DATA8004: Optimization for Statistical Learning

Subject Lecturer: Man-Chung YUE

Lecture 5
Convex Optimization
Fenchel duality

Convex optimization problems

Minimize a convex function, with constraints defined by convex sets.

Example:

$$\begin{aligned} \underset{x \in \mathbb{R}^2}{\text{Minimize}} \quad & x_1^2 + x_2^2 \\ \text{Subject to} \quad & x_1 + x_2 \leq 1. \end{aligned}$$

Convex optimization problems

Minimize a convex function, with constraints defined by convex sets.

Example:

$$\begin{array}{rl} \underset{x \in \mathbb{R}^2}{\text{Minimize}} & x_1^2 + x_2^2 \\ \text{Subject to} & x_1 + x_2 \leq 1. \end{array}$$

Recall that if we let $C := \{x \in \mathbb{R}^2 : \ x_1 + x_1 \leq 1\}$, then the above problem becomes

$$\underset{x \in \mathbb{R}^2}{\text{Minimize}} \quad x_1^2 + x_2^2 + \delta_C(x).$$

Convex optimization problems

Minimize a convex function, with constraints defined by convex sets.

Example:

$$\begin{array}{ll} \underset{x\in\mathbb{R}^2}{\text{Minimize}} & x_1^2 + x_2^2 \\ \text{Subject to} & x_1 + x_2 \le 1. \end{array}$$

Recall that if we let $C := \{x \in \mathbb{R}^2 : \ x_1 + x_1 \le 1\}$, then the above problem becomes

$$\underset{x\in\mathbb{R}^2}{\text{Minimize}} \ \ x_1^2 + x_2^2 + \delta_C(x).$$

Aim: Develop duality theory.

Convex optimization problems

Minimize a convex function, with constraints defined by convex sets.

Example:

$$\begin{array}{rl} \underset{x\in\mathbb{R}^2}{\text{Minimize}} & x_1^2 + x_2^2 \\ \text{Subject to} & x_1 + x_2 \leq 1. \end{array}$$

Recall that if we let $C := \{x \in \mathbb{R}^2 : \ x_1 + x_1 \leq 1\}$, then the above problem becomes

$$\underset{x\in\mathbb{R}^2}{\text{Minimize}} \ \ x_1^2 + x_2^2 + \delta_C(x).$$

Aim: Develop duality theory.
Topics outline:

- Fenchel conjugate (Legendre transform) & Young's inequality.
- Inf-projection (Value functions).
- Fenchel duality theorem.

Fenchel conjugate

Definition: Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and convex. Its (Fenchel) conjugate (or Legendre transform) is the function

$$f^*(x) := \sup_{y \in \mathbb{R}^n} \{x^T y - f(y)\}.$$

Fenchel conjugate

Definition: Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and convex. Its (Fenchel) conjugate (or Legendre transform) is the function

$$f^*(x) := \sup_{y \in \mathbb{R}^n} \{x^T y - f(y)\}.$$

Example:

(i) Let $f(x) = \frac{x^2}{2}$ for $x \in \mathbb{R}$. Then $f^*(x) = \frac{x^2}{2}$.

(ii) Let $f(x) = |x|$ for $x \in \mathbb{R}$. Then

$$f^*(x) = \sup_{y \in \mathbb{R}} \{xy - |y|\} = \delta_{[-1,1]}(x).$$

Fenchel conjugate

Definition: Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and convex. Its (Fenchel) conjugate (or Legendre transform) is the function

$$f^*(x) := \sup_{y \in \mathbb{R}^n} \{x^T y - f(y)\}.$$

Example:

(i) Let $f(x) = \frac{x^2}{2}$ for $x \in \mathbb{R}$. Then $f^*(x) = \frac{x^2}{2}$.

(ii) Let $f(x) = |x|$ for $x \in \mathbb{R}$. Then

$$f^*(x) = \sup_{y \in \mathbb{R}} \{xy - |y|\} = \delta_{[-1,1]}(x).$$

(iii) Let $f(x) = \|x\|_1$ for $x \in \mathbb{R}^n$. Then $f(x) = \sum_{i=1}^n |x_i|$. Hence

$$f^*(x) = \sup_{y \in \mathbb{R}^n} \left\{ x^T y - \sum_{i=1}^n |y_i| \right\} = \sum_{i=1}^n \sup_{y_i \in \mathbb{R}} \{x_i^T y_i - |y_i|\} = \delta_{\|\cdot\|_\infty \leq 1}(x).$$

Fenchel conjugate cont.

Theorem 5.1 (Properties of Fenchel conjugate)
Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and convex. Then the following statements hold.

(i) The function f^* is proper, closed and convex.

(ii) If g is proper convex and $f \geq g$, then $f^* \leq g^*$.

(iii) It holds that $f = f^{**}$ if and only if f is closed.

Fenchel conjugate cont.

Theorem 5.1 (Properties of Fenchel conjugate)
Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and convex. Then the following statements hold.

(i) The function f^* is proper, closed and convex.

(ii) If g is proper convex and $f \geq g$, then $f^* \leq g^*$.

(iii) It holds that $f = f^{**}$ if and only if f is closed.

Proof: We first prove item (i).

Notice that $x \mapsto x^T y - f(y)$ is an affine function for each $y \in \operatorname{dom} f \neq \emptyset$. Since

$$f^*(x) = \sup_{y \in \mathbb{R}^n} \{x^T y - f(y)\} = \sup_{y \in \operatorname{dom} f} \{x^T y - f(y)\},$$

we conclude that f^* is closed and convex as the supremum of affine functions.

Fenchel conjugate cont.

Proof of Theorem 5.1 cont.:
Next, since $\mathrm{dom}\, f \neq \emptyset$, we have $\mathrm{ri}(\mathrm{dom}\, f) \neq \emptyset$. Let $\hat{x} \in \mathrm{ri}(\mathrm{dom}\, f)$ and we know from Theorem 4.11 that there exists $\hat{v} \in \partial f(\hat{x})$. Hence

$$f(y) - f(\hat{x}) \geq \hat{v}^T(y - \hat{x}) \quad \forall y \in \mathbb{R}^n.$$

Thus,

$$f^*(\hat{v}) = \sup_{y \in \mathbb{R}^n} \{\hat{v}^T y - f(y)\} \leq \hat{v}^T \hat{x} - f(\hat{x}) < \infty.$$

This shows that $\mathrm{dom}\, f^* \neq \emptyset$. In addition, for any $v \in \mathbb{R}^n$, we have

$$f^*(v) = \sup_{y \in \mathbb{R}^n} \{v^T y - f(y)\} \geq v^T \hat{x} - f(\hat{x}) > -\infty$$

because $f(\hat{x}) < \infty$. These prove item (i).

