LaTeX4EI
Convex Optimization

1. Convex Sets

1.1. Combinations

$$\boldsymbol{z} = a\boldsymbol{x} + b\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$$

linear	$a, b \in \mathbb{R}$
conic	$a, b \geq 0$
affine	$a + b = 1$
convex	$a + b = 1, a, b \geq 0$

1.2. Sets

Set	affine	conic	convex	Examples
Linear Space	✓	✓	✓	$\mathbb{R}^n$
Convex Cone	✗	✓	✓	$\{\boldsymbol{p} \mid \boldsymbol{p}^\mathsf{T}(\boldsymbol{x} - \overline{\boldsymbol{x}}) \leq 0\}$
Affine Set	✓	✗	✓	$\{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}, \emptyset, \{x_0\}$
Convex Set	✗	✗	✓	$\{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}, \boldsymbol{b} \neq \boldsymbol{0}\}$

1.3. Cones

Polar Cone

$$\mathcal{X}_p = \{\boldsymbol{p} \mid \boldsymbol{x}^\mathsf{T}\boldsymbol{p} \leq 0, \forall \boldsymbol{x} \in \mathcal{X}\}$$

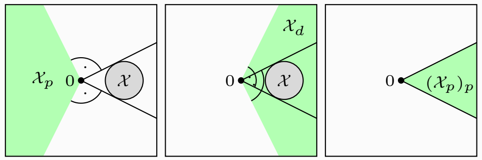
Dual Cone

$$\mathcal{X}_d = \{\boldsymbol{y} \mid \boldsymbol{x}^\mathsf{T}\boldsymbol{y} \geq 0, \forall \boldsymbol{x} \in \mathcal{X}\}$$

Conic facts

1. $\mathcal{X}_p$ is a closed convex cone
2. $\mathcal{X} \subset (\mathcal{X}_p)_p$
3. If $\mathcal{X}$ closed and convex: $(\mathcal{X}_d)_d = (\mathcal{X}_p)_p = \mathcal{X}$
4. Cones can be nonconvex

Geometric interpretation



Polar | Dual | Polar of Polar

1.4. Important Convex Sets

1.4.1. Levelsets, Epigraph, Hypograph

$f(x)$ convex $\iff$ epi$(f(x))$ convex
$f(x)$ concave $\iff$ hypo$(f(x))$ convex
$f(x)$ convex $\rightarrow \{\boldsymbol{x} \mid f(\boldsymbol{x}) \leq \alpha\}$ convex
$\{\boldsymbol{x} \mid f(\boldsymbol{x}) \leq \alpha\}$ convex $\not\rightarrow f(x)$ convex, example: $f(\boldsymbol{x}) = -e^x$

1.4.2. Convex Hull

$\text{conv}(\mathcal{X}) = \{\boldsymbol{x} \mid \boldsymbol{x} = \sum_{i=1}^K \lambda_i x_i, \lambda_i \geq 0, \sum_{i=1}^K \lambda_i = 1\}$

1.4.3. Conic Hull

$\text{cone}(\mathcal{X}) = \{\boldsymbol{x} \mid \boldsymbol{x} = \sum_{i=1}^K \lambda_i x_i, \lambda_i \geq 0\}$

1.4.4. Polytopes

$\mathcal{X} = \{\boldsymbol{x} \mid \boldsymbol{a}_1^\mathsf{T}\boldsymbol{x} \leq b1, \boldsymbol{a}_2^\mathsf{T}\boldsymbol{x} \leq b2 \dots\} = \{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\}$

1.4.5. Simplexes

a Polytope in $\mathbb{R}^n$ defined by $n+1$ inequalities

1.5. Caratheodory Theorem

Any $\boldsymbol{x} \in \text{conv}(\mathcal{X})$ with $\boldsymbol{x} \in \mathbb{R}^n$ can be represented by a convex combination of $n+1$ points $x_i \in \mathcal{X}$ as:

$$\boldsymbol{x} = \sum_{i=1}^{n+1} \lambda_i \boldsymbol{x_i}, \boldsymbol{x_i} \in \mathcal{X}, \lambda_i \geq 0, \sum_{i=1}^{n+1} \lambda_i = 1$$

1.6. Closest Point Theorem

If $\mathcal{X}$ **is convex** then there always exists a **unique** $\boldsymbol{x}^* \in \mathcal{X}$ with a minimum distance to $\boldsymbol{y} \in \mathbb{R}^n$:

$$(\boldsymbol{y} - \boldsymbol{x}^*)^\mathsf{T}(\boldsymbol{x} - \boldsymbol{x}^*) \leq 0, \forall \boldsymbol{x} \in \mathcal{X}$$

1.7. Fundamental Separation Theorem

(follows from Closest Point Theorem)
If $\mathcal{X}$ **is convex, non-empty and closed**, then there exists a normal vector $\boldsymbol{p}$ and scalar α such that:

$$\boldsymbol{p}^\mathsf{T}\boldsymbol{x} > \alpha \forall x \notin \mathcal{X}$$
$$\boldsymbol{p}^\mathsf{T}\boldsymbol{x} \leq \alpha \forall x \in \mathcal{X}$$

Corollary:
$\mathcal{X}$ is the intersection of all closed half-spaces containing $\mathcal{X}$

1.8. Supporting Hyperplane Theorem

If $\mathcal{X}$ **is convex, non-empty and closed** with $\bar{\boldsymbol{x}} \in \partial\mathcal{X}$ then there exists a hyperplane $\mathcal{H}$ defined by a normal vector $\boldsymbol{p}$ such that:

$$\mathcal{H} = \{\boldsymbol{x} \mid \boldsymbol{p}^\mathsf{T}(\boldsymbol{x} - \bar{\boldsymbol{x}}) \leq 0 \forall \boldsymbol{x} \in \mathcal{X}\}$$

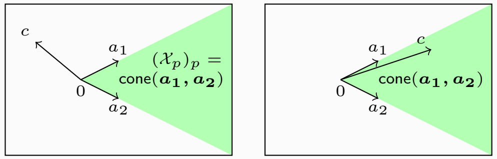
where $\mathcal{H}$ supports $\mathcal{X}$ at $\bar{\boldsymbol{x}}$

1.9. Farkas Theorem

(follows from Fundamental Separation Theorem)
Exactly one of the following statements is true:

1. $\exists \boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{0}, \boldsymbol{c}^\mathsf{T}\boldsymbol{x} > 0$
2. $\exists \boldsymbol{y} \in \mathbb{R}^m \mid \boldsymbol{A}^\mathsf{T}\boldsymbol{y} = \boldsymbol{c}, \boldsymbol{y} \leq \boldsymbol{0}$

Geometric interpretation



(1) has a solution
$c \notin \text{cone}(a_1, a_2)$

(2) has a solution
$c \in \text{cone}(a_1, a_2)$

1.9.1. Gordan's Theorem

Exactly one of the following statements is true:

1. $\exists \boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} < \boldsymbol{0}$
2. $\exists \boldsymbol{y} \in \mathbb{R}^m \mid \boldsymbol{A}^\mathsf{T}\boldsymbol{y} = \boldsymbol{c}, \boldsymbol{y} \geq \boldsymbol{0}, \boldsymbol{y} \neq \boldsymbol{0}$

2. Convex Functions

2.1. Convexity Preserving Operations

The function $f(\boldsymbol{x})$ is convex if it is a...

