Searching for Universal Truths
Convex Optimization

Sunghee Yun
sunghee.yun@gmail.com

Navigating Mathematical and Statistical Territories

- Notations & definitions & conventions
  - notations - 3
  - some definitions - 7
  - some conventions - 8
- Convex optimization - 9
  - convex sets - 10
  - convex functions - 31
  - convex optimization problems - 52
  - duality - 79
  - theorems of alternatives - 146
  - convex optimization with generalized inequalities - 155
  - unconstrained minimization - 170
  - equality constrained minimization - 198
  - barrier interior-point methods - 221
  - primal-dual interior-point methods - 235

- Proof & references & indices
  - selected proofs - 242
  - references - 248
  - index - 250

Notations

- sets of numbers
  - **N** - set of natural numbers
  - **Z** - set of integers
  - $\mathbf{Z}_+$ - set of nonnegative integers
  - **Q** - set of rational numbers
  - **R** - set of real numbers
  - $\mathbf{R}_+$ - set of nonnegative real numbers
  - $\mathbf{R}_{++}$ - set of positive real numbers
  - **C** - set of complex numbers
- sequences $\langle x_i \rangle$ and the like
  - finite $\langle x_i \rangle_{i=1}^n$, infinite $\langle x_i \rangle_{i=1}^\infty$ - use $\langle x_i \rangle$ whenever unambiguously understood
  - similarly for other operations, *e.g.*, $\sum x_i$, $\prod x_i$, $\cup A_i$, $\cap A_i$, $\bigtimes A_i$
  - similarly for integrals, *e.g.*, $\int f$ for $\int_{-\infty}^{\infty} f$
- sets
  - $\tilde{A}$ - complement of A

  - $A \sim B$ - $A \cap \tilde{B}$
  - $A \Delta B$ - $(A \cap \tilde{B}) \cup (\tilde{A} \cap B)$
  - $\mathcal{P}(A)$ - set of all subsets of A
- sets in metric vector spaces
  - $\overline{A}$ - closure of set A
  - A° - interior of set A
  - $\mathbf{relint}\, A$ - relative interior of set A
  - $\mathbf{bd}\, A$ - boundary of set A
- set algebra
  - $\sigma(\mathcal{A})$ - σ-algebra generated by $\mathcal{A}$, *i.e.*, smallest σ-algebra containing $\mathcal{A}$
- norms in $\mathbf{R}^n$
  - $\|x\|_p$ $(p \geq 1)$ - p-norm of $x \in \mathbf{R}^n$, *i.e.*, $(|x_1|^p + \cdots + |x_n|^p)^{1/p}$
  - *e.g.*, $\|x\|_2$ - Euclidean norm
- matrices and vectors
  - a_i - i-th entry of vector a
  - A_{ij} - entry of matrix A at position (i, j), *i.e.*, entry in i-th row and j-th column
  - $\mathbf{Tr}(A)$ - trace of $A \in \mathbf{R}^{n \times n}$, *i.e.*, $A_{1,1} + \cdots + A_{n,n}$

- symmetric, positive definite, and positive semi-definite matrices
  - $\mathbf{S}^n \subset \mathbf{R}^{n\times n}$ - set of symmetric matrices
  - $\mathbf{S}^n_+ \subset \mathbf{S}^n$ - set of positive semi-definite matrices; $A \succeq 0 \Leftrightarrow A \in \mathbf{S}^n_+$
  - $\mathbf{S}^n_{++} \subset \mathbf{S}^n$ - set of positive definite matrices; $A \succ 0 \Leftrightarrow A \in \mathbf{S}^n_{++}$
- sometimes, use Python script-like notations (with serious abuse of mathematical notations)
  - use $f : \mathbf{R} \to \mathbf{R}$ as if it were $f : \mathbf{R}^n \to \mathbf{R}^n$, *e.g.*,

$$\exp(x) = (\exp(x_1), \ldots, \exp(x_n)) \quad \text{for } x \in \mathbf{R}^n$$

    and

$$\log(x) = (\log(x_1), \ldots, \log(x_n)) \quad \text{for } x \in \mathbf{R}^n_{++}$$

    which corresponds to Python code `numpy.exp(x)` or `numpy.log(x)` where `x` is instance of `numpy.ndarray`, *i.e.*, `numpy` array
  - use $\sum x$ to mean $\mathbf{1}^T x$ for $x \in \mathbf{R}^n$, *i.e.*

$$\sum x = x_1 + \cdots + x_n$$

    which corresponds to Python code `x.sum()` where `x` is `numpy` array

- use x/y for $x, y \in \mathbf{R}^n$ to mean

$$\begin{bmatrix} x_1/y_1 & \cdots & x_n/y_n \end{bmatrix}^T$$

  which corresponds to Python code `x / y` where `x` and `y` are 1-d `numpy` arrays
- use X/Y for $X, Y \in \mathbf{R}^{m \times n}$ to mean

$$\begin{bmatrix} X_{1,1}/Y_{1,1} & X_{1,2}/Y_{1,2} & \cdots & X_{1,n}/Y_{1,n} \\ X_{2,1}/Y_{2,1} & X_{2,2}/Y_{2,2} & \cdots & X_{2,n}/Y_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m,1}/Y_{m,1} & X_{m,2}/Y_{m,2} & \cdots & X_{m,n}/Y_{m,n} \end{bmatrix}$$

  which corresponds to Python code `X / Y` where `X` and `Y` are 2-d `numpy` arrays

Some definitions

Definition 1. [infinitely often - i.o.] *statement P_n, said to happen* infinitely often *or* i.o. *if*

$$(\forall N \in \mathbf{N})\,(\exists n > N)\,(P_n)$$

Definition 2. [almost everywhere - a.e.] *statement $P(x)$, said to happen* almost everywhere *or* a.e. *or* almost surely *or* a.s. *(depending on context) associated with measure space $(X, \mathscr{B}, \mu)$ if*

$$\mu\{x|P(x)\} = 1$$

or equivalently

$$\mu\{x| \sim P(x)\} = 0$$

Some conventions

- (for some subjects) use following conventions
  - $0 \cdot \infty = \infty \cdot 0 = 0$
  - $(\forall x \in \mathbf{R}_{++})(x \cdot \infty = \infty \cdot x = \infty)$
  - $\infty \cdot \infty = \infty$

Convex Optimization

Convex Sets

Lines and line segmenets

Definition 3. [lines] *for some* $x, y \in \mathbf{R}^n$

$$\{\theta x + (1-\theta)y | \theta \in \mathbf{R}\}$$

called line going through x and y

Definition 4. [line segmenets] *for some* $x, y \in \mathbf{R}^n$

$$\{\theta x + (1-\theta)y | 0 \leq \theta \leq 1 \in \mathbf{R}\}$$

called line segment connecting x and y

Affine sets

Definition 5. [affine sets] *set, $C \subset \mathbf{R}^n$, every line going through any two points in which is contained in C, $i.e.$*

$$(\forall x, y \in C)\,(\{\theta x + (1-\theta)y | \theta \in \mathbf{R}\} \subset C)$$

called affine set

Definition 6. [affine hulls] *for set, $C \subset \mathbf{R}^n$, intersection of all affine sets containing C, called* affine hull of C*, denoted by* $\mathbf{aff}\, C$*, which is equal to set of all affine combinations of points in C, $i.e.$*

$$\bigcup_{n \in \mathbf{N}} \{\theta_1 x_1 + \cdots + \theta_n x_n | x_1, \ldots, x_n \in C, \theta_1 + \cdots + \theta_n = 1\}$$

Definition 7. [affine dimension] *for $C \subset \mathbf{R}^n$, dimension of $\mathbf{aff}\, C$, called* affine dimension

Relative interiors and boundaries

Definition 8. [relative interiors of sets] *for* $C \subset \mathbf{R}^n$,

$$\bigcup_{O:\text{open}, O\cap\mathbf{aff}\, C\subset C} O \cap \mathbf{aff}\, C$$

or equivalently

$$\{x|(\exists \epsilon > 0)(\forall y \in \mathbf{aff}\, C, \|y - x\| < \epsilon)(y \in C)\}$$

is called relative interior of C *or* interior relative to C, *denoted by* $\mathbf{relint}\, C$

Definition 9. [relative boundaries of sets] *for* $C \subset \mathbf{R}^n$, $\overline{C} \sim \mathbf{relint}\, C$, *called* relative boundary of C

Convex sets

Definition 10. [convex sets] *set, $C \subset \mathbf{R}^n$, every line segment connecting any two points in which is contained in C, $i.e.$*

$$(\forall x, y \in C)\,(\forall 0 \leq \theta \leq 1)\,(\theta x + (1-\theta)y \in C)$$

called convex set

Definition 11. [convex hulls] *for set, $C \subset \mathbf{R}^n$, intersection of all convex sets containing C, called* convex hull of C*, denoted by $\mathbf{Conv}\, C$, which is equal to set of all convex combinations of points in C, $i.e.$*

$$\bigcup_{n \in \mathbf{N}} \{\theta_1 x_1 + \cdots + \theta_n x_n | x_1, \ldots, x_n \in C, \theta_1 + \cdots + \theta_n = 1, \theta_1, \ldots, \theta_n > 0\}$$

- convex hull (of course) is convex set

Cones

Definition 12. [cones] *set,* $C \subset \mathbf{R}^n$*, for which*

$$(\forall x \in C, \theta \geq 0)\,(\theta x \in C)$$

called cone *or* nonnegative homogeneous

Definition 13. [convex cone] *set,* $C \subset \mathbf{R}^n$*, which is both convex and cone, called* convex cone*;* C *is convex cone* if and only if

$$(\forall x, y \in C, \theta, \xi \geq 0)\,(\theta x + \xi y \in C)$$

- convex cone (of course) is convex set
- examples of convex cones: $\mathbf{R}^n_+$, $\mathbf{R}^n_{++}$, $\mathbf{S}^n_+$, and $\mathbf{S}^n_{++}$

Hyperplanes and half spaces

Definition 14. [hyperplanes] *$n-1$ dimensional affine set in $\mathbf{R}^n$, called hyperplane; every hyperplane can be expressed as*

$$\{x \in \mathbf{R}^n | a^T = b\}$$

for some $a \neq 0 \in \mathbf{R}^n$ and $b \in \mathbf{R}$

Definition 15. [half spaces] *one of two sets divided by hyperplane, called half space; every half space can be expressed as*

$$\{x \in \mathbf{R}^n | a^T \leq b\}$$

for some $a \neq 0 \in \mathbf{R}^n$ and $b \in \mathbf{R}$

- hyperplanes and half spaces are convex sets

Euclidean balls and ellipsoids

Definition 16. [Euclidean ball] *set of all points distance of which from point, $x \in \mathbf{R}^n$, is no greater than $r > 0$, called (Euclidean) ball centered at x with radius, r, denoted by $B(x, r)$, $i.e.$*

$$B(x, r) = \{y \in \mathbf{R}^n | \|y - x\|_2 \leq r\}$$

Definition 17. [ellipsoids] *ball elongated along n orthogonal axes, called ellipsoid, $i.e.$,*

$$\{y \in \mathbf{R}^n | (y - x)^T P^{-1} (y - x) \leq 1\}$$

for some $x \in \mathbf{R}^n$ and $P \in \mathbf{S}^n_{++}$

- Euclidean balls and ellipsoids are convex sets

Norm balls and norm cones

Definition 18. [norm ball] *for norm,* $\|\cdot\| : \mathbf{R}^n \to \mathbf{R}_+$*, set of all points distance of which measured in the norm from point,* $x \in \mathbf{R}^n$*, is no greater than* $r > 0$*, called* norm ball centered at x with radius, r, associated with norm, $\|\cdot\|$*, i.e.*

$$\{y \in \mathbf{R}^n | \|y - x\| \leq r\}$$

Definition 19. [norm cone] *for norm,* $\|\cdot\| : \mathbf{R}^n \to \mathbf{R}_+$*,* $x \in \mathbf{R}^n$*, and* $r > 0$*,*

$$\{(x, y) \in \mathbf{R}^n \times \mathbf{R} | \|x\| \leq r\} \subset \mathbf{R}^{n+1}$$

called cone associated with norm, $\|\cdot\|$

Definition 20. [second-order cone] *norm cone associated with Euclidean norm, called* second-order cone

- norm balls and norm cones are convex sets

Polyhedra

Definition 21. [polyhedra] *intersection of finite number of hyperplanes and half spaces, called* polyhedron*; every polyhedron can be expressed as*

$$\{x \in \mathbf{R}^n | Ax \preceq b, Cx = d\}$$

for $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, $C \in \mathbf{R}^{p \times n}$, $d \in \mathbf{R}^p$

- polyhedron is convex set (by Proposition 1)

Convexity preserving set operations

Proposition 1. [convexity preserving set operations]

- *intersection preserves convexity*
  - *for (any) collection of convex sets, $\mathcal{C}$,*

$$\bigcap_{C \in \mathcal{C}} C$$

    is convex set (proof can be found in Proof 1)
- *scalar scaling preserves convexity*
  - *for convex set C*

$$\alpha C$$

    is convex set for any $\alpha \in \mathbf{R}$
- *sum preserves convexity*
  - *for convex sets C and D*

$$C + D$$

    is convex set

- *direct product preserves convexity*
  - *for convex sets C and D*

    $$C \times D$$

    is convex set
- *projection preserves convexity*
  - *for convex set $C \subset A \times B$*

    $$\{x \in A | (\exists y)((x, y) \in C)\}$$

    is convex
- *image and inverse image by affine function preserve convexity*
  - *for affine function $f : A \to B$ and convex sets $C \subset A$ and $D \subset B$*

    $$f(C) \ \& \ f^{-1}(D)$$

    are convex
- *image and inverse image by linear-fractional function preserve convexity*

- *for convex sets* $C \subset \mathbf{R}^n$, $D \subset \mathbf{R}^m$ *and linear-fractional function,* $g : \mathbf{R}^n \to \mathbf{R}^m$, $i.e.$, *function defined by* $g(x) = (Ax + b)/(c^T x + d)$ *for* $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, $c \in \mathbf{R}^n$, *and* $d \in \mathbf{R}$

  $$g(C) \ \& \ g^{-1}(D)$$

  are convex

Proper cones and generalized inequalities

Definition 22. [proper cones] *closed convex cone K which is*

- *solid, $i.e.$, $K^\circ \neq \emptyset$*
- *pointed, $i.e.$, $x \in vK$ and $-x \in K$ imply $x = 0$*

called proper cone

- examples of proper cones: $\mathbf{R}^n_+$ and $\mathbf{S}^n_+$

Definition 23. [generalized inequalities] *proper cone K defines* generalized inequalities

- *(nonstrict) generalized inequality*

$$x \preceq_K y \Leftrightarrow y - x \in K$$

- *strict generalized inequality*

$$x \prec_K y \Leftrightarrow y - x \in K^\circ$$

- $\preceq_K$ and $\prec_K$ are partial orderings

Convex sets induced by generalized inequalities

- for affine function $g : \mathbf{R}^n \to \mathbf{S}^m$, *i.e.*, $f(x) = A_0 + A_1x_1 + \cdots + A_nx_n$ for some $A_0, \ldots, A_n \in \mathbf{S}^m$, $f^{-1}(\mathbf{S}^n_+)$ is convex (by Proposition 1), *i.e.*,

$$\{x \in \mathbf{R}^n | A_0 + A_1x_1 + \cdots + A_nx_n \succeq 0\} \subset \mathbf{R}^n$$

  is convex

- can negate each matrix A_i and have same results, hence

$$\{x \in \mathbf{R}^n | A_0 + A_1x_1 + \cdots + A_nx_n \preceq 0\} \subset \mathbf{R}^n$$

  is (also) convex

Separating and supporting hyperplanes

Theorem 1. [separating hyperplane theorem] *for nonempty disjoint convex sets C and D, exists hyperplane which separates C and D, $i.e.*

$$(\exists a \neq 0 \in \mathbf{R}^n, b \in \mathbf{R})\,(\forall x \in C, y \in D)\left(a^T x + b \geq 0 \;\&\; a^T y + b \leq 0\right)$$

Definition 24. [separating hyperplanes] *for nonempty disjoint convex sets C and D, hyperplane satisfying property in Theorem 1, called separating hyperplane, said to separate C and D*

Theorem 2. [supporting hyperplane theorem] *for nonempty convex set C and $x \in \mathbf{bd}\, C$, exists hyperplane passing through x, $i.e.$,*

$$(\exists a \neq 0 \in \mathbf{R}^n)\,(\forall y \in C)\left(a^T(y - x) \leq 0\right)$$

Definition 25. [supporting hyperplanes] *for nonempty convex set C and $x \in \mathbf{bd}\, C$, hyperplane satisfied property in Theorem 2, called supporting hyperplane*

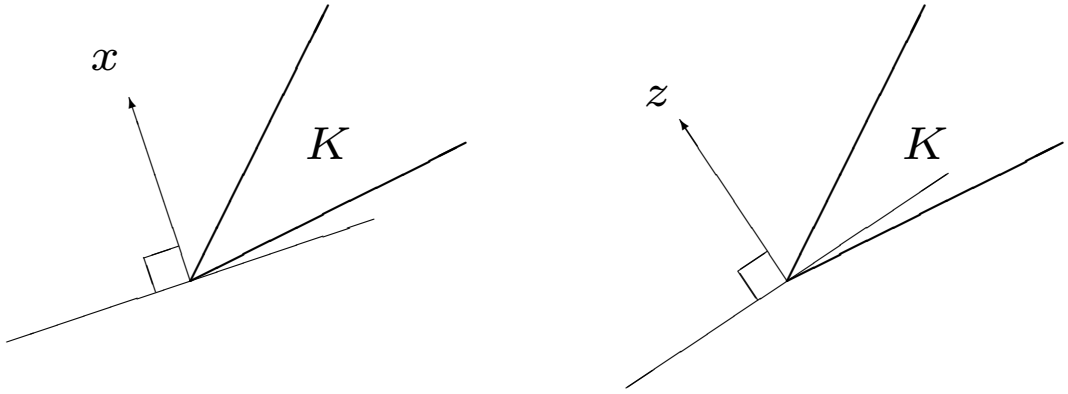
Dual cones

Definition 26. [dual cones] *for cone* K*,*

$$\{x | \forall y \in K, y^T x \geq 0\}$$

called dual cone of K*, denoted by* K^*

- the figure illustrates $x \in K^*$ while $z \notin K^*$



Dual norms

Definition 27. [dual norms] *for norm* $\|\cdot\|$*, fudnction defined by*

$$y \mapsto \sup\{y^T x | \|x\| \leq 1\}$$

called dual norm of $\|\cdot\|$*, denoted by* $\|\cdot\|_*$

- examples
  - dual cone of subspace $V \subset \mathbf{R}^n$ is orthogonal complement of V, $V^\perp$, where $V^\perp = \{y | \forall v \in V, v^T y = 0\}$
  - $\mathbf{R}^n_+$ and $\mathbf{S}^n_+$ are self-dual
  - *dual of norm cone* is *norm cone associated with dual norm*, $i.e.$, if $K = \{(x,t) \in \mathbf{R}^n \times \mathbf{R} | \|x\| \leq t\}$

$$K = \{(y,u) \in \mathbf{R}^n \times \mathbf{R} | \|y\|_* \leq u\}$$

Properties of dual cones

Proposition 2. [properties of dual cones] *for cones K, K_1, and K_2*

- K^* *is closed and convex*
- $K_1 \subset K_2 \Rightarrow K_2^* \subset K_1^*$
- *if* $K^\circ \neq \emptyset$, K^* *is pointed*
- *if* $\overline{K}$ *is pointed,* $(K^*)^\circ \neq \emptyset$
- $K^{**} = (K^*)^*$ *is closure of convex hull of* K,
- K^* *is closed and convex*

thus,

- *if* K *is closed and convex,* $K^{**} = K$
- *dual of proper cone is proper cone*
- *for proper cone* K, $K^{**} = K$

