Power System Analysis

Chapter 12 Nonsmooth convex optimization

Steven Low Caltech. Feb 7, 2025

Motivation

Consider

$$\min_{x \in \mathbb{R}^n} \quad f(x) \qquad \text{s.t.} \qquad x \in X$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is convex function and $X \subseteq \mathbb{R}^n$ is convex set

Develop basic theory to answer:

1. How to characterize optimal solutions?
   - Saddle point theorem, KKT theorem
2. When will optimal solutions exist and when will it be unique?
   - Primal optimality, Slater theorem

Motivation

Generalization of smooth convex optimization theory (Ch 7.3) to nonsmooth setting:

- Cost or constraint functions may not be differentiable
- Cost or constraint functions may take $\pm\infty$ values
- But cost and constraint functions are convex (hence subdifferentiable & continuous in ri(effective domain))

Nonsmoothness arises in:

- Dual problem: $\min_{\lambda,\mu} \; d(\lambda,\mu) \quad$ s.t. $\quad \mu \geq 0$ — nonsmooth $d(\lambda,\mu)$
- Robust optimization: $\min_{x\in\mathbb{R}^n} \; f(x) \quad$ s.t. $\quad \bar{h}(x) := \max_{\zeta\in Z} h(x,\zeta) \;\leq\; 0$ — nonsmooth $\bar{h}(x)$
- Two-stage optimization with recourse:

$$\inf_x \; f^1(x) + Q(x) \quad \text{s.t.} \quad h^1(x) \leq 0$$

nonsmooth $Q(x)$

$$Q(x) := E_\omega\left(\inf_{y(\omega)} \left\{ f^2(x, y(\omega)) : h^2(x, y(\omega)) \leq 0 \right\} \right)$$

Motivation

For convex opt, optimality conditions are based on linear approximations of cost function and feasible set, e.g.

$$-\nabla f(x^*) = \nabla g(x^*)\lambda^* + \nabla h(x^*)\mu^*$$

i.e. x^* is minimizer iff negative gradient $-\nabla f(x^*)$ points away from linear approximation of feasible set at x^*, defined by gradients $\nabla g(x^*), \nabla h(x^*)$ of constraint functions

For nonsmooth setting, how to generalize:

1. Linear approximation of feasible set
   - Tangent cone $T_X(x^*)$ or equivalently normal cone $N_X(x^*)$
2. Smooth real-valued functions
   - Extended real-valued CPC functions
3. Gradients
   - Subgradients $\partial f(x), \partial g(x), \partial h(x)$ which always exist for convex (extended real-valued) functions

Motivation

Turns out smoothness is unimportant for structural properties (important for computation)

- Fundamental property is convexity of cost and constraint functions

To generalize structural results to nonsmooth setting

- Generalize: linear approximations of feasible set, CPC functions, subgradients
- Express optimality conditions in terms of cost subgradient and normal cone
- Nonsmooth perspective is more abstract, but simpler, geometric and unifying

Outline

1. Normal cones of feasible sets
2. CPC functions
3. Gradient and subgradient
4. Characterization: saddle point = pd optimality + strong duality
5. Characterization: generalized KKT
6. Existence: primal optimum
7. Existence: dual optimum and strong duality
8. Special convex programs

Outline

1. Normal cones of feasible sets
    - Polar cone
    - Normal cone and tangent cone
    - Affine transformation
    - Second-order cones and SOC constraints
    - Proofs
2. CPC functions
3. Gradient and subgradient
4. Characterization: saddle point
5. Characterization: generalized KKT
6. Existence: primal optimum
7. Existence: dual optimum and strong duality
8. Special convex programs

Polar cone

Definition

Let $X \subseteq \mathbb{R}^n$ be a nonempty set

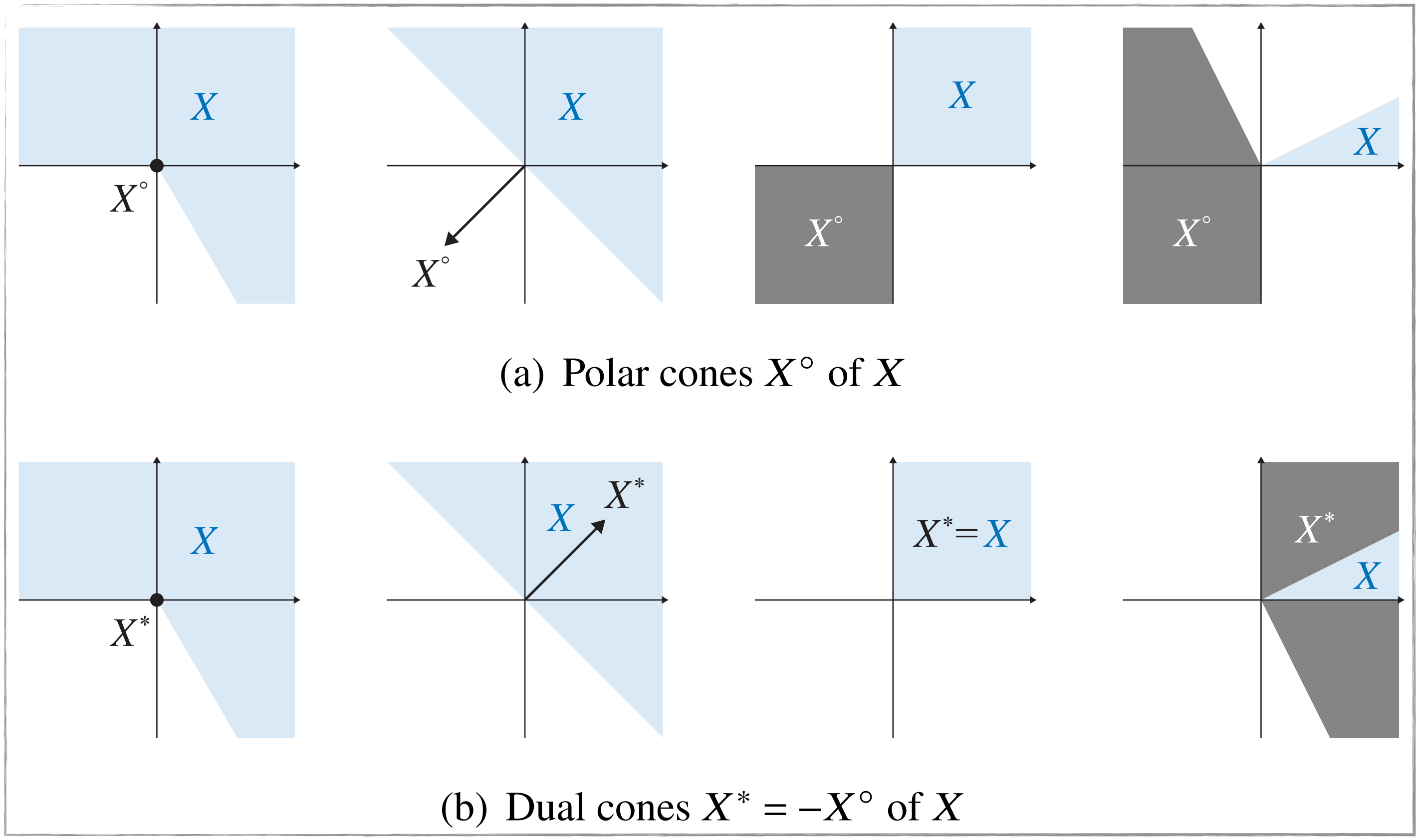
1. The polar cone of X is $X^\circ := \{y \in \mathbb{R}^n : y^\top x \le 0 \quad \forall x \in X\}$
2. The dual cone of X is $X^* := -X^\circ = \{y \in \mathbb{R}^n : y^\top x \ge 0 \quad \forall x \in X\}$
3. A cone K is called self-dual if $K^* = K$

Remarks

- X° is the set of points "most opposite to/away from" X
- X^* is the set of points "most aligned with/cloest to" X
- Dual cone K^* is used to define the dual problem of a conic program where the nonlinear constraint is $x \in K$ for a closed convex cone K

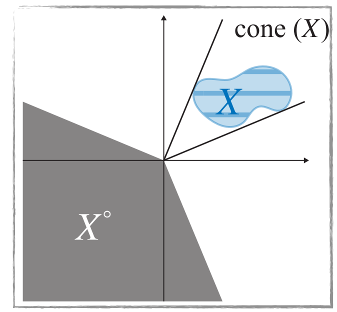
Polar cone

Examples



(a) Polar cones X° of X

(b) Dual cones $X^* = -X^\circ$ of X



$0 \notin X$

$X^\circ = (\text{cone}(X))^\circ$

Polar cone

Proposition

Let $X \subseteq \mathbb{R}^n$ be a nonempty set

1. X° is a closed convex cone
2. $X^\circ = [\mathrm{cl}(X)]^\circ = [\mathrm{conv}(X)]^\circ = [\mathrm{cone}(X)]^\circ$
3. If $X \subseteq Y$ then $Y^\circ \subseteq X^\circ$
4. If X is a cone then $(X^\circ)^\circ = \mathrm{cl}(\mathrm{conv}(X))$

Normal cone and tangent cone

Let $\bar{x} \in X \subseteq \mathbb{R}^n$. The feasible direction cone of X at $\bar{x}$ is

$$\text{cone}(X - \bar{x}) \;:=\; \left\{ \sum_{i=1}^{m} \alpha_i (x_i - \bar{x}) : x_i \in X,\ \alpha_i \geq 0,\ \text{integers } m > 0 \right\}$$

Remarks

- $\text{cone}(X - \bar{x})$ is set of directions and their convex combinations along which an infinitesimal step from $\bar{x}$ will stay in X
- $\text{cone}(X - \bar{x})$ is closed if and only if X is closed
- $\text{cl}(\text{cone}(X - \bar{x}))$ is a linear approximation of X at $\bar{x} \in X$: it is the smallest closed convex cone containing all feasible directions $x - \bar{x}$ at $\bar{x}$
- The feasible direction cone is sometimes defined as $\text{cone}(X - \bar{x}) := \{\gamma(x - \bar{x}) : x \in X,\ \gamma \geq 0\}$

Normal cone and tangent cone

Definition

Let $X \subseteq \mathbb{R}^n$ be a nonempty set and $\bar{x} \in X$

1. The tangent cone of X at $\bar{x}$ is $T_X(\bar{x}) := \text{cl}(\text{cone}(X - \bar{x}))$
2. The normal cone of X at $\bar{x}$ is

$$N_X(\bar{x}) := (\text{cone}(X - \bar{x}))^\circ = (X - \bar{x})^\circ = \{y \in \mathbb{R}^n : y^\top(x - \bar{x}) \leq 0 \quad \forall x \in X\}$$

Proposition

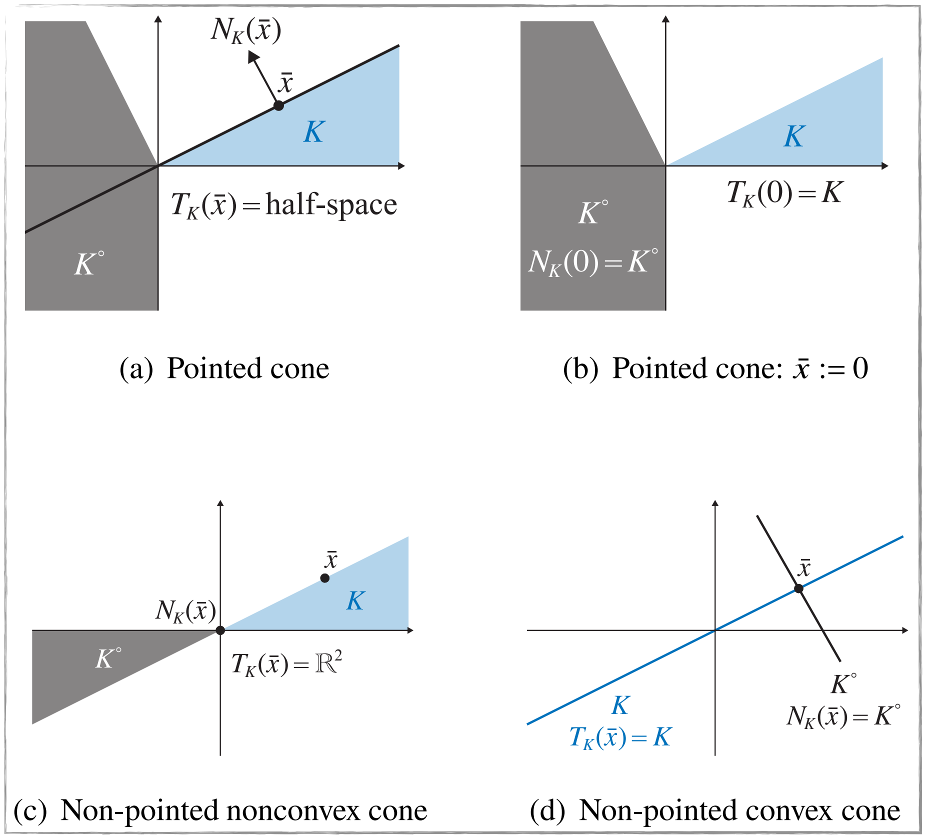
Let $X \subseteq \mathbb{R}^n$ be a nonempty set and $\bar{x} \in X$

1. $X^\circ, X^*, T_X(\bar{x}), N_X(\bar{x})$ are closed convex cones
2. $(T_X(\bar{x}))^\circ = N_X(\bar{x})$ and $T_X(\bar{x}) = (N_X(\bar{x}))^\circ$
3. If $\bar{x} \in \text{int}(X)$ then $N_X(\bar{x}) = \{0\}$ and $T_X(\bar{x}) = \mathbb{R}^n$

If $\bar{x} \in \text{ri}(X)$, then $N_X(\bar{x}) \supsetneq \{0\}$ in general

Normal cone and tangent cone

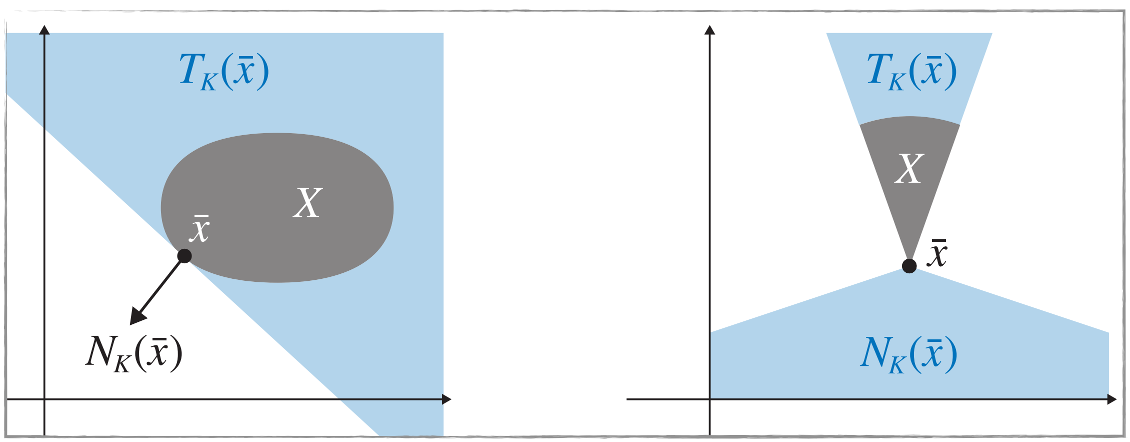
Examples



- While X° is a set, $N_K(\bar{x})$ and $T_X(\bar{x})$ are (set-valued) functions of $\bar{x}$, i.e., they depend on $\bar{x}$
- If $0 \in X$, then $X^\circ = N_X(0)$ and $T_X(\bar{x}) = (X^\circ)^\circ = \text{cl}(\text{conv}(X))$

Linear approximation & optimality

- $T_X(\bar{x})$ is a linear approximation of X at $\bar{x} \in X$
  - Smallest closed convex cone containing feasible directions at $\bar{x}$
- $N_X(\bar{x})$ is "most opposite" to $T_X(\bar{x})$ at $\bar{x}$
  - If X is smooth at $\bar{x}$, then $T_X(\bar{x})$ is a halfspace (supporting hyperplane) and $N_X(\bar{x})$ is singleton
- Optimality condition: x^* is optimal if direction of cost reduction at x^* aligns with $N_X(x^*)$, i.e., $-\nabla f(x^*) \in N_X(x^*)$
  - In smooth setting (KKT): $-\nabla f(x^*) = \nabla g(x^*)\lambda^* + \nabla h(x^*)\mu^*$



Normal cones

hyperplane:	$H_1 := \{x \in \mathbb{R}^n : Ax = b\}$
polyhedron:	$H_2 := \{x \in \mathbb{R}^n : Ax \leq b\}$
nonnegative cone:	$K_+ := \{x \in \mathbb{R}^n : x \geq 0\}$
convex cone:	$K \subseteq \mathbb{R}^n$

Theorem

1. $N_{H_1}(\bar{x}) \;=\;$ range$(A^\mathsf{T}) \;=\; \{A^\mathsf{T}\lambda \in \mathbb{R}^n : \lambda \in \mathbb{R}^m\}$
2. $N_{H_2}(\bar{x}) \;=\;$ cone $\left(A_I^\mathsf{T}\right) \;=\; \{A^\mathsf{T}\lambda \in \mathbb{R}^n : \lambda \in \mathbb{R}^m_+,\ \lambda^\mathsf{T}(A\bar{x} - b)\}$ $\quad I := I(\bar{x}) := \{i : a_i^\mathsf{T}\bar{x} = b_i\}$
3. $N_{K_+}(\bar{x}) \;=\; \{y \in \mathbb{R}^n : y \leq 0,\ y^\mathsf{T}\bar{x} = 0\}$
4. $N_K(\bar{x}) \;=\; \{y \in K^\circ : y^\mathsf{T}\bar{x} = 0\}$

Derivation of $N_{H_2}(\bar{x})$ uses Farkas Lemma (or Separating Hyperplane Thm)

Normal cone $N_C(\bar{x})$

convex set (non-polyhedral): $C := \{x \in \mathbb{R}^n : h(x) \leq 0\}$ with convex h

- Farkas Lemma-type proof for $N_{H_2}(\bar{x})$ inadequate
  - Due to second-order term in Taylor expansion of $h(x)$ around $\bar{x}$
- Need constraint qualification

LICQ (linear independence CQ):

columns of $\nabla h_I(\bar{x}) \in \mathbb{R}^{n \times |I|}$ are linearly independent

where $I := I(\bar{x}) := \{i : h_i(\bar{x}) = 0\}$ is the set of active constraints

Normal cone $N_C(\bar{x})$

convex set (non-polyhedral): $C := \{x \in \mathbb{R}^n : h(x) \leq 0\}$ with convex h

Let $I := I(\bar{x}) := \{i : h_i(\bar{x}) = 0\}$

Theorem

Suppose $h : \mathbb{R}^n \to \mathbb{R}^m$ is real-valued twice continuously differentiable function that is convex on $\mathbb{R}^n$. If $\bar{x} \in C$ satisfies LICQ, then