Fenchel conjugate cont.

Proof of Theorem 5.1 cont.:
Next, since $\operatorname{dom} f \neq \emptyset$, we have $\operatorname{ri}(\operatorname{dom} f) \neq \emptyset$. Let $\hat{x} \in \operatorname{ri}(\operatorname{dom} f)$ and we know from Theorem 4.11 that there exists $\hat{v} \in \partial f(\hat{x})$. Hence

$$f(y) - f(\hat{x}) \geq \hat{v}^T(y - \hat{x}) \quad \forall y \in \mathbb{R}^n.$$

Thus,

$$f^*(\hat{v}) = \sup_{y \in \mathbb{R}^n} \{\hat{v}^T y - f(y)\} \leq \hat{v}^T \hat{x} - f(\hat{x}) < \infty.$$

This shows that $\operatorname{dom} f^* \neq \emptyset$. In addition, for any $v \in \mathbb{R}^n$, we have

$$f^*(v) = \sup_{y \in \mathbb{R}^n} \{v^T y - f(y)\} \geq v^T \hat{x} - f(\hat{x}) > -\infty$$

because $f(\hat{x}) < \infty$. These prove item (i).

Item (ii) follows directly from the definition of Fenchel conjugate.

Fenchel conjugate cont.

Proof of Theorem 5.1 cont.:
We now prove item (iii). It is clear from item (i) that if $f = f^{**}$, then f is closed. It suffices to prove the converse implication.

Fenchel conjugate cont.

Proof of Theorem 5.1 cont.:
We now prove item (iii). It is clear from item (i) that if $f = f^{**}$, then f is closed. It suffices to prove the converse implication.

First, using the definition of conjugate, we have for all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$ that

$$f^*(y) = \sup_{u \in \mathbb{R}^n} \{y^T u - f(u)\} \geq y^T x - f(x).$$

Hence, $f(x) \geq \sup_{y \in \mathbb{R}^n} \{y^T x - f^*(y)\} = f^{**}(x)$.

Now, suppose that f is closed. Fix any $\hat{x} \in \mathbb{R}^n$ and let $\beta \in \mathbb{R}$ be such that $\beta < f(\hat{x})$. Then $(\hat{x}, \beta) \notin \operatorname{epi} f$.

Fenchel conjugate cont.

Proof of Theorem 5.1 cont.:
Since f is closed, Theorem 4.3 guarantees that there exists $(w, \alpha) \neq 0$ and $\gamma \in \mathbb{R}$ so that

$$w^T \hat{x} + \alpha\beta > \gamma > w^T x + \alpha r, \quad \forall (x, r) \in \operatorname{epi} f. \tag{1}$$

Take any $\tilde{x} \in \operatorname{dom} f$. Note that for all $t \geq 0$, $(\tilde{x}, f(\tilde{x}) + t) \in \operatorname{epi} f$. Then

$$w^T \hat{x} + \alpha\beta > w^T \tilde{x} + \alpha f(\tilde{x}) + \alpha t, \quad \forall t \geq 0.$$

Letting $t \to \infty$, we can see that $\alpha \leq 0$.

Fenchel conjugate cont.

Proof of Theorem 5.1 cont.:
Since f is closed, Theorem 4.3 guarantees that there exists $(w, \alpha) \neq 0$ and $\gamma \in \mathbb{R}$ so that

$$w^T \hat{x} + \alpha\beta > \gamma > w^T x + \alpha r, \quad \forall (x, r) \in \operatorname{epi} f. \tag{1}$$

Take any $\tilde{x} \in \operatorname{dom} f$. Note that for all $t \geq 0$, $(\tilde{x}, f(\tilde{x}) + t) \in \operatorname{epi} f$. Then

$$w^T \hat{x} + \alpha\beta > w^T \tilde{x} + \alpha f(\tilde{x}) + \alpha t, \quad \forall t \geq 0.$$

Letting $t \to \infty$, we can see that $\alpha \leq 0$.

If $\alpha < 0$, dividing $-\alpha$ on both sides of (1) and writing $v := -w/\alpha$ give

$$v^T \hat{x} - \beta > v^T x - f(x), \quad \forall x \in \operatorname{dom} f.$$

This implies $v^T \hat{x} - \beta \geq f^*(v)$, and hence

$$f^{**}(\hat{x}) \geq v^T \hat{x} - f^*(v) \geq \beta.$$

Since $\beta < f(\hat{x})$ is arbitrary, we must then have $f^{**}(\hat{x}) \geq f(\hat{x})$ and hence $f^{**}(\hat{x}) = f(\hat{x})$.

Fenchel conjugate cont.

Proof of Theorem 5.1 cont.: Finally, suppose that $\alpha = 0$. Then we have from (1) that

$$w^T \hat{x} \geq w^T x + c, \quad \forall x \in \operatorname{dom} f \tag{2}$$

where $c = w^T \hat{x} - \gamma > 0$.

Fenchel conjugate cont.

Proof of Theorem 5.1 cont.: Finally, suppose that $\alpha = 0$. Then we have from (1) that

$$w^T \hat{x} \geq w^T x + c, \quad \forall x \in \operatorname{dom} f \tag{2}$$

where $c = w^T \hat{x} - \gamma > 0$. Moreover, since $\operatorname{ri}(\operatorname{dom} f) \neq \emptyset$, we can take $\tilde{x} \in \operatorname{ri}(\operatorname{dom} f)$ and $\tilde{v} \in \partial f(\tilde{x})$ (thanks to Theorem 4.11) and obtain

$$f(x) \geq \tilde{v}^T x + [f(\tilde{x}) - \tilde{v}^T \tilde{x}] \tag{3}$$

Fenchel conjugate cont.