1. **conic combination** of convex functions:
$$f(\boldsymbol{x}) = \sum_{i=1}^K \lambda_i f_i(\boldsymbol{x}), \lambda_i \geq 0$$
2. composite of a **nondecreasing, univariate, convex** $g(\boldsymbol{x})$ and a **convex** $h(\boldsymbol{x})$:
$$f(\boldsymbol{x}) = g(h(\boldsymbol{x}))$$
3. **pointwise maximum** of convex functions:
$$f(\boldsymbol{x}) = \max_{i=1,\dots,K}\{f_1(\boldsymbol{x}), \dots, f_K(\boldsymbol{x})\}$$
4. **negative of a concave** function $g(\boldsymbol{x})$:
$$f(\boldsymbol{x}) = -g(\boldsymbol{x})$$

2.2. Checking for convexity

2.2.1. Definition of a convex function

A function $f(x)$ is convex if:

$$f(\lambda \boldsymbol{x_1} + (1-\lambda)\boldsymbol{x_2}) \leq \lambda f(\boldsymbol{x_1}) + (1-\lambda) f(\boldsymbol{x_2})$$
$$\forall \lambda \in [0, 1]$$

2.2.2. First Order criterion

A $f(x)$ which is **at least once differentiable** is convex if:

$$f(\bar{\boldsymbol{x}}) \nabla f^\mathsf{T}(\bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}}) \leq f(\boldsymbol{x}) \qquad \forall \boldsymbol{x}, \bar{\boldsymbol{x}}$$

2.2.3. Second Order criterion

A $f(x)$ which is **at least twice differentiable** is convex if:

$$0 \leq \nabla^2 f(\boldsymbol{x}) \qquad \forall \boldsymbol{x}$$

Remark: For concavity $\geq$ must hold

2.3. Quasiconvexity

A function $f(x)$ is quasiconvex if for a $\lambda \in (0, 1)$:

$$f(\lambda \boldsymbol{x_1} + (1-\lambda)\boldsymbol{x_2}) \leq \max\{f(\boldsymbol{x_1}), f(\boldsymbol{x_2})\}$$

and strictly quasiconvex if it hold with strict equality.
Interpretation: f remains nondecreasing in an increasing direction.
Quasiconvex function may not have a unique global minimizer

2.4. Pseudoconvexity

A function $f(\boldsymbol{x})$ is pseudoconvex if for $\boldsymbol{x_1}, \boldsymbol{x_2} \in \text{dom}(f)$:

$$\nabla f^\mathsf{T}(\boldsymbol{x_1})(\boldsymbol{x_2} - \boldsymbol{x_1}) \geq 0 \rightarrow f(\boldsymbol{x_2}) \geq f(\boldsymbol{x_1})$$

and strictly pseudoconvex if it holds with strict equality for $\boldsymbol{x_1} \neq \boldsymbol{x_2}$.
Interpretation: f is nondecreasing for all gradient directions.
Strict pseudoconvex functions have a unique global minimizer

3. Linear Programming

3.1. Extreme Points

$\boldsymbol{x}_{(\text{EP})} \in \mathcal{X}$ can't be represented by $\boldsymbol{x_1}, \boldsymbol{x_2} \in \mathcal{X}$ for any $\lambda \in (0, 1)$:

$$\boldsymbol{x}_{(\text{EP})} = \lambda \boldsymbol{x_1} + (1-\lambda)\boldsymbol{x_2} \quad \rightarrow \quad \boldsymbol{x}^{(\text{EP})} = \boldsymbol{x_1} = \boldsymbol{x_2}$$

Interpretation: Corner points or curve borders of a convex compact set.

3.2. Extreme Directions

$\boldsymbol{d}_{(\text{ED})}$ can't be represented by a positive linear combination of directions:

$$\boldsymbol{d}_{(\text{ED})} = \lambda_1 \boldsymbol{d_1} + \lambda_2 \boldsymbol{d_2} \quad \rightarrow \quad \boldsymbol{d_1} = \alpha \boldsymbol{d_2}$$
$$\lambda_1, \lambda_2, \alpha > 0$$

Interpretation: Direction in which the set is unbounded.

3.3. Polyhedral Sets (Polyhedron)

Intersection of half-spaces, **always closed and convex**.

Primal Standard Form

$$\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \geq 0\}$$

Dual Standard Form

$$\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\}$$

3.4. EPs and EDs of Polyhedra

Requirements for EPs: $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, rank$(\boldsymbol{A}) = m$ and $n \geq m$

$$\boldsymbol{A}\boldsymbol{x} = \begin{bmatrix}\boldsymbol{B} & \boldsymbol{N}\end{bmatrix}\begin{bmatrix}\boldsymbol{x_B} \\ \boldsymbol{x_N}\end{bmatrix}, \quad \boldsymbol{B} \in \mathbb{R}^{m \times m}, \boldsymbol{N} \in \mathbb{R}^{m \times (n-m)}$$

Extreme point:

$$\boldsymbol{x}_{(\text{EP})} = \begin{bmatrix}\boldsymbol{B}^{-1}\boldsymbol{b} \\ \boldsymbol{0}_{n-m}\end{bmatrix} \geq \boldsymbol{0}$$

Extreme direction:

$$\boldsymbol{d}_{(\text{ED})} = \begin{bmatrix}-\boldsymbol{B}^{-1}\boldsymbol{N}\boldsymbol{e_j} \\ \boldsymbol{e_j}\end{bmatrix} \geq \boldsymbol{0}$$

Finding all EPs and EDs:
There are **at most** $\binom{n}{m}$ EPs and **at most** $(n-m)\binom{n}{m}$ EDs.
Permute Problem with $\boldsymbol{\Pi}$ with $\boldsymbol{\Pi}^\mathsf{T}\boldsymbol{\Pi} = \boldsymbol{I}$:

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}\boldsymbol{\Pi}\boldsymbol{\Pi}^\mathsf{T}\boldsymbol{x} = \boldsymbol{A}^\Pi \boldsymbol{x}^\Pi$$

Example permutation: $\boldsymbol{a}_i$ is the i-th column of $\boldsymbol{A}$

$$\begin{bmatrix}\boldsymbol{a_1} & \boldsymbol{a_2} & \boldsymbol{a_3}\end{bmatrix}\begin{bmatrix}0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1\end{bmatrix} = \begin{bmatrix}\boldsymbol{a_2} & \boldsymbol{a_1} & \boldsymbol{a_3}\end{bmatrix}$$

3.5. Linear Programs

Standard Form:

$$\min_{\boldsymbol{x}} \boldsymbol{c}^\mathsf{T}\boldsymbol{x} \qquad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \quad \boldsymbol{x} \geq \boldsymbol{0}$$

LP is bounded if:

$$\boldsymbol{c}^\mathsf{T}\boldsymbol{d}_j^{(\text{ED})} \geq 0 \qquad \forall j$$

Optimizer of a LP is an EP:

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}^{(\text{EP})}}\{\boldsymbol{c}^\mathsf{T}\boldsymbol{x}_i^{(\text{EP})}\} \qquad \forall i$$

LPs are always **convex** OPs.

3.6. Simplex Algorithm

Solve Linear Program of the form

$$\boxed{\min_{\boldsymbol{x}} \boldsymbol{c}^\mathsf{T}\boldsymbol{x} \quad \text{s.t.} \; \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \quad \boldsymbol{x} \geq \boldsymbol{0}}$$

0 Find EP:

$$\boldsymbol{A}^\Pi = \boldsymbol{A}\boldsymbol{\Pi} = \begin{bmatrix}\boldsymbol{B} & \boldsymbol{N}\end{bmatrix}, \boldsymbol{x}_{\text{EP}}^\Pi = \begin{bmatrix}\boldsymbol{B}^{-1}\boldsymbol{b} \\ \boldsymbol{0}\end{bmatrix} \geq \boldsymbol{0}$$
$$\boldsymbol{c}^\Pi = \begin{bmatrix}\boldsymbol{c_B} \\ \boldsymbol{c_N}\end{bmatrix}$$