1. $N_C(\bar{x}) \;=\; \text{cone}\left(\nabla h_I(\bar{x})\right) \;=\; \{\nabla h(\bar{x})\lambda \in \mathbb{R}^n : \lambda \in \mathbb{R}^m_+, \lambda^\top h(\bar{x}) = 0\}$
2. For every $y \in N_C(\bar{x})$, there exists unique $\lambda_I \in \mathbb{R}^{|I|}_+$ such that $y = \nabla h_I(\bar{x})\lambda_I$

Remarks

- Constraint qualification is sufficient, but not necessary, for existence of λ
- LICQ in constrained optimization ensures existence and uniqueness of dual optimal solution
- Proof uses: Farkas Lemma (or Separating Hyperplane Thm), LP duality, LICQ

Normal cone $N_X(\bar{x})$

equality constrained: $X := \{x \in \mathbb{R}^n : g(x) = 0\}$ with convex g

Note: X is nonconvex unless g is affine

Theorem

Suppose $g : \mathbb{R}^n \to \mathbb{R}^m$ is real-valued twice continuously differentiable function that is convex on $\mathbb{R}^n$. If $\bar{x} \in X$ satisfies LICQ, then

1. $N_X(\bar{x}) = \text{range}\left(\nabla g(\bar{x})\right) = \{\nabla g(\bar{x})\lambda \in \mathbb{R}^n : \lambda \in \mathbb{R}^m\}$
2. For every $y \in N_X(\bar{x})$, there exists unique λ such that $y = \nabla g(\bar{x})\lambda$

Remark

- Proof: write $g(x) = 0$ as $g(x) \leq 0, -g(x) \leq 0$, and use previous theorem

Set intersection
Multiple constraints

Let $C_i \subseteq \mathbb{R}^n$ be

- polyhedral sets $C_i := \{x : A_i x \leq b_i\}$ for $i = 1,\ldots,\bar{m}$
- convex sets (e.g. $C_i := \{x : h(x) \leq 0\}$ for convex h) for $i = \bar{m}+1,\ldots,m$

Let $C := \bigcap_{i=1}^{m} C_i$

Theorem

If Slater-type CQ: $\left(\cap_{i=1}^{\bar{m}} C_i\right) \bigcap \left(\cap_{i=\bar{m}+1}^{m} \mathrm{ri}(C_i)\right) \neq \emptyset$ then

$$N_C(\bar{x}) = \sum_i N_{C_i}(\bar{x}), \qquad \forall \bar{x} \in C$$

Summary

Set $X \subseteq \mathbb{R}^n$	Normal cone $N_X(\bar{x}) \subseteq \mathbb{R}^n$
$\{x : Ax = b\}$	$\text{range}(A^\top) := \{A^\top \lambda : \lambda \in \mathbb{R}^m\}$
$\{x : \text{convex } h(x) = 0\}$	$\text{range}(\nabla h(\bar{x})) := \{\nabla h(\bar{x})\lambda : \lambda \in \mathbb{R}^m\}$
$\{x : Ax \leq b\}$	$\text{cone}\begin{pmatrix} A_I^\top \end{pmatrix} = \{A^\top \lambda \in \mathbb{R}^n : \lambda \in \mathbb{R}_+^m,\ \lambda^\top(A\bar{x} - b)\}$
$\{x : \text{convex } h(x) \leq 0\}$	$\text{cone}(\nabla h_I(\bar{x})) := \{\nabla h(\bar{x})\lambda : \lambda \in \mathbb{R}_+^m,\ \lambda^\top h(\bar{x}) = 0\}$
cone $\{x : x \geq 0\}$	$\{y \leq 0 : y^\top \bar{x} = 0\}$
cone K	$\{y \in K^\circ : y^\top \bar{x} = 0\}$

$\{x : h(x) = 0\}$ may be nonconvex

- KKT conditions in (smooth) convex optimization most encountered in applications are consequence of these results (more later): $-\nabla f(x^*) \in N_X(x^*)$
- Multiple constraints: $-\nabla f(x^*) \in \sum_i N_{C_i}(x^*) = \nabla g(x^*)\lambda^* + \nabla h(x^*)\mu^*$

Outline

1. Normal cones of feasible sets
    - Polar cone
    - Normal cone and tangent cone
    - Affine transformation
    - Second-order cones and SOC constraints
    - Proofs
2. CPC functions
3. Gradient and subgradient
4. Characterization: saddle point
5. Characterization: KKT
6. Existence: primal optimum
7. Existence: dual optimum and strong duality
8. Special convex programs

Linear transformation

Image

Given a nonempty set $X \subseteq \mathbb{R}^n$, its image under $A \in \mathbb{R}^{m \times n}$ is

$$Y \ := \ AX \ := \ \{Ax \in \mathbb{R}^m : x \in X\}$$

Let $\bar{x} \in X$ and $\bar{y} = A\bar{x} \in Y$

Theorem

1. The normal cone $N_Y(\bar{y})$ is pre-image of $N_X(\bar{x})$ under $A^\top$:

   $$N_Y(\bar{y}) \ = \ \{y \in \mathbb{R}^m : A^\top y \in N_X(\bar{x})\}$$

   Hence $A^\top N_Y(\bar{y}) \subseteq N_X(\bar{x})$

2. If $\text{rank}(A) = n$ (full column rank), then $A^\top N_Y(\bar{y}) = N_X(\bar{x})$

This result is used to derive normal cone of **rotated** second-order cone from that of standard SoC

Linear transformation

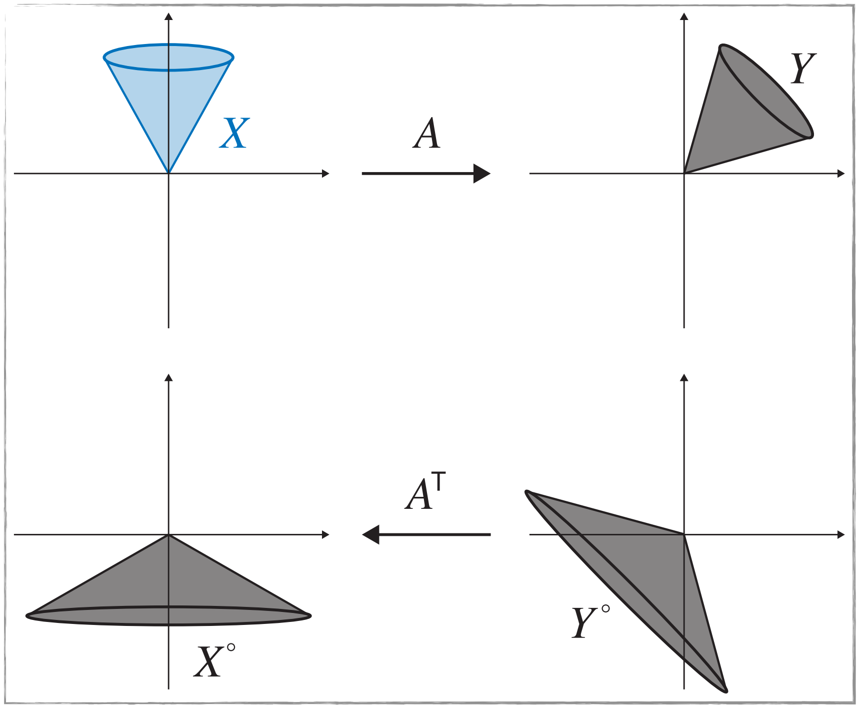
Image

Given a nonempty set $X \subseteq \mathbb{R}^n$, its image under $A \in \mathbb{R}^{m \times n}$ is

$$Y \;:=\; AX \;:=\; \{Ax \in \mathbb{R}^m : x \in X\}$$

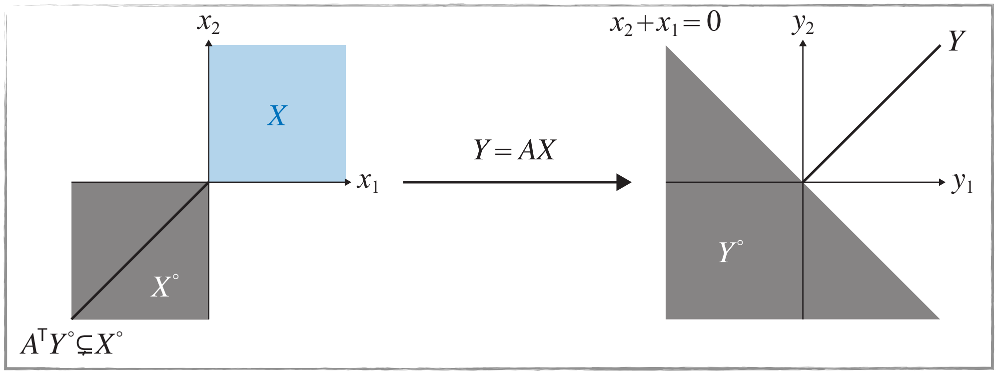
Let $\bar{x} \in X$ and $\bar{y} = A\bar{x} \in Y$

Example: $A^\top N_Y(\bar{y}) = N_X(\bar{x})$



Linear transformation

Example: $A^\top Y^\circ \subsetneq X^\circ$



Given X, $Y = AX$ is image

- $N_X(0) = X^\circ$ and $N_Y(0) = Y^\circ$
- A singular: $A^\top Y^\circ \subsetneq X^\circ$

Linear transformation

Pre-image

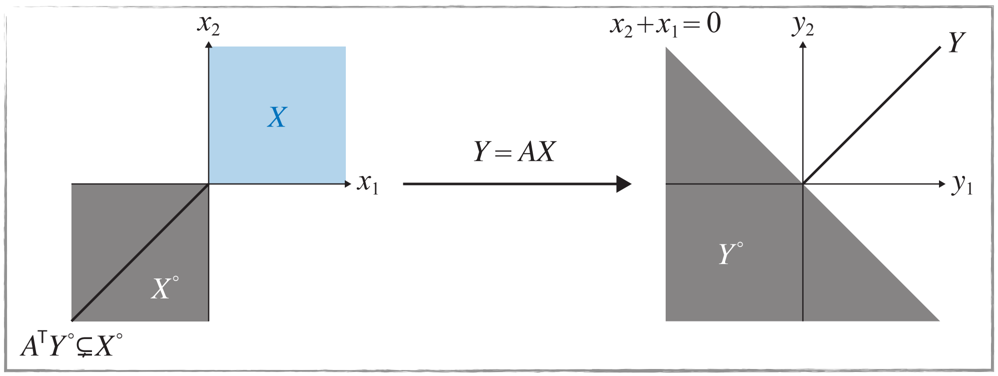
Given a nonempty set $Y \subseteq \mathbb{R}^m$, its pre-image under $A \in \mathbb{R}^{m \times n}$ is

$$X \; := \; \{x \in \mathbb{R}^n : Ax \in Y\}$$

Let $\bar{x} \in X$ and $\bar{y} = A\bar{x} \in Y$

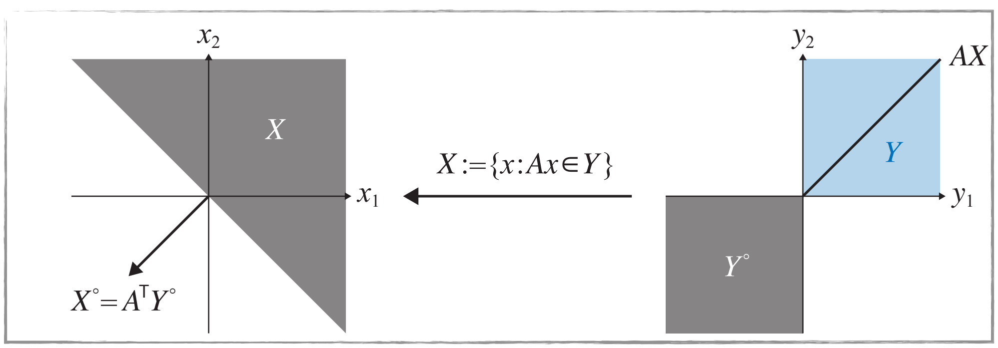
Linear transformation

Image vs pre-image



Given X, $Y = AX$ is image

- $N_X(0) = X^\circ$ and $N_Y(0) = Y^\circ$
- A singular: $A^\top Y^\circ \subsetneq X^\circ$



Given Y, $X := \{x : Ax \in Y\}$ is pre-image

- $N_X(0) = X^\circ$ and $N_Y(0) = Y^\circ$
- $A^\top Y^\circ = X^\circ$ (despite singular A)
- $X \supset$ null$(A) = \{x : x_1 + x_2 = 0\}$

Linear transformation

Pre-image

Given a nonempty set $Y \subseteq \mathbb{R}^m$, its pre-image under $A \in \mathbb{R}^{m \times n}$ is

$$X \ := \ \{x \in \mathbb{R}^n : Ax \in Y\}$$

Let $\bar{x} \in X$ and $\bar{y} = A\bar{x} \in Y$

Theorem

$$A^\top N_Y(\bar{y}) = N_X(\bar{x})$$

Affine transformation

Image

Affine transformation: $f(x) = Ax + b$

Given a nonempty set $X \subseteq \mathbb{R}^n$, its image under f is

$$Y_b \ := \ AX + b \ \subseteq \mathbb{R}^m$$

Let $\bar{y}_b = A\bar{x} + b \in Y$

Theorem

1. The normal cone $N_{Y_b}(\bar{y}_b)$ is pre-image of $N_X(\bar{x})$ under $A^\top$ and is independent of b:

   $$N_{Y_b}(\bar{y}_b) \ = \ N_{AX}(A\bar{x}) \ = \ \{y \in \mathbb{R}^m : A^\top y \in N_X(\bar{x})\}$$

   Hence $A^\top N_{Y_b}(\bar{y}_b) \subseteq N_X(\bar{x})$

2. If $\text{rank}(A) = n$ (full column rank), then $A^\top N_{Y_b}(\bar{y}_b) = N_X(\bar{x})$

Affine transformation

Pre-image

Given a nonempty set $Y \subseteq \mathbb{R}^m$, its pre-image under f is

$$X_b \;:=\; \{x \in \mathbb{R}^n : Ax + b \in Y\}$$

Let $\bar{x} \in X_b$ and $\bar{y}_b = A\bar{x} + b \in Y$

Theorem

$$A^\top N_Y(\bar{y}_b) = N_{X_b}(\bar{x})$$

Outline

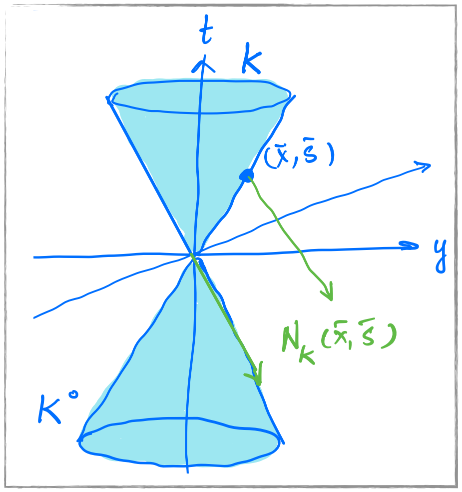
1. Normal cones of feasible sets
    - Polar cone
    - Normal cone and tangent cone
    - Affine transformation
    - Second-order cones and SOC constraints
    - Proofs
2. CPC functions
3. Gradient and subgradient
4. Characterization: saddle point
5. Characterization: KKT
6. Existence: primal optimum
7. Existence: dual optimum and strong duality
8. Special convex programs

Second-order cone

Standard K

Standard second-order cone is

$$K := \{(x, s) \in \mathbb{R}^{n+1} : \|x\|_2 \leq s\}$$



Theorem

1. K is closed convex cone
2. Polar cone is $K^\circ = \{(y, t) \in \mathbb{R}^{n+1} : \|y\|_2 \leq -t\}$
3. Normal cone is

$$N_K(\bar{x}, \bar{s}) = \begin{cases} K^\circ & \text{if } (\bar{x}, \bar{s}) = (0,0) \\ \{(0,0) \in \mathbb{R}^{n+1}\} & \text{if } \|\bar{x}\|_2 < \bar{s} \\ \left\{\mu(\bar{x}, -\bar{s}) \in \mathbb{R}^{n+1} : \mu \geq 0\right\} & \text{if } \|\bar{x}\|_2 = \bar{s} > 0 \end{cases}$$

Second-order cone

Rotated K_r

Rotated second-order cone is

$$K_r := \{x \in \mathbb{R}^{n+2} : \|x^n\|_2 \leq x_{n+1}x_{n+2},\ x_{n+1} \geq 0,\ x_{n+2} \geq 0\}$$

where $x^m := (x_1, \ldots, x_m)$

Theorem

1. K_r is closed convex cone
2. $K = AK_r$ where $A = \begin{bmatrix} 2\mathbb{I}_n & 0_n & 0_n \\ 0_n^\top & 1 & -1 \\ 0_n^\top & 1 & 1 \end{bmatrix}$

Second-order cone

Rotated K_r

Rotated second-order cone is

$$K_r := \{x \in \mathbb{R}^{n+2} : \|x^n\|_2 \le x_{n+1}x_{n+2},\ x_{n+1} \ge 0,\ x_{n+2} \ge 0\}$$

where $x^m := (x_1, \ldots, x_m)$

Theorem

3. Polar cone is $K_r^\circ = A^\top K^\circ = \{A^\top x \in \mathbb{R}^{n+2} : \|x^{n+1}\|_2 \le -x_{n+2}\}$
4. Normal cone is $N_{K_r}(\bar{x}) = A^\top N_K(A\bar{x})$ is

$$N_{K_r}(\bar{x}) = \begin{cases} A^\top K^\circ & \text{if } A\bar{x} = 0 \\ \{(0,0) \in \mathbb{R}^{n+2}\} & \text{if } \|[A\bar{x}]^{n+1}\|_2 < [A\bar{x}]_{n+2} \\ \left\{\mu\left([A\bar{x}]^{n+1}, -[A\bar{x}]_{n+2}\right) \in \mathbb{R}^{n+2} : \mu \ge 0\right\} & \text{if } \|[A\bar{x}]^{n+1}\|_2 = [A\bar{x}]_{n+2} > 0 \end{cases}$$

SOC constraint

Convex set C defined by SOC constraint:

$$C := \{x \in \mathbb{R}^n : (Ax + b, c^\top x + d) \in K\} = \{x \in \mathbb{R}^n : \|Ax + b\|_2 \leq c^\top x + d\}$$

Hence C is pre-image of standard second-order cone K under affine transformation:

$$C := \{x \in \mathbb{R}^n : \tilde{A}x + \tilde{b} \in K\} \quad \text{where} \quad \tilde{A} \; := \; \begin{bmatrix} A \\ c^\top \end{bmatrix}, \;\; \tilde{b} \; := \; \begin{bmatrix} b \\ d \end{bmatrix}$$

- $C = K$ standard second-order cone if $A = \begin{bmatrix} \mathbb{I}_{n-1} & 0 \\ 0 & 0 \end{bmatrix}$, $c = e_n$, $b = 0$, $d = 0$
- C may not be a cone, e.g., C is a hyperplane if $A = 0$

Theorem

$N_C(\bar{x}) = \tilde{A}^\top N_K(\bar{y})$ where K is standard second-order cone

Summary

First-order optimality condition for convex opt:

x^* is minimizer $\Leftrightarrow -\nabla f(x^*)$ points away from linear approximation of feasible set at x^*

(in smooth setting: $-\nabla f(x^*) = \nabla g(x^*)\lambda^* + \nabla h(x^*)\mu^*$)

Generalization to nonsmooth setting:

1. Linear approximation of feasible set
    - Tangent cone $T_X(x^*)$
2. Vectors most opposite to linear approximation
    - Normal cone $N_X(x^*)$
3. We have derived $N_X(\bar{x})$ for:
    - Feasible sets widely encountered in applications (hyperplanes, polyhedrons, cones, convex sets)
    - Affine transformation of these basic sets
    - Second-order cones, SOC constraints

Proofs

Normal cone $N_C(\bar{x})$

convex set (non-polyhedral): $C := \{x \in \mathbb{R}^n : h(x) \leq 0\}$ with convex h

Let $I := I(\bar{x}) := \{i : h_i(\bar{x}) = 0\}$

Theorem

Suppose $h : \mathbb{R}^n \to \mathbb{R}^m$ is real-valued twice continuously differentiable function that is convex on $\mathbb{R}^n$. If $\bar{x} \in C$ satisfies LICQ, then

1. $N_C(\bar{x}) \;=\;$ cone $\left(\nabla h_I(\bar{x})\right) \;=\; \{\nabla h(\bar{x})\lambda \in \mathbb{R}^n : \lambda \in \mathbb{R}^m_+, \lambda^\top h(\bar{x}) = 0\}$
2. For every $y \in N_C(\bar{x})$, there exists unique $\lambda_I \in \mathbb{R}^{|I|}_+$ such that $y = \nabla h_I(\bar{x})\lambda_I$

Proof uses: Farkas Lemma (or Separating Hyperplane Thm), LP duality, LICQ

Normal cone $N_C(\bar{x})$
Proof

Let

$$N_C(\bar{x}) := \{y \in \mathbb{R}^n : y^\mathsf{T}(x - \bar{x}) \le 0 \quad \forall x \text{ s.t. } h(x) \le 0\}$$

$$Y(\bar{x}) := \text{cone}\left(\nabla h_I(\bar{x})\right) = \{\nabla h(\bar{x})\lambda \in \mathbb{R}^n : \lambda \in \mathbb{R}^m_+, \lambda^\mathsf{T} h(\bar{x}) = 0\}$$

Suppose $y \in Y(\bar{x})$: easy (use convexity of h)

Suppose $y \in N_C(\bar{x})$: If $I(\bar{x}) = \emptyset$, then $\bar{x} \in \text{int}(C)$ and hence $N_C(\bar{x}) = Y(\bar{x}) = \{0\}$.