Proof of Theorem 5.1 cont.: Finally, suppose that $\alpha = 0$. Then we have from (1) that

$$w^T \hat{x} \geq w^T x + c, \quad \forall x \in \operatorname{dom} f \tag{2}$$

where $c = w^T \hat{x} - \gamma > 0$. Moreover, since $\operatorname{ri}(\operatorname{dom} f) \neq \emptyset$, we can take $\tilde{x} \in \operatorname{ri}(\operatorname{dom} f)$ and $\tilde{v} \in \partial f(\tilde{x})$ (thanks to Theorem 4.11) and obtain

$$f(x) \geq \tilde{v}^T x + [f(\tilde{x}) - \tilde{v}^T \tilde{x}] \tag{3}$$

Adding $t > 0$ times of (2) to (3) gives for all $x \in \operatorname{dom} f$ and all $t > 0$

$$\begin{aligned} & f(x) \geq \tilde{v}^T x + [f(\tilde{x}) - \tilde{v}^T \tilde{x}] + t(w^T x + c) - t w^T \hat{x} \\ \Longrightarrow \quad & t w^T \hat{x} \geq (\tilde{v} + tw)^T x - f(x) + [f(\tilde{x}) - \tilde{v}^T \tilde{x}] + tc. \end{aligned}$$

This implies that $t w^T \hat{x} \geq f^*(\tilde{v} + tw) + [f(\tilde{x}) - \tilde{v}^T \tilde{x}] + tc$. Thus,

$$f^{**}(\hat{x}) \geq \hat{x}^T (\tilde{v} + tw) - f^*(\tilde{v} + tw) \geq \hat{x}^T \tilde{v} + [f(\tilde{x}) - \tilde{v}^T \tilde{x}] + tc.$$

Since $c > 0$, we have $f^{**}(\hat{x}) = \infty \geq f(\hat{x})$. This completes the proof.

Young's inequality

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and convex. Then

$$f(x) + f^*(y) \geq x^T y, \quad \forall x,\ y \in \mathbb{R}^n. \tag{4}$$

Young's inequality

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and convex. Then

$$f(x) + f^*(y) \geq x^T y, \quad \forall x,\ y \in \mathbb{R}^n. \tag{4}$$

Moreover, for any x and $y \in \mathbb{R}^n$, we have:

$$\begin{aligned}
& f(u) - f(x) \geq y^T(u - x) \quad \forall u \in \mathbb{R}^n, \\
\iff\ & y^T x - f(x) \geq y^T u - f(u) \quad \forall u \in \mathbb{R}^n, \\
\iff\ & y^T x - f(x) \geq \sup_{u \in \mathbb{R}^n} \{y^T u - f(u)\}, \\
\iff\ & y^T x - f(x) \geq f^*(y), \\
\iff\ & y^T x \geq f^*(y) + f(x).
\end{aligned}$$

From the above, we obtain

$$\boxed{y \in \partial f(x) \iff y^T x \geq f^*(y) + f(x) \iff y^T x = f^*(y) + f(x).}$$

Inf-projection

Definition: Let $F : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ be a proper function. We define its inf-projection (w.r.t. y) as

$$f(x) := \inf_{y \in \mathbb{R}^m} F(x, y). \tag{5}$$

Remark: The inf-projection is not necessarily proper.

Inf-projection

Definition: Let $F : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ be a proper function. We define its inf-projection (w.r.t. y) as

$$f(x) := \inf_{y \in \mathbb{R}^m} F(x, y). \tag{5}$$

Remark: The inf-projection is not necessarily proper.

Example: Let $F : \mathbb{R}^n \times \mathbb{R}^n \to \overline{\mathbb{R}}$ be defined as

$$F(x, y) = \sum_{i=1}^{n} y_i + \delta_D(x, y),$$

with $D := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n : \ -y \leq x \leq y\}$. Then

$$\begin{aligned} f(x) := \inf_{y \in \mathbb{R}^n} \left\{ \sum_{i=1}^{n} y_i + \delta_D(x, y) \right\} &= \inf_{y \in \mathbb{R}^n} \left\{ \sum_{i=1}^{n} y_i : \ -y \leq x \leq y \right\} \\ = \inf_{y \in \mathbb{R}^n} \left\{ \sum_{i=1}^{n} y_i : \ |x| \leq y \right\} &= \sum_{i=1}^{n} |x_i| = \|x\|_1 . \end{aligned}$$

Thus, the ℓ_1 norm can be represented as an inf-projection.

Inf-projection cont.

Theorem 5.2 (Convexity preservation)
Let $F : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ be a proper convex function and consider its inf-projection f as in (5). Then f is a convex function.

Inf-projection cont.

Theorem 5.2 (Convexity preservation)
Let $F : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ be a proper convex function and consider its inf-projection f as in (5). Then f is a convex function.

Proof: Since f can be improper in general, we use the epigraph characterization of convexity.

Let (x_1, λ_1) and $(x_2, \lambda_2) \in \operatorname{epi} f$, and $t \in (0, 1)$. Fix any $\epsilon > 0$. Then there exist $y_{1,\epsilon}$ and $y_{2,\epsilon}$ such that

$$\lambda_1 + \epsilon > F(x_1, y_{1,\epsilon}) \text{ and } \lambda_2 + \epsilon > F(x_2, y_{2,\epsilon}).$$

Thus,

$$\begin{aligned} f(tx_1 + (1-t)x_2) &\leq F(tx_1 + (1-t)x_2, ty_{1,\epsilon} + (1-t)y_{2,\epsilon}) \\ &\leq t\,F(x_1, y_{1,\epsilon}) + (1-t)F(x_2, y_{2,\epsilon}) \leq t\lambda_1 + (1-t)\lambda_2 + \epsilon, \end{aligned}$$

Consequently, $t\lambda_1 + (1-t)\lambda_2 + \epsilon > f(tx_1 + (1-t)x_2)$. Since this is true for any $\epsilon > 0$, we have $t\lambda_1 + (1-t)\lambda_2 \geq f(tx_1 + (1-t)x_2)$, which means $(tx_1 + (1-t)x_2, t\lambda_1 + (1-t)\lambda_2) \in \operatorname{epi} f$.

Fenchel duality

Theorem 5.3 (Fenchel duality – interior version)
Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ be proper convex functions and let $A \in \mathbb{R}^{m\times n}$. Let V_p, $V_d \in \overline{\mathbb{R}}$ be the primal and dual values defined, respectively, by:

$$V_p := \inf_{x\in\mathbb{R}^n}\{f(x) + g(Ax)\}, \qquad V_d := \sup_{u\in\mathbb{R}^m}\{-f^*(A^T u) - g^*(-u)\}.$$

Then the following statements hold.

(i) It holds that $V_p \geq V_d$.

(ii) If we assume in addition that

$$0 \in \operatorname{int}(\operatorname{dom} g - A \operatorname{dom} f),$$

then $V_p = V_d$; moreover, V_d is attainable when finite.

Fenchel duality cont.

Proof of Theorem 5.3:
For any x and u, Young's inequality gives

$$\begin{aligned} f(x) + f^*(A^T u) &\geq x^T(A^T u) = (Ax)^T u, \\ g(Ax) + g^*(-u) &\geq -(Ax)^T u. \end{aligned}$$

Summing the above relations, rearranging terms and taking inf and sup suitably gives $V_p \geq V_d$. This proves item (i).

Fenchel duality cont.