1 Check criterium for optimality

$$\boxed{\boldsymbol{c}_{\boldsymbol{B}}^\mathsf{T}\boldsymbol{B}^{-1}\boldsymbol{N} - \boldsymbol{c}_{\boldsymbol{N}}^\mathsf{T} \leq \boldsymbol{0} \text{ ?}}$$

yes $\rightarrow \boldsymbol{x}^* = \boldsymbol{x}_{\text{EP}} = \boldsymbol{\Pi}\boldsymbol{x}_{\text{EP}}^\Pi$
no $\rightarrow$ continue

2 Find direction to next EP

$$\boxed{\begin{gathered} j = \arg\max\{[\boldsymbol{c}_{\boldsymbol{B}}^\mathsf{T}\boldsymbol{B}^{-1}\boldsymbol{N} - \boldsymbol{c}_{\boldsymbol{N}}^\mathsf{T}]_i\} \\ \boldsymbol{B}^{-1}\boldsymbol{N}\boldsymbol{e_j} \leq 0 \text{ ?} \end{gathered}}$$

yes $\rightarrow$ LP unbounded
no $\rightarrow$ continue

3 Compute new permuted EP

$$\boldsymbol{d}^\Pi = \begin{bmatrix}-\boldsymbol{B}^{-1}\boldsymbol{N}\boldsymbol{e_j} \\ \boldsymbol{e_j}\end{bmatrix}$$
$$\boldsymbol{x}_{\text{EP,new}}^\Pi = \boldsymbol{x}_{\text{EP}}^\Pi + \lambda_{\text{max}}\boldsymbol{d}^\Pi \geq \boldsymbol{0}$$
$$i = \arg\min\{[\lambda_{\text{max}}\boldsymbol{d}^\Pi]_j\}$$

(index in $\boldsymbol{x}_{EP}^\Pi$ that limits λ_{max})

4 Compute new Permutation Matrix and go to step 0

$$\boldsymbol{x}_{\text{EP, new}} = \boldsymbol{\Pi}\boldsymbol{x}_{\text{EP, new}}^\Pi$$

Choose $\boldsymbol{\Pi}_{\text{new}}$ such that i is nonbasic and j is basic.

3.6.1. Degeneracy

Theoretical infinite loop (*cycling*) if:

$$[\boldsymbol{x_B}]_i = 0 \quad \rightarrow \quad \lambda_{\text{max}} = 0$$

Binominalcoefficient:

$$\binom{n}{m} = \frac{n!}{m!(n-m)!}$$

4. Optimality Conditions

4.1. Unconstrained Optimization

4.1.1. First Order

Sufficient if $f(\boldsymbol{x})$ **pseudoconvex** at $\boldsymbol{x}^*$. $\boldsymbol{x}^*$ is **global** minimizer

$$\nabla f(\boldsymbol{x}^*) = 0$$

4.1.2. Second Order

Always sufficient. $\boldsymbol{x}^*$ is **local** minimizer

$$\nabla f(\boldsymbol{x}^*) = 0 \quad \text{and} \quad \nabla^2 f(\boldsymbol{x}^*) \quad \text{pdf}$$

4.2. Geometric Optimality Conditions

Cone of feasible directions

$$\mathcal{G}_0(\boldsymbol{x}^*) = \{\boldsymbol{d} \mid \nabla g_i^\mathsf{T}(\boldsymbol{x}^*)\boldsymbol{d} < 0 \forall i \in \mathcal{I}(\boldsymbol{x}^*)\}$$

Cone of improving directions

$$\mathcal{F}_0(\boldsymbol{x}^*) = \{\boldsymbol{d} \mid \nabla f^\mathsf{T}(\boldsymbol{x}^*)\boldsymbol{d} < 0\}$$

Sufficient condition for a *local* minimizer if $f(\boldsymbol{x})$ is pseudoconvex, $g_i(\boldsymbol{x})$ strictly pseudoconvex at $\boldsymbol{x}^*$

$$: \mathcal{G}_0(\boldsymbol{x}^*) \cap \mathcal{F}_0(\boldsymbol{x}^*) = \emptyset$$

Generic constrained OP:

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) \qquad \text{s.t.} \quad g_i(\boldsymbol{x}) \leq 0, \forall i \in \{1, \dots, l\}$$
$$h_j(\boldsymbol{x}) = 0, \forall j \in \{1, \dots, m\}$$

4.3. Fritz John Optimality Conditions

$$u_0 \nabla f(\boldsymbol{x}^*) + \sum_{i \in \mathcal{I}(\boldsymbol{x}^*)} u_i \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{m} v_j \nabla h_j(\boldsymbol{x}^*) = 0$$
$$u_0 \geq 0,\ u_i \geq 0$$

Necessary if:

- $f(\boldsymbol{x})$ and $g_i(\boldsymbol{x})\ \forall i \in \mathcal{I}(\boldsymbol{x}^*)$ differentiable at $\boldsymbol{x}^*$
- $g_i(\boldsymbol{x})\ \forall i \notin \mathcal{I}(\boldsymbol{x}^*)$ continuous at $\boldsymbol{x}^*$
- $h_j(\boldsymbol{x})\ \forall j \in \{1, \dots, m\}$ continuously differentiable at $\boldsymbol{x}^*$

Sufficient if:

- $h_j(\boldsymbol{x})\ \forall j \in \{1, \dots, m\}$ affine
- $\nabla h_j(\boldsymbol{x})\ \forall j \in \{1, \dots, m\}$ linearly independent
- $f(\boldsymbol{x})$ pseudoconvex over $\mathcal{U}(\boldsymbol{x}^*) \cap \mathcal{S}(\boldsymbol{x}^*)$
- $g_i(\boldsymbol{x})\ \forall i \notin \mathcal{I}(\boldsymbol{x}^*)$ strictly pseudoconvex over $\mathcal{U}(\boldsymbol{x}^*) \cap \mathcal{S}(\boldsymbol{x}^*)$

Problems :

- $u_0 = 0$ allowed $\to$ ignores objective
- $\mathcal{I}(\boldsymbol{x})$ requires determining active constraints
  $\to$ KKT resolves these issues with $u_0 = 1$ and CS

4.4. KKT Optimality Conditions

Primal Feasibility (PF):

$$g_i(\boldsymbol{x}) \leq 0, \forall i \in \{1, \dots, l\}$$
$$h_i(\boldsymbol{x}) = 0, \forall j \in \{1, \dots, m\}$$

Dual Feasibility (DF):

$$\nabla f(\boldsymbol{x}) + \sum_{i=1}^{l} u_i \nabla g_i(\boldsymbol{x}) + \sum_{j=1}^{m} v_j \nabla h_j(\boldsymbol{x}) = \boldsymbol{0}$$
$$u_i \geq 0\ \forall i \in \{1, \dots, l\}$$

Complementary Slackness (CS):

$$u_i g_i(\boldsymbol{x}) = 0\ \forall i \in \{1, \dots, l\}$$

Necessary if:

- $f(\boldsymbol{x})$ and $g_i(\boldsymbol{x})\ \forall i \in \mathcal{I}(\boldsymbol{x}^*)$ differentiable at $\boldsymbol{x}^*$
- $g_i(\boldsymbol{x})\ \forall i \notin \mathcal{I}(\boldsymbol{x}^*)$ continuous at $\boldsymbol{x}^*$
- $h_j(\boldsymbol{x})\ \forall j \in \{1, \dots, m\}$ continuously differentiable at $\boldsymbol{x}^*$
- $\nabla g_i(\boldsymbol{x}^*)\ \forall i \in \mathcal{I}(\boldsymbol{x}^*)$ and $\nabla h_j(\boldsymbol{x}^*)\ \forall j \in \{1, \dots, m\}$ linearly independent

Sufficient if:

- $f(\boldsymbol{x}^*)$ pseudoconvex and active $g_i(\boldsymbol{x}^*)$ quasiconvex
- $h_j(\boldsymbol{x}^*)$ quasiconvex $\forall j \mid v_j > 0$
- $h_j(\boldsymbol{x}^*)$ quasiconcave $\forall j \mid v_j < 0$