Dual generalized inequalities

- dual of proper cone is proper (Proposition 2), hence the dual also induces generalized inequalities

Proposition 3. *for proper cone K,*

- $x \preceq_K y$ if and only if $(\forall \lambda \succeq_{K^*} 0)(\lambda^T x \leq \lambda^T y)$
- $x \prec_K y$ if and only if $(\forall \lambda \succeq_{K^*} 0$ *with* $\lambda \neq 0)(\lambda^T x < \lambda^T y)$

$K^{**} = K$, hence above are equivalent to

- $x \preceq_{K^*} y$ if and only if $(\forall \lambda \succeq_K 0)(\lambda^T x \leq \lambda^T y)$
- $x \prec_{K^*} y$ if and only if $(\forall \lambda \succeq_K 0$ *with* $\lambda \neq 0)(\lambda^T x < \lambda^T y)$

Theorem of alternative for linear strict generalized inequalities

Theorem 3. [theorem of alternative for linear strict generalized inequalities] *for proper cone* $K \subset \mathbf{R}^m$, $A \in \mathbf{R}^{m \times n}$, *and* $b \in \mathbf{R}^m$,

$$Ax \prec_K b$$

is infeasible if and only if *exist nonzero* $\lambda \in \mathbf{R}^m$ *such that*

$$\lambda \neq 0,\ \lambda \succeq_{K^*} 0,\ A^T\lambda = 0,\ \lambda^T b \leq 0$$

Above two inequality systems are alternative, $i.e.$, *for any data,* A *and* b, *exactly one of them is feasible. (proof can be found in Proof 2)*

Convex Functions

Convex functions

Definition 28. [convex functions]

- *function* $f : \mathbf{R}^n \to \mathbf{R}$ *the domain of which is convex and which satisfies*

$$(\forall x, y \in \mathbf{dom}\, f, 0 \leq \theta \leq 1)\,(f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y))$$

  said to be convex
- *function* $f : \mathbf{R}^n \to \mathbf{R}$ *the domain of which is convex and which satisfies*

$$(\forall \text{ distinct } x, y \in \mathbf{dom}\, f, 0 < \theta < 1)\,(f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta)f(y))$$

  said to be strictly convex

Definition 29. [concave functions]

- *function* $f : \mathbf{R}^n \to \mathbf{R}$ *the domain of which is convex where* $-f$ *is convex, said to be* concave
- *function* $f : \mathbf{R}^n \to \mathbf{R}$ *the domain of which is convex where* $-f$ *is strictly convex, said to be* strictly concave

Extended real-value extensions of convex functions

Definition 30. [extended real-value extension of convex functions] *for convex function* f*, function* $\tilde{f} : \mathbf{R}^n \to \mathbf{R} \cup \{\infty\}$ *defined by*

$$\tilde{f}(x) = \begin{cases} f(x) & \text{if } x \in \mathbf{dom}\, f \\ \infty & \text{if } x \notin \mathbf{dom}\, f \end{cases}$$

called extended real-value extension of f

- using extended real-value extensions of convex functions, can drop "$\mathbf{dom}\, f$" in equations, $e.g.$,
  - f is convex *if and only if* its extended-value extension $\tilde{f}$ satisfies

$$(\forall x, y \in \mathbf{dom}\, f, 0 \leq \theta \leq 1)\,(f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y))$$

  - f is strictly convex *if and only if* its extended-value extension $\tilde{f}$ satisfies

$$(\forall \text{ distinct } x, y \in \mathbf{dom}\, f, 0 < \theta < 1)\,(f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta)f(y))$$

First-order condition for convexity

Theorem 4. [first-order condition for convexity] *differentiable f, i.e., $\mathbf{dom}\, f$ is open and gradient ∇f exists at every point in $\mathbf{dom}\, f$, is*

- *convex* if and only if $\mathbf{dom}\, f$ *is convex and*

$$(\forall x, y \in \mathbf{dom}\, f)\left(f(y) \geq f(x) + \nabla f(x)^T(y-x)\right)$$

- *strictly convex* if and only if $\mathbf{dom}\, f$ *is convex and*

$$(\forall \text{ distinct } x, y \in \mathbf{dom}\, f)\left(f(y) > f(x) + \nabla f(x)^T(y-x)\right)$$

- Theorem 4 implies that for convex function f
  - first-order Taylor approximation is *global underestimator*
  - can derive global information from local information
    - *e.g.*, if $\nabla f(x) = 0$, x is global minimizer
    - *explains remarkable properties of convex functions and convex optimization problems*

Second-order condition for convexity

Theorem 5. [second-order condition for convexity] *twice-differentiable* f, *i.e.*, $\mathbf{dom}\, f$ *is open and Hessian* $\nabla^2 f$ *exists at every point in* $\mathbf{dom}\, f$, *is convex* if and only if $\mathbf{dom}\, f$ *is convex and*

$$(\forall x \in \mathbf{dom}\, f)\left(\nabla^2 f(x) \succeq 0\right)$$

- *if* $\mathbf{dom}\, f$ *is convex and*

$$(\forall x \in \mathbf{dom}\, f)\left(\nabla^2 f(x) \succ 0\right)$$

it is strictly convex

Convex function examples

- assume function $f : \mathbf{R}^n \to \mathbf{R}$ and $\mathbf{dom}\, f = \mathbf{R}^n$ unlesss specified otherwise
- affine function, *i.e.*, $f(x) = a^T x + b$ for some $a \in \mathbf{R}^n$ and $b \in \mathbf{R}$, is convex
- quadratic functions - if $f(x) = x^T P x + q^T x$ for some $P \in \mathbf{S}^n$ and $q \in \mathbf{R}^n$
  - f is convex *if and only if* $P \succeq 0$
  - f is strictly convex *if and only if* $P \succ 0$
- exponential function, *i.e.*, $f(x) = \exp(a^T x + b)$ for some $a \in \mathbf{R}^n$ and $b \in \mathbf{R}$, is convex
- power, *i.e.*, $f(x) = x^a$ for some $a \geq 1$, is convex on $\mathbf{R}_{++}$
- power of absolute value, *i.e.*, $f(x) = |x|^a$ for some $a \geq 1$, is convex on $\mathbf{R}$
- logarithm function, *i.e.*, $f(x) = \log x$, is concave on $\mathbf{R}_{++}$
- negative entropy, *i.e.*,

$$f(x) = \begin{cases} x \log x & \text{if } x > 0 \\ 0 & \text{if } x = 0 \end{cases}$$

  is convex on $\mathbf{R}_+$

- norm as function is convex (by definition of norms, $i.e.$, triangle inequality & absolute homogeneity)
- max function, $i.e.$, $f(x) = \max(x_1, \ldots, x_n\}$, is convex
- quadratic-over-linear function, $f(x, y) = x^2/y$, is convex on $\mathbf{R} \times \mathbf{R}_{++}$
- log-sum-exp, $f(x) = \log(\exp(x_1) + \cdots + \exp(x_n))$, is convex
- geometric mean, $f(x) = (\prod_{i=1}^{n} x_i)^{1/n}$, is concave on $\mathbf{R}^n_{++}$
- log-determinant, $f(X) = \log\det X$, is concave on $\mathbf{S}^n_{++}$

Sublevel sets and superlevel sets

Definition 31. [sublevel sets] *for function* f *and* $\alpha \in \mathbf{R}$,

$$\{x \in \mathbf{dom}\, f | f(x) \leq \alpha\}$$

called α-sublevel set of f

Definition 32. [superlevel sets] *for function* f *and* $\alpha \in \mathbf{R}$,

$$\{x \in \mathbf{dom}\, f | f(x) \geq \alpha\}$$

called α-superlevel set of f

Proposition 4. [convexity of level sets]

- *every sublevel set of convex function is convex*
- *and every superlevel set of concave function is convex*

- note, however, converse is not true
  - $e.g.$, every sublevel set of $\log$ is convex, but $\log$ is concave

Epigraphs and hypographs

Definition 33. [epigraphs] *for function* f*,*

$$\{(x,t)|x \in \mathbf{dom}\, f, f(x) \leq t\}$$

called epigraph of f*, denoted by* $\mathbf{epi}\, f$

Definition 34. [hypographs] *for function* f*,*

$$\{(x,t)|x \in \mathbf{dom}\, f, f(x) \geq t\}$$

called hypograph of f*, denoted by* $\mathbf{hypo}\, f$

Proposition 5. [graphs and convexity]

- *function is convex* if and only if *its epigraph is convex*
- *function is concave* if and only if *its hypograph is convex*

Convexity preserving function operations

Proposition 6. [convexity preserving function operations]

- *nonnegative weighted sum preserves convexity*
  - *for convex functions* $f_1, \ldots, f_n$ *and nonnegative weights* $w_1, \ldots, w_n$

$$w_1 f_1 + \cdots w_n f_n$$

    is convex
- *nonnegative weighted integration preserves convexity*
  - *for measurable set* Y, $w : Y \to \mathbf{R}_+$, *and* $f : X \times Y$ *where* $f(x, y)$ *is convex in* x *for every* $y \in Y$ *and measurable in* y *for every* $x \in X$

$$\int_Y w(y) f(x, y) dy$$

    is convex
- *pointwise maximum preserves convexity*

  - *for convex functions* $f_1, \ldots, f_n$

$$\max\{f_1, \ldots, f_n\}$$

    is convex

- *pointwise supremum preserves convexity*
  - *for indexed family of convex functions* $\{f_\lambda\}_{\lambda \in \Lambda}$

$$\sup_{\lambda \in \Lambda} f_\lambda$$

    is convex (one way to see this is $\mathbf{epi} \sup_\lambda f_\lambda = \bigcap_\lambda \mathbf{epi}\, f_\lambda$*)*
- *composition*
  - *suppose* $g : \mathbf{R}^n \to \mathbf{R}^k$, $h : \mathbf{R}^k \to \mathbf{R}$, *and* $f = h \circ g$
    - f *convex if* h *convex & nondecreasing in each argument, and* g_i *convex*
    - f *convex if* h *convex & nonincreasing in each argument, and* g_i *concave*
    - f *concave if* h *concave & nondecreasing in each argument, and* g_i *concave*
    - f *concave if* h *concave & nonincreasing in each argument, and* g_i *convex*
- *minimization*

  - *for function $f(x, y)$ convex in (x, y) and convex set C*

$$\inf_{y \in C} f(x, y)$$

is convex provided it is bounded below where domain is $\{x | (\exists y \in C)((x, y) \in \mathbf{dom}\, f)\}$ (proof can be found in Proof 3)

- *perspective of convex function preserves convexity*
  - *for convex function $f : X \to \mathbf{R}$, function $g : X \times \mathbf{R} \to \mathbf{R}$ defined by*

$$g(x, t) = tf(x/t)$$

with $\mathbf{dom}\, g = \{(x, t) | x/t \in \mathbf{dom}\, f, t > 0\}$ is convex

Convex functions examples

Proposition 6 implies

- piecewise-linear function is convex, $i.e.$
  - $\max\{a_1^T x + b_1, \ldots, a_m^T x + b_m\}$ for some $a_i \in \mathbf{R}^n$ and $b_i \in \mathbf{R}$ is convex
- sum of k largest components is convex, $i.e.$
  - $x_{[1]} + \cdots + x_{[k]}$ where $x_{[i]}$ denotes i-th largest component, is convex (since $f(x) = \max\{x_{i_1} + \cdots + x_{i_r} | 1 \leq i_1 < i_2 < \cdots < i_r \leq n\}$)
- support function of set, $i.e.$,
  - $\sup\{x^T y | y \in A\}$ for $A \subset \mathbf{R}^n$ is convex
- distance (when measured by arbitrary norm) to farthest point of set
  - $\sup\{\|x - y\| | y \in A\}$ for $A \subset \mathbf{R}^n$ is convex
- least-squares cost as function of weights
  - $\inf_{x \in \mathbf{R}^n} \sum_{i=1}^n w_i (a_i^T x - b_i)^2$ for some $a_i \in \mathbf{R}^n$ and $b_i \in \mathbf{R}$ is concave
    - note that above function equals to $\sum_{i=1}^n w_i b_i^2 - \sum_{i=1}^n w_i^2 b_i^2 a_i^T \left(\sum_{j=1}^n w_j a_j a_j^T \right)^{-1} a_i$ but not clear whether it is concave

- maximum eigenvalue of symmetric matrix
  - $\lambda_{\max}(F(x)) = \sup\{y^T F(x) y | \|y\|_2 \leq 1\}$ where $F : \mathbf{R}^n \to \mathbf{S}^m$ is linear function in x
- norm of matrix
  - $\sup\{u^T G(x) v | \|u\|_2 \leq 1, \|v\|_2 \leq 1\}$ where $G : \mathbf{R}^n \to \mathbf{R}^{m \times n}$ is linear function in x
- distance (when measured by arbitrary norm) to convex set
  - for convex set C, $\inf\{\|x - y\| | y \in C\}$
- infimum of convex function subject to linear constraint
  - for convex function h, $\inf\{h(y) | Ay = x\}$ is convex (since it is $\inf_y(h(y) + I_{Ay=x}(x, y))$)
- perspective of Euclidean norm squared
  - map $(x, t) \mapsto x^T x / t$ induces convex function in (x, t) for $t > 0$
- perspective of negative log
  - map $(x, t) \mapsto -t \log(x/t)$ induces convex function in $(x, t) \in \mathbf{R}^2_{++}$

- perspective of convex function
  - for convex function $f : \mathbf{R}^n \to \mathbf{R}$, function $g : \mathbf{R}^n \to \mathbf{R}$ defined by

$$g(x) = (c^T x + d) f((Ax + b)/(c^T x + d))$$

from some $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, $c \in \mathbf{R}^n$, and $d \in \mathbf{R}$ with $\mathbf{dom}\, g = \{x | (Ax + b)/(c^T x + d) \in \mathbf{dom}\, f, c^T x + d > 0\}$ is convex

Conjugate functions

Definition 35. [conjugate functions] *for function* f

$$\sup_{y \in \mathbf{dom}\, f} (x^T y - f(y))$$

called conjugate function of f, *denoted by* f^*

- conjugate function is convex for any function f because it is supremum of linear (hence convex) functions (in x) (Proposition 6)

Inequality 1. [Fenchel's inequality] *definition of conjugate function implies*

$$f(x) + f^*(y) \geq x^T y$$

sometimes called Young's inequality

Proposition 7. [conjugate of conjugate] *for convex and closed function* f

$$f^{**} = f$$

where closed function f *is defined by function with closed* $\mathbf{epi}\, f$

Conjugate function examples

- strictly convex quadratic function
  - for $f : \mathbf{R}^n \to \mathbf{R}_+$ defined $f(x) = x^T Q x/2$ where $Q \in \mathbf{S}^n_{++}$,

$$f^*(x) = \sup_y (y^T x - y^T Q y/2) = (y^T x - y^T Q y/2)|_{y=Q^{-1}x} = x^T Q^{-1} x/2$$

    which is also strictly convex quadratic function
- log-determinant
  - for function $f : \mathbf{S}^n_{++} \to \mathbf{R}$ defined by $f(X) = \log\det X^{-1}$

$$f^*(X) = \sup_{Y \in \mathbf{S}^n_{++}} (\mathbf{Tr}\, XY + \log\det Y) = \log\det(-X)^{-1} - n$$

    where $\mathbf{dom}\, f^* = -\mathbf{S}^n_{++}$
- indicator function

  - for indicator function $I_A : \mathbf{R}^n \to \{0, \infty\}$ with $A \subset \mathbf{R}^n$

$$I_A^*(x) = \sup_y (y^T x - I_A(y)) = \sup\{y^T x | y \in A\}$$

    which is support function of A

- log-sum-exp function
  - for function $f : \mathbf{R}^n \to \mathbf{R}$ defined by $f(x) = \log(\sum_{i=1}^n \exp(x_i))$

$$f^*(x) = \sum_{i=1}^n x_i \log x_i + I_{x \succeq 0, \mathbf{1}^T x = 1}(x)$$

- norm
  - for norm function $f : \mathbf{R}^n \to \mathbf{R}_+$ defined by $f(x) = \|x\|$

$$f^*(x) = \sup_y (y^T x - \|y\|) = I_{\|x\|_* \leq 1}(x)$$

- norm squared

  - for function $f : \mathbf{R} \to \mathbf{R}_+$ defined by $f(x) = \|x\|^2/2$

$$f^*(x) = \|x\|_*^2/2$$

- differentiable convex function
  - for differentiable convex function $f : \mathbf{R}^n \to \mathbf{R}$

$$f^*(x) = (y^*)^T \nabla f(y^*) - f(y^*)$$

    where $y^* = \operatorname{argsup}_y(x^T y - f(y))$
- sum of independent functions
  - for function $f : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$ defined by $f(x, y) = f_1(x) + f_2(y)$ where $f_1 : \mathbf{R}^n \to \mathbf{R}$ and $f_2 : \mathbf{R}^m \to \mathbf{R}$

$$f^*(x, y) = f_1^*(x) + f_2^*(y)$$

Convex functions with respect to generalized inequalities

Definition 36. [K-convex functions] *for proper cone K,*

- *function f satisfying*

$$(\forall x, y \in \mathbf{dom}\, f, 0 \leq \theta \leq 1)\,(f(\theta x + (1-\theta)y) \preceq_K \theta f(x) + (1-\theta)f(y))$$

called K-convex

- *function f satisfying*

$$(\forall x \neq y \in \mathbf{dom}\, f, 0 < \theta < 1)\,(f(\theta x + (1-\theta)y) \prec_K \theta f(x) + (1-\theta)f(y))$$

called strictly K-convex

Proposition 8. [dual characterization of K-convexity] *for proper cone K*

- *function f is K-convex* if and only if *for every $w \succeq_{K^*} 0$, $w^T f$ is convex*
- *function f is strictly K-convex* if and only if *for every nonzero $w \succeq_{K^*} 0$, $w^T f$ is strictly convex*

Matrix convexity

Definition 37. [matrix convexity] *function of* $\mathbf{R}^n$ *into* $\mathbf{S}^m$ *which is* K*-convex where* $K = \mathbf{S}^m_+$*, called* matrix convex