Proof of Theorem 5.3:
For any x and u, Young's inequality gives

$$\begin{aligned} f(x) + f^*(A^T u) &\geq x^T(A^T u) = (Ax)^T u, \\ g(Ax) + g^*(-u) &\geq -(Ax)^T u. \end{aligned}$$

Summing the above relations, rearranging terms and taking inf and sup suitably gives $V_p \geq V_d$. This proves item (i).

Now suppose in addition that

$$0 \in \operatorname{int}(\operatorname{dom} g - A \operatorname{dom} f)$$

If $V_p = -\infty$, then $V_p \geq V_d$ forces $V_p = V_d$. Thus, it remains to consider the case $V_p > -\infty$.

Fenchel duality cont.

Proof of Theorem 5.3 cont.: Define

$$V(u) := \inf_{x \in \mathbb{R}^n} \{f(x) + g(Ax + u)\}.$$

Then $V(0) = V_p$ and V is convex thanks to Theorem 5.2.

Fenchel duality cont.

Proof of Theorem 5.3 cont.: Define

$$V(u) := \inf_{x \in \mathbb{R}^n} \{f(x) + g(Ax + u)\}.$$

Then $V(0) = V_p$ and V is convex thanks to Theorem 5.2.
We claim:

(a) $0 \in \operatorname{int}(\operatorname{dom} V)$.

(b) V is proper.

Fenchel duality cont.

Proof of Theorem 5.3 cont.: Define

$$V(u) := \inf_{x \in \mathbb{R}^n} \{f(x) + g(Ax + u)\}.$$

Then $V(0) = V_p$ and V is convex thanks to Theorem 5.2.
We claim:

(a) $0 \in \operatorname{int}(\operatorname{dom} V)$.

(b) V is proper.

Granting them, Theorem 4.10 shows that $\partial V(0) \neq \emptyset$. Let $\tilde{u} \in \partial V(0)$. Then Young's inequality gives $V(0) + V^*(\tilde{u}) = 0$. On the other hand,

$$\begin{aligned} V^*(\tilde{u}) &= \sup_{y \in \mathbb{R}^m} \{y^T \tilde{u} - \inf_{x \in \mathbb{R}^n} \{f(x) + g(Ax + y)\}\} \\ &= \sup_{x \in \mathbb{R}^n, y \in \mathbb{R}^m} \{(Ax + y)^T \tilde{u} - x^T A^T \tilde{u} - f(x) - g(Ax + y)\} \\ &= \sup_{w \in \mathbb{R}^m} \{w^T \tilde{u} - g(w)\} + \sup_{x \in \mathbb{R}^n} \{-x^T A^T \tilde{u} - f(x)\} = g^*(\tilde{u}) + f^*(-A^T \tilde{u}). \end{aligned}$$

Thus, $V_p = V(0) = -V^*(\tilde{u}) \leq V_d$. Consequently, $V_p = V_d$ and the value is attained at $-\tilde{u}$.

Fenchel duality cont.

Proof of Theorem 5.3 cont.: It now remains to prove claims (a) and (b).

For (a), it suffices to show that $\operatorname{dom} V = -A \operatorname{dom} f + \operatorname{dom} g$, and is left as an exercise.

For (b), recall that $V(0) > -\infty$ and $0 \in \operatorname{int}(\operatorname{dom} V)$. Suppose to the contrary that V is not proper. Then there exists $\hat{u}$ such that

$$V(\hat{u}) = -\infty.$$

Since $0 \in \operatorname{int}(\operatorname{dom} V)$, there exist $\epsilon \in (0, 1)$ and $w \in \operatorname{dom} V$ such that

$$0 = \epsilon \hat{u} + (1 - \epsilon) w.$$

Fenchel duality cont.

Proof of Theorem 5.3 cont.: It now remains to prove claims (a) and (b).

For (a), it suffices to show that $\operatorname{dom} V = -A \operatorname{dom} f + \operatorname{dom} g$, and is left as an exercise.

For (b), recall that $V(0) > -\infty$ and $0 \in \operatorname{int}(\operatorname{dom} V)$. Suppose to the contrary that V is not proper. Then there exists $\hat{u}$ such that

$$V(\hat{u}) = -\infty.$$

Since $0 \in \operatorname{int}(\operatorname{dom} V)$, there exist $\epsilon \in (0, 1)$ and $w \in \operatorname{dom} V$ such that

$$0 = \epsilon \hat{u} + (1 - \epsilon) w.$$

Fix any $\lambda \in \mathbb{R}$ such that $\lambda > V(w)$, which exists because $w \in \operatorname{dom} V$. Then $(w, \lambda) \in \operatorname{epi} V$.

Fenchel duality cont.

Proof of Theorem 5.3 cont.:
Take any $t > -\infty$. Then $(\hat{u}, t) \in \operatorname{epi} V$. Hence,

$$\begin{aligned} & \epsilon(\hat{u}, t) + (1-\epsilon)(w, \lambda) \in \operatorname{epi} V, \\ \Longrightarrow \; & (0, \epsilon t + (1-\epsilon)\lambda) \in \operatorname{epi} V, \\ \Longrightarrow \; & V(0) \leq \epsilon t + (1-\epsilon)\lambda. \end{aligned}$$

Since $t > -\infty$ is arbitrary, letting $t \to -\infty$, we obtain $V(0) = -\infty$, which is a contradiction.

This proves claim (b) and thus completes the proof.

Fenchel duality cont.

Notice that we used int to guarantee that V is proper and $\partial V(0) \neq \emptyset$. A close inspection of the proof reveals that one could have worked with ri instead. This gives:

Theorem 5.4 (Fenchel duality – relative interior version)
Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ be proper convex functions and let $A \in \mathbb{R}^{m \times n}$. Let $V_p, V_d \in \overline{\mathbb{R}}$ be the primal and dual values defined, respectively, by:

$$V_p := \inf_{x \in \mathbb{R}^n} \{f(x) + g(Ax)\}, \qquad V_d := \sup_{u \in \mathbb{R}^m} \{-f^*(A^T u) - g^*(-u)\}.$$

Then the following statements hold.

(i) It holds that $V_p \geq V_d$.

(ii) If we assume in addition that

$$0 \in \mathrm{ri}(\mathrm{dom}\, g - A\, \mathrm{dom}\, f),$$

then $V_p = V_d$; moreover, V_d is attainable when finite.