Hence suppose $I(\bar{x}) \neq \emptyset$ but $y \notin Y(\bar{x})$.

Will construct $x(t) := \bar{x} + t\Delta x$ such that

- $h(x(t)) \le 0$, $ty^\mathsf{T}\Delta x > 0$
- contradicting $y \in N_C(\bar{x})$, proving $y \in Y(\bar{x})$

in 3 steps

Normal cone $N_C(\bar{x})$
Proof

Step 1:

Farkas lemma implies that, if $y \notin Y(\bar{x})$ convex cone, then there exists nonzero $c \in \mathbb{R}^n$ with

$$c^\top \nabla h_i(\bar{x}) \ \le \ 0 \ < \ c^\top y, \qquad \forall i \in I$$

Step 2:

For each $i = 1, \ldots, m$,

$$h_i(x(t)) \ = \ h_i(\bar{x} + t\Delta x) \ = \ h_i(\bar{x}) \ + \ t\frac{\partial h_i}{\partial x}(\bar{x})\Delta x \ + \ \frac{t^2}{2}\Delta x^\top \frac{\partial^2 h_i}{\partial x^2}(x(s_i))\Delta x$$

for some $s_i \in [0,t]$. The second-order term can be upper bounded by

$$\alpha_i(\bar{x}, \Delta x) \ := \ \max_{s_i \in [0,1]} \Delta x^\top \frac{\partial^2 h_i}{\partial x^2}(x(s_i))\Delta x$$

which is finite (because h_i are twice continuously differentiable and s_i is in $[0,1]$), and independent of t.

Hence

$$h_i(x(t)) \ \le \ h_i(\bar{x}) + t\left(\frac{\partial h_i}{\partial x}(\bar{x})\Delta x + \frac{t}{2}\alpha_i(\bar{x}, \Delta x)\right) \qquad \text{for } t \in [0,1]$$

Normal cone $N_C(\bar{x})$

Proof

Step 2:

Hence we need to find Δx such that

1. $\frac{\partial h_i}{\partial x}(\bar{x})\Delta x < 0$ for all $i \in I$
2. $y^\mathsf{T}(x(t) - \bar{x}) \;=\; t y^\mathsf{T} \Delta x \;>\; 0$

Then 1 ⇒ there exists small enough $t > 0$ s.t.

- $i \in I: \;\; \frac{\partial h_i}{\partial x}(\bar{x})\Delta x + \frac{t}{2}\alpha_i(\bar{x}, \Delta x) \;\leq\; 0$
- $i \notin I: \;\; h_i(\bar{x}) + t\left(\frac{\partial h_i}{\partial x}(\bar{x})\Delta x + \frac{t}{2}\alpha_i(\bar{x}, \Delta x)\right) \;\leq\; 0$

Hence, for each $i = 1,\ldots,m$,

$$h_i(x(t)) \;\leq\; h_i(\bar{x}) + t\left(\frac{\partial h_i}{\partial x}(\bar{x})\Delta x + \frac{t}{2}\alpha_i(\bar{x}, \Delta x)\right) \;\leq\; 0$$

1 and 2 ⇒ $y \notin N_C(\bar{x})$, a contradiction

Normal cone $N_C(\bar{x})$

Proof

Step 3:

Consider LP ($\epsilon > 0$ to be chosen):

$$
\begin{aligned}
z^*(\epsilon) \quad &:= \quad \min_{(\Delta x, z) \in \mathbb{R}^{n+1}} \quad z \\
&\qquad \text{s.t.} \qquad \frac{\partial h_i}{\partial x}(\bar{x})\Delta x \leq z, \quad i = 1, \ldots, m \\
&\qquad \qquad \qquad y^\top \Delta x \geq \epsilon
\end{aligned}
$$

Then Δx^* satisfies 1 and 2 if and only if for some $\epsilon > 0$, $(\Delta x^*, z^*(\epsilon))$ is optimal for LP with $z^*(\epsilon) < 0$

Normal cone $N_C(\bar{x})$

Proof

Step 3:

- LP is feasible for sufficiently small $\epsilon > 0$, because

$$\Delta x \; := \; c, \qquad z \; := \; \max_{i \in I} \frac{\partial h_i}{\partial x}(\bar{x})c$$

  is a feasible point

- Its dual is infeasible, because

$$d^*(\epsilon) \quad := \quad \max_{(\lambda, \mu) \geq 0} \; \epsilon\mu$$

$$\text{s.t.} \quad \mathbf{1}^\mathsf{T}\lambda = 1, \quad \nabla h_I(\bar{x})\lambda = \mu y$$

  Suppose $(\lambda, \mu) \geq 0$ is feasible. Then $\lambda \neq 0$.

  LICQ ($\nabla h_I(\bar{x})$ has linearly independent columns) implies $\mu > 0$. Hence

$$y \; = \; \sum_{i \in I} \frac{\lambda_i}{\mu} \nabla h_i(\bar{x}), \quad \text{contradicting } y \notin Y(\bar{x})$$

Normal cone $N_C(\bar{x})$
Proof

Step 3:

- Feasible primal, infeasible dual, LP duality $\Rightarrow z^*(\epsilon) = d^*(\epsilon) = -\infty$.
- Therefore there exists finite Δx that satisfies 1 and 2

This completes the construction of $x(t) := \bar{x} + t\Delta x$ such that

- $h(x(t)) \leq 0$, $ty^\mathsf{T}\Delta x > 0$
- contradicting $y \in N_C(\bar{x})$, proving $y \in Y(\bar{x})$

Linear transformation

Pre-image

Given a nonempty set $Y \subseteq \mathbb{R}^m$, its pre-image under $A \in \mathbb{R}^{m \times n}$ is

$$X \;:=\; \{x \in \mathbb{R}^n : Ax \in Y\}$$

Let $\bar{x} \in X$ and $\bar{y} = A\bar{x} \in Y$

Theorem

$$A^\top N_Y(\bar{y}) = N_X(\bar{x})$$

Pre-image: $A^\top N_Y(\bar{y}) = N_X(\bar{x})$

Proof: SVD decomposition of $\mathbb{R}^n$

Singular value decomposition of A with rank$(A) = r$ is:

$$A \;=\; V\Sigma W^\top \;=\; V_r\Sigma_r W_r^\top$$

where

$$V = \begin{bmatrix} V_r & V_{m-r} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}, \quad W = \begin{bmatrix} W_r & W_{n-r} \end{bmatrix}$$

with range(W_r) = range$(A^\top)$, range(W_{n-r}) = null(A)

Pseudo-inverse of A is:

$$A^\dagger \;=\; W\Sigma^\dagger V^\top \;=\; W_r\Sigma_r^{-1}V_r^\top$$

Hence $A^\dagger A = W_r W_r^\top$

Pre-image: $A^\top N_Y(\bar{y}) = N_X(\bar{x})$

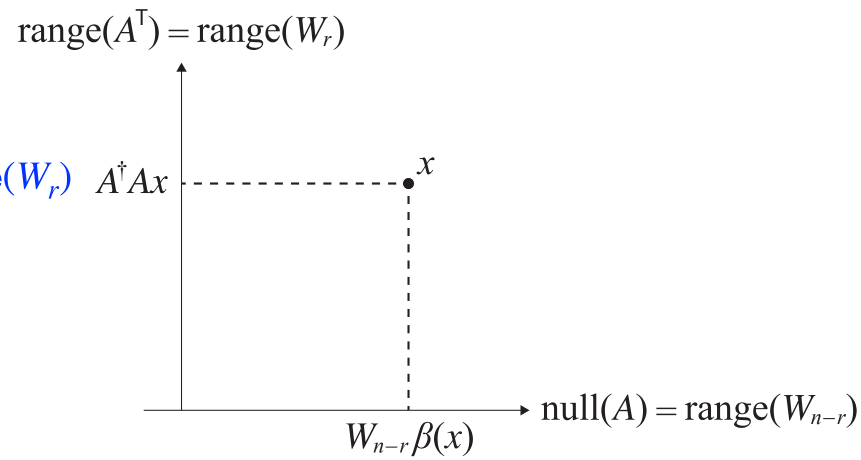
Proof: SVD decomposition of $\mathbb{R}^n$

Since columns of W form an orthonormal basis of $\mathbb{R}^n$, every $x \in \mathbb{R}^n$ can be written uniquely in terms of W_r and W_{n-r} as $x = W_r\left(W_r^\top x\right) + W_{n-r}\left(W_{n-r}^\top x\right)$:

$$x = \underbrace{W_r\left(W_r^\top x\right)}_{A^\dagger A} + W_{n-r}\underbrace{\left(W_{n-r}^\top x\right)}_{\beta(x)} = A^\dagger y + W_{n-r}\beta(x)$$

i.e., $\mathbb{R}^n$ can be decomposed uniquely into 2 orthogonal components in terms of W_r and W_{n-r} :

$A^\dagger A$: project x onto range(W_r)



Pre-image: $A^\top N_Y(\bar{y}) = N_X(\bar{x})$

Proof: SVD decomposition of $\mathbb{R}^n$

Hence any $x \in \mathbb{R}^n$ can be written as $x = A^\dagger y + W_{n-r}\beta$ for some $y = Ax$ and β (dependent on x)

Given $X \subseteq \mathbb{R}^n$: let its image under A be $Y := AX$. Then

$$X = A^\dagger Y + W_{n-r} B(X)$$

with $B(X) := \left\{ W_{n-r}^\top x : x \in X \right\} \subseteq \mathbb{R}^{n-r}$

Given $Y \subseteq \mathbb{R}^m$: let its pre-image under A be $X := \{x : Ax \in Y\}$. Then

$$X = A^\dagger Y + W_{n-r} \mathbb{R}^{n-r}$$

i.e., the pre-image of each $y \in Y$ consists of $A^\dagger y$ plus the entire null(A)

This is the key difference between image and pre-image under A

Pre-image: $A^\top N_Y(\bar{y}) = N_X(\bar{x})$

Proof

Suppose $\tilde{y} \in N_Y(\bar{y})$, i.e., $\tilde{y}^\top(y - \bar{y}) \le 0$ for all $y = Ax \in AX \subseteq Y$.

Then $\tilde{y}^\top A(x - \bar{x}) \le 0$ for all $x \in X$, i.e., $A^\top \tilde{y} \in N_X(\bar{x})$.

This shows $A^\top N_Y(\bar{y}) \subseteq N_X(\bar{x})$

Conversely, suppose $\tilde{x} \in N_X(\bar{x})$, i.e., $\tilde{x}^\top(x - \bar{x}) \le 0$ for all $x \in X$

Write

$$x = A^\dagger y + W_{n-r}\beta \qquad \text{for some } y \in Y, \beta \text{ (dependent on } x)$$

$$\bar{x} = A^\dagger \bar{y} + W_{n-r}\bar{\beta}$$

Then

$$\tilde{x}^\top A^\dagger(y - \bar{y}) + \tilde{x}^\top W_{n-r}(\beta - \bar{\beta}) \le 0, \quad \forall y \in Y, \forall \beta \in \mathbb{R}^{n-r}$$

Pre-image: $A^\top N_Y(\bar{y}) = N_X(\bar{x})$

Proof

Since this holds for all $y \in Y$, $\beta \in \mathbb{R}^{n-r}$, we must have (take $y = \bar{y}$ and $\beta = \bar{\beta}$):

$$
\begin{aligned}
\tilde{x}^\top A^\dagger (y - \bar{y}) \leq\ & 0 \qquad \forall y \in Y \\
\tilde{x}^\top W_{n-r}(\beta - \bar{\beta}) \leq\ & 0 \qquad \forall \beta \in \mathbb{R}^{n-r}
\end{aligned}
$$

Taking $\beta = \bar{\beta} \pm e_j$ implies $\tilde{x}^\top W_{n-r} = 0$, and hence $\tilde{x} \in$ range(W_r)

1st inequality implies $(A^\dagger)^\top \tilde{x} = \tilde{y}$ for some $\tilde{y} \in N_Y(\bar{y})$.

Multiplying both sides by $A^\top$ gives:

$$(A^\dagger A)^\top \tilde{x} \ = \ (A^\dagger A)\tilde{x} \ = \ \tilde{x} \ = \ A^\top \tilde{y}$$

where 1st equality follows because $A^\dagger A = W_r W_r^\top$ is symmetric, and the 2nd equality follows because $\tilde{x} \in$ range(W_r) and hence $W_r W_r^\top \tilde{x} = \tilde{x}$

Hence $N_X(\bar{x}) \subseteq A^\top N_Y(\bar{y})$

Outline

1. Normal cones of feasible sets
2. CPC functions
   - Extended real-valued functions
   - Indicator function, support function, polyhedral function
3. Gradient and subgradient
4. Characterization: saddle point
5. Characterization: generalized KKT
6. Existence: primal optimum
7. Existence: dual optimum and strong duality
8. Special convex programs

Extended real-valued function

A real-valued function $f : X \to \mathbb{R}$ with $X \subseteq \mathbb{R}^n$ maps a finite vector $x \in X$ to a finite value $f(x) \in \mathbb{R}$

An extended real-valued function $f : X \to [-\infty, \infty]$ can take a finite value in $\mathbb{R}$ or $\pm\infty$

- X : domain of f
- $\text{dom}(f) := \{x \in X : f(x) < \infty\}$: effective domain of f
- $\text{epi}(f) := \{(x, y) \in X \times \mathbb{R} : y \geq f(x)\} \subseteq R^{n+1}$: epigraph of f

Remarks

- If $(x, y) \in \text{epi}(f)$, then $y \notin \{-\infty, \infty\}$
- $x \in \text{dom}(f) \iff \exists y \in \mathbb{R} \text{ s.t. } (x, y) \in \text{epi}(f)$, i.e., $\text{dom}(f)$ is projection of $\text{epi}(f)$ onto $\mathbb{R}^n$

Extended real-valued function

Continuity

An extended real-valued function $f : X \to [-\infty, \infty]$ is lower semicontinuous (lsc) at $x \in X$ if

$$f(x) \leq \liminf_{k} f(x_k)$$

for every sequence $\{x_k\} \subseteq X$ with $x_k \to x$

f is lsc (on X) if it is lsc at every $x \in X$

f is upper semicontinuous (usc) if $-f$ is lsc

f is continuous if and only if it is both lsc and usc

Extended real-valued function

CPC functions

Definition

Consider $f : X \to [-\infty, \infty]$ with $X \subseteq \mathbb{R}^n$

1. f is closed if epi(f) is a closed set in $\mathbb{R}^{n+1}$
2. f is proper if $f(x) > -\infty$ for all $x \in X$ and $\exists \bar{x} \in X$ such that $f(\bar{x}) < \infty$ (so that epi$(f) \neq \emptyset$). In particular a real-valued function $f : X \to \mathbb{R}$ is proper
3. Suppose X is convex. Then f is convex if epi(f) is a convex set in $\mathbb{R}^{n+1}$

Remarks

- Convexity definition in terms of epi(f) reduces to usual definition for real-valued functions
- If a closed convex function is not proper, then $f(x) = -\infty$ if $x \in$ dom(f) and $f(x) = \infty$ if $x \notin$ dom(f). We therefore only consider proper functions with $f : X \to (-\infty, \infty)$
- A proper convex function is continuous, except possibly on its relative boundary. Moreover, it is Lipschitz continuous over a compact set

Extended real-valued function

CPC functions

Definition

Consider $f : X \to [-\infty, \infty]$ with $X \subseteq \mathbb{R}^n$

1. f is closed if epi(f) is a closed set in $\mathbb{R}^{n+1}$
2. f is proper if $f(x) > -\infty$ for all $x \in X$ and $\exists \bar{x} \in X$ such that $f(\bar{x}) < \infty$ (so that epi$(f) \neq \emptyset$). In particular a real-valued function $f : X \to \mathbb{R}$ is proper
3. Suppose X is convex. Then f is convex if epi(f) is a convex set in $\mathbb{R}^{n+1}$

Examples: Closed proper lsc functions $f : \mathbb{R} \to (-\infty, \infty]$

- Nonconvex function: $f(x) = \begin{cases} 0 & \text{if} \quad x \leq 0 \\ 1 & \text{if} \quad x > 0 \end{cases}$ (lsc)
- Convex function: $f(x) = \begin{cases} 0 & \text{if} \quad x \leq 0 \\ \infty & \text{if} \quad x > 0 \end{cases}$ (closed dom(f); convex)