- examples of matrix convexity
  - function of $\mathbf{R}^{n\times m}$ into $\mathbf{S}^n_+$ defined by $X \mapsto XX^T$ is matrix convex
  - function of $\mathbf{S}^n_{++}$ into itself defined by $X \mapsto X^p$ is matrix convex for $1 \leq p \leq 2$ or $-1 \leq p \leq 0$, and matrix concave for $0 \leq p \leq 1$
  - function of $\mathbf{S}^n$ into $\mathbf{S}^n_{++}$ defined by $X \mapsto \exp(X)$ is *not* matrix convex
  - quadratic matrix function of $\mathbf{R}^{m\times n}$ into $\mathbf{S}^n$ defined by $X \mapsto X^TAX + B^TX + X^TB + C$ for $A \in \mathbf{S}^m$, $B \in \mathbf{R}^{m\times n}$, and $C \in \mathbf{S}^n$ is matrix convex when $A \succeq 0$

Convex Optimization Problems

Optimization problems

Definition 38. [optimization problems] *for* $f : F \to \mathbf{R}$, $q : Q \to \mathbf{R}^m$, $h : H \to \mathbf{R}^p$ *where* F, Q, *and* H *are subsets of common set* X

$$\begin{array}{ll} \textit{minimize} & f(x) \\ \textit{subject to} & q(x) \preceq 0 \\ & h(x) = 0 \end{array}$$

called optimization problem *where* x *is* optimization variable

- f, q, *and* h *are* objective function, inequality & equality contraint function
- $q(x) \preceq 0$ *and* $h(x) = 0$ *are* inequality contraints *and* equality contraints
- $\mathcal{D} = F \cap Q \cap H$ *is* domain *of optimization problem*
- $\mathcal{F} = \{x \in \mathcal{D} | q(x) \preceq 0, h(x) = 0\}$, *called* feasible set, $x \in \mathcal{D}$, *said to be* feasible *if* $x \in \mathcal{F}$, *optimization problem, said to be* feasible *if* $\mathcal{F} \neq \emptyset$
- $p^* = \inf\{f(x) | x \in \mathcal{F}\}$, *called* optimal value *of optimization problem*
- *if optimization problem is* infeasible, $p^* = \infty$ *(following convention that infimum of empty set is* ∞*)*
- *if* $p^* = -\infty$, *optimization problem said to be* unbounded

Global and local optimalities

Definition 39. [global optimality] *for optimization problem in Definition 38*

- $x \in \mathcal{F}$ *with* $f(x) = p^*$, *called* (global) optimal point
- $X_{\text{opt}} = \{x \in \mathcal{F} | f(x) = p^*\}$, *called* optimal set
- *when* $X_{\text{opt}} \neq \emptyset$, *we say optimal value is* attained *or* achieved *and optimization problem is* solvable

- optimization problem is *not* solvable if $p^* = \infty$ or $p^* = -\infty$ (converse is not true)

Definition 40. [local optimality] *for optimization problem in Definition 38 where* X *is metric space,* $x \in \mathcal{F}$ *satisfying* $\inf\{f(z) | z \in \mathcal{F}, \rho(z, x) \leq r\}$ *where* $\rho : X \times X \to \mathbf{R}_+$ *is metric, for some* $r > 0$, *said to be* locally optimal, *i.e.,* x *solves*

$$\begin{array}{ll} \textit{minimize} & f(z) \\ \textit{subject to} & q(z) \preceq 0 \\ & h(z) = 0 \\ & \rho(z, x) \leq r \end{array}$$

Equivalent optimization problems

Definition 41. [equivalent optimization problems] *two optimization problems where solving one readily solve the other, said to be* equivalent

- below two optimization problems are equivalent
  - $$\begin{array}{ll} \text{minimize} & -x - y \\ \text{subject to} & 2x + y \leq 1 \\ & x + 2y \leq 1 \end{array}$$
  - $$\begin{array}{ll} \text{minimize} & -2u - v/3 \\ \text{subject to} & 4u + v/3 \leq 1 \\ & 2u + 2v/3 \leq 1 \end{array}$$

  since if (x^*, y^*) solves first, $(u, v) = (x^*/2, 3y^*)$ solves second, and if (u^*, v^*) solves second, $(x, y) = (2u^*, v^*/3)$ solves first

Change of variables

- given function $\phi : \mathcal{Z} \to X$, optimization problem in Definition 38 can be rewritten as

$$\begin{array}{ll} \text{minimize} & f(\phi(z)) \\ \text{subject to} & q(\phi(z)) \preceq 0 \\ & h(\phi(z)) = 0 \end{array}$$

  where $z \in \mathcal{Z}$ is optimization variable

- if ϕ is injective and $\mathcal{D} \subset \phi(\mathcal{Z})$, above optimization problem and optimization problem in Definition 38 are equivalent, *i.e.*
  - X_{opt} is optimal set of problem in Definition 38 $\Rightarrow$ $\phi^{-1}(X_{\text{opt}})$ is optimal set of above problem
  - Z_{opt} is optimal set of above problem $\Rightarrow$ $\phi(Z_{\text{opt}})$ is optimal set of problem in Definition 38
- two optimization problems said to be related by *change of variable or substitution of variable* $x = \phi(z)$

Convex optimization

Definition 42. [convex optimization] *optimization problem in Definition 38 where X is Banach space, $i.e.$, complete linear normed vector space, f & q are convex functions, and h is affine function, called* convex optimization problem

- *when $X = \mathbf{R}^n$, optimization problem can be formulated as*

$$\begin{array}{ll} \textit{minimize} & f(x) \\ \textit{subject to} & q(x) \preceq 0 \\ & Ax = b \end{array}$$

for some $A \in \mathbf{R}^{p \times n}$ and $b \in \mathbf{R}^p$

- domain of convex optimization problem is *convex*
  - since domains of f, q, and h are convex (by definition of convex functions) and intersection of convex sets is convex
- feasible set of convex optimization problem is *convex*
  - since sublevel sets of convex functions are convex, feasible sets for affine function is either empty set, singleton, or affine sets, all of which are convex sets

Optimality conditions for convex optimization problems

Theorem 6. [local optimality implies global optimality] *for convex optimization problem (in Definition 42), every local optimal point is global optimal point*

Theorem 7. [optimality conditions for convex optimality problems] *for convex optimization problem (in Definition 42), when f is differentiable ($i.e.$, $\mathbf{dom}\, f$ is open and ∇f exists everywhere in $\mathbf{dom}\, f$)*

- $x \in \mathcal{D}$ *is optimal* if and only if $x \in \mathcal{F}$ *and*

$$(\forall y \in \mathcal{F}) \left(\nabla f(x)^T (y - x) \geq 0 \right)$$

- *for unconstrained problems,* $x \in \mathcal{D}$ *is optimal* if and only if

$$\nabla f(x) = 0$$

Optimality conditions for some convex optimization problems

- unconstrained convex quadratic optimization

$$\text{minimize} \quad f(x) = (1/2)x^T Px + q^T x$$

where $F = \mathbf{R}^n$ and $P \in \mathbf{S}^n_+$

– x is optimal *if and only if*

$$\nabla f(x) = Px + q = 0$$

exist three cases

- if $P \in \mathbf{S}^n_{++}$, exists unique optimum $x^* = -P^{-1}q$
- if $q \in \mathcal{R}(P)$, $X_{\text{opt}} = -P^{\dagger}q + \mathcal{N}(P)$
- if $q \notin \mathcal{R}(P)$, $p^* = -\infty$

- analytic centering

$$\text{minimize} \quad f(x) = -\sum_{i=1}^m \log(b_i - a_i^T x)$$

where $F = \{x \in \mathbf{R}^n | Ax \prec b\}$

– x is optimal *if and only if*

$$\nabla f(x) = \sum_{i=1}^{m} \frac{1}{b_i - a_i^T x} a_i = 0$$

exist three cases

- exists unique optimum, which happens *if and only if* $\{x | b_i - a_i^T x\}$ is nonempty and bounded
- exist infinitely many optima, in which case, X_{opt} is affine set
- exists no optimum, which happens *if and only if* f is unbounded below

- convex optimization problem with equality constraints only

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \end{array}$$

where $X = \mathbf{R}^n$

– x is optimal *if and only if*

$$\nabla f(x) \perp \mathcal{N}(A)$$

or equivalently, exists $\nu \in \mathbf{R}^p$ such that

$$\nabla f(x) = A^T \nu$$

Linear programming

Definition 43. [linear programming] *convex optimization problem in Definition 42 with* $X = \mathbf{R}^n$ *and linear* f *&* g*, called* linear program (LP)*, which can be formulated as*

$$\begin{array}{ll} \textit{minimize} & c^T x \\ \textit{subject to} & Cx \preceq d \\ & Ax = b \end{array}$$

where $c \in \mathbf{R}^n$, $C \in \mathbf{R}^{m \times n}$, $d \in \mathbf{R}^m$, $A \in \mathbf{R}^{p \times n}$, $b \in \mathbf{R}^p$

- *can transform above LP into* standard form LP

$$\begin{array}{ll} \textit{minimize} & \tilde{c}^T \tilde{x} \\ \textit{subject to} & \tilde{A}\tilde{x} = \tilde{b} \\ & \tilde{x} \succeq 0 \end{array}$$

LP examples

- diet problem - find amount of n different food to minimize purchase cost while satisfying nutrition requirements
  - assume exist n food and m nutritions, c_i is cost of food i, A_{ji} is amount of nutrition j contained in unit quantity of food i, b_j is amount requirement for nutrition j
  - diet problem can be formulated as LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \succeq b \\ & x \succeq 0 \end{array}$$

- Chebyshev center of polyhedron - find largest Euclidean ball contained in polyhedron
  - assume polyhedron is $\{x \in \mathbf{R}^n | a_i^T x \leq b_i, i = 1, \ldots, m\}$
  - problem of finding Chebyshev center of polyhedron can be formulated as LP

$$\begin{array}{ll} \text{maximize} & r \\ \text{subject to} & a_i^T x + r\|a_i\|_2 \leq b_i \end{array}$$

where optimization variables are $x \in \mathbf{R}^n$ and $r \in \mathbf{R}$

- piecewise-linear minimization - minimize maximum of affine functions
  - assume m affine functions $a_i^T x + b_i$
  - piecewise-linear minimization problem can be formulated as LP

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & a_i^T x + b_i \leq t, \quad i = 1, \ldots, m \end{array}$$

- linear-fractional program

$$\begin{array}{ll} \text{minimize} & (c^T x + d)/(e^T x + f) \\ \text{subject to} & Gx \preceq h \\ & Ax = b \end{array}$$

– if feasible set is nonempty, can be formulated as LP

$$\begin{array}{ll} \text{minimize} & c^T y + dz \\ \text{subject to} & Gy - hz \preceq 0 \\ & Ay - bz = 0 \\ & e^T y + fz = 1 \\ & z \geq 0 \end{array}$$

Quadratic programming

Definition 44. [quadratic programming] *convex optimization problem in Definition 42 with $X = \mathbf{R}^n$ and convex quadratic f and linear q, called quadratic program (QP), which can be formulated as*

$$\begin{array}{ll} \textit{minimize} & (1/2)x^T Px + q^T x \\ \textit{subject to} & Gx \preceq h \\ & Ax = b \end{array}$$

where $P \in \mathbf{S}_+^n$, $q \in \mathbf{R}^n$, $G \in \mathbf{R}^{m \times n}$, $h \in \mathbf{R}^m$, $A \in \mathbf{R}^{p \times n}$, $b \in \mathbf{R}^p$

- when $P = 0$, QP reduces to LP, hence *LP is specialization of QP*

QP examples

- least-squares (LS) problems
  - LS can be formulated as QP

$$\text{minimize} \quad \|Ax - b\|_2^2$$

- distance between two polyhedra
  - assume two polyhedra $\{x \in \mathbf{R}^n | Ax \preceq b, Cx = d\}$ and $\{x \in \mathbf{R}^n | \tilde{A}x \preceq \tilde{b}, \tilde{C}x = \tilde{d}\}$
  - problem of finding distance between two polyhedra can be formulated as QP

$$\begin{array}{ll} \text{minimize} & \|x - y\|_2^2 \\ \text{subject to} & Ax \preceq b, \quad Cx = d \\ & \tilde{A}y \preceq \tilde{b}, \quad \tilde{C}y = \tilde{d} \end{array}$$

Quadratically constrained quadratic programming

Definition 45. [quadratically constrained quadratic programming] *convex optimization problem in Definition 42 with $X = \mathbf{R}^n$ and convex quadratic f & q, called quadratically constrained quadratic program (QCQP), which can be formulated as*

$$\begin{array}{ll} \textit{minimize} & (1/2)x^T P_0 x + q_0^T x \\ \textit{subject to} & (1/2)x^T P_i x + q_i^T x + r_i \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

where $P_i \in \mathbf{S}_+^n$, $q_i \in \mathbf{R}^n$, $r_i \in \mathbf{R}$, $A \in \mathbf{R}^{p \times n}$, $b \in \mathbf{R}^p$

- when $P_i = 0$ for $i = 1, \ldots, m$, QCQP reduces to QP, hence *QP is specialization of QCQP*

Second-order cone programming

Definition 46. [second-order cone programming] *convex optimization problem in Definition 42 with* $X = \mathbf{R}^n$ *and linear* f *and convex* q *of form*

$$\begin{array}{ll} \textit{minimize} & f^T x \\ \textit{subject to} & \|A_i x + b_i\|_2 \leq c_i^T x + d_i, \quad i = 1, \ldots, m \\ & Fx = g \end{array}$$

where $f \in \mathbf{R}^n$, $A_i \in \mathbf{R}^{n_i \times n}$, $b_i \in \mathbf{R}^{n_i}$, $c_i \in \mathbf{R}^n$, $d_i \in \mathbf{R}$, $F \in \mathbf{R}^{p \times n}$, $g \in \mathbf{R}^p$ *called* second-order cone program (SOCP)

- when $b_i = 0$, SOCP reduces to QCQP, hence *QCQP is specialization of SOCP*

SOCP examples

- robust linear program - minimize $c^T x$ while satisfying $\tilde{a}_i^T x \leq b_i$ for every $\tilde{a}_i \in \{a_i + P_i u \mid \|u\|_2 \leq 1\}$ where $P_i \in \mathbf{S}^n$
  - can be formulated as SOCP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x + \|P_i^T x\|_2 \leq b_i \end{array}$$

- linear program with random constraints - minimize $c^T x$ while satisfying $\tilde{a}_i^T x \leq b_i$ with probability no less than η where $\tilde{a} \sim \mathcal{N}(a_i, \Sigma_i)$
  - can be formulated as SOCP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x + \Phi^{-1}(\eta)\|\Sigma_i^{1/2} x\|_2 \leq b_i \end{array}$$

Geometric programming

Definition 47. [monomial functions] *function* $f : \mathbf{R}^n_{++} \to \mathbf{R}$ *defined by*

$$f(x) = cx_1^{a_1} \cdots x_n^{a_n}$$

where $c > 0$ *and* $a_i \in \mathbf{R}$, *called* monomial function *or simply* monomial

Definition 48. [posynomial functions] *function* $f : \mathbf{R}^n_{++} \to \mathbf{R}$ *which is finite sum of monomial functions, called* posynomial function *or simply* posynomial

Definition 49. [geometric programming] *optimization problem*

$$\begin{array}{ll} \textit{minimize} & f(x) \\ \textit{subject to} & q(x) \preceq 1 \\ & h(x) = 1 \end{array}$$

for posynomials $f : \mathbf{R}^n_{++} \to \mathbf{R}$ *&* $q : \mathbf{R}^n_{++} \to \mathbf{R}^m$ *and monomials* $h : \mathbf{R}^n_{++} \to \mathbf{R}^p$, *called* geometric program (GP)

Geometric programming in convex form

- geometric program in Definition 49 is not convex optimization problem (as it is)
- however, can be transformed to equivalent convex optimization problem by change of variables and transformation of functions

Proposition 9. [geometric programming in convex form] *geometric program (in Definition 49) can be transformed to equivalent convex optimization problem*

$$\begin{array}{ll} \textit{minimize} & \log\left(\sum_{k=1}^{K_0} \exp((a_k^{(0)})^T y + b_k^{(0)})\right) \\ \textit{subject to} & \log\left(\sum_{k=1}^{K_i} \exp((a_k^{(i)})^T y + b_k^{(i)})\right) \leq 0 \quad i = 1, \ldots, m \\ & Gy = h \end{array}$$

for some $a_k^{(i)} \in \mathbf{R}^n$, $b_k^{(i)} \in \mathbf{R}$, $G \in \mathbf{R}^{p \times n}$, $h \in \mathbf{R}^p$ *where optimization variable is* $y = \log(x) \in \mathbf{R}^n$

Convex optimization with generalized inequalities

Definition 50. [convex optimization with generalized inequality constraints] *convex optimization problem in Definition 42 with inequality constraints replaced by generalized inequality constraints, $i.e.$*

$$\begin{array}{ll} \textit{minimize} & f(x) \\ \textit{subject to} & q_i(x) \preceq_{K_i} 0 \quad i = 1, \ldots, q \\ & h(x) = 0 \end{array}$$

where $K_i \subset R^{k_i}$ are proper cones and $q_i : Q_i \to \mathbf{R}^{k_i}$ are K_i-convex, called convex optimization problem with generalized inequality constraints

- problem in Definition 50 reduces to convex optimization problem in Definition 42 when $q = 1$ and $K_1 = \mathbf{R}^m_+$, hence *convex optimization is specialization of convex optimization with generalized inequalities*
- like convex optimization
  - feasible set is $\mathcal{F} = \{x \in \mathcal{D} | q_i(x) \preceq_{K_i} 0, Ax = b\}$ is convex
  - local optimality implies global optimality
  - optimality conditions in Theorem 7 applies without modification

Conic programming

Definition 51. [conic programming] *convex optimization problem with generalized inequality constraints in Definition 50 with linear f and one affine q*

$$\begin{array}{ll} \textit{minimize} & f(x) \\ \textit{subject to} & q(x) \preceq_K 0 \\ & h(x) = 0 \end{array}$$

called conic program (CP)

- *can transform above CP to* standard form CP

$$\begin{array}{ll} \textit{minimize} & \tilde{f}(X) \\ \textit{subject to} & \tilde{h}(X) = 0 \\ & X \succeq_K 0 \end{array}$$

- cone program is one of simplest convex optimization problems with generalized inequalities

Semidefinite programming

Definition 52. [semidefinite programming] *conic program in Definition 51 with* $X = \mathbf{R}^n$ *and* $K = \mathbf{S}^n_+$

$$\begin{array}{ll} \textit{minimize} & c^T x \\ \textit{subject to} & x_1 F_1 + \cdots + x_n F_n + G \preceq 0 \\ & Ax = b \end{array}$$

where $F_1, \ldots, F_n, G \in \mathbf{S}^k$ *and* $A \in \mathbf{R}^{p \times n}$*, called* semidefinite program (SDP)