Subdifferential sum rule

Theorem 5.5 (Subdifferential sum rule)
Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ be proper convex functions, and let $A \in \mathbb{R}^{m \times n}$. Then for all x, we have:

$$\partial(f + g \circ A)(x) \supseteq \partial f(x) + A^T \partial g(Ax).$$

If we assume in addition that

$$0 \in \operatorname{ri}(\operatorname{dom} g - A \operatorname{dom} f),$$

then for all x, we have

$$\partial(f + g \circ A)(x) = \partial f(x) + A^T \partial g(Ax).$$

Subdifferential sum rule

Theorem 5.5 (Subdifferential sum rule)
Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ be proper convex functions, and let $A \in \mathbb{R}^{m\times n}$. Then for all x, we have:

$$\partial(f + g \circ A)(x) \supseteq \partial f(x) + A^T \partial g(Ax).$$

If we assume in addition that

$$0 \in \operatorname{ri}(\operatorname{dom} g - A \operatorname{dom} f),$$

then for all x, we have

$$\partial(f + g \circ A)(x) = \partial f(x) + A^T \partial g(Ax).$$

Remark:

- If $A = I$ and $f \in C^1(\mathbb{R}^n)$, then $\operatorname{dom} f = \mathbb{R}^n$ and hence the ri condition is trivially satisfied. Then we have for all x that

$$\partial(f + g)(x) = \nabla f(x) + \partial g(x).$$

Subdifferential sum rule cont.

Proof of Theorem 5.5:
Let $y_1 \in \partial f(x)$ and $y_2 \in \partial g(Ax)$. Then for all w, we have

$$f(w) \geq f(x) + y_1^T(w - x) \text{ and } g(Aw) \geq g(Ax) + y_2^T(Aw - Ax).$$

Summing the above inequalities gives

$$f(w) + g(Aw) \geq f(x) + g(Ax) + (y_1 + A^T y_2)^T(w - x).$$

Hence $y_1 + A^T y_2 \in \partial(f + g \circ A)(w)$. Note that ri condition is not needed for this inclusion.

Subdifferential sum rule cont.

Proof of Theorem 5.5:
Let $y_1 \in \partial f(x)$ and $y_2 \in \partial g(Ax)$. Then for all w, we have

$$f(w) \geq f(x) + y_1^T(w - x) \text{ and } g(Aw) \geq g(Ax) + y_2^T(Aw - Ax).$$

Summing the above inequalities gives

$$f(w) + g(Aw) \geq f(x) + g(Ax) + (y_1 + A^T y_2)^T(w - x).$$

Hence $y_1 + A^T y_2 \in \partial(f + g \circ A)(w)$. Note that ri condition is not needed for this inclusion.

Now, suppose in addition that $0 \in \mathrm{ri}(\mathrm{dom}\, g - A\,\mathrm{dom}\, f)$. We want to prove the converse inclusion.

Let $\varphi \in \partial(f + g \circ A)(x)$. Then

$$f(y) + g(Ay) - \varphi^T y \geq f(x) + g(Ax) - \varphi^T x$$

for all $y \in \mathbb{R}^n$. This means

$$f(x) + g(Ax) - \varphi^T x = \inf_y \{f(y) - \varphi^T y + g(Ay)\}.$$

Subdifferential sum rule cont.

Proof of Theorem 5.5 cont.:
Let $F(y) := f(y) - \varphi^T y$. Then $\operatorname{dom} F = \operatorname{dom} f$.

Apply Theorem 5.4 to F and g, we get

$$
\begin{aligned}
& f(x) + g(Ax) - \varphi^T x = \inf_y \{F(y) + g(Ay)\} = -F^*(A^T\psi) - g^*(-\psi) \\
& = - \sup_{u \in \mathbb{R}^n} \{u^T A^T \psi - f(u) + \varphi^T u\} - g^*(-\psi) \\
& = - \sup_{u \in \mathbb{R}^n} \{u^T (A^T \psi + \varphi) - f(u)\} - g^*(-\psi) \\
& = -f^*(A^T\psi + \varphi) - g^*(-\psi)
\end{aligned}
$$

for some $\psi \in \mathbb{R}^m$.

Subdifferential sum rule cont.

Proof of Theorem 5.5 cont.: Rearranging terms, we have

$$f(x) + f^*(A^T\psi + \varphi) + g(Ax) + g^*(-\psi) = \varphi^T x$$

Subdifferential sum rule cont.

Proof of Theorem 5.5 cont.: Rearranging terms, we have

$$f(x) + f^*(A^T\psi + \varphi) + g(Ax) + g^*(-\psi) = \varphi^T x$$
$$= (A^T\psi + \varphi)^T x + (-A^T\psi)^T x.$$

Subdifferential sum rule cont.

Proof of Theorem 5.5 cont.: Rearranging terms, we have

$$f(x) + f^*(A^T\psi + \varphi) + g(Ax) + g^*(-\psi) = \varphi^T x$$
$$= (A^T\psi + \varphi)^T x + (-A^T\psi)^T x.$$

On the other hand, recall from Young's inequality that

$$f(x)+f^*(A^T\psi+\varphi) \geq (A^T\psi+\varphi)^T x \quad \text{and} \quad g(Ax)+g^*(-\psi) \geq (-\psi)^T Ax.$$

Hence,

$$f(x)+f^*(A^T\psi+\varphi) = (A^T\psi+\varphi)^T x \quad \text{and} \quad g(Ax)+g^*(-\psi) = (-\psi)^T Ax.$$

Consequently, the equality case of Young's inequality gives

$$A^T\psi + \varphi \in \partial f(x) \quad \text{and} \quad -\psi \in \partial g(Ax).$$

Thus, we have

$$\varphi = A^T\psi + \varphi - A^T\psi \in \partial f(x) + A^T \partial g(Ax).$$

Example

Example: Note that the ri condition in Theorem 5.5 cannot be dropped in general.

To see this, consider

$$C = \{(x_1, x_2) \in \mathbb{R}^2 : \ x_2 \geq x_1^2\} \ \text{and} \ D = \{(x_1, 0) : \ x_1 \in \mathbb{R}\}.$$

Example

Example: Note that the ri condition in Theorem 5.5 cannot be dropped in general.