Examples

Constrained functions $f_X(x)$

A real-valued function $f : X \to \mathbb{R}$ constrained to a feasible set $X \subseteq \mathbb{R}^n$ can be extended to $\mathbb{R}^n$ as an extended real-valued function $f : \mathbb{R}^n \to [-\infty, \infty]$:

$$f_X(x) \;:=\; \begin{cases} f(x) & \text{if } \; x \in X \\ \infty & \text{if } \; x \in \mathbb{R}^n \backslash X \end{cases}$$

Remarks

- Constrained minimization $\min_{x \in X} f(x)$ is equivalent to unconstrained minimization $\min_{x \in \mathbb{R}^n} f_X(x)$
- Unified theory for unconstrained minimization

Examples

Indicator function $\delta_X(x)$

The indicator function of $X \subseteq \mathbb{R}^n$ is $\delta_X : \mathbb{R}^n \to (-\infty, \infty]$ defined by

$$\delta_X(x) \;:=\; \begin{cases} 0 & \text{if } \; x \in X \\ \infty & \text{if } \; x \notin X \end{cases}$$

- It is proper iff X is nonempty
- It is convex iff X is a convex set

Examples

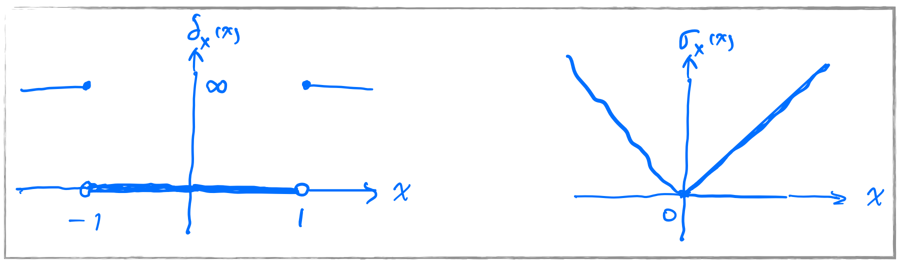
Support function $\sigma_X(x)$

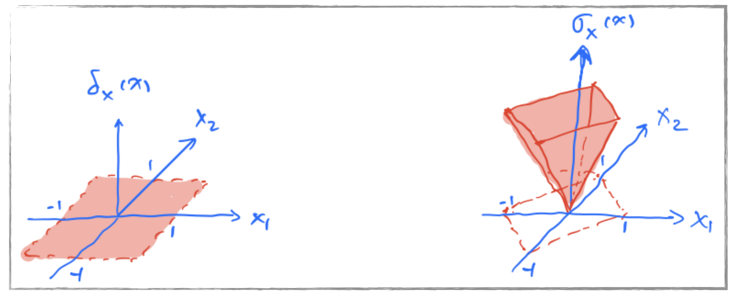
The support function of $X \subseteq \mathbb{R}^n$ is $\sigma_X : \mathbb{R}^n \to (-\infty, \infty]$ defined by

$$\sigma_X(x) \ := \ \sup_{y \in X} y^\top x$$

- It is proper iff X is nonempty and $\sup_{y \in X} y^\top x < \infty$ for some x
- $\sigma_X(x) \ = \ \sigma_{\text{cl}(X)}(x) \ = \ \sigma_{\text{conv}(X)}(x) \ = \ \sigma_{\text{cl}(\text{conv}(X))}(x) \ = \ \sigma_{\text{conv}(\text{cl}(X))}(x)$

Examples





$$\delta_{(-1,1)}(x) := \begin{cases} 0 & x \in (-1,1) \\ \infty & x \notin (-1,1) \end{cases}$$

$$\sigma_{(-1,1)}(x) := \sup_{y \in (-1,1)} yx = |x|$$

$$\delta_X(x) := \begin{cases} 0 & x_i \in (-1,1) \text{ for all } i \\ \infty & x_i \notin (-1,1) \text{ for some } i \end{cases}$$

$$\sigma_X(x) := \sum_i \sup_{y_i \in (-1,1)} y_i x_i = \sum_i |x_i|$$

Examples

Polyhedral function

A proper function $f : \mathbb{R}^n \to (-\infty, \infty]$ is a polyhedral function if $\mathrm{epi}(f)$ is a nonempty polyhedral set (polyhedron) in $\mathbb{R}^{n+1}$

- Hence a polyhedral function is closed proper convex

Lemma

Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a convex function. Then f is polyhedral iff $\mathrm{dom}(f)$ is a polyhedron and

$$f(x) = \max_{i \in \{1,\ldots,m\}} \left(a_i^\top x + b_i\right), \qquad x \in \mathrm{dom}(f)$$

for some $a_i \in \mathbb{R}^n,\ b_i \in \mathbb{R},$ integer $m > 0$

Outline

1. Normal cones of feasible sets
2. CPC functions
3. Gradient and subgradient
    - Derivative, directional derivative, partial derivative
    - Subgradient
    - Subdifferential calculus
4. Characterization: saddle point
5. Characterization: generalized KKT
6. Existence: primal optimum
7. Existence: dual optimum and strong duality
8. Special convex programs

Basic concepts

- Smooth functions: differentiable, partially differentiable, continuously differentiable
- … and their relationship

- Nonsmooth convex functions: subdifferentiable, subdifferential
- First-order optimality condition

Derivative

Consider a proper function $f : X \to (-\infty, \infty]$ with an open $X \subseteq \mathbb{R}^n$

f is differentiable at $x \in X$ if there exists $m \in \mathbb{R}^n$ s.t.

$$\lim_{\substack{h \in \mathbb{R}^n \\ h \to 0}} \frac{f(x+h) - f(x) - m^\top h}{\|h\|} = 0$$

The column vector m is called the gradient or derivative of f at x, denoted by $\nabla f(x)$

f is differentiable on X if f is differentiable at every $x \in X$

Directional derivative

The one-sided directional derivative of f at $x \in X$ in the direction $v \in \mathbb{R}^n$ is:

$$df(x;v) \;:=\; \lim_{\substack{t \in \mathbb{R} \\ t \downarrow 0}} \frac{f(x+tv)-f(x)}{t}$$

provided the limit exists, possibly $\pm\infty$

If $df(x;e_j) = df(x;-e_j)$ both exist, they are called partial derivative of f at $x \in X$ wrt x_j :

$$df(x;v) \;:=\; \lim_{\substack{t \in \mathbb{R} \\ t \to 0}} \frac{f(x+te_j)-f(x)}{t}$$

Then f is called partially differentiable at $x \in X$ wrt x_j

Partial derivative

The row vector of partial derivatives of f at $x \in X$ is:

$$\frac{\partial f}{\partial x}(x) \; := \; \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) & \cdots & \frac{\partial f}{\partial x_n}(x) \end{bmatrix}$$

f is called partially differentiable on X if f is partially differentiable at every $x \in X$

f is called continuously differentiable if f is partially differentiable and $\frac{\partial f}{\partial x}(x)$ is continuous

Remarks

- Derivative $\nabla f(x)$ describes behavior of f at x in all directions
- Partial derivative $\frac{\partial f}{\partial x}(x)$ describes behavior of f at x only along coordinate axes
- If f is differentiable, then it is partially differentiable (the converse may not hold)
- If f is continuously differentiable, then it is differentiable

Subgradient

Consider a proper convex function $f : \mathbb{R}^n \to (-\infty, \infty]$ (we can always extend f on X to $\mathbb{R}^n$)

A vector $y \in \mathbb{R}^n$ is a subgradient of f at $\bar{x} \in \text{dom}(f)$ if

$$f(x) \;\geq\; f(\bar{x}) + y^\mathsf{T}(x - \bar{x}) \qquad \forall x \in \mathbb{R}^n$$

The set of all subgradients is the subdifferential $\partial f(\bar{x})$ of f at $\bar{x}$

Remarks

- Inequality in subgradient definition must hold for all $x \in \mathbb{R}^n$, not just $x \in \text{dom}(f)$
- The affine function on RHS is supporting hyperplane (lower approximation) of f at $\bar{x}$ over $\mathbb{R}^n$
- Equivalent definition:
$$f(\bar{x}) - y^\mathsf{T}\bar{x} \;=\; \min_{x \in \mathbb{R}^n} \left(f(x) - y^\mathsf{T}x\right)$$
- If $\bar{x} \notin \text{dom}(f)$, then $\partial f(\bar{x}) := \emptyset$

Optimality condition

Consider

$$\inf_{x \in \mathbb{R}^n} f(x)$$

where $f : \mathbb{R}^n \to (-\infty, \infty]$ is a proper convex function

Corollary

$x^* \in \mathbb{R}^n$ is optimal if and only if $0 \in \partial f(x^*)$

Proof

Substitute $y = 0 \in \partial f(x^*)$ into $f(x^*) - y^\top x^* \;=\; \min_{x \in \mathbb{R}^n} \left(f(x) - y^\top x\right)$

Remark

- Optimality condition reduces to $\nabla f(x^*) = 0$ for smooth convex function f
- $0 \in \partial f(x^*)$ is a certificate of optimality of x^* (there may be other $y \in \partial f(x^*)$ with $y^\top(x - x^*) \neq 0$)

Optimality condition

For constrained optimization

$$\inf_{x \in \mathbb{R}^n} f_X(x)$$

where $X \subseteq \mathbb{R}^n$ is a convex set and $f_X : \mathbb{R}^n \to (-\infty, \infty]$ is a proper convex function, $x^* \in X$ is optimal if and only if there exists $y^* \in \partial f(x^*)$ s.t. $y^{*\top}(x - x^*) \geq 0$ for all $x \in X$ (i.e., $-y^* \in N_X(x^*)$) because then,

$$f(x^*) \ \leq \ f(x) - y^\top(x - x^*) \ \leq \ f(x) \ \text{ for all } x \in \mathbb{R}^n$$

(more formal statement later)

Subdifferentiabilty & continuity

A proper convex function $f : \mathbb{R}^n \to (-\infty, \infty]$ is subdifferentiable at any interior point $x \in \text{int}(\text{dom}(f))$

Lemma

Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a proper convex function

1. For $x \in \text{ri}(\text{dom}(f))$, $f(x)$ is continuous at x
2. For $x \in \text{int}(\text{dom}(f))$, $\partial f(x)$ is nonempty, convex and compact
3. If $X \subseteq \text{dom}(f)$ is nonempty and compact, then $\partial_X f := \cup_{x \in X} \partial f(x)$ is nonempty and bounded. Moreover f is Lipschitz continuous over X with Lipschitz constant $L := \sup_{\xi \in \partial_X f} \|\xi\|_2$

Remark

- If $f : \mathbb{R}^n \to \mathbb{R}$ is real-valued, then $\partial f(x)$ is always nonempty convex compact.
- If f is extended real-valued, then $\partial f(x)$ can be unbounded or empty at the boundary of or outside of $\text{dom}(f)$
- If $X \subseteq \mathbb{R}^n$ is nonempty convex, then $\partial \delta_X(x) = N_X(x)$ important for constrained opt (later)

Subdifferential calculus

Theorem

Let $f_i : \mathbb{R}^n \to (-\infty, \infty]$, $i = 1, \ldots, m$, be convex functions. Suppose $F(x) := \sum_i f_i(x)$ is proper.

If f_i, $i = 1, \ldots, \bar{m}$ for some $\bar{m}$, are polyhedral (i.e., epi(f_i) are polyhedrons) and

$$\left(\cap_{i=1}^{\bar{m}} \operatorname{dom}(f_i) \right) \bigcap \left(\cap_{i=\bar{m}+1}^{m} \operatorname{ri}(\operatorname{dom}(f_i)) \right) \neq \emptyset$$

then

1. F is convex
2. $\partial F(x) = \sum_i \partial f_i(x), \qquad x \in \operatorname{dom}(F)$

differentiable f_i: $\nabla F(x) = \sum_i \nabla f_i(x)$

Subdifferential calculus

Theorem

Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a convex function and $A \in \mathbb{R}^{m \times n}$. Suppose $F(x) := f(Ax)$ is proper.

If

- f is polyhedra, or
- there exists $\tilde{x} \in \mathbb{R}^n$ s.t. $A\tilde{x} \in \text{ri}(\text{dom}(f))$

then

1. F is convex
2. $\partial F(x) = A^\top \partial f(Ax)$ for all $x \in \mathbb{R}^n$

differentiable f: $\nabla F(x) = A^\top \nabla f(Ax)$

Outline

1. Normal cones of feasible sets
2. CPC functions
3. Gradient and subgradient
4. Characterization: saddle point
   - Saddle point = pd optimality + strong duality
5. Characterization: generalized KKT
6. Existence: primal optimum
7. Existence: dual optimum and strong duality
8. Special convex programs

Saddle point theorem

Consider

$$f^* \quad := \quad \min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in X',\ g(x) = 0,\ h(x) \leq 0$$

where $X' \subseteq \mathbb{R}^n$ is nonempty, $f : \mathbb{R}^n \to (-\infty, \infty]$, $g : \mathbb{R}^n \to (-\infty, \infty]^m$, $h : \mathbb{R}^n \to (-\infty, \infty]^l$

- X' may be nonconvex set, f, g, h may be nonconvex functions

Saddle point theorem

Consider

$$f^* \quad := \quad \min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in X',\ g(x) = 0,\ h(x) \le 0$$

where $X' \subseteq \mathbb{R}^n$ is nonempty, $f : \mathbb{R}^n \to (-\infty, \infty]$, $g : \mathbb{R}^n \to (-\infty, \infty]^m$, $h : \mathbb{R}^n \to (-\infty, \infty]^l$

- X' may be nonconvex set, f, g, h may be nonconvex functions

Lagrangian:

$$L(x, \lambda, \mu) \ := \ f(x) + \lambda^\top g(x) + \mu^\top h(x), \quad x \in \mathbb{R}^n, \lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^l$$

Dual function:

$$d(\lambda, \mu) \ := \ \inf_{x \in X'} L(x, \lambda, \mu), \qquad \lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^l$$

partial dualization

Dual problem:

$$d^* \ := \ \sup_{\lambda, \mu \ge 0} \ d(\lambda, \mu)$$

Saddle point theorem

Let primal and dual feasible sets be

$$X := \{x \in X',\ g(x) = 0,\ h(x) \leq 0\}, \qquad Y := \{(\lambda, \mu) \in \mathbb{R}^{m+l} : \mu \geq 0\}$$

Definition

A point $(x^*, \lambda^*, \mu^*) \in X' \times Y$ is a saddle point if

$$\max_{(\lambda,\mu)\in Y} L(x^*, \lambda, \mu) \ \leq \ L(x^*, \lambda^*, \mu^*) \ \leq \min_{x \in X'} L(x, \lambda^*, \mu^*) \quad \in \quad \mathbb{R}$$

In particular, $L(x^*, \lambda^*, \mu^*)$ is finite

Saddle point theorem

Let primal and dual feasible sets be

$$X := \{x \in X',\ g(x) = 0,\ h(x) \le 0\}, \qquad Y := \{(\lambda, \mu) \in \mathbb{R}^{m+l} : \mu \ge 0\}$$

Definition

A point $(x^*, \lambda^*, \mu^*) \in X' \times Y$ is a saddle point if

$$\max_{(\lambda,\mu)\in Y} L(x^*, \lambda, \mu) \;\le\; L(x^*, \lambda^*, \mu^*) \;\le \min_{x\in X'} L(x, \lambda^*, \mu^*) \quad \in \quad \mathbb{R}$$

In particular, $L(x^*, \lambda^*, \mu^*)$ is finite

Theorem

A point $(x^*, \lambda^*, \mu^*) \in X' \times Y$ is a saddle point if and only if

1. It is primal-dual optimal
2. The duality gap is zero at (x^*, λ^*, μ^*), i.e.,

$$d(\lambda^*, \mu^*) \;=\; d^* \;=\; f^* \;=\; f(x^*)$$

   In particular, $L(x^*, \lambda^*, \mu^*)$ is finite

same as in smooth case except allowing functions to be extended real-valued

Outline

1. Normal cones of feasible sets
2. CPC functions
3. Gradient and subgradient
4. Characterization: saddle point
5. Characterization: generalized KKT
6. Existence: primal optimum
7. Existence: dual optimum and strong duality
8. Special convex programs

Generalized KKT

Consider

$$f^* \quad := \quad \min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in P \cap C$$

where $P \subseteq \mathbb{R}^n$ is a nonempty polyhedral set, $C \subseteq \mathbb{R}^n$ is a nonempty convex set, and $f : \mathbb{R}^n \to (-\infty, \infty]$ is a proper convex extended real-valued function

Generalized KKT

Consider

$$f^* \quad := \quad \min_{x \in \mathbb{R}^n} \; f(x) \quad \text{s.t.} \quad x \in P \cap C$$

where $P \subseteq \mathbb{R}^n$ is a nonempty polyhedral set, $C \subseteq \mathbb{R}^n$ is a nonempty convex set, and $f : \mathbb{R}^n \to (-\infty, \infty]$ is a proper convex extended real-valued function

Recall

- Constrained opt can be written as unconstrained opt using indicator function

$$\min_{x \in X} f(x) \;=\; \min_{x \in \mathbb{R}^n} \; f(x) + \delta_X(x)$$

- Optimality condition for unconstrained convex optimization is $0 \in \partial f(x^*)$ $0 = \nabla f(x^*)$ in smooth case
- If f_i are convex, $F = \sum_i f_i$ is proper and Slater condition, then $\partial F(x) = \sum_i \partial f_i(x)$
- If $X \subseteq \mathbb{R}^n$ is nonempty convex, then $\partial \delta_X(x) = N_X(x)$ this is why normal cones are important

Generalized KKT

Theorem

Suppose one of the following Slater conditions holds:

1. f is polyhedral and dom$(f) \cap P \cap$ ri$(C) \neq \emptyset$; or
2. ri(dom$(f)) \cap P \cap$ ri$(C) \neq \emptyset$;

Then $x^* \in P \cap C$ is optimal if and only if

$$0 \in \partial f(x^*) + N_P(x^*) + N_C(x^*)$$

Remarks

- Equivalent to: $\exists y^* \in \partial f(x^*)$ s.t. $-y^* \in N_P(x^*) + N_C(x^*)$
- Equivalent to: $\exists y^* \in \partial f(x^*)$ s.t. $y^{*\mathsf{T}}(x - x^*) \geq 0$ for all $x \in P \cap C$
- If f is real-valued, Slater condition reduces to: $P \cap$ ri$(C) \neq \emptyset$
- In f is differentiable, KKT condition reduces to: $-\nabla f(x^*) \in N_P(x^*) + N_C(x^*)$

Generalized KKT

Theorem

Suppose one of the following Slater conditions holds:

1. f is polyhedral and $\text{dom}(f) \cap P \cap \text{ri}(C) \neq \emptyset$; or
2. $\text{ri}(\text{dom}(f)) \cap P \cap \text{ri}(C) \neq \emptyset$;

Then $x^* \in P \cap C$ is optimal if and only if

$$0 \in \partial f(x^*) + N_P(x^*) + N_C(x^*)$$

Remarks

- When the feasible sets P, C are explicitly specified by equality and inequality constraints that enable the computation of the normal cones $N_P(x^*)$, $N_C(x^*)$, the condition reduces to usual KKT conditions (later)
- This theorem illustrates the conceptual simplicity based on set theoretic concepts nonsmooth opt