4.5. Constraint Qualifications

KKT conditions are **necessary** if at least one CQ is satisfied

4.5.1. Linear Independence Constraint Qualification

- inactive g_i are continuous at $\boldsymbol{x}^*$
- $h_j\ \forall j \in \{1, \dots, m\}$ are continuously differentiable at $\boldsymbol{x}^*$
- active $\nabla g_i(\boldsymbol{x}^*)$ and $\nabla h_j(\boldsymbol{x}^*)$ are linearly independent

4.5.2. Slater's Constraint Qualification

- active g_i are pseudoconvex at $\boldsymbol{x}^*$
- inactive g_i are continuous at $\boldsymbol{x}^*$
- h_j are pseudoconvex, pseudoconcave and continuously differentiable at $\boldsymbol{x}^*$
- ∇h_j are linearly independent
- $\mathcal{X}$ **has an interior point**

KKT conditions are **sufficient** if:
... the **OP is convex** (e.g. LPs)
... OR Slater CQs + $f(\boldsymbol{x})$ **pseudoconvex**

5. Lagrangian Duality

5.1. Lagrangian Dual Problem

Lagrangian function

$$\phi(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}) = f(\boldsymbol{x}) + \boldsymbol{u}^\mathsf{T}\boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{v}^\mathsf{T}\boldsymbol{h}(\boldsymbol{x})$$

Dual function

$$\theta(\boldsymbol{u}, \boldsymbol{v}) = \inf_{\boldsymbol{x} \in \mathcal{S}} \{\phi(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v})\}$$

Dual Problem

$$\max_{\boldsymbol{u}, \boldsymbol{v}} \quad \theta(\boldsymbol{u}, \boldsymbol{v}) \qquad \text{s.t.} \quad \boldsymbol{u} \geq \boldsymbol{0}$$

The D-OP is **always convex and not unique**.
The dual function θ is always concave and may not be differentiable.

5.2. Geometric Interpretation of the D-OP

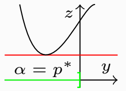
D-OP for Problem with single IEQ:

$$\max_{u \leq 0} \inf_{\boldsymbol{x}} \{f(\boldsymbol{x}) + u g(\boldsymbol{x})\}$$

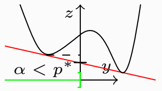
Reformulate to get a linear expression:

$$\alpha := f(\boldsymbol{x}) + u g(\boldsymbol{x}) := z + uy$$
$$z = \alpha - uy$$

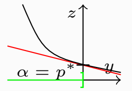
where $-u$ defines the slope which is always ≤ 0.



$g(x)$ inactive
zero gap



$g(x)$ inactive
nonzero gap



$g(x)$ active
zero gap

5.3. Weak and Strong Duality

Weak Duality

$$d^* \leq p^*$$

Weak duality always holds

Strong Duality

$$d^* = p^*$$

Strong duality holds if:

- OP is linear
- or OP is convex and QC hold

5.4. Saddle Point Property

Fulfilled by $(\boldsymbol{x}^*, \boldsymbol{u}^*, \boldsymbol{v}^*)$ if strong duality holds:

$$\sup_{\boldsymbol{x}} \inf_{\boldsymbol{y}} f(\boldsymbol{x}, \boldsymbol{y}) \leq \inf_{\boldsymbol{y}} \sup_{\boldsymbol{x}} f(\boldsymbol{x}, \boldsymbol{y})$$

5.5. Boundedness of D-OPs

P-OP is never unbounded from above
D-OP is never unbounded from below

Primal / Dual	finite	unbounded	infeasible
finite	✓	✗	✗
unbounded	✗	✗	✓
infeasible	✗	✓	✓

6. Concepts of Algorithms

6.1. Algorithmic Map

A point-to-point or point-to-set map to generate a new iterate $\boldsymbol{x}^{(k+1)}$ from $\boldsymbol{x}^{(k)}$

$$\boldsymbol{A} : \mathcal{S} \to \mathcal{S}$$

6.2. Solution Sets

The algorithm terminates if $\boldsymbol{x}^{(k+1)} \in \Omega \subset \mathcal{S}$. Possible solution sets Ω:

- $\Omega = \{\bar{\boldsymbol{x}} \mid \bar{\boldsymbol{x}} \text{is a local miminizer}\}$
- $\Omega = \{\bar{\boldsymbol{x}} \in \mathcal{X} \mid f(\bar{\boldsymbol{x}}) \leq c\}$ with threshold c
- $\Omega = \{\bar{\boldsymbol{x}} \mid \bar{\boldsymbol{x}} \text{is a KKT point}\}$

6.3. Stopping Criteria

Error-based:

$$\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}\| < \epsilon$$

Residual-based:

$$\alpha(\boldsymbol{x}^{(k)}) - \alpha(\boldsymbol{x}^{(k+1)}) < \epsilon$$

with descent function α

6.4. Convergence Analysis

Order of Convergence

$$\lim_{k \to \infty} \sup \frac{\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(\infty)}\|_2}{\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^{(\infty)}\|_2^p} = \beta < \infty$$

with convergence ratio β and order of convergence $\sup p$.

1. $p = 1$ and $\beta = 1$: sublinear convergence
2. $p = 1$ and $\beta \in (0, 1)$: linear convergence
3. $p = 1$ and $\beta = 0$: superlinear convergence
4. $p = 2$ and $\beta < \inf$: superlinear convergence with quadratic order

7. Unconstrained Optimization

7.1. Bisection Algorithm

- The function $f(\boldsymbol{x})$ **must be pseudoconvex** and once differentiable
- Requires initial guess for interval $\boldsymbol{x}^* \in [a^{(1)}, b^{(1)}]$
- Requires n steps for accuracy ϵ such that:

$$2^n \leq \frac{b^{(1)} - a^{(1)}}{\epsilon}$$

0 Initialization:

$$\mathcal{I}^{(k)} = [a^{(k)}, b^{(k)}] \leftarrow a, b \mid x^* \in [a, b], \quad a < b$$

1 Interval Center:

$$x^{(k)} \leftarrow \frac{a^{(k)} - b^{(k)}}{2}$$

2 Derivative at Center:

$$f'(x^{(k)}) \leftarrow \left.\frac{\partial f(x)}{\partial x}\right|_{x = x^{(k)}}$$

3 Check derivative sign:

$$f'(x^{(k)}) < 0 \to \mathcal{I}^{(k+1)} = [x^k, b^k]$$
$$f'(x^{(k)}) > 0 \to \mathcal{I}^{(k+1)} = [a^k, x^k]$$
$$f'(x^{(k)}) = 0 \to x^* = x^{(k)}$$

4 Repeat until $b^{(k)} - a^{(k)} < \epsilon$

7.2. Newton Algorithm

- The function $f(\boldsymbol{x})$ must be **twice differentiable**
- **Convergence not guaranteed** for arbitrary $\boldsymbol{x}^{(0)}$
- Good convergence if $\boldsymbol{x}^{(0)}$ close to $\boldsymbol{x}^*$
- Numerical problems if $\nabla^2 f(\boldsymbol{x}^{(k)}) = 0$ (no inverse of hessian)
- **Convergence Independent of linear transformations**
- Finds $\boldsymbol{x}^*$ in one step if $f(\boldsymbol{x})$ is quadratic

Newton Update:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \boldsymbol{H}(\boldsymbol{x}^{(k)})^{-1} \nabla f(\boldsymbol{x}^{(k)})$$

Newton Step:

$$\Delta \boldsymbol{x}^{(k)} = -\boldsymbol{H}(\boldsymbol{x}^{(k)})^{-1} \nabla f(\boldsymbol{x}^{(k)})$$

with

$$\nabla f^\mathsf{T}(\boldsymbol{x}^{(k)}) \Delta \boldsymbol{x}^{(k)} < 0 \to \text{descent direction}$$

