m.$$

If λ is any solution of the Problem (D):

$$\text{maximize } G(\mu)$$
$$\text{subject to } \mu \in \mathbb{R}^m_+,$$

then the solution u_λ of the corresponding problem (P_λ) is a solution to the Problem (P).

(2) Assume Problem (P) has some solution $u \in U$, and that Ω is convex, φ_i and J are convex and differentiable at u, and that

(2) Assume Problem (P) has some solution $u \in U$, and that Ω is convex, φ_i and J are convex and differentiable at u, and that the constraints are qualified. Then the Problem (D) has a solution