Optimality characterization

Remarks

- Saddle point theorem and generalized KKT theorem characterize optimal points
- They do not ensure existence of optimal points
- Examples exist where primal optimal solutions do not exist, even though
  - Slater condition is satisfied
  - f^* is finite
  - Dual optimal solutions exist and strong duality holds

  because the primal feasible set is not compact

We next study sufficient conditions for existence of primal and dual optimal solutions (and strong duality)

Outline

1. Normal cones of feasible sets
2. CPC functions
3. Gradient and subgradient
4. Characterization: saddle point
5. Characterization: generalized KKT
6. Existence: primal optimum
7. Existence: dual optimum and strong duality
8. Special convex programs

Nonexistence of primal optimum

Example: feasible set not closed

Consider

$$f^* \quad := \quad \inf_{x \in \mathbb{R}} \quad f(x) := x^2 \quad \text{s.t.} \quad x > 1$$

feasible set **not** closed

- Primal optimal value $f^* = 1$
- No primal optimal x^* with $f(x^*) = f^*$

Nonexistence of primal optimum

Example: feasible set not closed

Consider

$$f^* \quad := \quad \inf_{x \in \mathbb{R}} \quad f(x) := x^2 \quad \text{s.t.} \quad x > 1$$

feasible set **not** closed

- Primal optimal value $f^* = 1$
- No primal optimal x^* with $f(x^*) = f^*$

Lagrangian and dual function:

$$L(x, \mu) := x^2 - \mu x + \mu, \qquad d(\mu) := \inf_x L(x, \mu) = -\frac{\mu^2}{4} + \mu$$

Dual optimal value:

$$d^* \;:=\; \sup_{\mu \geq 0} \; d(\mu) \;=\; d(2) \;=\; 1 \;=\; f^*$$

- Strong duality holds
- Dual optimality is attained at $\mu^* = 2$

Nonexistence of primal optimum

Example: feasible set not closed

Conclusions:

- (Primal) feasible set is not closed (hence not compact)
- Primal optimal value $f^* = 1$ is finite, but not attained
- Strong duality holds $f^* = d^*$, and dual optimality is attained at $\mu^* = 2$

KKT condition cannot be satisfied:

- Stationarity and complementary slackness are

$$2x^* = \mu^*, \qquad \mu^*(1 - x^*) = 0$$

which cannot be satisfied when $\mu^* = 2$ and $x^* > 1$

Nonexistence of primal optimum

Example: feasible set not bounded

Consider

$$f^* \quad := \quad \inf_{x \in \mathbb{R}} \; f(x) := e^{-x} \quad \text{s.t.} \quad x \geq 0$$

feasible set **not** bounded

- Primal optimal value $f^* = 0$
- No primal optimal x^* with $f(x^*) = f^*$

Nonexistence of primal optimum

Example: feasible set not bounded

Consider

$$f^* \;:=\; \inf_{x\in\mathbb{R}} \; f(x) := e^{-x} \quad \text{s.t.} \quad x \geq 0$$

feasible set **not** bounded

- Primal optimal value $f^* = 0$
- No primal optimal x^* with $f(x^*) = f^*$

Lagrangian and dual function:

$$L(x,\mu) := e^{-x} - \mu x, \qquad d(\mu) := \min_x \; e^{-x} - \mu x \;=\; \begin{cases} 0, & \mu = 0 \\ -\infty, & \mu > 0 \end{cases}$$

Dual optimal value:

$$d^* \;:=\; \sup_{\mu\geq 0} \; d(\mu) \;=\; d(0) \;=\; 0 \;=\; f^*$$

- Strong duality holds
- Dual optimality is attained at $\mu^* = 0$

Nonexistence of primal optimum

Example: feasible set not bounded

Conclusions:

- (Primal) feasible set is not bounded (hence not compact)
- Primal optimal value $f^* = 0$ is finite, but not attained
- Strong duality holds $f^* = d^*$, and dual optimality is attained at $\mu^* = 0$

KKT condition cannot be satisfied:

- Stationarity condition is

$$e^{-x^*} = -\mu^*$$

which cannot be satisfied by any finite x^* when $\mu^* = 0$

Primal optimality

Consider

$$f^* \quad := \quad \min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in X \subseteq \mathbb{R}^n$$

where $f : X \to (-\infty, \infty]$ and $X \cap \text{dom}(f) \neq \emptyset$

- $f : \mathbb{R}^n \to (-\infty, \infty]$ is called radially unbounded if $\lim_k f(x_k) = \infty$ for every sequence $\{x_k\}$ with $\|x_k\| \to \infty$
- All nonempty level sets $V_\gamma := \{x \in \mathbb{R}^n : f(x) \leq \gamma\}$ of a radially unbounded function are bounded

Primal optimality

Consider

$$f^* \quad := \quad \min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in X \subseteq \mathbb{R}^n$$

where $f : X \to (-\infty, \infty]$ and $X \cap \text{dom}(f) \neq \emptyset$

Theorem (Weierstrass theorem)

If X is closed, f is lower semicontinuous at every $x \in X$, and one of the following holds:

1. X is bounded; or
2. There exists $\gamma \in \mathbb{R}$ s.t. the level set $V_\gamma := \{x \in \mathbb{R}^n : f(x) \leq \gamma\}$ is nonempty and bounded; or
3. f is radially unbounded.

then the set $X^* \subseteq X$ of minima is nonempty and compact ((X^* is convex if X and f are convex)

Remark

- Essentially restrict minimization to a compact subset of X

Exact optimality condition

CPF function f

Remark

If X is nonempty closed and convex, f is closed proper and convex, and $X \cap \text{dom}(f) \neq \emptyset$

- Under the conditions of the theorem (lsc of f and boundedness): X^* is nonempty, compact and convex if and only if X and f have no common nonzero direction of recession (details in textbook)

Outline

1. Normal cones of feasible sets
2. CPC functions
3. Gradient and subgradient
4. Characterization: saddle point
5. Characterization: generalized KKT
6. Existence: primal optimum
7. Existence: dual optimum and strong duality
    - Slater Theorem
    - MC/MC problems
    - Slater Theorem: proof
8. Special convex programs

Slater Theorem

Consider

$$f^* \quad := \quad \min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in X',\ h(x) \le 0$$

where $X' \subseteq \mathbb{R}^n$ is a nonempty convex set, $f : \mathbb{R}^n \to (-\infty, \infty]$ and $h : \mathbb{R}^n \to (-\infty, \infty]^l$ are proper convex functions.

Lagrangian:

$$L(x, \mu) \ := \ f(x) + \mu^\top h(x), \quad x \in \mathbb{R}^n,\ \mu \in \mathbb{R}^l$$

Dual function:

$$d(\mu) \ := \ \inf_{x \in X'} L(x, \mu), \qquad \mu \in \mathbb{R}^l$$

Dual problem:

$$d^* \ := \ \sup_{\mu \ge 0} \ d(\mu)$$

Slater Theorem

Consider

$$f^* \quad := \quad \min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in X',\ h(x) \leq 0$$

where $X' \subseteq \mathbb{R}^n$ is a nonempty convex set, $f : \mathbb{R}^n \to (-\infty, \infty]$ and $h : \mathbb{R}^n \to (-\infty, \infty]^l$ are proper convex functions.

Set of dual optimal solutions:

$$Q^* := \left\{ \mu^* \geq 0 : d(\mu^*) = \inf_{x \in X'} f(x) + \mu^{*\top} h(x) = f^* \right\}$$

Slater Theorem

Theorem

Suppose

- *Finite primal value*: $f^* > -\infty$
- *Convexity*: X' is a nonempty convex set; f, h are proper convex functions;
- *Slater condition*: one of the following holds:
  - CQ1: $\exists \bar{x} \in \text{dom}(f) \cap X'$ s.t. $h(\bar{x}) < 0$; or
  - CQ2: $\exists \bar{x} \in \text{ri}(\text{dom}(f)) \cap \text{ri}(X')$ s.t. $h(\bar{x}) := A\bar{x} + b \leq 0$

Then

1. $f^* = d^*$
2. If CQ1 holds, then Q^* is nonempty, convex and compact
3. If CQ2 holds, then Q^* is nonempty, convex and closed

Slater Theorem

Variant

Equivalent formulation:

$$f^* \quad := \quad \min_{x \in \mathbb{R}^n} \; f(x) \quad \text{s.t.} \quad x \in P \cap C, \; Ax = b, \; h(x) \le 0$$

polyhedron (P) — convex (C) — polyhedral $h_i, i = 1, \ldots, \bar{l}$

Lagrangian:

$$L(x, \lambda, \mu) \;\; := \;\; f(x) + \lambda^{\mathsf{T}} g(x) + \mu^{\mathsf{T}} h(x), \quad x \in \mathbb{R}^n, \; \lambda \in \mathbb{R}^m, \; \mu \in \mathbb{R}^l$$

Dual function:

$$d(\lambda, \mu) \;\; := \;\; \inf_{x \in X'} L(x, \lambda, \mu), \qquad \lambda \in \mathbb{R}^m, \; \mu \in \mathbb{R}^l$$

Dual problem:

$$d^* \;\; := \;\; \sup_{\lambda, \mu \ge 0} \; d(\lambda, \mu)$$

Slater Theorem

Variant

Theorem

Suppose

- *Finite primal value*: $f^* > -\infty$
- *Convexity*: P is nonempty polyhedron, C is nonempty convex; f, h are proper convex functions;
- *Slater condition*: $\exists \bar{x} \in \text{ri}(\text{dom}(f)) \cap P \cap \text{ri}(C)$ s.t. $A\bar{x} = b$, polyhedral $h_i(\bar{x}) \leq 0, i = 1, \ldots, \bar{l}$, $h_i(\bar{x}) < 0, i = \bar{l} + 1, \ldots, l$

Then

1. $f^* = d^*$
2. The set of dual optimal solutions is nonempty, convex and closed

Slater Theorem

Theorem

Suppose

- *Finite primal value*: $f^* > -\infty$
- *Convexity*: X' is a nonempty convex set; f, h are proper convex functions;
- *Slater condition*: one of the following holds:
  - CQ1: $\exists \bar{x} \in \text{dom}(f) \cap X'$ s.t. $h(\bar{x}) < 0$; or
  - CQ2: $\exists \bar{x} \in \text{ri}(\text{dom}(f)) \cap \text{ri}(X')$ s.t. $h(\bar{x}) := A\bar{x} + b \leq 0$

Prove CQ1

Then

1. $f^* = d^*$
2. If CQ1 holds, then Q^* is nonempty, convex and compact
3. If CQ2 holds, then Q^* is nonempty, convex and closed

MC/MC problems

Let $M \subseteq \mathbb{R}^{l+1}$ be a nonempty set

Primal (min common): $$w^* := \inf_{(0,w)\in M} w$$

Dual (max crossing): $$d^* := \sup_{\mu\in\mathbb{R}^l} \left(d(\mu) := \inf_{(u,w)\in M} \mu^\mathsf{T} u + w \right)$$

MC/MC problems

Let $M \subseteq \mathbb{R}^{l+1}$ be a nonempty set

Primal (min common): $$w^* := \inf_{(0,w)\in M} w$$

Dual (max crossing): $$d^* := \sup_{\mu\in\mathbb{R}^l} \left(d(\mu) := \inf_{(u,w)\in M} \mu^\mathsf{T} u + w \right)$$

Easier to work with positive extension of M:

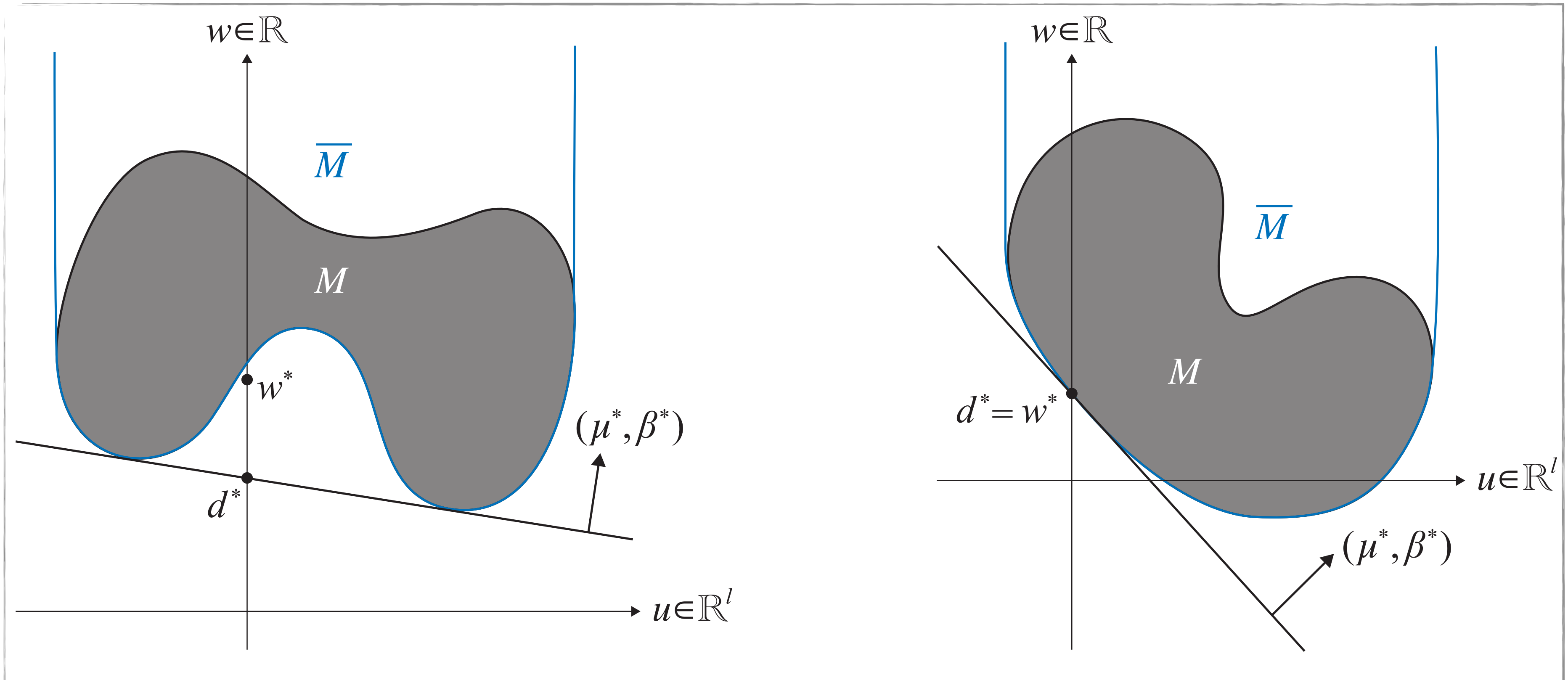
$$\overline{M} := M + \{(0,w) : w \geq 0\} = \{(u,w) : w \geq \bar{w} \text{ for some } (u,\bar{w}) \in M\}$$

Then

Primal (min common): $$w^* := \inf_{(0,w)\in \overline{M}} w$$

Dual (max crossing): $$d^* := \sup_{\mu\in\mathbb{R}^l} \left(d(\mu) := \inf_{(u,w)\in \overline{M}} \mu^\mathsf{T} u + w \right)$$

MC/MC problems



(a) Nonconvex $\overline{M}$ (b) Convex $\overline{M}$

MC/MC problems

Primal (min common): $$w^* := \inf_{(0,w)\in\overline{M}} w$$

Dual (max crossing): $$d^* := \sup_{\mu\in\mathbb{R}^l} \left(d(\mu) := \inf_{(u,w)\in\overline{M}} \mu^\mathsf{T} u + w \right)$$

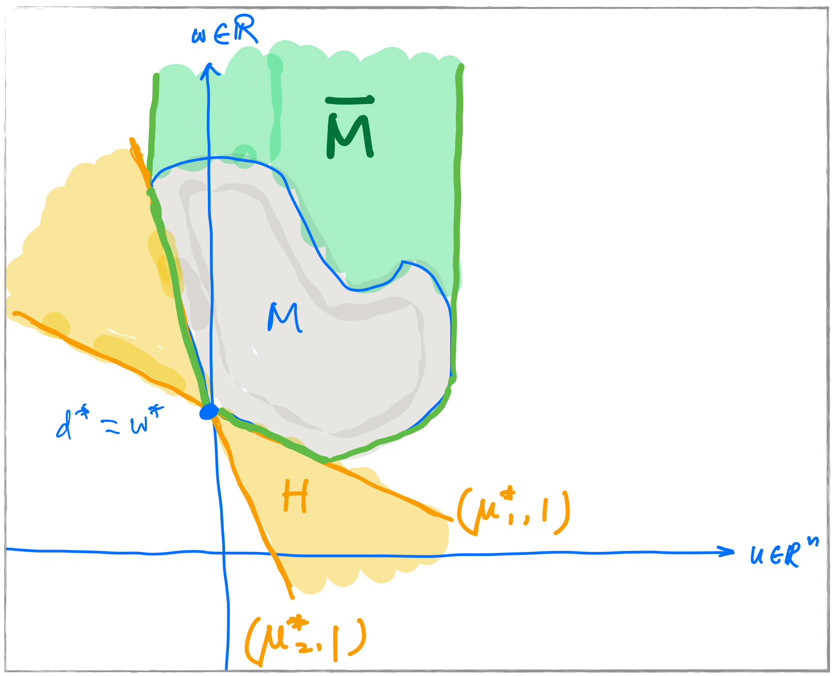
Remarks

1. Dual relaxation: $d(\mu) := \inf_{(u,w)\in\overline{M}} \mu^\mathsf{T} u + w$ relaxes $u = 0$ but adds penalty $\mu^\mathsf{T} u$
2. Strong duality: there exists a nonvertical hyperplane that contains $\overline{M}$ in its ``upper" closed halfspace, i.e., there exist a normal $(\mu, 1) \in \mathbb{R}^{l+1}$ and an w-intercept $\xi \in \mathbb{R}$ s.t.
   $$\mu^\mathsf{T} u + w \geq \xi, \qquad \forall (u, w) \in \overline{M}$$
2. Given μ, $d(\mu)$ is the smallest w-intercept of the hyperplane that touches (supports) $\overline{M}$
3. Dual problem: find a normal $(\mu^*, 1)$ s.t. the smallest w-intercept $d(\mu^*)$ is the max over $\mu \in \mathbb{R}^l$

Dual optimal solution set Q^*

Let the set of dual optimal solutions be:

$$Q^* \;=\; \left\{ \mu^* \in \mathbb{R}^l \;:\; d(\mu^*) \;:=\; \inf_{(u,w)\in\overline{M}} \mu^{*\mathsf{T}} u + w \;=\; w^* \right\}$$



Every dual optimal $\mu^* \in Q^*$ defines a supporting hyperplane

$$H := \{(u,w) \in \mathbb{R}^{l+1} : \mu^{*\mathsf{T}} u + w = w^*\}$$

at $(0, w^*) \in$ cl($\overline{M}$), with cl($\overline{M}$) in ``upper" halfspace of H

In this example, Q^* is nonempty, convex and compact

MC/MC strong duality

Let projection of $\overline{M}$ onto u-space

$$D_{\overline{M}} \ := \ \{u \in \mathbb{R}^n : (u, w) \in \overline{M} \text{ for some } w \in \mathbb{R}\}$$

Lemma

Suppose

- *Finite primal value*: $w^* > -\infty$
- *Convexity*: $\overline{M}$ is convex
- *Constraint qualification:* $0 \in$ *ri* $\left(D_{\overline{M}}\right)$

Then

1. $w^* = d^*$
2. Q^* is nonempty, convex and closed
3. If $0 \in$ int $\left(D_{\overline{M}}\right)$, then Q^* is nonempty, convex and compact

MC/MC strong duality

Proof sketch

1. $(0,w^*) \notin \text{ri}(\overline{M})$: Otherwise, $\exists (0,\bar{w}) \in \overline{M}$ s.t. $w^* > \bar{w}$
2. H separating $(0,w^*)$ from $\overline{M}$: $\exists (\mu,\beta) \in \mathbb{R}^{l+1}$ s.t.

$$\beta w^* \leq \mu^\top u + \beta w, \qquad \forall (u,w) \in \overline{M}$$

   $(0,w^*) \notin \text{ri}(\overline{M})$ implies

$$\beta w^* \leq \inf_{(u,w)\in\overline{M}} \mu^\top u + \beta w < \sup_{(u,w)\in\overline{M}} \mu^\top u + \beta w$$

3. $\beta > 0$:
   - β cannot be negative, for otherwise $\inf_{(u,w)\in\overline{M}} \mu^\top u + \beta w \to -\infty$
   - β cannot be 0, for otherwise $0 \leq \inf_{(u,w)\in\overline{M}} \mu^\top u = \inf_{u\in D_{\overline{M}}} \mu^\top u$. Since $0 \in \text{ri}(D_{\overline{M}})$, this infimum is attained at $u = 0$ over the convex set $D_{\overline{M}}$.