Newton Decrement: Useful as termination criterion

$$\lambda(\boldsymbol{x}^{(k)}) = (\Delta \boldsymbol{x}^{(k)\mathsf{T}} \boldsymbol{H}(\boldsymbol{x}^{(k)}) \Delta \boldsymbol{x}^{(k)})$$

7.3. Armijo's Stepsize Rule

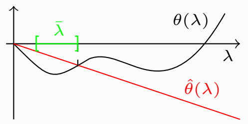
- *Choose stepsize $\lambda^{(k)}$ not too small or too large*
- Tuning parameters $\epsilon \in (0, 1)$ and $\alpha > 1$

Line Search Function:

$$\theta(\lambda) = f(\bar{\boldsymbol{x}} + \lambda \boldsymbol{d})$$

First Order Approximation at $\lambda = 0$:

$$\hat{\theta}(\lambda) = \theta(0) + \lambda \epsilon \theta'(0)$$



Not too large:
$\theta(\bar{\lambda}) \leq \bar{\theta}(\bar{\lambda})$
Not too small:
$\theta(\alpha\bar{\lambda}) > \bar{\theta}(\alpha\bar{\lambda})$

8. Solution Methods for Dual Problems

8.1. Possible Advantages

- reduction in problem complexity
- decoupling into smaller problems

8.2. Subgradients

Subgradient ξ of <u>convex</u> $f(\boldsymbol{x})$ at $\bar{\boldsymbol{x}}$:
$$f(\boldsymbol{x}) \geq f(\bar{\boldsymbol{x}}) + \xi^\top(\boldsymbol{x} - \bar{\boldsymbol{x}})$$

Subgradient ξ of <u>concave</u> $\theta(\boldsymbol{u})$ at $\bar{\boldsymbol{u}}$:
$$\theta(\boldsymbol{u}) \leq \theta(\bar{\boldsymbol{u}}) + \xi^\top(\boldsymbol{u} - \bar{\boldsymbol{u}})$$

8.3. Subdifferential

Convex set of all subgradients for <u>**concave**</u> $\theta(\bar{\boldsymbol{u}})$:
$$\partial\theta(\bar{\boldsymbol{u}}) = \operatorname{conv}(\{\xi \mid \theta(\boldsymbol{u}) \leq \theta(\bar{\boldsymbol{u}}) + \xi^\top(\boldsymbol{u} - \bar{\boldsymbol{u}}) \forall \boldsymbol{u} \in \mathcal{U}\})$$
is a singleton if $\theta(\boldsymbol{u})$ is differentiable at $\bar{\boldsymbol{u}}$.

8.4. Subgradient Methods

- $\theta(\boldsymbol{u}, \boldsymbol{v})$ not always differentiable $\rightarrow$ use subgradients
- subgradients not always ascent directions
- Stepsize control necessary

0 Start with Candidate Point $\boldsymbol{u}^{(k)}$

1 Get primal variables $\boldsymbol{x}^{(k)}$
$$\boldsymbol{x}^{(k)} = \arg\min_{\boldsymbol{x}} \phi(\boldsymbol{x}, \boldsymbol{u}^{(k)})$$

2 Calculate subgradient $\boldsymbol{\xi}^{(k)}$
$$\boldsymbol{\xi}^{(k)} = \boldsymbol{g}(\boldsymbol{x}^{(k)})$$

3 Update dual variables

$$\boldsymbol{u}^{(k+1)} = \boldsymbol{P}_{\mathcal{U}}(\boldsymbol{u}^{(k)} \pm s^{(k)}\boldsymbol{\xi}^{(k)})$$
with $+$ **for concave** and $-$ **for convex**
and $\boldsymbol{P}_{\mathcal{U}}$ projects new iterate into feasible set $\mathcal{U}$.

Optimal Stepsize (practically unknown, requires $\boldsymbol{u}^*$):
$$s^{(k)} = \frac{\theta(\boldsymbol{u}^*) - \theta(\boldsymbol{u}^{(k)})}{\|\boldsymbol{\xi}^{(k)}\|_2^2}$$

Constant Stepsize
$$s^{(k)} = s > 0 \quad \forall k$$
Bounded norm of subgradient $\|g(\boldsymbol{x}^{(k)})\|_2 < C$
At best, $\theta(\boldsymbol{u}^{(k)}) \leq \theta(\boldsymbol{u}^*) - s\frac{C^2}{2} = d^* - s\frac{C^2}{2}$

Diminishing Stepsize
$$s^{(k)} > 0, \quad \lim_{k\to\infty} s^{(k)} = 0, \quad \sum_{k}^{\infty} s^{(k)} = \infty$$
Convergence of $\boldsymbol{u}^{(k)} \rightarrow \boldsymbol{u}^*$ only guaranteed if $\boldsymbol{g}(\boldsymbol{x}^{(k)})$ bounded and additionally:
$$\sum_{k=1}^{\infty} (s^{(k)})^2 < \infty$$

8.5. Cutting Plane Algorithm

- approximate $\theta(\boldsymbol{u}, \boldsymbol{v})$ by finite number of hyperplanes
- Outer approximation: $\theta(\boldsymbol{u}^*, \boldsymbol{v}^*) = d^* \leq z^{(k)}$
- First MP may be unbounded if $\boldsymbol{x}^{(0)}$ is not strictly feasible

Lower Bound on d^*
$$y^{(k)} = \max_{j=1,\dots,k} \theta(\boldsymbol{u}^{(j)}) \quad \leq \quad d^*$$
Smallest value among all current supporting points $\boldsymbol{u}^{(j)}$

Upper Bound on d^*
$$z^{(k)} \quad \geq \quad d^*$$
Solution to current master program, outer approximation

0 Initialization:
Interior point $\boldsymbol{x}^{(0)} \iff \boldsymbol{g}(\boldsymbol{x}^{(0)}) < 0 \qquad (\rightarrow \boldsymbol{u}^{(0)} = 0)$

1 Master Program (MP): Add new constraint at $\boldsymbol{x}^{(k-1)}$
$$\max_{z,\boldsymbol{u}} \quad z \qquad \text{s.t.} \quad z \leq f(\boldsymbol{x}^{j}) + \boldsymbol{u}^\top \boldsymbol{g}(\boldsymbol{x}^{(j)}) \quad \forall j \in \{0, \dots, k-1\}$$
$$\boldsymbol{u} \geq \boldsymbol{0}$$
Is a LP which approximates the D-OP

2 Solve MP with Simplex: Get $(z^{(k)}, \boldsymbol{u}^{(k)})$
$\rightarrow$ **See Section 3.6**

3 Evaluate Dual Function:
$$\boldsymbol{x}^{(k)} = \arg\min_{\boldsymbol{x}}\{\phi(\boldsymbol{x}, \boldsymbol{u}^k)\}$$

4 Check termination Criterion:
$$z^{(k)} - \theta(\boldsymbol{u}^{(k)}) < \epsilon$$

Disadvantages:
- only linear convergence
- larger MP in each iteration
- initial MP must be bounded $\rightarrow$ bounded solution set

8.6. Primal Reconstruction

- $\boldsymbol{u}^*$ not unique if $\theta(\boldsymbol{u}^*)$ not differentiable
- $\lim_{k\to\infty} \boldsymbol{u}^k \rightarrow \boldsymbol{u}^* \not\Rightarrow \lim_{k\to\infty} \boldsymbol{x}^{(k)} \rightarrow \boldsymbol{x}^*$
- Sequence $\{\boldsymbol{x}^{(k)}\}$ may even converge to infeasibility

strong duality must hold for primal reconstruction

8.6.1. Reconstruction for Subgradient Methods

Create weighted averaged solution sequence $\{\hat{\boldsymbol{x}}^{(k)}\}$ to avoid oscillations (only for diminishing stepsize rule):
$$c^{(k)} = \sum_{j=1}^{k} s^{(j)}, \quad \hat{\boldsymbol{x}}^{(k)} = \frac{1}{c^{(k)}} \sum_{j=1}^{k} s^{(j)}\boldsymbol{x}^{(j)}$$