- *above inequality, called* linear matrix inequality (LMI)
- *can transform SDP to standard form SDP*

$$\begin{array}{ll} \textit{minimize} & \mathbf{Tr}(CX) \\ \textit{subject to} & \mathbf{Tr}(A_i X) = b_i \quad i = 1, \ldots, p \\ & X \succeq 0 \end{array}$$

where $X = \mathbf{S}^n_+$ *and* $C, A_1, \ldots, A_p \in \mathbf{S}^n$ *and* $b_i \in \mathbf{R}$

SDP examples

- LP
  - if $k = m$, $F_i = \mathbf{diag}(C_{1,i}, \ldots, C_{m,i})$, $G = -\mathbf{diag}(d_1, \ldots, d_m)$ in Definition 52, SDP reduces to LP in Definition 43
  - hence, LP is specialization of SDP
- SOCP
  - SOCP in Definition 46 is equivalent to

$$\begin{array}{ll} \text{minimize} & f^T x \\ \text{subject to} & Fx = g \\ & \begin{bmatrix} c_i^T x + d_i & x^T A_i^T + b_i^T \\ A_i x + b_i & (c_i^T x + d_i) I_{n_i} \end{bmatrix} \succeq 0 \quad i = 1, \ldots, m \end{array}$$

    which can be transformed to SDP in Definition 52, thus, SDP reduces to SOCP
  - hence, SOCP is specialization of SDP

Determinant maximization problems

Definition 53. [determinant maximization problems] *convex optimization problem with generalized inequality constraints in Definition 50 with* $X = \mathbf{R}^n$ *of form*

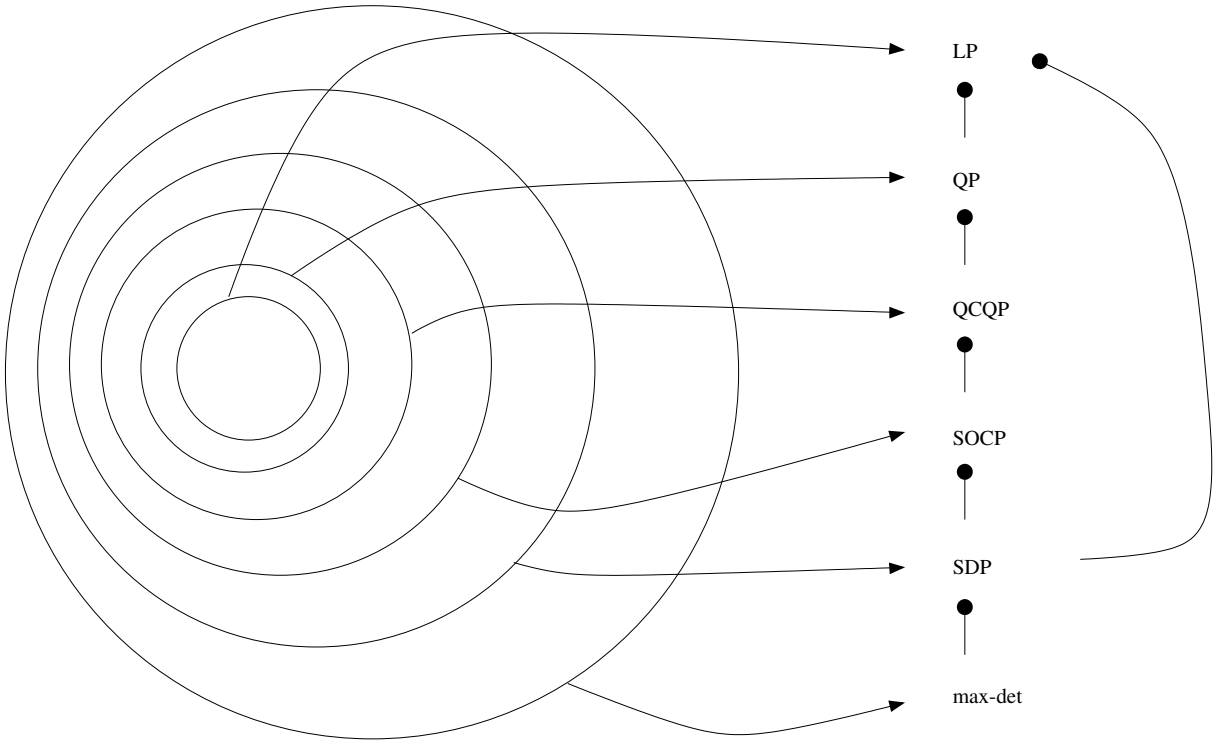
$$\begin{array}{ll} \textit{minimize} & -\log\det(x_1C_1 + \cdots + x_nC_n + D) + c^Tx \\ \textit{subject to} & x_1F_1 + \cdots + x_nF_n + G \preceq 0 \\ & -x_1C_1 - \cdots - x_nC_n - D \prec 0 \\ & Ax = b \end{array}$$

where $c \in \mathbf{R}^n$, $C_1, \ldots, C_n, D \in \mathbf{S}^l$, $F_1, \ldots, F_n, G \in \mathbf{S}^k$, *and* $A \in \mathbf{R}^{p\times n}$, *called* determinant maximization problem *or simply* max-det problem *(since it maximizes determinant of (positive definite) matrix with constraints)*

- if $l = 1$, $C_1 = \cdots = C_n = 0$, $D = 1$, max-det problem reduces to SDP, hence *SDP is specialization of max-det problem*

Diagrams for containment of convex optimization problems

- the figure shows containment relations among convex optimization problems
- vertical lines ending with filled circles indicate existence of direct reductions, $i.e.$, optimization problem transformations to special cases



Duality

Lagrangian

Definition 54. [Lagrangian] *for optimization problem in Definition 38 with nonempty domain $\mathcal{D}$, function $L : \mathcal{D} \times \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$ defined by*

$$L(x, \lambda, \nu) = f(x) + \lambda^T q(x) + \nu^T h(x)$$

called Lagrangian associated with the optimization problem where

- λ, *called Lagrange multiplier associated inequality constraints* $q(x) \preceq 0$
- λ_i, *called Lagrange multiplier associated* i*-th inequality constraint* $q_i(x) \leq 0$
- ν, *called Lagrange multiplier associated equality constraints* $h(x) = 0$
- ν_i, *called Lagrange multiplier associated* i*-th equality constraint* $h_i(x) = 0$
- λ *and* ν, *called dual variables or Lagrange multiplier vectors associated with the optimization problem*

Lagrange dual functions

Definition 55. [Lagrange dual functions] *for optimization problem in Definition 38 for which Lagrangian is defined, function* $g : \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R} \cup \{-\infty\}$ *defined by*

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} \left(f(x) + \lambda^T q(x) + \nu^T h(x) \right)$$

called Lagrange dual function *or just* dual function *associated with the optimization problem*

- g is *(always) concave function* (even when optimization problem is not convex)
  - since is pointwise infimum of linear (hence concave) functions is concave
- $g(\lambda, \nu)$ provides lower bound for optimal value of associated optimization problem, $i.e.$,

$$g(\lambda, \nu) \leq p^*$$

  for every $\lambda \succeq 0$ (proof can be found in Proof 4)
- $(\lambda, \nu) \in \{(\lambda, \nu) | \lambda \succeq 0, g(\lambda, \nu) > -\infty\}$, said to be *dual feasible*

Dual function examples

- LS solution of linear equations

$$\begin{array}{ll} \text{minimize} & x^T x \\ \text{subject to} & Ax = b \end{array}$$

  - Lagrangian - $L(x, \nu) = x^T x + \nu^T (Ax - b)$
  - Lagrange dual function

$$g(\nu) = -\frac{1}{4}\nu^T A A^T \nu - b^T \nu$$

- standard form LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \succeq 0 \end{array}$$

  - Lagrangian - $L(x, \lambda, \nu) = c^T x - \lambda^T x + \nu^T (Ax - b)$

  - Lagrange dual function

$$g(\lambda, \nu) = \begin{cases} -b^T\nu & A^T\nu - \lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases}$$

    - hence, set of dual feasible points is $\{(A^T\nu + c, \nu) | A^T\nu + c \succeq 0\}$

- *maximum cut*, sometimes called *max-cut*, problem, which is NP-hard

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1 \end{array}$$

  where $W \in \mathbf{S}^n$
  - Lagrangian - $L(x, \nu) = x^T(W + \mathbf{diag}(\nu))x - \mathbf{1}^T x$
  - Lagrange dual function

$$g(\nu) = \begin{cases} -\mathbf{1}^T\nu & W + \mathbf{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise} \end{cases}$$

    - hence, set of dual feasible points is $\{\nu | W + \mathbf{diag}(\nu) \succeq 0\}$

- some trivial problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x = 0 \end{array}$$

  - Lagrangian - $L(x, \nu) = f(x) + \nu^T x$
  - Lagrange dual function

$$g(\nu) = \inf_{x \in \mathbf{R}^n} (f(x) + \nu^T x) = - \sup_{x \in \mathbf{R}^n} ((-\nu)^T x - f(x)) = -f^*(-\nu)$$

    - hence, set of dual feasible points is $-\mathbf{dom}\, f^*$, and for every $f : \mathbf{R}^n \to \mathbf{R}$ and $\nu \in \mathbf{R}^n$

$$-f^*(-\nu) \leq f(0)$$

- minimization with linear inequality and equality constraints

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax \preceq b \\ & Cx = d \end{array}$$

  - Lagrangian - $L(x, \lambda, \nu) = f(x) + \lambda^T (Ax - b) + \nu^T (Cx - d)$

- Lagrange dual function

$$g(\nu) = -b^T\lambda - d^T\nu - f^*(-A^T\lambda - C^T\nu)$$

  - hence, set of dual feasible points is $\{(\lambda, \nu) | -A^T\lambda - C^T\nu \in \mathbf{dom}\, f^*, \lambda \succeq 0\}$

- equality constrained norm minimization

$$\begin{array}{ll} \text{minimize} & \|x\| \\ \text{subject to} & Ax = b \end{array}$$

- Lagrangian - $L(x, \nu) = \|x\| + \nu^T(Ax - b)$
- Lagrange dual function

$$g(\nu) = -b^T\nu - \sup_{x \in \mathbf{R}^n}((-A^T\nu)^T x - \|x\|) = \begin{cases} -b^T\nu & \|A^T\nu\|_* \leq 1 \\ -\infty & \text{otherwise} \end{cases}$$

  - hence, set of dual feasible points is $\{\nu | \|A^T\nu\|_* \leq 1\}$

- entropy maximization

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^n x_i \log x_i \\ \text{subject to} & Ax \preceq b \\ & \mathbf{1}^T x = 1 \end{array}$$

  where domain of objective function is $\mathbf{R}^n_{++}$
  - Lagrangian - $L(x, \lambda, \nu) = \sum_{i=1}^n x_i \log x_i + \lambda^T (Ax - b) + \nu(\mathbf{1}^T x - 1)$
  - Lagrange dual function

$$g(\lambda, \nu) = -b^T \lambda - \nu - \exp(-\nu - 1) \sum_{i=1}^n \exp(a_i^T \lambda)$$

    obtained using $f^*(y) = \sum_{i=1}^n \exp(y_i - 1)$ where a_i is i-th column vector of A

- minimum volume covering ellipsoid

$$\begin{array}{ll} \text{minimize} & -\log\det X \\ \text{subject to} & a_i^T X a_i \leq 1 \quad i = 1, \ldots, m \end{array}$$

  where domain of objective function is $\mathbf{S}^n_{++}$
  - Lagrangian - $L(X, \lambda) = -\log\det X + \sum_{i=1}^m \lambda_i (a_i^T X a_i - 1)$

- Lagrange dual function

$$g(\lambda) = \begin{cases} \log\det(\sum_{i=1}^m \lambda_i a_i a_i^T) - \mathbf{1}^T\lambda + n & \sum_{i=1}^m \lambda_i a_i a_i^T \succ 0 \\ -\infty & \text{otherwise} \end{cases}$$

obtained using $f^*(Y) = -\log\det(-Y) - n$

Best lower bound

- for every (λ, ν) with $\lambda \succeq 0$, Lagrange dual function $g(\lambda, \nu)$ (in Definition 55) provides lower bound for optimal value p^* for optimization problem in Definition 38

- natural question to ask is
  - how good is the lower bound?
  - what is best lower bound we can achieve?

- these questions lead to definition of *Lagrange dual problem*

Lagrange dual problems

Definition 56. [Lagrange dual problems] *for optimization problem in Definition 38, optimization problem*

$$\begin{array}{ll} \textit{maximize} & g(\lambda, \nu) \\ \textit{subject to} & \lambda \succeq 0 \end{array}$$

called Lagrange dual problem associated with problem in Definition 38

- *original problem in Definition 38, (somestime) called primal problem*
- *domain is* $\mathbf{R}^m \times \mathbf{R}^p$
- *dual feasibility defined in page 81, i.e.,* (λ, ν) *satisfying* $\lambda \succeq 0 \quad g(\lambda, \nu) > -\infty$ *indeed means feasibility for Lagrange dual problem*
- $d^* = \sup\{g(\lambda, \nu) | \lambda \in \mathbf{R}^m,\ \nu \in \mathbf{R}^p,\ \lambda \succeq 0\}$*, called dual optimal value*
- $(\lambda^*, \nu^*) = \operatorname{argsup}\{g(\lambda, \nu) | \lambda \in \mathbf{R}^m,\ \nu \in \mathbf{R}^p,\ \lambda \succeq 0\}$*, said to be dual optimal or called optimal Lagrange multipliers (if exists)*

- Lagrange dual problem in Definition 56 is convex optimization (even though original problem is not) since $g(\lambda, \nu)$ is always convex

Making dual constraints explicit dual problems

- (out specific) way we define Lagrange dual function in Definition 55 as function g of $\mathbf{R}^m \times \mathbf{R}^p$ into $\mathbf{R} \cup \{-\infty\}$, *i.e.*, $\mathbf{dom}\, g = \mathbf{R}^n \times \mathbf{R}^p$

- however, in many cases, feasible set $\{(\lambda, \nu) | \lambda \succeq 0 \quad g(\lambda, \nu) > -\infty\}$ is proper subset of $\mathbf{R}^n \times \mathbf{R}^p$

- can make this implicit feasibility condition explicit by adding it as constraint (as shown in following examples)

Lagrange dual problems associated with LPs

- standard form LP
  - primal problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \succeq 0 \end{array}$$

  - Lagrange dual problem

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) = \begin{cases} -b^T \nu & A^T \nu - \lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases} \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

(refer to page 83 for Lagrange dual function)

- can make dual feasibility explicit by adding it to constraints as mentioned on page 90

$$\begin{array}{ll} \text{maximize} & -b^T \nu \\ \text{subject to} & \lambda \succeq 0 \\ & A^T \nu - \lambda + c = 0 \end{array}$$

- can further simplify problem

$$\begin{array}{ll} \text{maximize} & -b^T\nu \\ \text{subject to} & A^T\nu + c \succeq 0 \end{array}$$

  - last problem is *inequality form LP*
  - all three problems are equivalent, but *not* same problems
  - will, however, with abuse of terminology, refer to all three problems as Lagrange dual problem

- inequality form LP
  - primal problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \end{array}$$

  - Lagrangian

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

  - Lagrange dual function

$$g(\lambda) = -b^T\lambda + \inf_{x \in \mathbf{R}^n} (c + A^T\lambda)^T x = \begin{cases} -b^T\lambda & A^T\lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases}$$

- Lagrange dual problem

$$\begin{array}{ll} \text{maximize} & g(\lambda) = \begin{cases} -b^T\lambda & A^T\lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases} \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

  - can make dual feasibility explicit by adding it to constraints as mentioned on page 90

$$\begin{array}{ll} \text{maximize} & -b^T\nu \\ \text{subject to} & A^T\lambda + c = 0 \\ & \lambda \succeq 0 \end{array}$$

- dual problem is *standard form LP*

- thus, dual of standard form LP is inequality form LP and vice versa

- also, for both cases, dual of dual is same as primal problem

Lagrange dual problem of equality constrained optimization problem

- equality constrained optimization problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \end{array}$$

- dual function

$$\begin{aligned} g(\nu) &= \inf_{x \in \mathbf{dom}\, f} (f(x) + \nu^T (Ax - b)) = -b^T \nu - \sup_{x \in \mathbf{dom}\, f} (-\nu^T A x - f(x)) \\ &= -b^T \nu - f^*(-A^T \nu) \end{aligned}$$

- dual problem

$$\text{maximize} \quad -b^T \nu - f^*(-A^T \nu)$$

Lagrange dual problem associated with equality constrained quadratic program

- strictly convex quadratic problem

$$\begin{array}{ll} \text{minimize} & f(x) = x^T P x + q^T x + r \\ \text{subject to} & Ax = b \end{array}$$

  - conjugate function of objective function

$$f^*(x) = (x-q)^T P^{-1}(x-q)/4 - r = x^T P^{-1} x/4 - q^T P^{-1} x/2 + q^T P^{-1} q/4 - r$$

  - dual problem

$$\text{maximize} \quad -\nu^T (AP^{-1}A^T)\nu/4 - (b + AP^{-1}q/2)^T \nu - q^T P^{-1} q/4 + r$$

Lagrange dual problems associated with nonconvex quadratic problems

- primal problem

$$\begin{array}{ll} \text{minimize} & x^T A x + 2b^T x \\ \text{subject to} & x^T x \leq 1 \end{array}$$

  where $A \in \mathbf{S}^n$, $A \notin \mathbf{S}^n_+$, and $b \in \mathbf{R}^n$
  - since $A \not\succeq 0$, not convex optimization problem
  - sometimes called *trust region problem* arising minimizing second-order approximation of function over bounded region

- Lagrange dual function

$$g(\lambda) = \begin{cases} -b^T(A + \lambda I)^\dagger b - \lambda & A + \lambda I \succeq 0,\ b \in \mathcal{R}(A + \lambda I) \\ -\infty & \text{otherwise} \end{cases}$$

  where $(A + \lambda I)^\dagger$ is pseudo-inverse of $A + \lambda I$

- Lagrange dual problem

$$\begin{array}{ll} \text{maximize} & -b^T(A+\lambda I)^\dagger b - \lambda \\ \text{subject to} & A+\lambda I \succeq 0,\ b \in \mathcal{R}(A+\lambda I) \end{array}$$

  where optimization variable is $\lambda \in \mathbf{R}$
  - note we do not need constraint $\lambda \geq 0$ since it is implied by $A + \lambda I \succeq 0$
  - though not obvious from what it appears to be, it is (of course) convex optimization problem (by definition of Lagrange dual function, $i.e.$, Definition 55)
  - can be expressed ar

$$\begin{array}{ll} \text{maximize} & -\sum_{i=1}^n (q_i^T b)^2/(\lambda_i+\lambda) - \lambda \\ \text{subject to} & \lambda \geq -\lambda_{\min}(A) \end{array}$$

    where λ_i and q_i are eigenvalues and corresponding orthogormal eigenvectors of A, when $\lambda_i + \lambda = 0$ for some i, we interpret $(q_i^T b)^2/0$ as 0 if $q_i^T 0$ and ∞ otherwise