     This is possible only if $\mu^\top u = 0$ is constant over $D_{\overline{M}}$, a contradiction

MC/MC strong duality

Proof sketch

4. Strong duality : Since $\beta > 0$, can renormalize to $\mu^* := \mu/\beta$, $\beta^* = 1$

$$w^* \;\leq\; \inf_{(u,w)\in\overline{M}} \mu^{*\mathsf{T}} u + w \;=:\; d(\mu^*) \;\leq\; d^*$$

5. Q^* convex and closed : The dual function $d(\mu)$ is concave $\Rightarrow$ Q^* is convex.

   The dual function $d(\mu)$ is upper semicontinuous $\Rightarrow$ Q^* is closed.

   If $0 \in \text{int}(D_{\overline{M}})$, then Q^* is bounded, and hence compact.

Slater Theorem

Proof: CQ1

Theorem

Suppose

- *Finite primal value*: $f^* > -\infty$
- *Convexity*: X' is a nonempty convex set; f, h are proper convex functions;
- *Slater condition*: one of the following holds:
    - CQ1: $\exists \bar{x} \in \mathrm{dom}(f) \cap X'$ s.t. $h(\bar{x}) < 0$; or

Then

1. $f^* = d^*$
2. If CQ1 holds, then Q^* is nonempty, convex and compact

Idea: specify $\overline{M}$ in terms of cost and constraint functions f, h, and use Lemma

Slater Theorem

Proof: CQ1

Consider

$$\text{Primal} \qquad f^* \;:=\; \inf_{x\in\mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in X',\ h(x) \le 0$$

$$\text{Dual} \qquad d^* \;:=\; \sup_{\mu\ge 0} \left(d(\mu) := \inf_{x\in X'} f(x) + \mu^\top h(x) \right)$$

Let

$$M \;:=\; \{(h(x), f(x)) \in \mathbb{R}^{l+1} : x \in \text{dom}(f) \cap X'\}$$

Its positive extension:

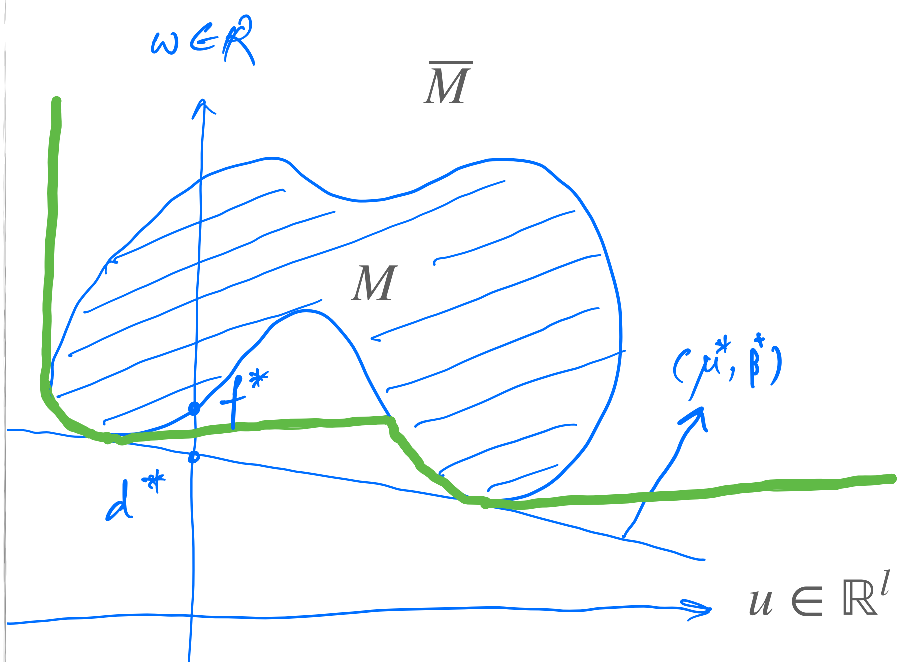
$$\overline{M} \;:=\; \{(u, w) \in \mathbb{R}^{l+1} : u \ge h(x),\ w \ge f(x) \text{ for some } x \in \text{dom}(f) \cap X'\}$$

Projection onto u-space

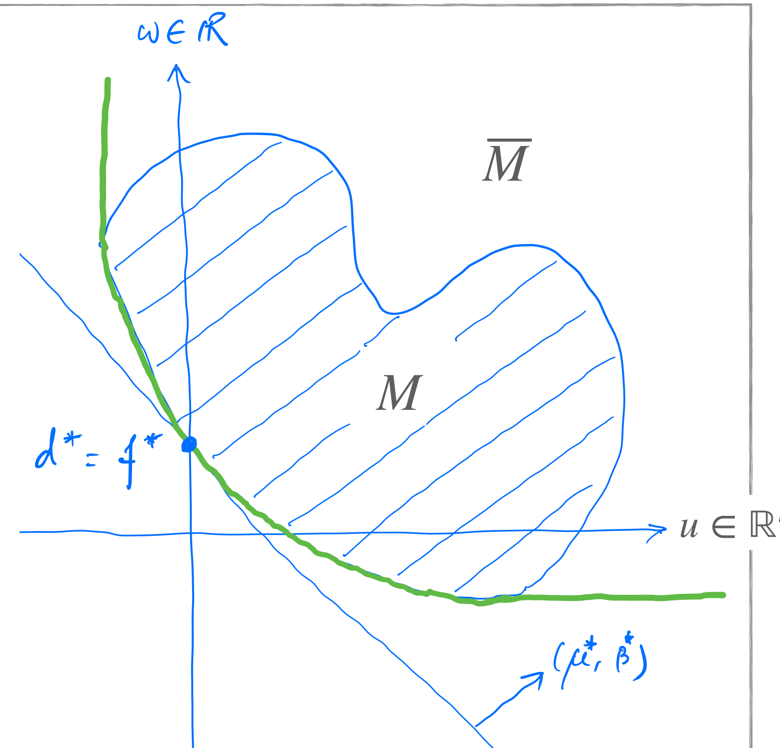
$$D_{\overline{M}} \;=\; \{u \in \mathbb{R}^l : u \ge h(x) \text{ for some } x \in \text{dom}(f) \cap X'\}$$

Slater Theorem

Proof: CQ1



(a) Nonconvex $\overline{M}$



(b) Convex $\overline{M}$

Slater Theorem

Proof: CQ1

Equivalent MC/MC formulation in terms of $\overline{M}$:

Primal: $f^* := \inf_{(0,w)\in\overline{M}} w$

Dual: $d^* := \sup_{\mu\in\mathbb{R}^l} \left(d(\mu) := \inf_{(u,w)\in\overline{M}} \mu^\top u + w \right)$

Set of dual optimal solutions:

$$Q^* = \left\{ \mu^* \in \mathbb{R}^l \; : \; d(\mu^*) := \inf_{(u,w)\in\overline{M}} \mu^{*\top} u + w = w^* \right\}$$

MC/MC strong duality

Let projection of $\overline{M}$ onto u-space

$$D_{\overline{M}} := \{u \in \mathbb{R}^n : (u, w) \in \overline{M} \text{ for some } w \in \mathbb{R}\}$$

Lemma

Suppose

- *Finite primal value*: $w^* > -\infty$
- *Convexity*: $\overline{M}$ is convex
- *Constraint qualification:* $0 \in$ *ri* $\left(D_{\overline{M}}\right)$

Then

1. $w^* = d^*$
2. Q^* is nonempty, convex and closed
3. If $0 \in$ int $\left(D_{\overline{M}}\right)$, then Q^* is nonempty, convex and compact

Slater Theorem

Proof: CQ1

1. $f^* > -\infty$: By assumption
2. Convex $\overline{M}$: Suppose $(u_1, w_1), (u_2, w_2) \in \overline{M}$, i.e., $\exists x_1, x_2 \in \text{dom}(f) \cap X'$ s.t.

$$u_i \geq h(x_i),\ w_i \geq f(x_i) \qquad i = 1,2$$

   Convexity of h implies

$$\alpha u_1 + (1-\alpha)u_2 \ \geq\ h(\alpha x_1 + (1-\alpha)x_2)$$

   Convexity of f implies

$$\alpha w_1 + (1-\alpha)w_2 \geq f(\alpha x_1 + (1-\alpha)x_2)$$

   i.e., $\alpha(u_1, w_1) + (1-\alpha)(u_2, w_2) \in \overline{M}$

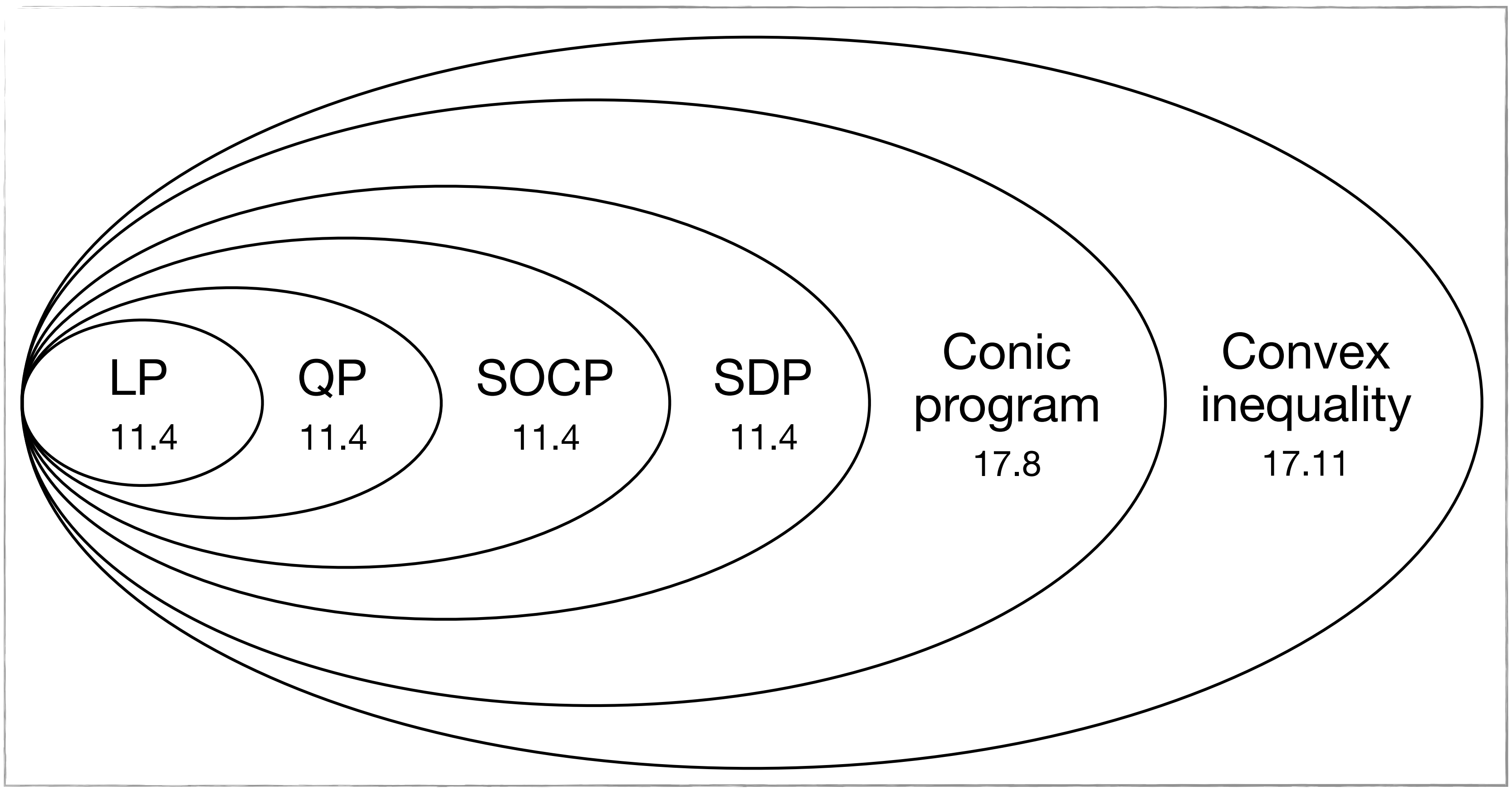
3. $0 \in \text{int}\left(D_{\overline{M}}\right)$: CQ1 gives $\bar{x} \in \text{dom}(f) \cap X'$ with $h(\bar{x}) < 0 \ \Rightarrow 0 \in \text{int}\left(D_{\overline{M}}\right)$ where

$$D_{\overline{M}} \ = \ \{u \in \mathbb{R}^l : u \geq h(x) \text{ for some } x \in \text{dom}(f) \cap X'\}$$

Outline

1. Normal cones of feasible sets
2. CPC functions
3. Gradient and subgradient
4. Characterization: saddle point
5. Characterization: generalized KKT
6. Existence: primal optimum
7. Existence: dual optimum and strong duality
8. Special convex programs
   - General method
   - Linear program, convex quadratic program
   - Second-order cone program
   - Conic program

Special convex programs



Special convex programs

	$f(x)$	$h(x) \leq 0$	sufficient condition	$f^* = d^* = d(\lambda^*, \mu^*)$ KKT, saddle pt
LP	linear	affine: $Bx + d \in \mathbb{R}^l_+$	finite f^*	Th 7.21
QP	quadratic	affine: $Bx + d \in \mathbb{R}^l_+$	feasibility (if $Q \succ 0$)	Th 7.22, 7.23
SOCP	convex	$h(x) \in K_{\text{soc}}$ $h(x) := \tilde{B}x + \tilde{d}$	finite f^*, $A\bar{x} = b$ $h(\bar{x}) \in \text{ri}(K_{\text{soc}})$	Th 7.24, 7.25
SDP	convex	$h(x) \in K_{\text{psd}}$ $h(x) := B_0 + \sum_{i=1}^n x_i B_i$	finite f^*, $A\bar{x} = b$ $h(\bar{x}) \in \text{ri}(K_{\text{psd}})$	Th 7.26
Conic prog.	convex	$h(x) \in K$ $h(x) := Bx + d$	finite f^*, $A\bar{x} = b$ $h(\bar{x}) \in \text{ri}(K)$	Th 12.31, 12.32
Convex prog.	convex	convex	finite f^*, $A\bar{x} = b$ $h(\bar{x}) < 0$	Exercise 12.22

Table 7.3 Summary: strong duality, dual optimality and KKT condition.

General method

Smooth setting

Consider

$$f^* \quad := \quad \min_{x \in \mathbb{R}^n} \quad f(x) \quad \text{s.t.} \quad Ax = b,\ h(x) \le 0$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, and $h : \mathbb{R}^n \to \mathbb{R}^l$ is a convex function

General method

1. *Dual problem*

   Define Lagrangian $L(x, \lambda, \mu) : \mathbb{R}^{n+m+l} \to \mathbb{R}$:

   $$L(x, \lambda, \mu) \ := \ f(x) - \lambda^\mathsf{T}(Ax - b) + \mu h(x), \qquad x \in \mathbb{R}^n,\ (\lambda, \mu) \in \mathbb{R}^{m+l}$$

   Then dual function is $d(\lambda, \mu) := \min_{x \in \mathbb{R}^n} L(x, \lambda, \mu)$ and dual problem is :

   $$d^* \quad := \quad \max_{(\lambda,\mu) \in \mathbb{R}^{m+l}} d(\lambda, \mu) \qquad \text{s.t.} \qquad \mu \ge 0$$

General method

Smooth setting

2. *Strong duality and dual optimality*

   If (i) f^* is finite, (ii) f, h are convex, and (iii) Slater condition is satisfied, then the Slater Theorem implies that strong duality holds and dual optimality is attained (do not guaranteed primal optimality is attained)

3. *KKT condition and primal optimality*

   If (i) f, h are convex, and (ii) Slater condition is satisfied, then the KKT Theorem implies that a feasible x^* is optimal if and only if $\exists(\lambda^*, \mu^*) \in \mathbb{R}^{m+l}$ s.t.

$$\nabla f(x^*) = A^\top \lambda^* - \nabla h(x^*)\mu, \qquad {\mu^*}^\top h(x^*) = 0, \qquad \mu^* \geq 0$$

   Moreover, a KKT point is a saddle point that attains both primal and dual optimality and closes duality gap, i.e.,

$$f^* = f(x^*) = d(\lambda^*, \mu^*) = d^*$$

General method

Nonsmooth setting

Consider

$$f^* \quad := \quad \min_{x\in\mathbb{R}^n} \quad f(x) \quad \text{s.t.} \quad Ax = b,\ x \in X \subseteq \mathbb{R}^n$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, X is a nonempty closed convex set that may be specified explicitly as $X := \{x : h(x) \leq 0\}$ for a convex function $h : \mathbb{R}^n \to \mathbb{R}^l$

General method

1. *Dual problem*
   - If $X := \{x : h(x) \leq 0\}$, then same as in smooth setting, except using subgradients
   - If $X := \{x : Bx + d \in K\}$ for a closed convex cone K, then define Lagrangian $L(x, \lambda, \mu) : \mathbb{R}^{n+m+l} \to \mathbb{R}$:

$$L(x,\lambda,\mu) \ := \ f(x) - \lambda^\mathsf{T}(Ax - b) + \mu(Bx + d), \quad x \in \mathbb{R}^n,\ \lambda \in \mathbb{R}^m,\ \mu \in K^* \subseteq \mathbb{R}^l$$

Then dual function is $d(\lambda,\mu) := \min_{x\in\mathbb{R}^n} L(x,\lambda,\mu)$ and dual problem is :

$$d^* \quad := \quad \max_{(\lambda,\mu)\in\mathbb{R}^{m+l}} d(\lambda,\mu) \qquad \text{s.t.} \qquad \mu \in K^*$$

General method

Nonsmooth setting

2. *Strong duality and dual optimality*

   Same as in smooth setting (does not require differentiability)

3. *KKT condition and primal optimality*

   Suppose (i) f, h are convex, and (ii) Slater condition is satisfied. Stationarity condition in KKT may no longer be derived from $\nabla_x L(x^*, \mu^*) = 0$.