8.6.2. Reconstruction for Cutting Plane Algorithm

Primal feasible point $\hat{\boldsymbol{x}}^{(k)}$ must lie in convex hull of iterates $\boldsymbol{x}^{(k-1)}$ which is obtained from the dual of the MP
Dual of MP: Yields factors $\hat{\lambda}_j$ for convex combination of $\boldsymbol{x}^{(k-1)}$

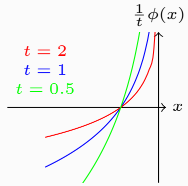
$$\min_{\boldsymbol{\lambda}} \sum_{j=0}^{k-1} \lambda_j f(\boldsymbol{x}^{(j)}) \qquad \text{s.t.} \sum_{j=0}^{k-1} \lambda_j \boldsymbol{g}(\boldsymbol{x}^{(j)}) \leq 0, \sum_{j=0}^{k-1} \lambda_j = 1$$
$$\lambda_j \geq 0$$

Then $\hat{\boldsymbol{x}}^{(k)}$ is a feasible solution:
$$\hat{\boldsymbol{x}}^{(k)} = \sum_{j=0}^{k-1} \hat{\lambda}_j \boldsymbol{x}^{(j)}$$
With optimality gap:
$$z^{(k)} - \theta(\boldsymbol{u}) \leq \epsilon \quad \rightarrow \quad f(\hat{\boldsymbol{x}}^{(k)}) - p^* \leq \epsilon$$
When the **MP** is solved with a primal-dual solver, $\hat{\boldsymbol{x}}^{(k)}$ is obtained for *free*.

9. Interior-Point Methods

9.1. Log Barrier Function

$$\phi(\boldsymbol{x}) = -\sum_{i=1}^{l} \log(-g_i(\boldsymbol{x}))$$



Approximated P-OP:
$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \sum_{i=1}^{l} -\frac{1}{t}\log(-g_i(\boldsymbol{x})) \quad \text{s.t.} \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

Central Path: $\hat{\boldsymbol{x}}(t), t > 0$ that satisfy DF
$$0 = t\nabla f(\hat{\boldsymbol{x}}(t)) + \nabla\phi(\hat{\boldsymbol{x}}(t)) + \boldsymbol{A}^\top\hat{\boldsymbol{v}}$$

Duality Gap:
$$f(\hat{\boldsymbol{x}}(t)) - p^* \leq \frac{l}{t}$$

9.2. KKT Interpretation of P-OP

Primal Feasibility:
$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \qquad g_i(\boldsymbol{x}) < 0$$

Dual Feasibility:
$$\nabla f(\boldsymbol{x}) + \sum_{i=1}^{l} u_i \nabla g_i(\boldsymbol{x}) + \boldsymbol{A}^\top\boldsymbol{v} = \boldsymbol{0}$$
$$\boldsymbol{u} \geq \boldsymbol{0}$$

Complementary Slackness:
$$-u_i g_i(\boldsymbol{x}) = \frac{1}{t}$$
$\rightarrow$ CS is relaxed such that no g_i are active

9.3. Primal Interior Point Algorithm

P-OP:
$$\min_{\boldsymbol{x}} \boldsymbol{c}^\top\boldsymbol{x} - \tau\sum_{j=1}^{n}\log(x_j) \qquad \text{s.t.} \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \geq 0$$

PF:
$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \qquad \boldsymbol{x} \geq \boldsymbol{0}$$

DF:
$$\boldsymbol{c} - \tau\sum_{j=1}^{n}\frac{1}{x_j}e_j - \boldsymbol{A}^\top\boldsymbol{y} = \boldsymbol{0}$$

Resulting System:
$$\begin{aligned} \boldsymbol{c} - \tau\boldsymbol{X}^{-1}\boldsymbol{1} - \boldsymbol{A}^\top\boldsymbol{y} &= \boldsymbol{0} \\ \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} &= \boldsymbol{0} \\ \boldsymbol{x} &\geq \boldsymbol{0} \end{aligned} \iff \boldsymbol{F}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{0}, \boldsymbol{x} \geq \boldsymbol{0}$$

Newton Step:
$$\begin{bmatrix} \Delta\boldsymbol{x} \\ \Delta\boldsymbol{y} \end{bmatrix} = (\nabla_{\boldsymbol{x},\boldsymbol{y}}\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{y}))^{-1}\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{y})$$

Newton Update:
$$\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix}^{(k+1)} = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix}^{(k)} + \alpha\begin{bmatrix} \Delta\boldsymbol{x} \\ \Delta\boldsymbol{y} \end{bmatrix}$$
With α such that $\boldsymbol{x} \geq \boldsymbol{0}$ is not violated!

9.4. Primal-Dual Interior Point Algorithm

P-OP:
$$\min_{\boldsymbol{x}} \boldsymbol{c}^\top\boldsymbol{x} - \tau\sum_{j=1}^{n}\log(x_j) \qquad \text{s.t.} \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \geq 0$$

D-OP:
$$\max_{\boldsymbol{y}} \boldsymbol{b}^\top\boldsymbol{y} \qquad \text{s.t.} \boldsymbol{A}^\top\boldsymbol{y} + \boldsymbol{u} = \boldsymbol{0}, \boldsymbol{u} \geq 0$$

PF:
$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \qquad \boldsymbol{x} \geq \boldsymbol{0}$$

DF:
$$\boldsymbol{c}^\top - \boldsymbol{A}^\top\boldsymbol{y} - \boldsymbol{u} = \boldsymbol{0} \qquad \text{s.t.} \boldsymbol{u} \geq \boldsymbol{0}$$

CS: Is relaxed for better numerical behaviour
$$\boldsymbol{X}\boldsymbol{U}\boldsymbol{1} = 0 \quad \rightarrow \quad \boldsymbol{X}\boldsymbol{U}\boldsymbol{1} = \tau\boldsymbol{1} \iff \boldsymbol{u} = \boldsymbol{X}^{-1}\tau\boldsymbol{1}$$

Resulting System:
$$\begin{aligned} \boldsymbol{A}^\top\boldsymbol{y} + \tau\boldsymbol{X}^{-1}\boldsymbol{1} - \boldsymbol{c} &= \boldsymbol{0} \\ \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} &= \boldsymbol{0} \\ \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{u} \end{bmatrix} &\geq \boldsymbol{0} \end{aligned} \iff \boldsymbol{F}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}) = \boldsymbol{0}, \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{u} \end{bmatrix} \geq \boldsymbol{0}$$

Solve for Newton Step:
$$\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}) + \nabla_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{u}}\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u})\begin{bmatrix} \Delta\boldsymbol{x} \\ \Delta\boldsymbol{y} \\ \Delta\boldsymbol{u} \end{bmatrix} = \boldsymbol{0}$$

Newton Update:
$$\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \\ \boldsymbol{u} \end{bmatrix}^{(k+1)} = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \\ \boldsymbol{u} \end{bmatrix}^{(k)} + \alpha\begin{bmatrix} \Delta\boldsymbol{x} \\ \Delta\boldsymbol{y} \\ \Delta\boldsymbol{u} \end{bmatrix}$$
With α such that $\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{u} \end{bmatrix} \geq \boldsymbol{0}$ is not violated!