Weak duality

- since $g(\lambda, \nu) \leq p^*$ for every $\lambda \succeq 0$, we have

$$d^* = \sup\{g(\lambda, \nu) | \lambda \in \mathbf{R}^m,\ \nu \in \mathbf{R}^p,\ \lambda \succeq 0\} \leq p^*$$

Definition 57. [weak duality] *property that that optimal value of optimization problem (in Definition 38) is always no less than optimal value of Lagrange daul problem (in Definition 56)*

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

called weak duality

- *d^* is best lower bound for primal problem that can be obtained from Lagrange dual function (by definition)*
- *weak duality holds even when d^* or/and p^* are not finite, $e.g.$*
  - if primal problem is unbounded below *so that $p^* = -\infty$, must have $d^* = -\infty$,* $i.e.$, dual problem is infeasible
  - *conversely,* if dual problem is unbounded above *so that $d^* = \infty$, must have $p^* = \infty$,* $i.e.$, primal problem is infeasible

Optimal duality gap

Definition 58. [optimal duality gap] *difference between optimal value of optimization problem (in Definition 38) and optimal value of Lagrange daul problem (in Definition 56),* $i.e.$

$$p^* - d^*$$

called optimal duality gap

- sometimes used for lower bound of optimal value of problem which is difficult to solve
  - for example, dual problem of max-cut problem (on page 83), which is NP-hard, is

$$\begin{array}{ll} \text{minimize} & -\mathbf{1}^T\nu \\ \text{subject to} & W + \mathbf{diag}(\nu) \succeq 0 \end{array}$$

    where optimization variable is $\nu \in \mathbf{R}^n$
    - the dual problem can be solved very efficiently using polynomial time algorithms while primal problme *cannot* be solved unless n is very small

Strong duality

Definition 59. [strong duality] *if optimal value of optimization problem (in Definition 38) equals to optimal value of Lagrange daul problem (in Definition 56),* $i.e.$

$$p^* = d^*$$

strong duality *said to hold*

- strong duality does *not* hold in general
  - if it held always, max-cut problem, which is NP-hard, can be solved in polynomial time, which would be one of biggest breakthrough in field of theoretical computer science
  - may mean some of strongest cryptography methods, $e.g.$, homeomorphic cryptography, can be broken

Slater's theorem

- exist many conditions which guarantee strong duality, which are called *constraint qualifications* - one of them is Slater's condition

Theorem 8. [Slater's theorem] *if optimization problem is convex (Definition 42), and exists feasible $x \in \mathcal{D}$ contained in $\mathbf{relint}\,\mathcal{D}$ such that*

$$q(x) \prec 0 \quad h(x) = 0$$

strong duality *holds (and* dual optimum is attained when $d^* > -\infty$*)*

- *such condition, called* Slater's condition
- *such point, (sometimes) said to be* strictly feasible

when there are affine inequality constraints, can refine Slater's condition - if first k inequality constraint functions $q_1, \ldots, q_k$ are affine, Slater's condition can be relaxed to

$$q_i(x) \leq 0 \;\; i = 1, \ldots, k \quad q_i(x) < 0 \;\; i = k+1, \ldots, m \quad h(x) = 0$$

Strong duality for LS solution of linear equations

- primal problem

$$\begin{array}{ll} \text{minimize} & x^T x \\ \text{subject to} & Ax = b \end{array}$$

- dual problem

$$\text{maximize} \quad g(\nu) = -\tfrac{1}{4}\nu^T A A^T \nu - b^T \nu$$

(refer to page 82 for Lagrange dual function)

- "dual is always feasible" and "primal is feasible $\Rightarrow$ Slater's condition holds", thus Slater's theorem (Theorem 8) implies, exist only three cases
  - $(d^* = p^* \in \mathbf{R})$ or $(d^* \in \mathbf{R}\ \&\ p^* = \infty)$ or $(d^* = p^* = \infty)$
- if primal is infeasible, though, $b \notin \mathcal{R}(A)$, thus exists z, such that $A^T z = 0$ and $b^T z \neq 0$, then line $\{tz | t \in \mathbf{R}\}$ makes dual problem unbounded above, hence $d^* = \infty$
- hence, *strong duality always holds*, *i.e.*, $(d^* = p^* \in \mathbf{R})$ or $(d^* = p^* = \infty)$

Strong duality for LP

- every LP either is infeasible or satisfies Slater's condition

- dual of LP is LP, hence, Slater's theorem (Theorem 8) implies
  - if primal is feaisble, either $(d^* = p^* = -\infty)$ or $(d^* = p^* \in \mathbf{R})$
  - if dual is feaisble, either $(d^* = p^* = \infty)$ or $(d^* = p^* \in \mathbf{R})$
  - only other case left is $(d^* = -\infty \ \& \ p^* = \infty)$
    - indeed, this pathological case can happen

Strong duality for entropy maximization

- primal problem

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^n x_i \log x_i \\ \text{subject to} & Ax \preceq b \\ & \mathbf{1}^T x = 1 \end{array}$$

- dual problem (refer to page 86 for Lagrange dual function)

$$\begin{array}{ll} \text{maximize} & -b^T\lambda - \nu - \exp(-\nu - 1)\sum_{i=1}^n \exp(a_i^T\lambda) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

- dual problem is feasible, hence, Slater's theorem (Theorem 8) implies, if exists $x \succ 0$ with $Ax \preceq b$ and $\mathbf{1}^T x = 1$, strong duality holds, and indeed $d^* = p^* \in \mathbf{R}$
- by the way, can simplify dual problem by maximizing dual objective function over ν

$$\begin{array}{ll} \text{maximize} & -b^T\lambda - \log\left(\sum_{i=1}^n \exp(a_i^T\lambda)\right) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

  which is geometry program in convex form (Proposition 9) with nonnegativity contraint

Strong duality for minimum volume covering ellipsoid

- primal problem

$$\begin{array}{ll} \text{minimize} & -\log\det X \\ \text{subject to} & a_i^T X a_i \leq 1 \quad i = 1, \ldots, m \end{array}$$

  where $\mathcal{D} = \mathbf{S}_{++}^n$

- dual problem

$$\begin{array}{ll} \text{maximize} & \begin{cases} \log\det(\sum_{i=1}^m \lambda_i a_i a_i^T) - \mathbf{1}^T\lambda + n & \sum_{i=1}^m \lambda_i a_i a_i^T \succ 0 \\ -\infty & \text{otherwise} \end{cases} \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

  (refer to page 87 for Lagrange dual function)

- $X = \alpha I$ with large enough $\alpha > 0$ satisfies primal's constraints, hence Slater's condition *always* holds, thus, *strong duality always holds*, $i.e.$, $(d^* = p^* \in \mathbf{R})$ or $(d^* = p^* = -\infty)$
- in fact, $\mathcal{R}(a_1, \ldots, a_m) = \mathbf{R}^n$ *if and only if* $d^* = p^* \in \mathbf{R}^n$

Strong duality for trust region nonconvex quadratic problems

- one of rare occasions in which *strong duality obtains for nonconvex problems*
- primal problem

$$\begin{array}{ll} \text{minimize} & x^T A x + 2b^T x \\ \text{subject to} & x^T x \leq 1 \end{array}$$

  where $A \in \mathbf{S}^n$, $A \notin \mathbf{S}_+^n$, and $b \in \mathbf{R}^n$
- Lagrange dual problem (page 97)

$$\begin{array}{ll} \text{maximize} & -b^T(A + \lambda I)^\dagger b - \lambda \\ \text{subject to} & A + \lambda I \succeq 0,\ b \in \mathcal{R}(A + \lambda I) \end{array}$$

- *strong duality always holds* and $d^* = p^* \in \mathbf{R}$ (since dual problem is feasible - large enough λ satisfies dual constraints)
- in fact, exists stronger result - *strong dual holds* for optimization problem with quadratic objective and *one* quadratic inequality constraint, provided Slater's condition holds

Matrix games using mixed strategies

- matrix game - consider game with two players A and B
  - player A makes choice $1 \leq a \leq n$, player B makes choice $1 \leq b \leq m$, then player A makes payment of P_{ab} to player B
  - matrix $P \in \mathbf{R}^{n \times m}$, called *payoff matrix*
  - player A tries to pay as little as possible & player B tries to received as much as possible
  - players use *randomized or mixed strategies*, $i.e.$, each player makes choice randomly and independently of other player's choice according to probability distributions

$$\mathbf{Prob}(a = i) = u_i \; i = 1 \leq i \leq n \quad \mathbf{Prob}(b = j) = v_j \; i = 1 \leq j \leq m$$

- expected payoff (from player A to player B)

$$\sum_i \sum_j u_i v_j P_{ij} = u^T P v$$

- assume player A's strategy is known to play B
  - player B will choose v to maximize $u^T P v$

$$\sup\{u^T P v | v \succeq 0,\ \mathbf{1}^T v = 1\} = \max_{1 \le j \le m} (P^T u)_j$$

  - player A (assuming that player B will employ above strategy to maximize payment) will choose u to minimize payment

$$\begin{array}{ll} \text{minimize} & \max_{1 \le j \le m} (P^T u)_j \\ \text{subject to} & u \succeq 0 \quad \mathbf{1}^T u = 1 \end{array}$$

- assume player B's strategy is known to play A
  - then player B will do same to maximize payment (assuming that player A will employ such strategy to minimize payment)

$$\begin{array}{ll} \text{maximize} & \min_{1 \le i \le n} (P v)_i \\ \text{subject to} & v \succeq 0 \quad \mathbf{1}^T v = 1 \end{array}$$

Strong duality for matrix games using mixed strategies

- in matrix game, can guess in frist came, player B has advantage over player A because A's strategy's exposed to B, and vice versa, hence optimal value of first problem is greater than that of second problem
- surprising, no one has advantage over the other, $i.e.$, optimal values of two problems are *same* - will show this
- first observe both problems are (convex) piecewise-linear optimization problems
- formulate first problem as LP

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & u \succeq 0 \quad \mathbf{1}^T u = 1 \quad P^T u \preceq t\mathbf{1} \end{array}$$

  - Lagrangian

$$L(u, t, \lambda_1, \lambda_2, \nu) = \nu + (1 - \mathbf{1}^T \lambda_1)t + (P\lambda_1 - \nu\mathbf{1} - \lambda_2)^T u$$

– Lagrange dual function

$$g(\lambda_1, \lambda_2, \nu) = \begin{cases} \nu & \mathbf{1}^T\lambda_1 = 1 \;\&\; P\lambda_1 - \nu\mathbf{1} = \lambda_2 \\ -\infty & \text{otherwise} \end{cases}$$

- Lagrange dual problem

$$\begin{array}{ll} \text{maximize} & \nu \\ \text{subject to} & \mathbf{1}^T\lambda_1 = 1 \quad P\lambda_1 - \nu\mathbf{1} = \lambda_2 \\ & \lambda_1 \succeq 0 \quad \lambda_2 \succeq 0 \end{array}$$

- eliminating λ_2 gives below Lagrange dual problem

$$\begin{array}{ll} \text{maximize} & \nu \\ \text{subject to} & \lambda_1 \succeq 0 \quad \mathbf{1}^T\lambda_1 = 1 \quad P\lambda_1 \succeq \nu\mathbf{1} \end{array}$$

  which is equivalent to second problem in matrix game
- weak duality confirms "player who knows other player's strategy has advantage or on par"

- moreoever, primal problem satisfies Slater's condition, hence *strong duality always holds*, and dual is feasible, hence $d^* = p^* \in \mathbf{R}$, $i.e.$, regardless of who knows other player's strategy, no player has advantage

Geometric interpretation of duality

- assume (not necessarily convex) optimization problem in Definition 38
- define graph

$$G = \{(q(x), h(x), f(x))|x \in \mathcal{D}\} \subset \mathbf{R}^m \times \mathbf{R}^p \times \mathbf{R}$$

- for every $\lambda \succeq 0$ and ν

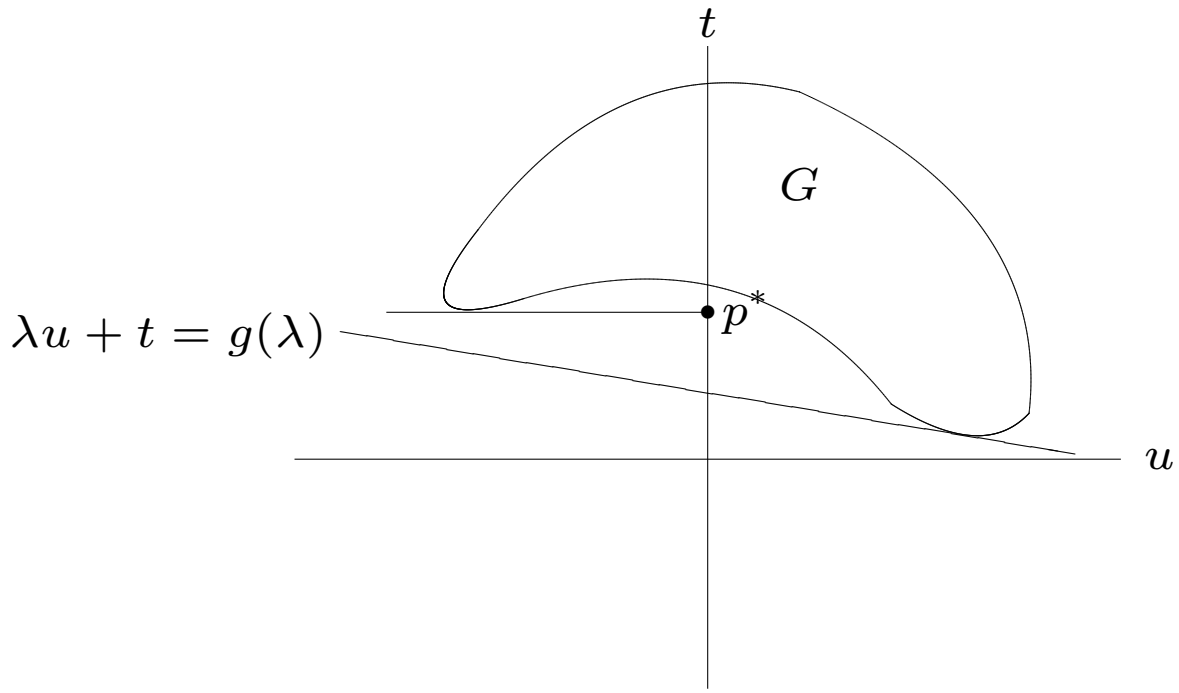
$$\begin{aligned} p^* &= \inf\{t|(u,v,t) \in G, u \preceq 0, v = 0\} \\ &\geq \inf\{t + \lambda^T u + \nu^T v|(u,v,t) \in G, u \preceq 0, v = 0\} \\ &\geq \inf\{t + \lambda^T u + \nu^T v|(u,v,t) \in G\} = g(\lambda, \nu) \end{aligned}$$

  where second inequality comes from $\{(u,v,t)|(u,v,t) \in G, u \preceq 0, v = 0\} \subset G$
- above establishes *weak duality using graph*
- last equality implies that

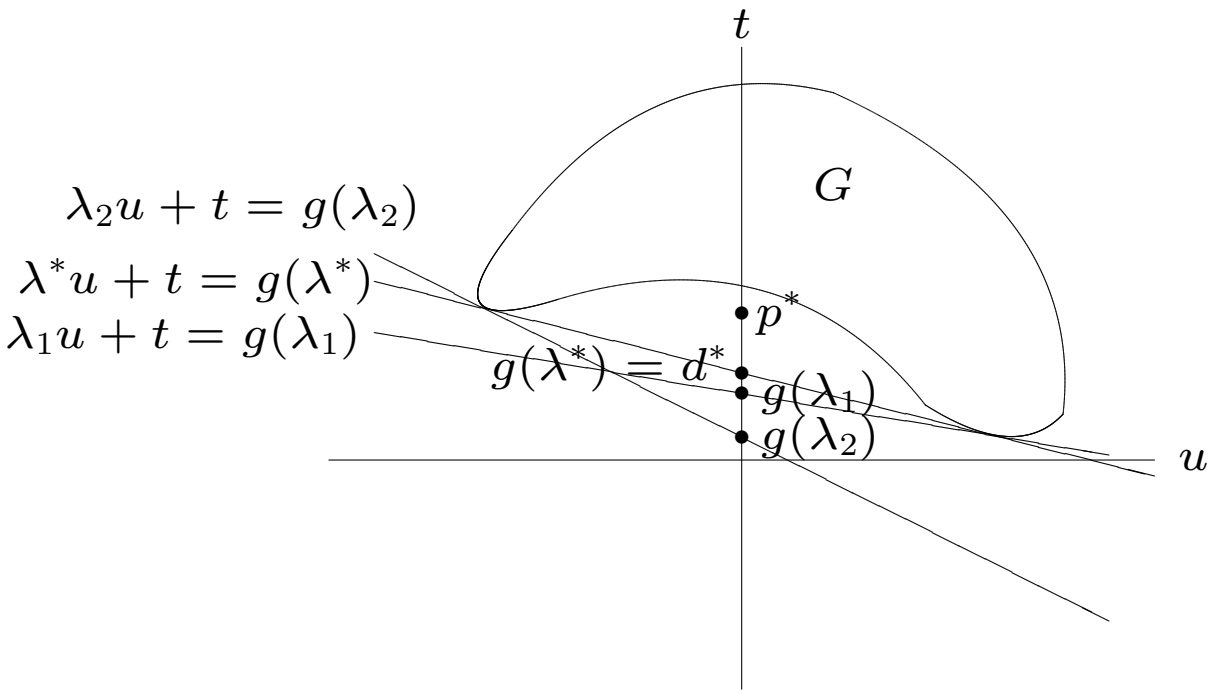
$$(\lambda, \nu, 1)^T (u, v, t) \geq g(\lambda, \nu)$$

hence if $g(\lambda, \nu) > -\infty$, $(\lambda, \nu, 1)$ and $g(\lambda, \nu)$ define *nonvertical supporting hyperplane* for G - nonvertical because third component is nonzero

- the figure shows G as area inside closed curve contained in $\mathbf{R}^m \times \mathbf{R}^p \times \mathbf{R}$ where $m = 1$ and $p = 0$ as primal optimal value p^* and supporting hyperplane $\lambda u + t = g(\lambda)$



- the figure shows three hyperplanes determined by three values for λ, one of which λ^* is optimal solution for dual problem



Epigraph interpretation of duality

- define extended graph over G - sort of epigraph of G

$$\begin{aligned} H &= G + \mathbf{R}_+^m \times \{0\} \times \mathbf{R}_+ \\ &= \{(u, v, t) | x \in \mathcal{D}, q(x) \preceq u, h(x) = v, f(x) \leq t\} \end{aligned}$$