   Example Suppose $X := \{x : Bx + d \in K\}$. Convert to unconstrained optimization:

$$f^* \quad := \quad \min_{x \in \mathbb{R}^n} \; f(x) + \delta_H(x) + \delta_K(Bx + d)$$

   where $H := \{x \in \mathbb{R}^n : Ax = b\}$. Generalized KKT Theorem implies: x^* is optimal if and only if $\exists \xi^* \in \partial f(x^*), \lambda^* \in \mathbb{R}^m, \mu^* \in \mathbb{R}^l$ such that

$$\xi^* \;\in -N_H(x^*) - B^\top N_K(Bx^* + d)$$

   or: $\quad \xi^* = A^\top \lambda^* + B^\top \mu^*, \qquad {\mu^*}^\top (Bx^* + d) = 0, \qquad \mu^* \in K^*$

Linear program

Consider

$$f^* \quad := \quad \min_{x \in \mathbb{R}^n} c^\top x \qquad \text{s.t.} \qquad Ax \geq b$$

where $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$

Linear program

Consider

$$f^* \ := \ \min_{x \in \mathbb{R}^n} c^\top x \qquad \text{s.t.} \qquad Ax \geq b$$

where $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$

Lagrangian:

$$L(x, \mu) \ := \ \left(c - A^\top \mu\right)^\top x + b^\top \mu \qquad x \in \mathbb{R}^n, \, \mu \in \mathbb{R}^m$$

Dual function:

$$d(\mu) \ := \ \min_{x \in \mathbb{R}^n} L(x, \mu) \ = \ \begin{cases} b^\top \mu & \text{if } A^\top \mu = c \\ -\infty & \text{if } A^\top \mu \neq c \end{cases}$$

Dual problem:

$$d^* \ := \ \max_{\mu \geq 0} d(\mu) \ = \ \max_{\mu \geq 0} b^\top \mu \qquad \text{s.t.} \qquad A^\top \mu = c$$

Let $X := \{x \in \mathbb{R}^n : Ax \geq b\}$, $Y := \{\mu \in \mathbb{R}^m : A^\top \mu = c, \mu \geq 0\}$ be feasible sets

Linear program

Theorem

1. *Strong duality and primal-dual optimality*. Exactly one of the following holds:

   (a) If $-\infty < f^* < \infty$ or $-\infty < d^* < \infty$, then $\exists (x^*, \mu^*) \in X \times Y$ such that
   $$c^\top x^* = f^* = d^* = b^\top \mu^*$$

   (b) If primal is feasible but unbounded, then $f^* = -\infty = d^*$, i.e., dual infeasible

   (c) If dual is feasible but unbounded, then $f^* = \infty = d^*$, i.e., primal infeasible

   (d) Both are infeasible, i.e., $f^* = \infty$ and $d^* = -\infty$

2. *KKT characterization*. A feasible $x^* \in X$ is optimal if and only if there is a feasible $\mu^* \in Y$ s.t.
   $$\mu^{*\top}(Ax^* - b) = 0$$

   Such a point (x^*, μ^*) is a saddle point and a KKT point, and hence is primal-dual optimal with $c^\top x^* = b^\top \mu^*$

Linear program

		primal		
		bounded feasible	unbounded feasible	infeasible
dual	bounded feasible	(x^*, λ^*, μ^*)	$\times$ (sd)	$\times$ (sd)
	unbounded feasible	$\times$ (sd)	$\times$ (wd)	$f^* = d^* = \infty$
	infeasible	$\times$ (sd)	$f^* = d^* = -\infty$	$d^* = -\infty < \infty = f^*$

Table 7.4 Four possibilities: Strong duality in Theorem 7.21 excludes 4 possibilities labeled "$\times$(sd)". The 5th impossibility, labeled "$\times$(wd)", violates weak duality. Optimal values are attained only in one case.

Linear program

Example: *Infeasible LP pair*

Consider

$$\min_{x} \ x \qquad \text{s.t.} \qquad \begin{bmatrix} 1 \\ -1 \end{bmatrix} x \geq \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Its dual is

$$\max_{\mu \geq 0} \ \mu_2 \qquad \text{s.t.} \qquad -\mu_2 = 1$$

Linear program

Example: *Unbounded primal, infeasible dual*

Consider $(\alpha < 1)$

$$f^* \quad := \quad \min_{x \geq 0} \; - x_1 + \alpha x_2 \quad \text{s.t.} \quad x_1 - x_2 = 0$$

Then $f^* = -\infty$.

Its dual function is:

$$d(\lambda, \mu) \; := \; \begin{cases} 0 & \text{if} \;\; \begin{bmatrix} -1 \\ \alpha \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \lambda + \mu, \qquad \lambda \in \mathbb{R}, \; \mu \in \mathbb{R}^2_+ \\ -\infty & \text{otherwise} \end{cases}$$

The constraint implies $\mu_1 + \mu_2 = -(1 - \alpha) < 0$, and hence there is no (λ, μ) with $\mu \geq 0$

Quadratic program

Consider

$$f_1^* \;:=\; \min_{x \in \mathbb{R}^n} f(x) \;:=\; x^\top Q x \;+\; 2c^\top x$$

where $Q \in \mathbb{R}^{n \times n}$ with $Q \succeq 0$, $c \in \mathbb{R}^n$

Quadratic program

Consider

$$f_1^* \;:=\; \min_{x \in \mathbb{R}^n} f(x) \;:=\; x^\top Q x \;+\; 2c^\top x$$

where $Q \in \mathbb{R}^{n \times n}$ with $Q \succeq 0,\ c \in \mathbb{R}^n$

Since $Q \succeq 0$ we have the spectral decomposition

$$Q \;=\; U \Lambda U^\top \;=\; \begin{bmatrix} U_r & U_{n-r} \end{bmatrix} \begin{bmatrix} \Lambda_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_r^\top \\ U_{n-r}^\top \end{bmatrix} \;=\; U_r \Lambda_r U_r^\top$$

where $r = \operatorname{rank}(Q)$ and columns of U form an orthonormal basis

Hence

$$\text{range}(Q) = \text{span}(U_r), \qquad \text{null}(Q) = \text{span}(U_{n-r}), \qquad Q^\dagger := U_r \Lambda_r^{-1} U_r^\top, \quad r \le n$$

Quadratic program

Theorem

1. If $c \in \text{range}(Q)$, then a minimizer x^* is
$$x^* = -Q^{\dagger}c, \qquad f_1^* = -c^{\mathsf{T}}Q^{\dagger}c$$
   The set set of minimizers is $x^* \in -Q^{\dagger}c + \text{null}(Q)$

2. If $c \notin \text{range}(Q)$, then $c_1^* = -\infty$

3. If $Q \succ 0$, then the unique minimizer x^* is
$$x^* = -Q^{-1}c, \qquad f_1^* = -c^{\mathsf{T}}Q^{-1}c$$

Quadratic program

Consider

$$f_2^* \quad := \quad \min_{x \in \mathbb{R}^n} f(x) \;:=\; x^\top Q x \;+\; 2c^\top x \quad \text{s.t.} \quad Ax = b,\; Bx + d \geq 0$$

where $Q \in \mathbb{R}^{n \times n}$ with $Q \succeq 0$, $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $B \in \mathbb{R}^{k \times n}$, $d \in \mathbb{R}^k$

Quadratic program

Consider

$$f_2^* \quad := \quad \min_{x \in \mathbb{R}^n} f(x) \;:=\; x^\top Q x \;+\; 2c^\top x \quad \text{s.t.} \quad Ax = b,\; Bx + d \geq 0$$

where $Q \in \mathbb{R}^{n \times n}$ with $Q \succeq 0$, $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $B \in \mathbb{R}^{k \times n}$, $d \in \mathbb{R}^k$

Theorem: strong duality, dual optimality, KKT characterization

Suppose $Q \succ 0$. (can be generalized to $Q \succeq 0$ in Exercises)

1. Strong duality holds and dual optimality is attained
2. A feasible x^* is optimal if and only if $\exists(\lambda^*, \mu^*)$ such that $\mu^* \geq 0$ and

$$x^* = Q^{-1}(A^\top \lambda^* + B^\top \mu^* - c), \qquad \mu^{*\top}(Bx^* + d) = 0$$

Such a point is a saddle point and a KKT point, (x^*, λ^*, μ^*) is primal-dual optimal and $f_2^* = f(x^*) = d(\lambda^*, \mu^*) = d^*$

Outline

1. Normal cones of feasible sets
2. CPC functions
3. Gradient and subgradient
4. Characterization: saddle point
5. Characterization: generalized KKT
6. Existence: primal optimum
7. Existence: dual optimum and strong duality
8. Special convex programs
    - General method
    - Linear program, convex quadratic program
    - Second-order cone program
    - Conic program

Second-order cone program (SOCP)

Consider

$$f^* \quad := \quad \min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad Ax = b,\ x \in K$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ and K is the standard second-order cone:

$$K \ := \ \{x \in \mathbb{R}^n : \|x^{n-1}\|_2 \ \le \ x_n\}$$

Second-order cone program (SOCP)

Consider

$$f^* \;:=\; \min_{x\in\mathbb{R}^n} f(x) \quad \text{s.t.} \quad Ax = b,\, x \in K$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ and K is the standard second-order cone:

$$K \;:=\; \{x \in \mathbb{R}^n : \|x^{n-1}\|_2 \;\le\; x_n\}$$

Lagrangian is

$$L(x,\lambda,\mu) \;:=\; f(x) \,-\, \lambda^\mathsf{T}(Ax-b) \,+\, \mu\left(\|x^{n-1}\|_2 - x_n\right), \quad x \in \mathbb{R}^n,\, \lambda \in \mathbb{R}^m, \mu \in \mathbb{R}$$

Dual function is $d(\mu) \;:=\; \min_{x\in\mathbb{R}^n} L(x,\mu)$ and dual problem is

$$d^* \;:=\; \max_{\mu\ge 0} d(\mu)$$

Let $X := \left\{x \in \mathbb{R}^n : Ax = b, \|x^{n-1}\|_2 \le x_n\right\}$, $Y := \left\{(\lambda,\mu) \in \mathbb{R}^{m+1} : \mu \ge 0\right\}$ be feasible sets

Second-order cone program (SOCP)

Consider

$$f^* \quad := \quad \min_{x \in \mathbb{R}^n} \; f(x) \quad \text{s.t.} \quad Ax = b, \; x \in K$$

Theorem: SOCP duality and KKT

1. Suppose f^* is finite and $\exists \bar{x}$ such that $A\bar{x} = b$ and $\|\bar{x}^{n-1}\|_2 < \bar{x}_n$. Then strong duality holds and dual optimality is attained, i.e., $\exists(\lambda^*, \mu^*) \in Y$ with $f^* = d^* = d(\lambda^*, \mu^*)$

2. Suppose $[x^*]^{n-1} \neq 0$. A feasible $(x^*, \lambda^*, \mu^*) \in X \times Y$ is primal-dual optimal and closes duality gap if and only if

(constraint function $h(x) := \|x^{n-1}\|_2 - x_n$ differentiable at x^*)

$$\nabla f(x^*) \;=\; A^\mathsf{T}\lambda^* + \mu^* \begin{bmatrix} -[x^*]^{n-1} \\ \|[x^*]^{n-1}\|_2 \end{bmatrix}, \qquad \mu^* \left(\|[x^*]^{n-1}\|_2 - x_n^* \right) \;=\; 0$$

Such a point is a saddle point and a KKT point, (x^*, λ^*, μ^*) is primal-dual optimal and $f_2^* = f(x^*) = d(\lambda^*, \mu^*) = d^*$

Second-order cone program (SOCP)

Derivation

Part 1: Slater theorem

Part 2: Constraint function $h(x) := \|x^{n-1}\|_2 - x_n$ is differentiable at x^*. Hence KKT condition can be derived by setting

$$\nabla_x L(x^*, \lambda^*, \mu^*) \ = \ 0$$

for stationarity, in addition to complementary slackness $\mu^* \left(\|[x^*]^{n-1}\|_2 - x_n^* \right) = 0$

Second-order cone program (SOCP)

Consider

$$f^* \quad := \quad \min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad Ax = b,\, x \in K$$

If $[x^*]^{n-1} = 0$, then the constraint function $h(x) := \|x^{n-1}\|_2 - x_n$ is not differentiable at x^*

Need nonsmooth analysis to derive KKT condition

Theorem: SOCP duality and KKT

3. Suppose $[x^*]^{n-1} = 0$. Suppose $\exists \bar{x}$ s.t. $A\bar{x} = b,\ \|\bar{x}^{n-1}\|_2 < \bar{x}_n$ (Slater condition).

   (a) Case $x_n^* > \|[x^*]^{n-1}\|_2 = 0$: x^* is optimal iff $\partial f(x^*) \ni A^\mathsf{T}\lambda^*$ for some $\lambda^* \in \mathbb{R}^m$

   (b) Case $x_n^* = \|[x^*]^{n-1}\|_2 = 0$: $x^* = 0$ is optimal iff

   $$\partial f(0) \ni A^\mathsf{T}\lambda^* + \eta^* \text{ for some } \lambda^* \in \mathbb{R}^m,\ \eta^* \in K$$

Second-order cone program (SOCP)
Derivation

Part 3: Constraint function $h(x) := \|x^{n-1}\|_2 - x_n$ is nondifferentiable at x^*. Use generalized KKT Theorem which requires the Slater condition.

(a) Rewrite SOCP as unconstrained optimization:

$$\min_{x \in \mathbb{R}^n} \; f(x) \; + \; \delta_H(x) \; + \; \delta_K(x)$$

where $H := \{x \in \mathbb{R}^n : Ax = b\}$ and $K := \{x : \|x^{n-1}\|_2 \leq x_n\}$

(b) If Slater condition holds ($\exists \bar{x}$ s.t. $A\bar{x} = b, \; \|\bar{x}^{n-1}\|_2 < \bar{x}_n$), then x^* is optimal if and only if $0 \in \partial f(x^*) + N_H(x^*) + N_K(x^*)$, i.e., $\exists \xi^* \in \partial f(x^*)$ such that

$$\xi^* \in - N_H(x^*) - N_K(x^*)$$

Second-order cone program (SOCP)

Derivation

(c) Normal cones:

$$N_H(x^*) = \{A^\top \lambda \in \mathbb{R}^n : \lambda \in \mathbb{R}^m\}$$

$$N_K(x^*) = \begin{cases} \{\xi \in \mathbb{R}^n : \|\xi^{n-1}\|_2 \leq -\xi_n\} & \text{if } x^* = 0 \\ \{0 \in \mathbb{R}^n\} & \text{if } \|[x^*]^{n-1}\|_2 < x_n^* \\ \left\{\mu([x^*]^{n-1}, -x_n^*) \in \mathbb{R}^n : \mu \geq 0\right\} & \text{if } \|[x^*]^{n-1}\|_2 = x_n^* > 0 \end{cases}$$

Substitute into optimality condition: $\exists \xi^* \in \partial f(x^*)$ such that $\xi^* \in -N_H(x^*) - N_K(x^*)$

Second-order cone program (SOCP)

Derivation

(d) A feasible x^* is optimal if and only if $\exists \xi^* \in \partial f(x^*),\ \lambda^* \in \mathbb{R}^m$ and

(i) Case $x_n^* > \|[x^*]^{n-1}\|_2 \geq 0$: such that $\xi^* = A^\top \lambda^*$

(constraint function $h(x) := \|x^{n-1}\|_2 - x_n$ nondifferentiable if $[x^*]^{n-1} = 0$)

Second-order cone program (SOCP)

Derivation

(d) A feasible x^* is optimal if and only if $\exists \xi^* \in \partial f(x^*),\ \lambda^* \in \mathbb{R}^m$ and

- (i) Case $x_n^* > \|[x^*]^{n-1}\|_2 \geq 0$: such that $\xi^* = A^\top \lambda^*$
- (ii) Case $x_n^* = \|[x^*]^{n-1}\|_2 > 0$: $\exists \mu^* \in \mathbb{R}_+$ such that $\xi^* = A^\top \lambda^* + \mu^* \begin{bmatrix} -[x^*]^{n-1} \\ x_n^* \end{bmatrix}$

(this is the smooth case)

Second-order cone program (SOCP)

Derivation

(d) A feasible x^* is optimal if and only if $\exists \xi^* \in \partial f(x^*),\ \lambda^* \in \mathbb{R}^m$ and

(i) Case $x_n^* > \|[x^*]^{n-1}\|_2 \geq 0$: such that $\xi^* = A^\top \lambda^*$

(ii) Case $x_n^* = \|[x^*]^{n-1}\|_2 > 0$: $\exists \mu^* \in \mathbb{R}_+$ such that $\xi^* = A^\top \lambda^* + \mu^* \begin{bmatrix} -[x^*]^{n-1} \\ x_n^* \end{bmatrix}$

(iii) Case $x_n^* = \|[x^*]^{n-1}\|_2 = 0$: $\exists \eta^* \in \mathbb{K}$ such that $\xi^* = A^\top \lambda^* + \eta^*$

(constraint function $h(x) := \|x^{n-1}\|_2 - x_n$ nondifferentiable at $x^* = 0$)

Remark

In all 3 cases, conditions are of the form $\xi^* = A^\top \lambda^* + \eta^*$ for some $\eta^* \in \mathbb{K}$

$\left(\eta^* = 0 \text{ or } \eta^* = \mu^* \begin{bmatrix} -[x^*]^{n-1} \\ x_n^* \end{bmatrix} \in K \right)$

Second-order cone program (SOCP)

SOC constraint

Consider

$$f^* \quad := \quad \min_{x \in \mathbb{R}^n} f(x) \qquad \text{s.t.} \qquad Ax = b, \ \|Bx + d\|_2 \le \beta^\top x + \delta$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $B \in \mathbb{R}^{(l-1) \times n}$, $d \in \mathbb{R}^{l-1}$, $\beta \in \mathbb{R}^n$, $\delta \in \mathbb{R}$

Second-order cone program (SOCP)

SOC constraint

Consider

$$f^* \quad := \quad \min_{x \in \mathbb{R}^n} \; f(x) \quad \text{s.t.} \quad Ax = b, \; \|Bx + d\|_2 \leq \beta^\mathsf{T} x + \delta$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $B \in \mathbb{R}^{(l-1) \times n}$, $d \in \mathbb{R}^{l-1}$, $\beta \in \mathbb{R}^n$, $\delta \in \mathbb{R}$