To see this, consider

$$C = \{(x_1, x_2) \in \mathbb{R}^2 : \; x_2 \geq x_1^2\} \text{ and } D = \{(x_1, 0) : \; x_1 \in \mathbb{R}\}.$$

Then

$$\begin{aligned}\partial(\delta_C + \delta_D)(0) &= \partial\delta_{C\cap D}(0) = \partial\delta_{\{(0,0)\}}(0) \\ &= \{u \in \mathbb{R}^2 : \; u^T(y - 0) \leq 0, \;\; \forall y \in \{(0,0)\}\} = \mathbb{R}^2.\end{aligned}$$

On the other hand,

$$\partial\delta_C(0) = \{u \in \mathbb{R}^2 : \; u^T(y - 0) \leq 0 \;\; \forall y \in C\} = \{t(0, -1) : \; t \geq 0\},$$

$$\partial\delta_D(0) = \{u \in \mathbb{R}^2 : \; u^T(y - 0) \leq 0 \;\; \forall y \in D\} = \{t(0, -1) : \; t \in \mathbb{R}\}.$$

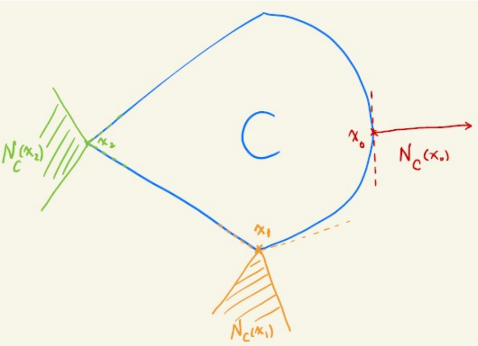
Hence, $\partial\delta_C(0) + \partial\delta_D(0) = \{t(0, -1) : \; t \in \mathbb{R}\} \neq \mathbb{R}^2$.

Normal cones

Definition: Let C be a nonempty closed convex set. The normal cone at $x \in C$ is defined as

$$N_C(x) := \partial\delta_C(x) = \{y \in \mathbb{R}^n : \; y^T(u - x) \le 0, \quad \forall u \in C\}.$$

Remark: Geometrically, $N_C(x)$ is the collection of all vectors making an obtuse angle with vectors $u - x$ for any $u \in C$.



Example

Example: Let $A \in \mathbb{R}^{m \times n}$ with $m \leq n$. Then $C := \ker A$ is closed and convex (and is a subspace). Moreover, at any $x \in C$, we have

$$\begin{aligned} N_C(x) &= \{y \in \mathbb{R}^n : \ y^T(u - x) \leq 0, \ \ \forall u \in C\} \\ &= \{y \in \mathbb{R}^n : \ y^T w \leq 0, \ \ \forall w \in C\} \quad \text{(since } x \in C\text{, it holds that } C - x = C\text{)} \\ &= (\ker A)^{\perp} = \mathrm{Range}(A^T) \\ &= \{A^T \lambda : \ \lambda \in \mathbb{R}^m\}. \end{aligned}$$

Lagrangian

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and convex. Let $A \in \mathbb{R}^{m \times n}$ and $b \in \text{Range}(A)$. We consider the following optimization problem:

$$\begin{array}{rl} \underset{x \in \mathbb{R}^n}{\text{Minimize}} & f(x) \\ \text{Subject to} & Ax = b. \end{array} \tag{6}$$

Theorem 5.6 (Lagrange multipliers)

Consider (6). Suppose that the optimal value is finite and

$$0 \in \text{ri}(b - A\,\text{dom}\, f).$$

Then there exists $\lambda \in \mathbb{R}^m$ so that

$$\inf_{Ax=b} f(x) = \inf_{x \in \mathbb{R}^n} \{f(x) - \lambda^T (Ax - b)\}.$$

Lagrangian cont.

Proof of Theorem 5.6:

Since the optimal value is finite, we can apply Theorem 5.4 (with $g = \delta_{\{b\}}$) to conclude the existence of $\lambda \in \mathbb{R}^m$ such that

$$\inf_{Ax=b} f(x) = -f^*(A^T\lambda) - g^*(-\lambda).$$

Now, a direct computation shows

$$\begin{aligned} -f^*(A^T\lambda) &= -\sup_{x\in\mathbb{R}^n}\{x^TA^T\lambda - f(x)\} \\ &= \inf_{x\in\mathbb{R}^n}\{f(x) - \lambda^TAx\}, \\ -g^*(-\lambda) &= b^T\lambda. \end{aligned}$$

This completes the proof.

A projection problem

As an application of Theorem 5.6, we discuss the problem of projecting a y onto the unit simplex:

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{Minimize}} & \frac{1}{2}\|x-y\|_2^2 \\ \text{Subject to} & \sum_{i=1}^n x_i = 1, \quad x \geq 0. \end{array}$$

A projection problem

As an application of Theorem 5.6, we discuss the problem of projecting a y onto the unit simplex:

$$\begin{array}{rl} \underset{x\in\mathbb{R}^n}{\text{Minimize}} & \frac{1}{2}\|x-y\|_2^2 \\ \text{Subject to} & \sum_{i=1}^n x_i = 1, \quad x \geq 0. \end{array}$$

- A unique minimizer $\bar{x}$ of the above problem exists in view of Theorem 4.1.
- By Theorem 5.6 and Theorem 5.5, there exists $\lambda \in \mathbb{R}$ so that

  $$0 \in \bar{x} - y + N_{\mathbb{R}^n_+}(\bar{x}) - \lambda e,$$

  where e is the vector of ones.

A projection problem

As an application of Theorem 5.6, we discuss the problem of projecting a y onto the unit simplex:

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{Minimize}} & \frac{1}{2}\|x-y\|_2^2 \\ \text{Subject to} & \sum_{i=1}^n x_i = 1, \quad x \geq 0. \end{array}$$

- A unique minimizer $\bar{x}$ of the above problem exists in view of Theorem 4.1.
- By Theorem 5.6 and Theorem 5.5, there exists $\lambda \in \mathbb{R}$ so that
  $$0 \in \bar{x} - y + N_{\mathbb{R}^n_+}(\bar{x}) - \lambda e,$$
  where e is the vector of ones.
- Observe that exercise
  $$N_{\mathbb{R}^n_+}(\bar{x}) = \left\{ y \in \mathbb{R}^n : \sum_{i=1}^n y_i(x_i - \bar{x}_i) \leq 0 \ \ \forall x \in \mathbb{R}^n_+ \right\} = \prod_{i=1}^n N_{\mathbb{R}_+}(\bar{x}_i).$$

A projection problem cont.

As an application of Theorem 5.6, we discuss the problem of projecting a y onto the unit simplex:

$$\begin{array}{rl} \underset{x\in\mathbb{R}^n}{\text{Minimize}} & \frac{1}{2}\|x-y\|_2^2 \\ \text{Subject to} & \sum_{i=1}^n x_i = 1, \quad x \geq 0. \end{array}$$

- Note that for each i, we have $0 \in \bar{x}_i - y_i + N_{\mathbb{R}_+}(\bar{x}_i) - \lambda$, i.e., $\bar{x}_i$ minimizes
$$t \mapsto \frac{1}{2}(t - y_i - \lambda)^2 + \delta_{\mathbb{R}_+}(t).$$
This means that $\bar{x}_i = (y_i + \lambda)_+$.

A projection problem cont.