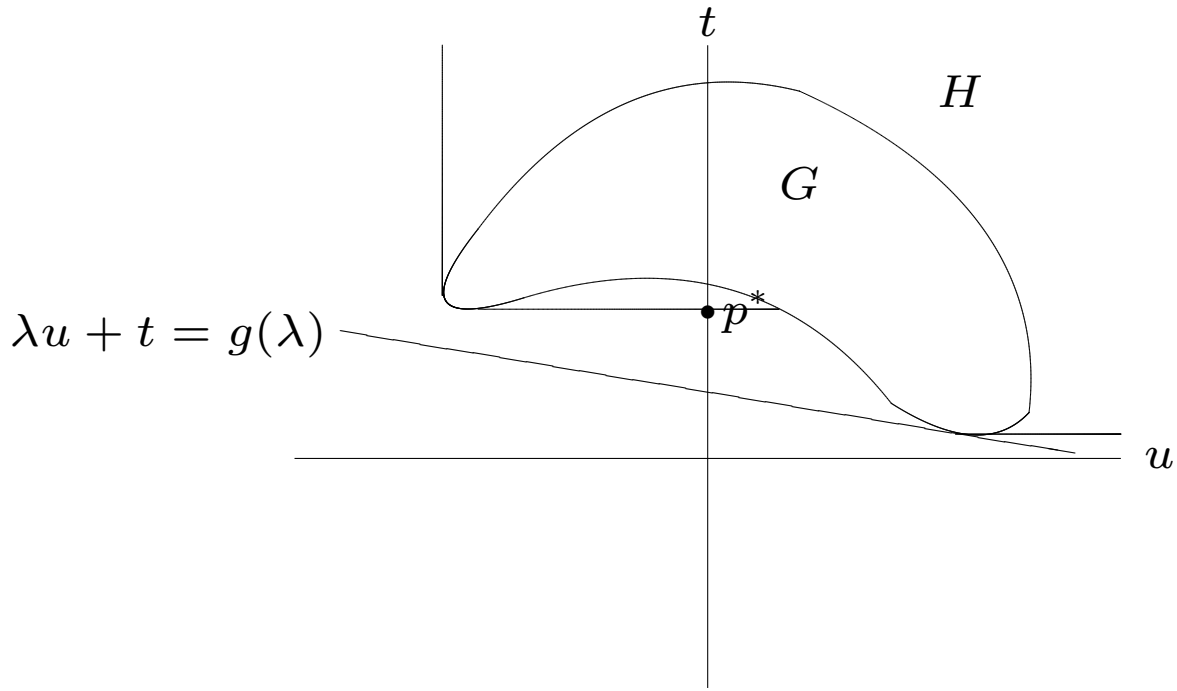
- if $\lambda \succeq 0$, $g(\lambda, \nu) = \inf\{(\lambda, \nu, 1)^T(u, v, t) | (u, v, t) \in H\}$, thus

$$(\lambda, \nu, 1)^T(u, v, t) \geq g(\lambda, \nu)$$

  defines nonvertical supporting hyperplane for H
- now $p^* = \inf\{t | (0, 0, t) \in H\}$, hence $(0, 0, p^*) \in \mathbf{bd}\, H$, hence

$$p^* = (\lambda, \nu, 1)^T(0, 0, p^*) \geq g(\lambda, \nu)$$

- once again establishes *weak duality*
- the figure shows epigraph interpretation


t
H
G
$\lambda u + t = g(\lambda)$
p^*
u

Proof of strong duality under constraint qualification

- now we show proof of strong duality - this is one of rare cases where proof is shown in main slides instead of "selected proofs" section like Galois theory since - (I hope) it will give you some good intuition about why strong duality holds for (most) convex optimization problems
- assume Slater's condition holds, *i.e.*, f and q are convex, h is affine, and exists $x \in \mathcal{D}$ such that $q(x) \prec 0$ and $h(x) = 0$
- further assume $\mathcal{D}$ has interior (hence, $\mathbf{relint}\, \mathcal{D} = \mathcal{D}^\circ$ and $\mathbf{rank}\, A = p$
- assume $p^* \in \mathbf{R}$ - since exists feasible x, the other possibility is $p^* = -\infty$, but then, $d^* = -\infty$, hence strong duality holds
- H is convex (proof can be found in Proof 6)
- now define

$$B = \{(0, 0, s) \in \mathbf{R}^m \times \mathbf{R}^p \times \mathbf{R} | s < p^*\}$$

- then $B \cap H = \emptyset$, hence Theorem 1 implies exists separable hyperplane with $(\tilde{\lambda}, \tilde{\nu}, \mu) \neq$

0 and α such that

$$\begin{aligned}(u,v,t) \in H \quad &\Rightarrow \quad \tilde{\lambda}^T u + \tilde{\nu}^T v + \mu t \geq \alpha \\ (u,v,t) \in B \quad &\Rightarrow \quad \tilde{\lambda}^T u + \tilde{\nu}^T v + \mu t \leq \alpha\end{aligned}$$

- then $\tilde{\lambda} \succeq 0$ & $\mu \geq 0$ - assume $\mu > 0$
  - can prove when $\mu = 0$, but kind of tedius, plus, whole purpose is provide good intuition, so will not do it here
- above second inequality implies $\mu p^* \leq \alpha$ and for some $x \in \mathcal{D}$

$$\mu L(x, \tilde{\lambda}/\mu, \tilde{\nu}/\mu) = \tilde{\lambda}^T q(x) + \tilde{\nu}^T h(x) + \mu f(x) \geq \alpha \geq \mu p^*$$

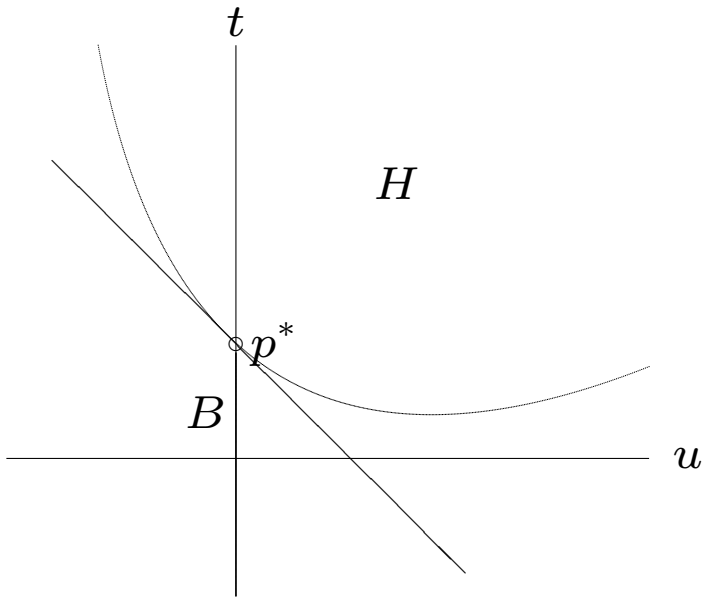
thus,

$$g(\tilde{\lambda}/\mu, \tilde{\nu}/\mu) \geq p^*$$

- finally, weak duality implies

$$g(\lambda, \nu) = p^*$$

where $\lambda = \tilde{\lambda}/\mu$ & $\nu = \tilde{\nu}/\mu$


t
H
p*
B
u

Max-min characterization of weak and strong dualities

- note

$$\begin{aligned}\sup_{\lambda\succeq 0,\nu} L(x,\lambda,\nu) &= \sup_{\lambda\succeq 0,\nu}\Big(f(x)+\lambda^T q(x)+\nu^T h(x)\Big)\\ &= \begin{cases} f(x) & x\in\mathcal{F}\\ \infty & \text{otherwise}\end{cases}\end{aligned}$$

- thus $p^*=\inf_{x\in\mathcal{D}}\sup_{\lambda\succeq 0,\nu} L(x,\lambda,\nu)$ whereas $d^*=\sup_{\lambda\succeq 0,\nu}\inf_{x\in\mathcal{D}} L(x,\lambda,\nu)$
- weak duality means

$$\sup_{\lambda\succeq 0,\nu}\inf_{x\in\mathcal{D}} L(x,\lambda,\nu)\le\inf_{x\in\mathcal{D}}\sup_{\lambda\succeq 0,\nu} L(x,\lambda,\nu)$$

- strong duality means

$$\sup_{\lambda\succeq 0,\nu}\inf_{x\in\mathcal{D}} L(x,\lambda,\nu)=\inf_{x\in\mathcal{D}}\sup_{\lambda\succeq 0,\nu} L(x,\lambda,\nu)$$

Max-min inequality

- indeed, inequality $\sup_{\lambda \succeq 0} \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \leq \inf_{x \in \mathcal{D}} \sup_{\lambda \succeq 0} L(x, \lambda, \nu)$ holds for general case

Inequality 2. [max-min inequality] *for* $f : X \times Y \to \mathbf{R}$

$$\sup_{y \in Y} \inf_{x \in X} f(x, y) \leq \inf_{x \in X} \sup_{y \in Y} f(x, y)$$

(proof can be found in Proof 5)

Definition 60. [strong max-min property] *if below equality holds, we say* f *(and* X *and* Y*) satisfies* strong max-min property *or* saddle point property

$$\sup_{y \in Y} \inf_{x \in X} f(x, y) = \inf_{x \in X} \sup_{y \in Y} f(x, y)$$

- this happens, *e.g.*, $X = \mathcal{D}$, $Y = \mathbf{R}^m_+ \times \mathbf{R}^p$, f is Lagrangian of optimization problem (in Definition 38) for which strong duality holds

Saddle-points

Definition 61. [saddle-points] *for $f : X \times Y \to \mathbf{R}$, pair $x^* \in X$ and $y^* \in Y$ such that*

$$(\forall x \in X, y \in Y)\left(f(x^*, y) \leq f(x^*, y^*) \leq f(x, y^*)\right)$$

called saddle-point for f (and X and Y)

- if assumption in Definition 61 holds, x^* minimizes $f(x, y^*)$ over X and y^* maximizes $f(x^*, y)$ over Y

$$\sup_{y \in Y} f(x^*, y) = f(x^*, y^*) = \inf_{x \in X} f(x, y^*)$$

  - strong max-min property (in Definition 60) holds with $f(x^*, y^*)$ as common value

Saddle-point interpretation of strong duality

- for primal optimum x^* and dual optimum (λ^*, ν^*)

$$g(\lambda^*, \nu^*) \leq L(x^*, \lambda^*, \nu^*) \leq f(x^*)$$

- if strong duality holds, for every $x \in \mathcal{D}$, $\lambda \succeq 0$, and ν

$$L(x^*, \lambda, \nu) \leq f(x^*) = L(x^*, \lambda^*, \nu^*) = g(\lambda^*, \nu^*) \leq L(x, \lambda^*, \nu^*)$$

  - thus x^* and (λ^*, ν^*) form saddle-point of Lagrangian

- conversely, if $\tilde{x}$ and $(\tilde{\lambda}, \tilde{\nu})$ are saddle-point of Lagrangian, *i.e.*, for every $x \in \mathcal{D}$, $\lambda \succeq 0$, and ν

$$L(\tilde{x}, \lambda, \nu) \leq L(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) \leq L(x, \tilde{\lambda}, \tilde{\nu})$$

  - hence $g(\tilde{\lambda}, \tilde{\nu}) = \inf_{x \in \mathcal{D}} L(x, \tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) = \sup_{\lambda \succeq 0, \nu} L(\tilde{x}, \lambda, \nu) = f(\tilde{x})$, thus $g(\lambda^*, \nu^*) \leq g(\tilde{\lambda}, \tilde{\nu})$ & $f(\tilde{x}) \leq f(x^*)$
  - thus $\tilde{x}$ and $(\tilde{\lambda}, \tilde{\nu})$ are primal and dual optimal

Game interpretation

- assume two players play zero-sum game with payment function $f : X \times Y \to \mathbf{R}$ where player A pays player B amount equal to $f(x, y)$ when player A chooses x and player B chooses y

- player A will try to minimize $f(x, y)$ and player B will try to maximize $f(x, y)$

- assume player A chooses first then player B chooses after learning opponent's choice
  - if player A chooses x, player B will choose $\operatorname{argsup}_{y \in Y} f(x, y)$
  - knowing that, player A will first choose $\operatorname{arginf}_{x \in X} \sup_{y \in Y} f(x, y)$
  - hence payment will be $\inf_{x \in X} \sup_{y \in Y} f(x, y)$

- if player B makes her choise first, opposite happens, *i.e.*, payment will be $\sup_{y \in Y} \inf_{x \in X} f(x, y)$

- max-min inequality of Ineq 2 says

$$\sup_{y \in Y} \inf_{x \in X} f(x, y) \leq \inf_{x \in X} \sup_{y \in Y} f(x, y)$$

  $i.e.$, whoever chooses later has advantage, which is similar or rather same as matrix games using mixed strategies on page 107

- saddle-point for f (and X and Y), (x^*, y^*), called *solution of game* - x^* is optimal choice for player A and x^* is optimal choice for player B

Game interpretation for weak and strong dualities

- assume payment function in zero-sum game on page 124 is Lagrangian of optimization problem in Definition 38
- assume that $X = X$ and $Y = \mathbf{R}_+^n \times \mathbf{R}^p$
- if player A chooses first, knowing that player B will choose $\operatorname{argsup}_{(\lambda,\nu)\in Y} L(x, \lambda, \nu)$, she will choose $x^* = \operatorname{arginf}_{x\in X} \sup_{(\lambda,\nu)\in Y} L(x, \lambda, \nu)$
- likewise, player B will choose $(\lambda^*, \nu^*) = \operatorname{argsup}_{(\lambda,\nu)\in Y} \inf_{x\in X} L(x, \lambda, \nu)$
- optimal dualtiy gap $p^* - d^*$ equals to advantage player who goes second has
- if strong dualtiy holds, (x^*, λ^*, ν^*) is solution of game, in which case no one has advantage

Certificate of suboptimality

- dual feasible point (λ, ν) degree of suboptimality of current solution
- assume x is feasible solution, then

$$f(x) - p^* \leq f(x) - g(\lambda, \nu)$$

  guarantees that $f(x)$ is no further than $\epsilon = f(x) - g(\lambda, \nu)$ from optimal point point x^* (even though we do not know optimal solution)

- for this reason, (λ, ν), called *certificate of suboptimality*
- x is ϵ-suboptimal for primal problem and (λ, ν) is ϵ-suboptimal for dual problem
- strong duality means we *could* find arbitrarily small certificate of suboptimality

Complementary slackness

- assume strong duality holds for optimization problem in Definition 38 and assume x^* is primal optimum and (λ^*, ν^*) is dual optimum, then

$$f(x^*) = L(x^*, \lambda^*, \nu^*) = f(x^*) + {\lambda^*}^T q(x^*) + {\nu^*}^T h(x^*)$$

- $h(x^*) = 0$ implies ${\lambda^*}^T q(x^*) = 0$
- then $\lambda^* \succeq 0$ and $q(x^*) \preceq 0$ imply

$$\lambda_i^* q_i(x^*) = 0 \quad i = 1, \ldots, m$$

Proposition 10. [complementary slackness] *when strong duality holds, for primal and dual optimal points x^* and (λ^*, ν)*

$$\lambda_i^* q_i(x^*) = 0 \quad i = 1, \ldots, m$$

this property, called complementary slackness

KKT optimality conditions

Definition 62. [KKT optimality conditions] *for optimization problem in Definition 38 where f, q, and h are all differentiable, below conditions for $x \in \mathcal{D}$ and $(\lambda, \nu) \in \mathbf{R}^m \times \mathbf{R}^p$*

$$
\begin{aligned}
q(x) &\preceq 0 && \text{- primal feasibility} \\
h(x) &= 0 && \text{- primal feasibility} \\
\lambda &\succeq 0 && \text{- dual feasibility} \\
\lambda^T q(x) &= 0 && \text{- complementary slackness} \\
\nabla_x L(x, \lambda, \nu) &= 0 && \text{- vanishing gradient of Lagrangian}
\end{aligned}
$$

called Karush-Kuhn-Tucker (KKT) optimality conditions

KKT necessary for optimality with strong duality

Theorem 9. [KKT necessary for optimality with strong duality] *for optimization problem in Definition 38 where f, q, and h are all differentiable, if strong duality holds, primal and dual optimal solutions x^* and (λ^*, ν) satisfy KKT optimality conditions (in Definition 62), $i.e.$, for every optimization problem*

- *when strong duality holds, KKT optimality conditions are necessary for primal and dual optimality*

  or equivalently

- *primal and dual optimality with strong duality imply KKT optimality conditions*

KKT and convexity sufficient for optimality with strong duality

- assume convex optimization problem where f, q, and h are all differentiable and $x \in \mathcal{D}$ and $(\lambda, \nu) \in \mathbf{R}^m \times \mathbf{R}^p$ satisfying KKT conditions, $i.e.$

$$q(x) \preceq 0,\ h(x) = 0,\ \lambda \succeq 0,\ \lambda^T q(x) = 0,\ \nabla_x L(x, \lambda, \nu) = 0$$

- since $L(x, \lambda, \nu)$ is convex for $\lambda \succeq 0$, $i.e.$, each of $f(x)$, $\lambda^T q(x)$, and $\nu^T h(x)$ is convex, vanishing gradient implies x achieves infimum for Lagrangian, hence

$$g(\lambda, \nu) = L(x, \lambda, \nu) = f(x) + \lambda^T q(x) + \nu^T h(x) = f(x)$$

- thus, strong duality holds, $i.e.$, x and (λ, ν) are primal and dual optimal solutions with zero duality gap

Theorem 10. [KKT and convexity sufficient for optimality with strong duality] *for convex optimization problem in Definition 42 where f, q, and h are all differentiable, if $x \in \mathcal{D}$ and $(\lambda, \nu) \in \mathbf{R}^m \times \mathbf{R}^p$ satisfy KKT optimality conditions (in Definition 62), they are primal and dual optimal solutions having zero duality gap $i.e.$*

- *for convex optimization problem, KKT optimality conditions are sufficient for primal and dual optimality with strong duality*

  or equivalently

- *KKT optimality conditions and convexity imply primal and dual optimality and strong duality*

- Theorem 9 together with Theorem 10 implies that for convex optimization problem
  - *KKT optimality conditions are necessary and sufficient for primal and dual optimality with strong duality*

Solving primal problems via dual problems

- when strong duality holds, can retrieve primal optimum from dual optimum since primal optimal solution is minimize of

$$L(x, \lambda^*, \nu^*)$$

  where (λ^*, ν^*) is dual optimum
- example - entropy maximization ($\mathcal{D} = \mathbf{R}^n_{++}$)
  - primal problem - min. $f(x) = \sum_{i=1}^n x_i \log x_i$ s.t. $Ax \preceq b$, $\sum x = 1$
  - dual problem - max. $-b^T\lambda - \nu - \exp(-\nu - 1)\sum \exp(A^T\lambda)$ s.t. $\lambda \succeq 0$
  - provided dual optimum (λ^*, ν^*), primal optimum is

$$x^* = \underset{x \in \mathcal{D}}{\operatorname{argmin}} \left(\sum x_i \log x_i + {\lambda^*}^T (Ax - b) + \nu^* (\mathbf{1}^T x - 1) \right)$$