Rewrite as standard SOCP:

$$f^* \quad := \quad \min_{(x,z) \in \mathbb{R}^{n+l}} \; f(x) \quad \text{s.t.} \quad Ax = b, \; z = \tilde{B}x + \tilde{d}, \; z \in K$$

where $z = (z^{l-1}, z_l) \in \mathbb{R}^l$, $K := \{x : \|x^{n-1}\|_2 \leq x_n\}$ and

$$\tilde{B} := \begin{bmatrix} B \\ \beta^\mathsf{T} \end{bmatrix}, \qquad \tilde{d} := \begin{bmatrix} d \\ \delta \end{bmatrix}$$

Second-order cone program (SOCP)

SOC constraint

Rewrite as standard SOCP:

$$f^* \;:=\; \min_{(x,z)\in\mathbb{R}^{n+l}} \; f(x) \quad \text{s.t.} \quad Ax = b,\; z = \tilde{B}x + \tilde{d},\; z \in K$$

Lagrangian:

$$L(x, z, \lambda, \gamma, \mu) \;:=\; f(x) \,-\, \lambda^\mathsf{T}(Ax - b) \,-\, \gamma^\mathsf{T}(\tilde{B}x + \tilde{d} - z) \,+\, \mu \left(\|z^{l-1}\|_2 - z_l\right)$$

Dual problem:

$$d^* \;:=\; \max_{\lambda,\gamma} \left(b^\mathsf{T}\lambda - \tilde{d}^\mathsf{T}\gamma\right) \,+\, d_0(\lambda, \gamma) \qquad \text{s.t.} \qquad \gamma \in K$$

where $d_0(\lambda, \gamma) := \min_{x\in\mathbb{R}^n} \left(f(x) - (A^\mathsf{T}\lambda + \tilde{B}^\mathsf{T}\gamma)^\mathsf{T} x\right)$

E.g. If $f(x) := c^\mathsf{T} x$, then $d^* \;:=\; \max_{(\lambda,\gamma)\in\mathbb{R}^{m+l}} \; b^\mathsf{T}\lambda - \tilde{d}^\mathsf{T}\gamma \quad \text{s.t.} \quad A^\mathsf{T}\lambda + \tilde{B}^\mathsf{T}\gamma = c,\; \|\gamma^{l-1}\|_2 \le \gamma_l$

Second-order cone program (SOCP)

SOC constraint

Rewrite as standard SOCP:

$$f^* \quad := \quad \min_{(x,z)\in\mathbb{R}^{n+l}} \; f(x) \quad \text{s.t.} \quad Ax = b,\; z = \tilde{B}x + \tilde{d},\; z \in K$$

Lagrangian:

$$L(x, z, \lambda, \gamma, \mu) \;:=\; f(x) \;-\; \lambda^\mathsf{T}(Ax - b) \;-\; \gamma^\mathsf{T}(\tilde{B}x + \tilde{d} - z) \;+\; \mu\left(\|z^{l-1}\|_2 - z_l\right)$$

Dual problem:

$$d^* \quad := \quad \max_{\lambda,\gamma} \left(b^\mathsf{T}\lambda - \tilde{d}^\mathsf{T}\gamma\right) \;+\; d_0(\lambda, \gamma) \qquad \text{s.t.} \qquad \gamma \in K$$

where $d_0(\lambda, \gamma) := \min_{x\in\mathbb{R}^n} \left(f(x) - (A^\mathsf{T}\lambda + \tilde{B}^\mathsf{T}\gamma)^\mathsf{T} x\right)$

Let $X := \left\{x \in \mathbb{R}^n : Ax = b, \|Bx + d\|_2 \le \beta^\mathsf{T}x + \delta\right\},\; Y := \left\{(\lambda, \mu) \in \mathbb{R}^{m+1} : \mu \ge 0\right\}$

Second-order cone program (SOCP)

SOC constraint

Rewrite as standard SOCP:

$$f^* \quad := \quad \min_{(x,z)\in\mathbb{R}^{n+l}} \; f(x) \quad \text{s.t.} \quad Ax = b, \; z = \tilde{B}x + \tilde{d}, \; z \in K$$

Theorem: SOCP duality and KKT

Suppose $\exists \bar{x}$ s.t. $A\bar{x} = b, \; \|B\bar{x} + d\|_2 < \beta^\mathsf{T} x + \delta$ (Slater condition).

1. Suppose f^* is finite. Then strong duality holds and dual optimality is attained.
2. Suppose $Bx^* + d \neq 0$. A point $x^* \in X$ is optimal if and only if $\exists(\lambda^*, \mu^*) \in Y$ such that

$$\nabla f(x^*) = \; A^\mathsf{T}\lambda^* + \mu^* \left(-B^\mathsf{T}(Bx^* + d) \; + \; \beta\|Bx^* + d\|_2\right)$$

$$0 = \; \mu^* \left(\|Bx^* + d\|_2 - (\beta^\mathsf{T} x^* + \delta)\right)$$

Such a point is a saddle point and a KKT point that closes duality gap.

(constraint function $h(z) := \|z^{l-1}\|_2 - z_l$ differentiable at x^*)

Second-order cone program (SOCP)

SOC constraint: derivation

Part 1: Slater theorem

Part 2: Constraint function $h(z) := \|z^{l-1}\|_2 - z_l$ is differentiable at (x^*, z^*). Hence KKT condition can be derived by setting

$$\nabla_{x,z} L(x^*, z^*, \lambda^*, \gamma^*, \mu^*) \ = \ 0$$

for stationarity, in addition to complementary slackness $\mu^* \left(\|[z^*]^{l-1}\|_2 - z_l^* \right) = 0$

Second-order cone program (SOCP)

SOC constraint

Rewrite as standard SOCP:

$$f^* \quad := \quad \min_{(x,z)\in\mathbb{R}^{n+l}} \; f(x) \quad \text{s.t.} \quad Ax = b,\ z = \tilde{B}x + \tilde{d},\ z \in K$$

If $z^{l-1} := Bx^* + d = 0$, then the constraint function $h(z) := \|z^{l-1}\|_2 - z_l$ is not differentiable at (x^*, z^*). Need nonsmooth analysis to derive KKT condition.

Second-order cone program (SOCP)

SOC constraint

Rewrite as standard SOCP:

$$f^* \quad := \quad \min_{(x,z)\in\mathbb{R}^{n+l}} \; f(x) \quad \text{s.t.} \quad Ax = b,\; z = \tilde{B}x + \tilde{d},\; z \in K$$

If $z^{l-1} := Bx^* + d = 0$, then the constraint function $h(z) := \|z^{l-1}\|_2 - z_l$ is not differentiable at (x^*, z^*). Need nonsmooth analysis to derive KKT condition.

Theorem: SOCP duality and KKT

3. Suppose $Bx^* + d = 0$. Suppose $\exists \bar{x}$ s.t. $A\bar{x} = b,\; \|B\bar{x} + d\|_2 < \beta^\mathsf{T}\bar{x} + \delta$ (Slater condition).
   (a) Case $\beta^\mathsf{T}x^* + \delta > 0$: x^* is optimal iff $\partial f(x^*) \ni A^\mathsf{T}\lambda^*$ for some $\lambda^* \in \mathbb{R}^m$
   (b) Case $\beta^\mathsf{T}x^* + \delta = 0$: $x^* = 0$ is optimal iff

$$\partial f(0) \ni A^\mathsf{T}\lambda^* + \tilde{B}^\mathsf{T}\eta^* \text{ for some } \lambda^* \in \mathbb{R}^m,\; \eta^* \in K$$

(constraint function $h(z) := \|z^{l-1}\|_2 - z_l$ nondifferentiable at (x^*, z^*))

Second-order cone program (SOCP)

SOC constraint: derivation

Part 3: Constraint function $h(z) := \|z^{l-1}\|_2 - z_l$ is nondifferentiable at (x^*, z^*). Use generalized KKT Theorem which requires the Slater condition.

(a) Rewrite as unconstrained optimization:

$$\min_{(x,z)\in\mathbb{R}^{n+l}} \quad f(x) \ + \ \delta_{\tilde{H}_1}(x,z) \ + \ \delta_{\tilde{K}}(x,z) \ + \ \delta_{H_2}(x,z)$$

where

$$\tilde{H}_1 := \{(x,z)\in\mathbb{R}^{n+l} : Ax = b\} \ =: \ H_1 \times \mathbb{R}^l, \qquad H_1 := \{x \in \mathbb{R}^n : Ax = b\}$$

$$\tilde{K} := \{(x,z)\in\mathbb{R}^{n+l} : \|z^{l-1}\|_2 \le z_l\} \ =: \ \mathbb{R}^n \times K, \qquad K := \{z \in \mathbb{R}^l : \|z^{l-1}\|_2 \le z_l$$

$$H_2 := \{(x,z)\in\mathbb{R}^{n+l} : z = \tilde{B}x + \tilde{d}\}$$

and

$$N_{\tilde{H}_1}(x,z) = N_{H_1}(x) \times \{0 \in \mathbb{R}^l\}, \quad N_{\tilde{K}}(x,z) = \{0 \in \mathbb{R}^n\} \times N_K(z)$$

Second-order cone program (SOCP)

SOC constraint: derivation

(b) If Slater condition holds ($\exists \bar{x}$ s.t. $A\bar{x} = b,\ \|B\bar{x} + d\|_2 < \beta^\top \bar{x} + \delta$), then a feasible (x^*, z^*) is optimal if and only if $\exists \xi^* \in \partial f(x^*)$ such that

$$\begin{bmatrix} \xi^* \\ 0 \end{bmatrix} \in -N_{\tilde{H}_1}(x^*, z^*) - N_{\tilde{K}}(x^*, z^*) - N_{H_2}(x^*, z^*)$$

$$= -\begin{bmatrix} N_{H_1}(x^*) \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ N_K(z^*) \end{bmatrix} - N_{H_2}(x^*, z^*)$$

Second-order cone program (SOCP)

Derivation

(c) Normal cones:

$$N_{H_1}(x^*) = \{A^\top \lambda \in \mathbb{R}^n : \lambda \in \mathbb{R}^m\}$$

$$N_K(z^*) = \begin{cases} \{\eta \in \mathbb{R}^l : \|\eta^{l-1}\|_2 \le -\eta_l\} & \text{if } z^* = 0 \\ \{0 \in \mathbb{R}^l\} & \text{if } \|[z^*]^{l-1}\|_2 < z_l^* \\ \left\{\mu([z^*]^{l-1}, -z_l^*) \in \mathbb{R}^l : \mu \ge 0\right\} & \text{if } \|[z^*]^{l-1}\|_2 = z_l^* > 0 \end{cases}$$

$$N_{H_2}(x^*, z^*) = \left\{(\tilde{B}^\top \gamma, -\gamma) \in \mathbb{R}^{n+l} : \gamma \in \mathbb{R}^l\right\}$$

Substitute into optimality condition: $\exists \xi^* \in \partial f(x^*)$ such that

$$\begin{bmatrix} \xi^* \\ 0 \end{bmatrix} \in -\begin{bmatrix} N_{H_1}(x^*) \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ N_K(z^*) \end{bmatrix} - N_{H_2}(x^*, z^*)$$

Second-order cone program (SOCP)

Derivation

(d) A feasible x^* is optimal if and only if $\exists \xi^* \in \partial f(x^*),\ \lambda^* \in \mathbb{R}^m$ and

(i) Case $\beta^\top x^* + \delta > \|Bx^* + d\|_2 \geq 0$: such that $\xi^* = A^\top \lambda^*$ $(\gamma^* = 0)$

(constraint function $h(z) := \|z^{l-1}\|_2 - z_l$ nondifferentiable if $[z^*]^{l-1} = 0$)

Second-order cone program (SOCP)

Derivation

(d) A feasible x^* is optimal if and only if $\exists \xi^* \in \partial f(x^*),\ \lambda^* \in \mathbb{R}^m$ and

(i) Case $\beta^\top x^* + \delta > \|Bx^* + d\|_2 \geq 0$: such that $\xi^* = A^\top \lambda^*$ $(\gamma^* = 0)$

(ii) Case $\beta^\top x^* + \delta = \|Bx^* + d\|_2 > 0$: $\exists \mu^* \in \mathbb{R}_+$ such that

$$\xi^* \ = \ A^\top \lambda^* \ + \ \mu^* \left(-B^\top (Bx^* + d) + \beta(\beta^\top x^* + \delta) \right)$$

(this is the smooth case)

Second-order cone program (SOCP)

Derivation

(d) A feasible x^* is optimal if and only if $\exists \xi^* \in \partial f(x^*),\ \lambda^* \in \mathbb{R}^m$ and

(i) Case $\beta^\mathsf{T} x^* + \delta > \|Bx^* + d\|_2 \geq 0$: such that $\xi^* = A^\mathsf{T}\lambda^*\ (\gamma^* = 0)$

(ii) Case $\beta^\mathsf{T} x^* + \delta = \|Bx^* + d\|_2 > 0$: $\exists \mu^* \in \mathbb{R}_+$ such that

(i) $\xi^* \ = \ A^\mathsf{T}\lambda^* \ + \ \mu^* \left(-B^\mathsf{T}(Bx^* + d) + \beta(\beta^\mathsf{T} x^* + \delta)\right)$

(iii) Case $\beta^\mathsf{T} x^* + \delta = \|Bx^* + d\|_2 = 0$: $\exists \eta^* \in \mathbb{K}$ such that $\xi^* \ = \ A^\mathsf{T}\lambda^* + \tilde{B}^\mathsf{T}\eta^*$

(constraint function $h(z) := \|z^{l-1}\|_2 - z_l$ nondifferentiable at $z^* = 0$)

Remark

In all 3 cases, conditions are of the form $\xi^* = A^\mathsf{T}\lambda^* + \tilde{B}^\mathsf{T}\eta^*$ for some $\eta^* \in \mathbb{K}$

$$\left(\eta^* = 0 \text{ or } \eta^* = \mu^* \begin{bmatrix} -[z^*]^{l-1} \\ z_k^* \end{bmatrix} \in K\right)$$

Outline

1. Normal cones of feasible sets
2. CPC functions
3. Gradient and subgradient
4. Characterization: saddle point
5. Characterization: generalized KKT
6. Existence: primal optimum
7. Existence: dual optimum and strong duality
8. Special convex programs
    - General method
    - Linear program, convex quadratic program
    - Second-order cone program
    - Conic program

Conic program

Conic feasible set

Consider

$$f^* \quad := \quad \min_{x \in \mathbb{R}^n} \; f(x) \quad \text{s.t.} \quad Ax = b,\ x \in K$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $K \subseteq \mathbb{R}^n$ is a closed convex cone

Conic program

Conic feasible set

Consider

$$f^* \quad := \quad \min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad Ax = b,\ x \in K$$

where $f : \mathbb{R}^n \to \mathbb{R},\ A \in \mathbb{R}^{m \times n},\ b \in \mathbb{R}^m,\ K \subseteq \mathbb{R}^n$ is a closed convex cone

Lagrangian:

$$L(x, \lambda, \mu) \ := \ f(x) - \lambda^\top (Ax - b) - \mu^\top x, \quad x \in \mathbb{R}^n,\ \lambda \in \mathbb{R}^m,\ \mu \in K^* \subseteq \mathbb{R}^n$$

Dual function:

$$d(\lambda, \mu) \ := \ \min_{x \in \mathbb{R}^n} L(x, \lambda, \mu) \ = \ \lambda^\top b + d_0(\lambda, \mu), \quad \lambda \in \mathbb{R}^m,\ \mu \in K^* \subseteq R^n$$

where $d_0(\lambda, \mu) := \min_{x \in \mathbb{R}^n} \left(f(x) - (A^\top \lambda + \mu)^\top x \right)$

Dual problem:

$$d^* \quad := \quad \max_{\lambda \in \mathbb{R}^m,\, \mu \in K^*} \ \lambda^\top b + d_0(\lambda, \mu)$$

Conic program

Conic feasible set

To derive KKT condition, rewrite as unconstrained optimization

$$\min_{x \in \mathbb{R}^n} \quad f(x) + \delta_H(x) + \delta_K(x)$$

where $H := \{x \in \mathbb{R}^n : Ax = b\}$

Under Slater condition ($\exists \bar{x} \in \text{ri}(K)$ s.t. $A\bar{x} = b$), generalized KKT theorem implies: x^* is optimal if and only if $\exists \xi^* \in \partial f(x^*)$ such that

$$-\xi^* \;\in\; N_H(x^*) + N_K(x^*)$$

where

$$N_H(x^*) = \;\{A^\top \lambda \in \mathbb{R}^n : \lambda \in \mathbb{R}^m\}$$

$$N_K(x^*) = \;\{\tilde{\mu} \in K^\circ \subseteq \mathbb{R}^n : \tilde{\mu}^\top x^* = 0\}$$

Conic program

Conic feasible set

Consider $f^* \quad := \quad \min_{x \in \mathbb{R}^n} f(x) \quad$ s.t. $\quad Ax = b,\, x \in K$

Theorem: Conic duality and KKT

Suppose $\exists \bar{x} \in \mathrm{ri}(K)$ s.t. $A\bar{x} = b$ (Slater condition).

1. Suppose f^* is finite. Then strong duality holds and dual optimality is attained.
2. A feasible x^* is optimal if and only if $\exists \xi^* \in \partial f(x^*)$ and $(\lambda^*, \mu^*) \in \mathbb{R}^m \times K^*$ such that

$$\xi^* \;=\; A^\mathsf{T}\lambda^* + \mu^*, \qquad \mu^{*\mathsf{T}} x^* = 0$$

Such a point (x^*, λ^*, μ^*) is a saddle point and a KKT point that closes the duality gap.

Conic program

Conic constraint

Consider

$$f^* \quad := \quad \min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad Ax = b, \; Bx + d \in K$$

feasible may not be cone

where $f : \mathbb{R}^n \to \mathbb{R}$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $B \in \mathbb{R}^{l \times n}$, $d \in \mathbb{R}^l$, $K \subseteq \mathbb{R}^l$ is a closed convex cone

Dual problem:

$$d^* \; := \max_{(\lambda,\mu) \in \mathbb{R}^{m+l}} d(\lambda, \mu) \; := \; \left(b^\mathsf{T}\lambda - d^\mathsf{T}\mu\right) + d_0(\lambda, \mu) \quad \text{s.t.} \quad \mu \in K^* \subseteq \mathbb{R}^l$$

Conic program

Conic constraint

Consider $f^* \quad := \quad \min_{x \in \mathbb{R}^n} f(x) \quad$ s.t. $\quad Ax = b,\ Bx + d \in K$

Theorem: Conic duality and KKT

Suppose $\exists \bar{x}$ s.t. $A\bar{x} = b,\ B\bar{x} + d \in$ ri(K) (Slater condition).

1. Suppose f^* is finite. Then strong duality holds and dual optimality is attained.
2. A feasible x^* is optimal if and only if $\exists \xi^* \in \partial f(x^*)$ and $(\lambda^*, \mu^*) \in \mathbb{R}^m \times K^*$ such that

$$\xi^* \ = \ A^\top \lambda^* + B^\top \mu^*, \qquad \mu^{*\top}(Bx^* + d) = 0$$

   Such a point (x^*, λ^*, μ^*) is a saddle point and a KKT point that closes the duality gap.