10. Matrix Calculus

Denominator layout convention

$$\boldsymbol{x},\boldsymbol{a} \in \mathbb{R}^n, \quad \boldsymbol{u} \in \mathbb{R}^m \quad \boldsymbol{A} \in \mathbb{R}^{m\times n}, \quad \boldsymbol{C} \in \mathbb{R}^{n\times n}$$

$$\frac{\partial \boldsymbol{a}^\top \boldsymbol{x}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{x}^\top \boldsymbol{a}}{\partial \boldsymbol{x}} = \boldsymbol{a} \tag{1}$$

$$\frac{\partial \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{A}^\top \tag{2}$$

$$\frac{\partial \boldsymbol{u}^\top \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{A}^\top \boldsymbol{u} \tag{3}$$

$$\frac{\partial \boldsymbol{x}^\top \boldsymbol{C}\boldsymbol{x}}{\partial \boldsymbol{x}} = (\boldsymbol{C}+\boldsymbol{C}^\top)\boldsymbol{x} \tag{4}$$

$$\frac{\partial}{\partial \boldsymbol{x}}\|\boldsymbol{x}-\boldsymbol{a}\|_2 = \frac{\boldsymbol{x}-\boldsymbol{a}}{\|\boldsymbol{x}-\boldsymbol{a}\|_2} \tag{5}$$

$$\frac{\partial}{\partial \boldsymbol{x}}\frac{\boldsymbol{x}-\boldsymbol{a}}{\|\boldsymbol{x}-\boldsymbol{a}\|_2} = \frac{\boldsymbol{I}}{\|\boldsymbol{x}-\boldsymbol{a}\|_2} - \frac{(\boldsymbol{x}-\boldsymbol{a})(\boldsymbol{x}-\boldsymbol{a})^\top}{\|\boldsymbol{x}-\boldsymbol{a}\|_2^3} \tag{6}$$

$$\frac{\partial \|\boldsymbol{x}\|_2^2}{\partial \boldsymbol{x}} = \frac{\partial \|\boldsymbol{x}^\top \boldsymbol{x}\|_2}{\partial \boldsymbol{x}} = 2\boldsymbol{x} \tag{7}$$

11. Polyhedral Set Conversions

Primal Standard Form → Dual Standard Form

$$\mathcal{S}_p = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \geq \boldsymbol{0}\}$$

$$\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}, \quad -\boldsymbol{A}\boldsymbol{x} \leq -\boldsymbol{b}, \quad -\boldsymbol{x} \leq \boldsymbol{0}$$

$$\mathcal{S}_d = \{\boldsymbol{x} \in \mathbb{R}^n \mid \begin{bmatrix}\boldsymbol{A}\\ -\boldsymbol{A}\\ -\boldsymbol{I}\end{bmatrix}\boldsymbol{x} \leq \begin{bmatrix}\boldsymbol{b}\\ -\boldsymbol{b}\\ \boldsymbol{0}\end{bmatrix}\}$$

Dual Standard Form → Primal Standard Form

$$\mathcal{S}_d = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\}$$

$$\boldsymbol{A}\boldsymbol{x}^+ - \boldsymbol{A}\boldsymbol{x}^- + \boldsymbol{s} = \boldsymbol{b}, \qquad \boldsymbol{x}^+, \boldsymbol{x}^-, \boldsymbol{s} \geq \boldsymbol{0}$$

$$\mathcal{S}_p =$$

$$\{\begin{bmatrix}\boldsymbol{x}^+\\ \boldsymbol{x}^-\\ \boldsymbol{s}\end{bmatrix} \in \mathbb{R}^{2n+m} \mid \begin{bmatrix}\boldsymbol{A} & -\boldsymbol{A} & \boldsymbol{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{x}^+\\ \boldsymbol{x}^-\\ \boldsymbol{s}\end{bmatrix} = \boldsymbol{b}, \begin{bmatrix}\boldsymbol{x}^+\\ \boldsymbol{x}^-\\ \boldsymbol{s}\end{bmatrix} \geq \boldsymbol{0}\}$$

12. Properties of some Sets

The following assumptions are made for the statements below:

$$\boldsymbol{x} \in \mathbb{R}^n, \quad \boldsymbol{y} \in \mathbb{C}^n, \quad \boldsymbol{A} \in \mathbb{R}^{m\times n}, \quad \boldsymbol{b} \in \mathbb{R}^m$$
$$\boldsymbol{Q} \in \mathbb{C}^{n\times n}, \quad \boldsymbol{P} \in \mathbb{C}^{m\times n}, \quad \boldsymbol{p} \in \mathbb{R}^n$$

$\mathcal{X} \subset \mathbb{R}^n$ is closed and convex

Set	linear	affine	conic	convex
$\{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}$		✓		✓
$\{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\}$				✓
$\{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{0}, \boldsymbol{1}^\top \boldsymbol{A}\boldsymbol{x} = 0\}$	✓	✓	✓	✓
$\{\boldsymbol{Q} \mid \boldsymbol{y}^H \boldsymbol{Q}\boldsymbol{y} \geq 0, \forall \boldsymbol{y}\mathbb{C}^n\}$			✓	✓
$\{\boldsymbol{P} \mid \mathrm{tr}(\boldsymbol{P}\boldsymbol{P}^H) \leq 1\}$				✓
$\{\boldsymbol{P} \mid \det(\boldsymbol{P}^{-1}) \neq 0\}$			✓	
$\{\boldsymbol{p} \mid \boldsymbol{p}^\top(\boldsymbol{x}-\bar{\boldsymbol{x}}) \leq 0, \forall \boldsymbol{x} \in \mathcal{X}\}$			✓	✓
$\{\emptyset\}$	✓	✓	✓	✓
$\{\boldsymbol{x}_0\}$		✓		✓
$\{\boldsymbol{x}_0, \boldsymbol{x}_1 \mid \boldsymbol{x}_0 \neq \boldsymbol{x}_1\}$				

13. Properties of a Norm

1. $\|\boldsymbol{x}\| \geq 0, \ \|\boldsymbol{x}\| = 0 \iff \boldsymbol{x} = 0$
2. $\|\alpha\boldsymbol{x}\| = \mid \alpha \mid \|\boldsymbol{x}\|$
3. $\|\boldsymbol{x}+\boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$

Every Norm is a convex function!

14. Infeasible OP Example

P-OP and D-OP are infeasible simultaneously for this example.

P-OP:

$$\inf_{\boldsymbol{x}\in\mathbb{R}^n} \boldsymbol{c}^\top \boldsymbol{x} \qquad \text{s.t. } \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \geq \boldsymbol{0}$$

D-OP:

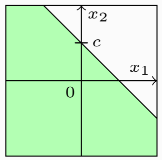
$$\sup_{\boldsymbol{y}\in\mathbb{R}^m} \boldsymbol{b}^\top \boldsymbol{y} \qquad \text{s.t. } \boldsymbol{A}^\top \boldsymbol{y} \leq \boldsymbol{c}$$

$$\boldsymbol{A} = \begin{bmatrix}1 & -1 & -1\\ 1 & 0 & -1\end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix}1\\ 0\end{bmatrix}, \quad \boldsymbol{c} < \boldsymbol{0}$$

15. Deriving IEQs graphically

Green Area is feasible area

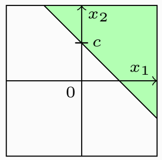
α is slope of line



$x_2 \leq \alpha x_1 + c$

$g(\boldsymbol{x}) = x_2 - \alpha x_1 - c \leq 0$

All x_2 below line

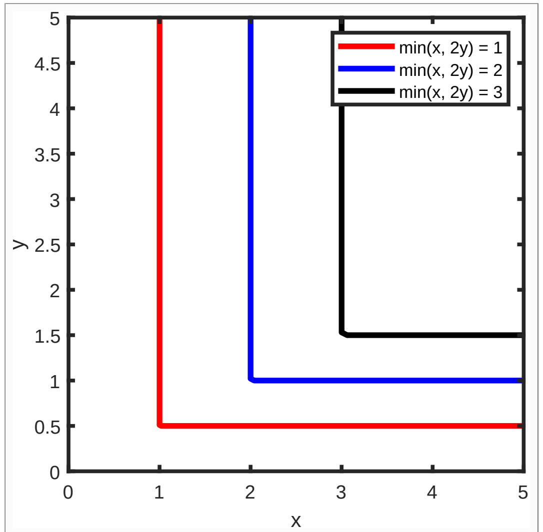


$x_2 \geq \alpha x_1 + c$

$g(\boldsymbol{x}) = \alpha x_1 + c - x_2 \leq 0$

All x_2 above line

16. Weighted Min Function Contour Lines



17. Proofs

17.1. Proof of the Saddle Point Property

$$\sup_x \inf_y f(x,y) \leq \inf_y \sup_x f(x,y)$$

$$\sup_x \inf_y f(x,y) = \inf_y f(\bar{x},y) \leq f(\bar{x},y) \quad \forall y \text{ (including } \bar{y})$$

$$\inf_y \sup_x f(x,y) = \sup_x f(x,\bar{y}) \geq f(\bar{x},\bar{y}) \quad \forall x \text{ (including } \bar{x})$$