  - $\nabla_x L(x, \lambda^*, \nu^*) = \log x + A^T\lambda^* + (1 + \nu^*)\mathbf{1}$, hence

$$x^* = \exp(-(A^T\lambda^* + (1 + \nu^*)\mathbf{1}))$$

Perturbed optimization problems

- original problem in Definition 38 with perturbed constraints

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & q(x) \preceq u \\ & h(x) = v \end{array}$$

  where $u \in \mathbf{R}^m$ and $v \in \mathbf{R}^p$
- define $p^*(u, v)$ as optimal value of above *perturbed* problem, $i.e.$

$$p^*(u, v) = \inf\{f(x) | x \in \mathcal{D}, q(x) \preceq u, h(x) = v\}$$

  which is convex when problem is convex optimization problem (proof can be found in Proof 6) - note $p^*(0, 0) = p^*$
- assume and dual optimum (λ^*, ν^*), if strong duality holds, for every feasible x for perturbed problem

$$p^*(0, 0) = g(\lambda^*, \nu^*) \leq f(x) + {\lambda^*}^T q(x) + {\nu^*}^T h(x) \leq f(x) + {\lambda^*}^T u + {\nu^*}^T v$$

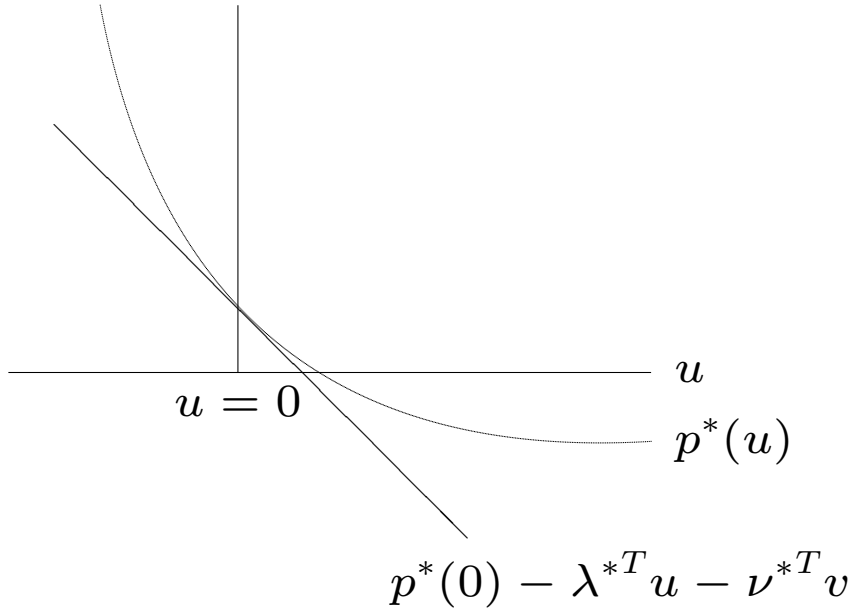
thus

$$p^*(0,0) \leq p^*(u,v) + \lambda^{*T} u + \nu^{*T} v$$

hence

$$p^*(u,v) \geq p^*(0,0) - \lambda^{*T} u - \nu^{*T} v$$

- the figure shows this for optimization problem with one inequality constraint and no equality constraint



Global sensitivity analysis via perturbed problems

- recall

$$p^*(u, v) \geq p^*(0, 0) - {\lambda^*}^T u - {\nu^*}^T v$$

- interpretations
  - if λ_i^* is large, when i-th inequality constraint is tightened, optimal value increases a lot
  - if λ_i^* is small, when i-th inequality constraint is relaxed, optimal value decreases not a lot
  - if $|\nu_i^*|$ is large, reducing v_i when $\nu_i^* > 0$ or increasing v_i when $\nu_i^* < 0$ increases optimval value a lot
  - if $|\nu_i^*|$ is small, increasing v_i when $\nu_i^* > 0$ or decreasing v_i when $\nu_i^* < 0$ decreases optimval value not a lot

- it only gives lower bounds - will explore local behavior

Local sensitivity analysis via perturbed problems

- assume $p^*(u, v)$ is differentiable with respect to u and v, *i.e.*, $\nabla_{(u,v)} p^*(u, v)$ exist
  - then

$$\frac{\partial}{\partial u_i} p^*(0,0) = \lim_{h \to 0^+} \frac{p^*(he_i, 0) - p^*(0,0)}{h} \geq \lim_{h \to 0^+} \frac{-\lambda^{*T}(he_i)}{h} = -\lambda_i$$

    and

$$\frac{\partial}{\partial u_i} p^*(0,0) = \lim_{h \to 0^-} \frac{p^*(he_i, 0) - p^*(0,0)}{h} \leq \lim_{h \to 0^-} \frac{-\lambda^{*T}(he_i)}{h} = -\lambda_i$$

  - obtain same result for v_i, hence

$$\nabla_u \, p^*(0,0) = -\lambda \quad \nabla_v \, p^*(0,0) = -\nu$$

- so larger λ_i or $|\nu_i|$ means larger change in optimal value of perturbed problem when u_i or v_i change a bit and vice versa quantitatively, - λ_i an ν_i provide exact ratio and direction

Different dual problems for equivalent optimization problems - 1

- introducing new variables and equality constraints for unconstrained problems
  - unconstrained optimization problem

$$\text{minimize} \quad f(Ax+b)$$

    - dual Lagrange function is $g = p^*$, hence strong duality holds, which, however, does not provide useful information
  - reformulate as equivalent optimization problem

$$\begin{array}{ll} \text{minimize} & f(y) \\ \text{subject to} & Ax+b=y \end{array}$$

    - Lagrangian - $L(x,y,\nu) = f(y) + \nu^T(Ax+b-y)$
    - Lagrange dual function - $g(\nu) = -I(A^T\nu = 0) + b^T\nu - f^*(\nu)$
    - dual optimization problem

$$\begin{array}{ll} \text{maximize} & b^T\nu - f^*(\nu) \\ \text{subject to} & A^T\nu = 0 \end{array}$$

- examples
  - unconstrained geometric problem

$$\text{minimize} \quad \log\left(\sum_{i=1}^{m} \exp(a_i^T x + b_i)\right)$$

    - reformulation

$$\begin{array}{ll} \text{minimize} & \log\left(\sum_{i=1}^{m} \exp(y_i)\right) \\ \text{subject to} & Ax + b = y \end{array}$$

    - dual optimization problem

$$\begin{array}{ll} \text{maximize} & b^T\nu - \sum_{i=1}^{m} \nu_i \log \nu_i \\ \text{subject to} & \mathbf{1}^T\nu = 1 \\ & A^T\nu = 0 \\ & \nu \succeq 0 \end{array}$$

      which is entropy maximization problem
  - norm minimization problem

$$\text{minimize} \quad \|Ax - b\|$$

- reformulation

$$\begin{array}{ll} \text{minimize} & \|y\| \\ \text{subject to} & Ax - b = y \end{array}$$

- dual optimization problem

$$\begin{array}{ll} \text{maximize} & b^T \nu \\ \text{subject to} & \|\nu\|_* \leq 1 \\ & A^T \nu = 0 \end{array}$$

Different dual problems for equivalent optimization problems - 2

- introducing new variables and equality constraints for constrained problems
  - inequality constrained optimization problem

$$\begin{array}{ll} \text{minimize} & f_0(A_0x + b_0) \\ \text{subject to} & f_i(A_ix + b_i) \leq 0 \quad i = 1, \ldots, m \end{array}$$

  - reformulation

$$\begin{array}{ll} \text{minimize} & f_0(y_0) \\ \text{subject to} & f_i(y_i) \leq 0 \quad i = 1, \ldots, m \\ & A_ix + b_i = y_i \quad i = 0, \ldots, m \end{array}$$

  - dual optimization problem

$$\begin{array}{ll} \text{maximize} & \sum_{i=0}^m \nu_i^T b_i - f_0^*(\nu_0) - \sum_{i=1}^m \lambda_i f_i^*(\nu_i/\lambda_i) \\ \text{subject to} & \sum_{i=0}^m A_i^T \nu_i = 0 \\ & \lambda \succeq 0 \end{array}$$

- examples
  - inequality constrained geometric program

$$\begin{array}{ll} \text{minimize} & \log\left(\sum \exp(A_0 x + b_0)\right) \\ \text{subject to} & \log\left(\sum \exp(A_i x + b_i)\right) \leq 0 \quad i = 1, \ldots, m \end{array}$$

where $A_i \in \mathbf{R}^{K_i \times n}$ and $\exp(z) := (\exp(z_1), \ldots, \exp(z_k))) \in \mathbf{R}^n$ and $\sum z := \sum_{i=1}^k z_i \in \mathbf{R}$ for $z \in \mathbf{R}^k$

- reformulation

$$\begin{array}{ll} \text{minimize} & \log\left(\sum \exp(y_0)\right) \\ \text{subject to} & \log\left(\sum \exp(y_i)\right) \leq 0 \quad i = 1, \ldots, m \\ & A_i x + b_i = y_i \quad i = 0, \ldots, m \end{array}$$

- dual optimization problem

$$\begin{array}{ll} \text{maximize} & \sum_{i=0}^m b_i^T \nu_i - \nu_0^T \log(\nu_0) - \sum_{i=1}^m \nu_i^T \log(\nu_i/\lambda_i) \\ \text{subject to} & \nu_i \succeq 0 \quad i = 0, \ldots, m \\ & \mathbf{1}^T \nu_0 = 1, \ \mathbf{1}^T \nu_i = \lambda_i \quad i = 1, \ldots, m \\ & \lambda_i \geq 0 \quad i = 1, \ldots, m \\ & \sum_{i=0}^m A_i^T \nu_i = 0 \end{array}$$

where and $\log(z) := (\log(z_1), \ldots, \log(z_k))) \in \mathbf{R}^n$ for $z \in \mathbf{R}^k_{++}$

- simplified dual optimization problem

$$\begin{array}{ll} \text{maximize} & \sum_{i=0}^m b_i^T \nu_i - \nu_0^T \log(\nu_0) - \sum_{i=1}^m \nu_i^T \log(\nu_i/\mathbf{1}^T \nu_i) \\ \text{subject to} & \nu_i \succeq 0 \quad i = 0, \ldots, m \\ & \mathbf{1}^T \nu_0 = 1 \\ & \sum_{i=0}^m A_i^T \nu_i = 0 \end{array}$$

Different dual problems for equivalent optimization problems - 3

- transforming objectives
  - norm minimization problem

$$\text{minimize} \quad \|Ax - b\|$$

  - reformulation

$$\begin{array}{ll} \text{minimize} & (1/2)\|y\|^2 \\ \text{subject to} & Ax - b = y \end{array}$$

  - dual optimization problem

$$\begin{array}{ll} \text{maximize} & -(1/2)\|\nu\|_*^2 + b^T\nu \\ \text{subject to} & A^T\nu = 0 \end{array}$$

Different dual problems for equivalent optimization problems - 4

- making contraints implicit
  - LP with box constraints

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b,\ l \preceq x \preceq u \end{array}$$

  - dual optimization problem

$$\begin{array}{ll} \text{maximize} & -b^T \nu - \lambda_1^T u + \lambda_2^T l \\ \text{subject to} & A^T \nu + \lambda_1 - \lambda_2 + c = 0,\ \lambda_1 \succeq 0,\ \lambda_2 \succeq 0 \end{array}$$

  - reformulation

$$\begin{array}{ll} \text{minimize} & c^T x + I(l \preceq x \preceq u) \\ \text{subject to} & Ax = b \end{array}$$

  - dual optimization problem for reformulated primal problem

$$\text{maximize} \quad -b^T \nu - u^T (A^T \nu + c)^- + l^T (A^T \nu + c)^+$$

Theorems of Alternatives

Weak alternatives

Theorem 11. [weak alternatives of two systems] *for $q : Q \to \mathbf{R}^m$ & $h : H \to \mathbf{R}^p$ where Q and H are subsets of common set X, which is subset of Banach space, assuming $\mathcal{D} = Q \cap H \neq \emptyset$, and $\lambda \in \mathbf{R}^m$ & $\nu \in \mathbf{R}^p$, below two systems of inequalities and equalities are weak alternatives, $i.e.$, at most one of them is feasible*

$$q(x) \preceq 0 \quad h(x) = 0$$

and

$$\lambda \succeq 0 \quad \inf_{x \in \mathcal{D}} \left(\lambda^T q(x) + \nu^T h(x) \right) > 0$$

- can prove Theorem 11 using duality of optimization problems
- consider primal and dual problems
  - primal problem

$$\begin{array}{ll} \text{minimize} & 0 \\ \text{subject to} & q(x) \preceq 0 \\ & h(x) = 0 \end{array}$$

– dual problem

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

where

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} \left(\lambda^T q(x) + \nu^T h(x) \right)$$

- then $p^*,\ d^* \in \{0, \infty\}$
- now assume *first system of Theorem 11 is feasible,* then $p^* = 0$, hence weak duality applies $d^* = 0$, thus there exist no λ and ν such that $\lambda \succeq 0$ and $g(\lambda, \nu) > 0$ $i.e.$, *second system is infeasible,* since otherwise there exist λ and ν making $g(\lambda, \nu)$ arbitrarily large; if $\tilde{\lambda} \succeq 0$ and $\tilde{\nu}$ satisfy $g(\lambda, \nu) > 0$, $g(\alpha\tilde{\lambda}, \alpha\tilde{\nu}) = \alpha g(\tilde{\lambda}, \tilde{\nu})$ goes to ∞ when $\alpha \to \infty$
- assume *second system is feasible,* then $g(\lambda, \nu)$ can be arbitrarily large for above reasons, thus $d^* = \infty$, hence weak duality implies $p^* = \infty$, which implies *first system is infeasible*
- therefore two systems are weak alternatives; at most one of them is feasible

  (actually, not hard to prove it without using weak duality)

Weak alternatives with strict inequalities

Theorem 12. [weak alternatives of two systems with strict inequalities] *for* $q : Q \to \mathbf{R}^m$ *&* $h : H \to \mathbf{R}^p$ *where* Q *and* H *are subsets of common set* X*, which is subset of Banach space, assuming* $\mathcal{D} = Q \cap H \neq \emptyset$*, and* $\lambda \in \mathbf{R}^m$ *&* $\nu \in \mathbf{R}^p$*, below two systems of inequalities and equalities are weak alternatives,* $i.e.$*, at most one of them is feasible*

$$q(x) \prec 0 \quad h(x) = 0$$

and

$$\lambda \succeq 0 \quad \lambda \neq 0 \quad \inf_{x \in \mathcal{D}} \left(\lambda^T q(x) + \nu^T h(x) \right) \geq 0$$

Strong alternatives

Theorem 13. [strong alternatives of two systems] *for convex $q : Q \to \mathbf{R}^m$ & affine $h : H \to \mathbf{R}^p$ where Q and H are subsets $\mathbf{R}^n$ assuming $\mathcal{D} = Q \cap H \neq \emptyset$ and $\lambda \in \mathbf{R}^m$ & $\nu \in \mathbf{R}^p$, if exists $x \in \mathbf{relint}\, \mathcal{D}$ with $h(x) = 0$, below two systems of inequalities and equalities are strong alternatives, $i.e.$, exactly one of them is feasible*

$$q(x) \preceq 0 \quad h(x) = 0$$

and

$$\lambda \succeq 0 \quad \inf_{x \in \mathcal{D}} \left(\lambda^T q(x) + \nu^T h(x) \right) > 0$$

Strong alternatives with strict inequalities

Theorem 14. [strong alternatives of two systems with strict inequalities] *for convex* $q : Q \to \mathbf{R}^m$ *&* *affine* $h : H \to \mathbf{R}^p$ *where* Q *and* H *are subsets* $\mathbf{R}^n$ *assuming* $\mathcal{D} = Q \cap H \neq \emptyset$ *and* $\lambda \in \mathbf{R}^m$ *&* $\nu \in \mathbf{R}^p$*, if exists* $x \in \mathbf{relint}\, \mathcal{D}$ *with* $h(x) = 0$*, below two systems of inequalities and equalities are strong alternatives,* $i.e.$*, exactly one of them is feasible*

$$q(x) \prec 0 \quad h(x) = 0$$

and

$$\lambda \succeq 0 \quad \lambda \neq 0 \quad \inf_{x \in \mathcal{D}} \left(\lambda^T q(x) + \nu^T h(x) \right) \geq 0$$

- proof - consider convex optimization problem and its dual
  - primal problem

$$\begin{array}{ll} \text{minimize} & s \\ \text{subject to} & q(x) - s\mathbf{1} \preceq 0 \\ & h(x) = 0 \end{array}$$

– dual problem

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \quad \mathbf{1}^T \lambda = 1 \end{array}$$

where $g(\lambda, \nu) = \inf_{x \in \mathcal{D}} \left(\lambda^T q(x) + \nu^T h(x) \right)$

- first observe Slater's condition holds for primal problem since by hypothesis of Theorem 14, exists $y \in \mathbf{relint}\, \mathcal{D}$ with $h(y) = 0$, hence $(y, q(y)) \in Q \times \mathbf{R}$ is primal feasible satisifying Slater's condition
- hence Slater's theorem (Theorem 8) implies $d^* = p^*$
- assume first system is feasible, then primal problem is strictly feasible and $d^* = p^* < 0$, hence second system infeasible since otherwise feasible point for second system is feasible point of dual problem, hence $d^* \geq 0$
- assume first system is infeasible, then $d^* = p^* \geq 0$, hence Slater's theorem (Theorem 8) implies exists dual optimal (λ^*, ν^*) (whether or not $d^* = \infty$), hence (λ^*, ν^*) is feasible point for second system of Theorem 14
- therefore two systems are strong alternatives; each is feasible *if and only if* the other is infeasible

Strong alternatives for linear inequalities

- dual function of feasibility problem for $Ax \preceq b$ is

$$g(\lambda) = \inf_{x \in \mathbf{R}^n} \lambda^T (Ax - b) = \begin{cases} -b^T\lambda & A^T\lambda = 0 \\ -\infty & \text{otherwise} \end{cases}$$

- hence alternative system is $\lambda \succeq 0,\ b^T\lambda < 0,\ A^T\lambda = 0$
- thus Theorem 13 implies below systems are strong alternatives

$$Ax \preceq b \quad \& \quad \lambda \succeq 0 \quad b^T\lambda < 0 \quad A^T\lambda = 0$$

- similarly alternative system is $\lambda \succeq 0,\ b^T\lambda < 0,\ A^T\lambda = 0$ and Theorem 13 implies below systems are strong alternatives

$$Ax \prec b \quad \& \quad \lambda \succeq 0 \quad \lambda \neq 0 \quad b^T\lambda \leq 0 \quad A^T\lambda = 0$$

Farkas' lemma

Theorem 15. [Farkas' lemma] *below systems of inequalities and equalities are strong alternatives*

$$Ax \preceq 0 \quad c^T x < 0 \quad \& \quad A^T y + c = 0 \quad y \succeq 0$$

- will prove Theorem 15 using LP and its dual
- consider LP $\big(\text{minimize } c^T x \quad \text{subject to } Ax \preceq 0\big)$
- dual function is $g(y) = \inf_{x \in \mathbf{R}^n} \left(c^T x + y^T A x\right) = \begin{cases} 0 & A^T y + c = 0 \\ -\infty & \text{otherwise} \end{cases}$
- hence dual problem is $\big(\text{maximize } 0 \quad \text{subject to } A^T y + c = 0,\ y \succeq 0\big)$
- assume first system is feasible, then homogeneity of primal problem implies $p^* = -\infty$, thus d^*, *i.e.*, dual is infeasible, hence second system is infeasible
- assume first system is infeasible, since primal is always feasible, $p^* = 0$, hence strong duality implies $d^* = 0$, thus second system is feasible