As an application of Theorem 5.6, we discuss the problem of projecting a y onto the unit simplex:

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{Minimize}} & \frac{1}{2}\|x-y\|_2^2 \\ \text{Subject to} & \sum_{i=1}^n x_i = 1, \quad x \geq 0. \end{array}$$

- Note that for each i, we have $0 \in \bar{x}_i - y_i + N_{\mathbb{R}_+}(\bar{x}_i) - \lambda$, i.e., $\bar{x}_i$ minimizes
$$t \mapsto \frac{1}{2}(t - y_i - \lambda)^2 + \delta_{\mathbb{R}_+}(t).$$
This means that $\bar{x}_i = (y_i + \lambda)_+$.
- Since $e^T\bar{x} = 1$, we now find a solution λ_* of the following 1-dimensional piecewise-linear function:
$$\sum_{i=1}^n (y_i + \lambda)_+ = 1.$$
- The optimal solution $\bar{x}$ is then obtained via $\bar{x}_i = (y_i + \lambda_*)_+$ for all i.

Lagrange duality

We consider the following convex optimization problem:

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{Minimize}} & f(x) \\ \text{Subject to} & g_i(x) \leq 0,\ i = 1, \ldots, m, \end{array} \tag{7}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$, are convex.

Theorem 5.7 (Lagrange duality)
Consider (7). Suppose that

- the optimal value is finite; and
- there exists $\hat{x}$ satisfying $g_i(\hat{x}) < 0$ for all i.

Then there exists $\lambda^* \in \mathbb{R}^m_+$ such that

$$\inf_{x \in \mathbb{R}^n} \{f(x) :\ g_i(x) \leq 0, i = 1, \ldots, m\} = \inf_{x \in \mathbb{R}^n} \left\{ f(x) + \sum_{i=1}^m \lambda_i^* g_i(x) \right\}.$$

Lagrange duality

We consider the following convex optimization problem:

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{Minimize}} & f(x) \\ \text{Subject to} & g_i(x) \le 0,\ i = 1, \ldots, m, \end{array} \tag{7}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$, are convex.

Theorem 5.7 (Lagrange duality)
Consider (7). Suppose that

- the optimal value is finite; and
- there exists $\hat{x}$ satisfying $g_i(\hat{x}) < 0$ for all i.

Then there exists $\lambda^* \in \mathbb{R}^m_+$ such that

$$\inf_{x \in \mathbb{R}^n} \{f(x) :\ g_i(x) \le 0, i = 1, \ldots, m\} = \inf_{x \in \mathbb{R}^n} \left\{ f(x) + \sum_{i=1}^m \lambda_i^* g_i(x) \right\}.$$

Remark: The above $\hat{x}$ is called a Slater point.

Lagrange duality cont.

Proof sketch of Theorem 5.7:
Let V_p denote the optimal value of (7). Define the value function $V : \mathbb{R}^m \to \overline{\mathbb{R}}$ by

$$V(u) := \inf_{x \in \mathbb{R}^n} \{f(x) : \; g_i(x) \leq u_i, i = 1, \dots, m\}.$$

Then $V(0) = V_p$ and V is convex in virtue of Theorem 5.2. Exercise

Similar to the arguments on slides 14 and 15, one can show that

- $0 \in \text{int}(\text{dom}\, V)$;
- V is proper.

Using the above and Theorem 4.10, we conclude that $\partial V(0) \neq \emptyset$. Let $-\lambda^* \in \partial V(0)$. Then Young's inequality gives

$$V(0) + V^*(-\lambda^*) = 0. \tag{8}$$

Lagrange duality cont.

Proof sketch of Theorem 5.7 cont.: We next compute $V^*(y)$. From the definition, we have (upon writing $g(x) = \begin{bmatrix} g_1(x) & \cdots & g_m(x) \end{bmatrix}^T$)

$$V^*(y) = \sup_{u \in \mathbb{R}^m} \{u^T y - V(u)\} = \sup_{u \in \mathbb{R}^m} \sup_{g(x) \leq u} \{u^T y - f(x)\}.$$

Lagrange duality cont.

Proof sketch of Theorem 5.7 cont.: We next compute $V^*(y)$. From the definition, we have (upon writing $g(x) = \begin{bmatrix} g_1(x) & \cdots & g_m(x) \end{bmatrix}^T$)

$$V^*(y) = \sup_{u \in \mathbb{R}^m} \{u^T y - V(u)\} = \sup_{u \in \mathbb{R}^m} \sup_{g(x) \leq u} \{u^T y - f(x)\}.$$

Note that $g(x) \leq u$ if and only if $\exists w \in \mathbb{R}^m_+$ such that $u = g(x) + w$.

Then

$$\begin{aligned} V^*(y) &= \sup_{w \in \mathbb{R}^m_+} \sup_{x \in \mathbb{R}^n} \{(g(x) + w)^T y - f(x)\} \\ &= \sup_{w \in \mathbb{R}^m_+} \sup_{x \in \mathbb{R}^n} \{y^T g(x) - f(x) + w^T y\} \\ &= \sup_{x \in \mathbb{R}^n} \{y^T g(x) - f(x)\} + \delta_{\mathbb{R}^m_-}(y) \end{aligned}$$

Lagrange duality cont.

Proof sketch of Theorem 5.7 cont.: In particular, Young's inequality gives that for all $z \in \mathbb{R}^m_+$,

$$V(0) \geq -V^*(-z) = \inf_{x \in \mathbb{R}^n} \left\{ f(x) + \sum_{i=1}^{m} z_i g_i(x) \right\}.$$

Finally, combining the expression of V^* with (8) and the finiteness of $V(0)$, we have $\lambda^* \in \mathbb{R}^m_+$ and

$$V(0) = -V^*(-\lambda^*) = \inf_{x \in \mathbb{R}^n} \left\{ f(x) + \sum_{i=1}^{m} \lambda_i^* g_i(x) \right\}.$$

Remarks

Remarks

- The proof actually shows that under the assumptions of Theorem 5.7,

$$\inf_{x\in\mathbb{R}^n}\left\{f(x)+\sum_{i=1}^{m}\lambda_i^* g_i(x)\right\} = \sup_{\lambda\in\mathbb{R}^m_+}\inf_{x\in\mathbb{R}^n}\left\{f(x)+\sum_{i=1}^{m}\lambda_i g_i(x)\right\}.$$

Remarks

Remarks

- The proof actually shows that under the assumptions of Theorem 5.7,

$$\inf_{x\in\mathbb{R}^n}\left\{f(x)+\sum_{i=1}^{m}\lambda_i^* g_i(x)\right\} = \sup_{\lambda\in\mathbb{R}^m_+}\inf_{x\in\mathbb{R}^n}\left\{f(x)+\sum_{i=1}^{m}\lambda_i g_i(x)\right\}.$$

- Notice that we have

$$\sup_{\lambda\in\mathbb{R}^m_+}\lambda^T w = \delta_{\mathbb{R}^m_-}(w)$$

for $w \in \mathbb{R}^m$.