17.2. Proof of Farkas Theorem

Farkas: Only one system can be true at a time:

1. $\exists \boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{0}, \boldsymbol{c}^\top \boldsymbol{x} > 0$
2. $\exists \boldsymbol{y} \in \mathbb{R}^m \mid \boldsymbol{A}^\top \boldsymbol{y} = \boldsymbol{c}, \boldsymbol{y} \leq \boldsymbol{0}$

If (2) has a solution: Then there exists $\boldsymbol{y} \geq \boldsymbol{0}$ that fulfills $\boldsymbol{A}^\top \boldsymbol{y} = \boldsymbol{c}$. If **(1)** would also be true, then $\boldsymbol{x}$ has to satisfy $\boldsymbol{c}^\top \boldsymbol{x} = \boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{x} > \boldsymbol{0}$ which is a contradiction.

If (2) has no solution: Then $\boldsymbol{c}$ is **not** in the cone $\mathcal{A}$ spanned by the columns of $\boldsymbol{A}$. Since the cone is closed and convex, there must exist a separating hyperplane that separates all $\boldsymbol{x} \in \mathcal{A}$ and $\boldsymbol{c}$, which is defined by the normal vector $\boldsymbol{p}$. We then have $\boldsymbol{p}^\top \boldsymbol{c} > 0$ and $\boldsymbol{p}^\top \boldsymbol{x} \leq 0$ for all $\boldsymbol{x} \in \mathcal{A}$ since $\boldsymbol{0} \in \mathcal{A}$. Define $\boldsymbol{x} := \boldsymbol{A}^\top \boldsymbol{y}$ yields $\boldsymbol{p}^\top \boldsymbol{A}^\top \boldsymbol{y} = \boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{p} \leq 0$ for $\boldsymbol{y} \geq \boldsymbol{0}$ which implies $\boldsymbol{A}\boldsymbol{p} \leq \boldsymbol{0}$ with $\boldsymbol{c}^\top \boldsymbol{p} > 0$. This is equal to **(1)**.

17.3. Proof of Gordons Theorem via Farkas Theorem

F1 and **F2** are System 1 and 2 of Farkas, **G1** and **G2** are system 1 and 2 of Gordon.

$$\textbf{G1: } \boldsymbol{A}\boldsymbol{x} < \boldsymbol{0} \iff \textbf{F1: } \begin{bmatrix}\boldsymbol{A} & \boldsymbol{1}\end{bmatrix}\begin{bmatrix}\boldsymbol{x}\\ s\end{bmatrix} \leq \boldsymbol{0} \begin{bmatrix}\boldsymbol{0}^\top & 1\end{bmatrix}\begin{bmatrix}\boldsymbol{x}\\ s\end{bmatrix} > 0$$

$$\textbf{F2: } \begin{bmatrix}\boldsymbol{A}^\top\\ \boldsymbol{1}^\top\end{bmatrix}\boldsymbol{y} = \begin{bmatrix}\boldsymbol{0}\\ 1\end{bmatrix} \iff \textbf{G2: } \boldsymbol{A}^\top \boldsymbol{y} = \boldsymbol{0}, \quad \boldsymbol{1}^\top \boldsymbol{y} = 1, \quad \boldsymbol{y} \geq \boldsymbol{0}$$

17.4. Proof that the epigraph of a convex function is convex

Given two points of the epigraph of a convex function f:

$$[x_1, y_1]^\top, [x_2, y_2]^\top \in \text{epi}(f)$$

The definition of convexity yields:

$\lambda y_1 + (1-\lambda)y_2 \geq \lambda f(x_1) + (1-\lambda)f(x_2) \geq f(\lambda x_1 + (1-\lambda)x_2)$

This implies:

$$\begin{bmatrix}\lambda x_1 + (1-\lambda)x_2\\ \lambda y_1 + (1-\lambda)y_2\end{bmatrix} = \lambda\begin{bmatrix}x_1\\ y_1\end{bmatrix} + (1-\lambda)\begin{bmatrix}x_2\\ y_2\end{bmatrix} \in \text{epi}(f)$$

17.5. Proof that the set of subgradients is convex

Given two subgradients of f at $\bar{x}$:

$$f(x) \geq f(\bar{x}) + \xi_1^\top (x-\bar{x})$$
$$f(x) \geq f(\bar{x}) + \xi_2^\top (x-\bar{x})$$

The definition of convexity yields:

$$\lambda f(x) + (1-\lambda)f(x) \geq f(\bar{x}) + (\lambda\xi_1^\top + (1-\lambda)\xi_2^\top)(x-\bar{x})$$

Hence the convex combination of two subgradients is also a subgradient

17.6. Proof of Weak Duality

$$\begin{aligned} d^* &= \sup_{v,u\geq 0} \inf_{x\in\mathcal{S}} \{f(x) + u^\top g(x) + v^\top h(x)\} \\ &\leq \sup_{v,u\geq 0} \inf_{x\in\mathcal{X}} \{f(x) + u^\top g(x) + v^\top h(x)\} \\ &\leq \sup_{v,u\geq 0} \{f(x) + u^\top g(x) + v^\top h(x)\} \\ &\leq \sup_{v,u\geq 0} f(x) \\ &\leq f(x) \\ &\leq p^* \end{aligned}$$

with $\mathcal{S} \subset \mathcal{X}$

17.7. Proof that the optimizer of a LP is an extremal point

From representation theorem, we know that every point $\boldsymbol{x}$ in a polyhedron can be represented by the EPs $\boldsymbol{x}_i$ and EDs $\boldsymbol{d}_j$:

$$\boldsymbol{x} = \sum_{i=1}^K \lambda_i \boldsymbol{x}_i + \sum_{j=1}^L \mu_j \boldsymbol{d}_j, \qquad \sum_{i=1}^K \lambda_i = 1$$
$$\lambda_i \geq 0, \mu_j \geq 0$$

A LP can thus be written as:

$$\min_{\boldsymbol{\lambda},\boldsymbol{\mu}} \boldsymbol{c}^\top \Big(\sum_{i=1}^K \lambda_i \boldsymbol{x}_i + \sum_{j=1}^L \mu_j \boldsymbol{d}_j\Big), \quad \text{s.t. } \sum_{i=1}^K \lambda_i = 1$$
$$\lambda_i \geq 0, \mu_j \geq 0$$

where the polyhedral constraint set is enforced by the representation with EPs and EDs. Due to the boundedness requirement for LPs, we have $\boldsymbol{c}^\top \boldsymbol{d}_j \geq 0$ which leads to $\boldsymbol{\mu} = \boldsymbol{0}$. Therefore, the minimizer is the EP $\boldsymbol{x}_i$ with the smallest inner product with $\boldsymbol{c}$:

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x_i}} \{\boldsymbol{c}^\top \boldsymbol{x}_i\} \qquad \forall i \in \{1,\ldots,K\}$$