Remarks

Remarks

- The proof actually shows that under the assumptions of Theorem 5.7,

$$\inf_{x\in\mathbb{R}^n}\left\{f(x)+\sum_{i=1}^m \lambda_i^* g_i(x)\right\} = \sup_{\lambda\in\mathbb{R}^m_+}\inf_{x\in\mathbb{R}^n}\left\{f(x)+\sum_{i=1}^m \lambda_i g_i(x)\right\}.$$

- Notice that we have

$$\sup_{\lambda\in\mathbb{R}^m_+} \lambda^T w = \delta_{\mathbb{R}^m_-}(w)$$

for $w \in \mathbb{R}^m$. Hence, we can write

$$\inf_{x\in\mathbb{R}^n}\{f(x):\ g_i(x)\le 0, i=1,\dots,m\} = \inf_{x\in\mathbb{R}^n}\sup_{\lambda\in\mathbb{R}^m_+}\left\{f(x)+\sum_{i=1}^m \lambda_i g_i(x)\right\}.$$

- Lagrange duality can be interpreted as a minimax duality.

Remarks

Define,

$$d(\lambda) = \inf_{x \in \mathbb{R}^n} \left\{ f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) \right\}.$$

Then the Lagrange duality result for (7) can be restated as

Corollary 5.1 (Lagrange duality alternative form)
Consider (7). Suppose that

- the optimal value is finite; and
- there exists $\hat{x}$ satisfying $g_i(\hat{x}) < 0$ for all i.

Define d as above.

$$\inf_{x \in \mathbb{R}^n} \{ f(x) : \ g_i(x) \le 0, i = 1, \dots, m \} = \sup_{\lambda \in \mathbb{R}^m_+} d(\lambda)$$

and the supremum is attainable.

Remark: $\text{Maximize}_{\lambda \in \mathbb{R}^m_+} d(\lambda)$ is the Lagrange dual problem of (7).

A normal cone formula

Proposition 5.1

Let $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$, be convex functions, and consider

$$C := \{x : \; g_i(x) \leq 0, \; i = 1, \ldots, m\}.$$

Suppose that there exists $\hat{x}$ satisfying $g_i(\hat{x}) < 0$ for all i. Then for any $\bar{x} \in C$, it holds that

$$N_C(\bar{x}) = \left\{ \sum_{i \in I(\bar{x})} \lambda_i \partial g_i(\bar{x}) : \; \lambda_i \geq 0 \; \forall i \in I(\bar{x}) \right\},$$

where $I(\bar{x}) := \{j : \; g_j(\bar{x}) = 0\}$.

Remark: Note that if $I(\bar{x}) = \emptyset$, then $N_C(\bar{x}) = \{0\}$.

A normal cone formula cont.

Proof of Proposition 5.1: Write, for notational simplicity,

$$\Omega(\bar{x}) := \left\{ \sum_{i \in I(\bar{x})} \lambda_i \partial g_i(\bar{x}) : \ \lambda_i \geq 0 \ \forall i \in I(\bar{x}) \right\}.$$

We show that $N_C(\bar{x}) = \Omega(\bar{x})$.

A normal cone formula cont.

Proof of Proposition 5.1: Write, for notational simplicity,

$$\Omega(\bar{x}) := \left\{ \sum_{i \in I(\bar{x})} \lambda_i \partial g_i(\bar{x}) : \ \lambda_i \geq 0 \ \forall i \in I(\bar{x}) \right\}.$$

We show that $N_C(\bar{x}) = \Omega(\bar{x})$.

Let $z \in \Omega(\bar{x})$. Then there exist $\xi_i \in \partial g_i(\bar{x})$ and $\lambda_i \geq 0$ for each $i \in I(x)$ such that $z = \sum_{i \in I(\bar{x})} \lambda_i \xi_i$. Then for any $y \in C$, it holds that

$$\begin{aligned} z^T(y - \bar{x}) &= \sum_{i \in I(\bar{x})} \lambda_i \xi_i^T (y - \bar{x}) \leq \sum_{i \in I(\bar{x})} \lambda_i (g_i(y) - g_i(\bar{x})) \\ &= \sum_{i \in I(\bar{x})} \lambda_i g_i(y) \leq 0 \end{aligned}$$

since $g_i(\bar{x}) = 0$ when $i \in I(\bar{x})$. Thus, $z \in N_C(\bar{x})$.

A normal cone formula cont.

Proof of Proposition 5.1: Conversely, suppose that $z \in N_C(\bar{x})$. Then

$$z^T(y - \bar{x}) \leq 0 \ \ \forall y \in C.$$

In particular, $\bar{x}$ solves $\underset{x \in C}{\text{Minimize}} -z^T x$. Since a Slater point exists, by Theorem 5.7, there exists $\lambda^* \in \mathbb{R}^m_+$ such that

$$-z^T\bar{x} = \inf_{x \in \mathbb{R}^n} \left\{ -z^T x + \sum_{i=1}^m \lambda_i^* g_i(x) \right\}$$

A normal cone formula cont.

Proof of Proposition 5.1: Conversely, suppose that $z \in N_C(\bar{x})$. Then

$$z^T(y - \bar{x}) \leq 0 \;\; \forall y \in C.$$

In particular, $\bar{x}$ solves $\underset{x \in C}{\text{Minimize}} -z^T x$. Since a Slater point exists, by Theorem 5.7, there exists $\lambda^* \in \mathbb{R}^m_+$ such that

$$-z^T\bar{x} = \inf_{x \in \mathbb{R}^n} \left\{ -z^T x + \sum_{i=1}^m \lambda_i^* g_i(x) \right\} \leq -z^T\bar{x} + \sum_{i=1}^m \lambda_i^* g_i(\bar{x}) \leq -z^T\bar{x}.$$

The equality together with Proposition 4.4 and Theorem 5.5 gives

$$0 \in -z + \sum_{i=1}^m \partial(\lambda_i^* g_i)(\bar{x}). \tag{9}$$

Moreover, the violet part gives $\sum_{i=1}^m \lambda_i^* g_i(\bar{x}) = 0$, which implies $\lambda_i^* = 0$ whenever $i \notin I(\bar{x})$. Thus, it suffices to sum over $I(\bar{x})$ in (9). This proves $z \in \Omega(\bar{x})$.