Convex Optimization with Generalized Inequalities

Optimization problems with generalized inequalities

Definition 63. [optimization problems with generalized inequalities] *for* $f : F \to \mathbf{R}$, $q : Q \to \bigtimes_{i=1}^{m} \mathbf{R}^{k_i}$, $h : H \to \mathbf{R}^p$ *where* F, Q, *and* H *are subsets of common set* X

$$\begin{array}{ll} \textit{minimize} & f(x) \\ \textit{subject to} & q(x) \preceq_{\mathcal{K}} 0 \\ & h(x) = 0 \end{array}$$

called optimization problem with generalized inequalities *where* $\mathcal{K} = \bigtimes K_i$ *is proper cone with* m *proper cones* $K_1 \subset \mathbf{R}^{k_1}, \ldots, K_n \subset \mathbf{R}^{k_m}$

- *every terminology and associated notation is same as of optimization problem in Definition 38 such as objective & inequality & equality contraint functions, domain of optimization problem* $\mathcal{D}$, *feasible set* $\mathcal{F}$, *optimal value* p^*
- *note that when* $K_i = \mathbf{R}_+$ *(hence* $\mathcal{K} = \mathbf{R}_+^m$*), above optimization problem coincides with that in Definition 38,* $i.e.$*, optimization problems with generalized inequalities subsume (normal) optimization problems*

Lagrangian for generalized inequalities

Definition 64. [Lagrangian for generalized inequalities] *for optimization problem in Definition 63 with nonempty domain $\mathcal{D}$, function $L : \mathcal{D} \times \bigtimes_{i=1}^{m} \mathbf{R}^{k_i} \times \mathbf{R}^p \to \mathbf{R}$ defined by*

$$L(x, \lambda, \nu) = f(x) + \lambda^T q(x) + \nu^T h(x)$$

called Lagrangian *associated with the optimization problem where*

- *every terminology and associated notation is same as of optimization problem in Definition 54 such as dual variables or Lagrange multipliers λ and ν.*
- *Lagrangian for generalized inequalities subsumes (normal) Lagrangian (Definition 54)*

Lagrange dual functions for generalized inequalities

Definition 65. [Lagrange dual functions for generalized inequalities] *for optimization problem in Definition 63 for which Lagrangian is defined, function* $g : \bigtimes \mathbf{R}^{k_i} \times \mathbf{R}^p \to \mathbf{R} \cup \{-\infty\}$ *defined by*

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} \left(f(x) + \lambda^T q(x) + \nu^T h(x) \right)$$

called Lagrange dual function *or just* dual function *associated with optimization problem*

- *Lagrange dual functions for generalized inequalities subsume (normal) Lagrange dual functions (Definition 55)*

- g is *concave function*
- $g(\lambda, \nu)$ is lower bound for optimal value of associated optimization problem $i.e.$,

$$g(\lambda, \nu) \leq p^*$$

  for every $\lambda \succeq_{\mathcal{K}}^* 0$ where $\mathcal{K}^*$ denotes dual cone of $\mathcal{K}$, $i.e.$, $\mathcal{K}^* = \bigtimes K_i^*$ where $K_i^* \subset \mathbf{R}^{k_i}$ is dual cone of $K_i \subset \mathbf{R}^{k_i}$
- (λ, ν) with $\lambda \succeq_{\mathcal{K}} 0$ and $g(\lambda, \nu) > -\infty$ said to be *dual feasible*

Lagrange dual problems for generalized inequalities

Definition 66. [Lagrange dual problems for generalized inequalities] *for optimization problem in Definition 63, optimization problem*

$$\begin{array}{ll} \textit{maximize} & g(\lambda, \nu) \\ \textit{subject to} & \lambda \succeq_{\mathcal{K}^*} 0 \end{array}$$

where $\mathcal{K}^*$ denotes dual cone of $\mathcal{K}$, i.e., $\mathcal{K}^* = \bigtimes K_i^*$ where $K_i^* \subset \mathbf{R}^{k_i}$ is dual cone of $K_i \subset \mathbf{R}^{k_i}$, called Lagrange dual problem associated with problem in Definition 63

- *every terminology and related notation is same as that in Definition 56 such as dual feasibility, dual optimal value d^*, optimal Lagrange multipliers (λ^*, ν^*)*
- *Lagrange dual problems for generalized inequalities subsume (normal) Lagrange dual problems (Definition 56)*

- Lagrange dual problem in Definition 66 is convex optimization since $g(\lambda, \nu)$ is convex

Slater's theorem for generalized inequalities

Theorem 16. [Slater's theorem for generalized inequalities] *if optimization problem in Definition 63 is convex, i.e., f is convex, q is $\mathcal{K}$-convex (i.e., every q_i is K_i-convex) (Definition 36), and exists feasible $x \in \mathcal{D}$ contained in $\mathbf{relint}\,\mathcal{D}$ such that*

$$q(x) \prec_{\mathcal{K}} 0 \quad h(x) = 0$$

strong duality *holds (and* dual optimal value is attained when $d^* > -\infty$*)*

- *such condition, called* Slater's condition
- *such point, (sometimes) said to be* strictly feasible
- *note resemblance with Slater's theorem in Theorem 8*

Duality for SDP

- (inequality form) SDP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x_1 F_1 + \cdots + x_n F_n + G \preceq 0 \end{array}$$

  where $F_1, \ldots, F_n, G \in \mathbf{S}^k$ and $\mathcal{K} = \mathbf{S}^k_+$

- Lagrangian

$$L(x, Z) = c^T x + (x_1 F_1 + \cdots + x_n F_n + G) \bullet Z = \sum x_i (F_i \bullet Z + c_i) + G \bullet Z$$

  where $X \bullet Y = \mathbf{Tr}\, XY$ for $X, Y \in \mathbf{S}^k$

- Lagrange dual function

$$g(Z) = \inf_{x \in \mathbf{R}^n} L(x, Z) = \begin{cases} G \bullet Z & F_i \bullet Z + c_i = 0 \quad i = 1, \ldots, n \\ -\infty & \text{otherwise} \end{cases}$$

- Lagrange dual problem

$$\begin{array}{ll} \text{maximize} & G \bullet Z \\ \text{subject to} & F_i \bullet Z + c_i = 0 \quad i = 1, \ldots, n \\ & Z \succeq 0 \end{array}$$

  where fact that $\mathbf{S}_+^k$ is self-dual, *i.e.*, $\mathcal{K}^* = \mathcal{K}$

- Slater's theorem (Theorem 16) implies if primal problem is strictly feasible, *i.e.*, exists $x \in \mathbf{R}^n$ such that $\sum x_i F_i + G \prec 0$, strong duality holds

KKT optimality conditions for generalized inequalities

Definition 67. [KKT optimality conditions for generalized inequalities] *for optimization problem in Definition 63 where f, q, and h are all differentiable, below conditions for $x \in \mathcal{D}$ and $(\lambda, \nu) \in \bigtimes \mathbf{R}^{k_i} \times \mathbf{R}^p$*

$$
\begin{aligned}
q(x) &\preceq_{\mathcal{K}} 0 && \text{- primal feasibility} \\
h(x) &= 0 && \text{- primal feasibility} \\
\lambda &\succeq_{\mathcal{K}^*} 0 && \text{- dual feasibility} \\
\lambda^T q(x) &= 0 && \text{- complementary slackness} \\
\nabla_x L(x, \lambda, \nu) &= 0 && \text{- vanishing gradient of Lagrangian}
\end{aligned}
$$

called Karush-Kuhn-Tucker (KKT) optimality conditions

- *note KKT optimality conditions for generalized inequalities subsume (normal) KKT optimality conditions (Definition 62)*

KKT conditions and optimalities for generalized inequalities

- for every optimization problem with generalized inequalities (Definition 63), every statement for normal optimization problem (Definition 38), regarding relations among KKT conditions, optimality, primal and dual optimality, and strong duality, is *exactly the same*
  - for every optimization problem with generalized inequalities (Definition 63)
    - if strong duality holds, primal and dual optimal points satisfy KKT optimality conditions in Definition 67 (same as Theorem 9)
    - if optimization problem is convex and primal and dual solutions satisfy KKT optimality conditions in Definition 67, the solutions are optimal with strong duality (same as Theorem 10)
    - therefore, for convex optimization problem, *KKT optimality conditions are necessary and sufficient for primal and dual optimality with strong duality*

Perturbation and sensitivity analysis for generalized inequalities

- original problem in Definition 63 with perturbed constraints

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & q(x) \preceq_{\mathcal{K}} u \\ & h(x) = v \end{array}$$

  where $u \in \mathbf{R}^m$ and $v \in \mathbf{R}^p$
- define $p^*(u,v) = p^*(u,v) = \inf\{f(x) | x \in \mathcal{D}, q(x) \preceq u, h(x) = v\}$, which is convex when problem is convex optimization problem - note $p^*(0,0) = p^*$
- as for normal optimization problem case (page 134), if and dual optimum (λ^*, ν^*), if strong duality holds,

$$p^*(u,v) \geq p^*(0,0) - {\lambda^*}^T u - {\nu^*}^T v$$

  and

$$\nabla_u \, p^*(0,0) = -\lambda \quad \nabla_v \, p^*(0,0) = -\nu$$

Sensitivity analysis for SDP

- assume inequality form SDP and its dual problem on page 161 and page 162
- consider perturbed SDP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x_1 F_1 + \cdots + x_n F_n + G \preceq U \end{array}$$

  for some $U \in \mathbf{S}^k$

  – define $p^* : \mathbf{S}^k \to \mathbf{R}$ such that $p^*(U)$ is optimal value of above problem
- assume $x^* \in \mathbf{R}^n$ and $Z^* \in \mathbf{S}^k_+$ are primal and dual optimum with zero dualty gap
- then

$$p^*(U) \geq p^* - Z^* \bullet U$$

- if $\nabla_U p^*$ exists at $U = 0$

$$\nabla_U p^*(0) = -Z^*$$

Weak alternatives for generalized inequalities

Theorem 17. [weak alternatives for generalized inequalities] *for* $q : Q \to \bigtimes \mathbf{R}^{k_i}$ *&* $h : H \to \mathbf{R}^p$ *where* Q *and* H *are subsets of common Banach space assuming* $\mathcal{D} = Q \cap H \neq \emptyset$*, and* $\lambda \in \bigtimes \mathbf{R}^{k_i}$ *&* $\nu \in \mathbf{R}^p$*, below pairs of systems are strong alternatives*

$$q(x) \preceq_{\mathcal{K}} 0 \quad h(x) = 0 \qquad \& \qquad \lambda \succeq_{\mathcal{K}^*} 0 \quad g(\lambda, \nu) > 0$$

$$q(x) \prec_{\mathcal{K}} 0 \quad h(x) = 0 \qquad \& \qquad \lambda \succeq_{\mathcal{K}^*} 0 \quad \lambda \neq 0 \quad g(\lambda, \nu) \geq 0$$

where $\mathcal{K} = \bigtimes K_i$ *with proper cones* $K_i \subset \mathbf{R}^{k_i}$ *and function* $g : \bigtimes \mathbf{R}^{k_i} \times \mathbf{R}^p \to \mathbf{R}$ *defined by*

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} \left(\lambda^T q(x) + \nu^T h(x) \right)$$

note this theorem subsumes Theorem 11 and Theorem 12

Strong alternatives for generalized inequalities

Theorem 18. [strong alternatives for generalized inequalities] *for $\mathcal{K}$-convex $q : Q \to \bigtimes \mathbf{R}^{k_i}$ & affine $h : H \to \mathbf{R}^p$ where Q and H are subsets of $\mathbf{R}^n$ assuming $\mathcal{D} = Q \cap H \neq \emptyset$, and $\lambda \in \bigtimes \mathbf{R}^{k_i}$ & $\nu \in \mathbf{R}^p$, if exists $x \in \mathbf{relint}\, \mathcal{D}$ with $h(x) = 0$, below pairs of systems are strong alternatives*

$$q(x) \preceq_{\mathcal{K}} 0 \quad h(x) = 0 \qquad \& \qquad \lambda \succeq_{\mathcal{K}^*} 0 \quad g(\lambda, \nu) > 0$$

$$q(x) \prec_{\mathcal{K}} 0 \quad h(x) = 0 \qquad \& \qquad \lambda \succeq_{\mathcal{K}^*} 0 \quad \lambda \neq 0 \quad g(\lambda, \nu) \geq 0$$

where $\mathcal{K} = \bigtimes K_i$ with proper cones $K_i \subset \mathbf{R}^{k_i}$ and function $g : \bigtimes \mathbf{R}^{k_i} \times \mathbf{R}^p \to \mathbf{R}$ defined by

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} \left(\lambda^T q(x) + \nu^T h(x) \right)$$

note this theorem subsumes Theorem 13 and Theorem 14

Strong alternatives for SDP

- for $F_1, \ldots, F_n, G \in \mathbf{S}^k$, $x \in \mathbf{R}^n$, and $Z \in \mathbf{S}^k$
  - below systems are strong alternatives

$$x_1 F_1 + \cdots + x_n F_n + G \prec 0$$

and

$$Z \succeq 0 \quad Z \neq 0 \quad G \bullet Z \geq 0 \quad F_i \bullet Z = 0 \; i = 1, \ldots, n$$

  - if $\sum v_i F_i \succeq 0 \Rightarrow \sum v_i F_i = 0$, below systems are strong alternatives

$$x_1 F_1 + \cdots + x_n F_n + G \preceq 0$$

and

$$Z \succeq 0 \quad G \bullet Z > 0 \quad F_i \bullet Z = 0 \; i = 1, \ldots, n$$

Unconstrained Minimization

Unconstrained minimization

- consider unconstrained convex optimization problem, $i.e.$, $m = p = 0$ in Definition 42

$$\text{minimize} \quad f(x)$$

  where domain of optimization problem is $\mathcal{D} = F \subset \mathbf{R}^n$

- assume
  - f is twice-differentiable (hence by definition F is open)
  - optimal solution x^* exists, $i.e.$, $p^* = \inf_{x \in \mathcal{D}} f(x) = f(x^*)$
- Theorem 4 implies x^* is optimal solution *if and only if*

$$\nabla f(x^*) = 0$$

- can solve above equation directly for few cases, but usually depend on iterative method, $i.e.$, find sequence of points $x^{(0)}, x^{(1)}, \ldots \in F$ such that $\lim_{k \to \infty} f(x^{(k)}) = p^*$

Requirements for iterative methods

- requirements for iterative methods
  - initial point $x^{(0)}$ should be in domain of optimization problem, $i.e.$

$$x^{(0)} \in F$$

  - sublevel set of $f(x^{(0)})$

$$S = \left\{ x \in F \,\middle|\, f(x) \leq f(x^{(0)}) \right\}$$

    should be closed
- $e.g.$
  - sublevel set of $f(x^{(0)})$ is closed for all $x^{(0)} \in F$ if f is closed, $i.e.$, all its sublevel sets are closed
  - f is closed if $F = \mathbf{R}^n$ and f is continuous
  - f is closed if f is continuous, F is open, and $f(x) \to \infty$ as $x \to \mathbf{bd}\, F$

Unconstrained minimization examples

- convex quadratic problem

$$\text{minimize} \quad f(x) = (1/2)x^T Px + q^T x$$

  where $P \in \mathbf{S}_+^n$ and $q \in \mathbf{R}^n$

  – solution obtained by solving

$$\nabla f(x^*) = Px^* + q = 0$$

    - if solution exists, $x^* = -P^\dagger q$ (thus $p^* > -\infty$)
    - otherwise, problem is unbounded below, *i.e.*, $p^* = -\infty$

  – *ability to analytically solve quadratic minimization problem is basis for Newton's method, power method for unconstrained minimization*

– least-squares (LS) is special case of convex quadratic problem

$$\text{minimize} \quad (1/2)\|Ax - b\|_2^2 = (1/2)x^T(A^TA)x - b^TAx + (1/2)\|b\|_2^2$$

- optimal always exists, can be obtained via normal equations

$$A^TAx^* = b$$

- unconstrained GP

$$\text{minimize} \quad f(x) = \log\left(\sum \exp(Ax + b)\right)$$

for $A \in \mathbf{R}^{m\times n}$ and $b \in \mathbf{R}^m$

– solution obtained by solving

$$\nabla f(x^*) = \frac{\sum A^T \exp(Ax^* + b)}{\sum \exp(Ax^* + b)} = 0$$

- need to resort to iterative method - since $F = \mathbf{R}^n$ and f is continuous, f is closed, hence every point in $\mathbf{R}^n$ can be initial point

- analytic center of linear inequalities

$$\text{minimize} \quad f(x) = -\sum \log(b - Ax)$$

where $F = \{x \in \mathbf{R}^n | b - Ax \succ 0\}$

- need to resort to iterative method - since F is open, f is continuous, and $f(x) \to \infty$ as $x \to \mathbf{bd}\, F$, f is closed, hence every point in F can be initial point

- f, called *logarithmic barrier* for inequalities $Ax \preceq b$

- analytic center of LMI

$$\text{minimize} \quad f(x) = -\log\det F(x) = \log\det F(x)^{-1}$$

where $F : \mathbf{R}^n \to \mathbf{S}^k$ is defined by

$$F(x) = x_1 F_1 + \cdots + x_n F_n$$

where $F_i \in \mathbf{S}^k$ and $F = \{x \in \mathbf{R}^n | F(x) \succ 0\}$

- need to resort to iterative method - since F is open, f is continuous, and $f(x) \to \infty$ as $x \to \mathbf{bd}\, F$, f is closed, hence every point in F can be initial point

- f, called *logarithmic barrier* for LMI

Strong convexity and implications

- function f is strongly convex on S

$$(\exists m > 0)\,(\forall x \in S)\left(\nabla^2 f(x) \succeq mI\right)$$

- strong convexity implies for every $x, y \in S$

$$f(y) \geq f(x) + \nabla f(x)^T(y - x) + (m/2)\|y - x\|_2^2$$

  – which implies gradient provides optimality certificate and tells us how far current point is from optimum, $i.e.$

$$f(x) - p^* \leq (1/2m)\|\nabla f(x)\|_2^2 \quad \|x - x^*\|_2 \leq (2/m)\|\nabla f(x)\|_2$$

- first equation implies sublevel sets contained in S is bounded, hence continuous function $\nabla^2 f(x)$ is also bounded, $i.e.$, $(\exists M > 0)\left(\nabla^2 f(x) \preceq MI\right)$, then

$$f(x) - p^* \geq \frac{1}{2M}\|\nabla f(x)\|_2^2$$

Iterative methods

Definition 68. [iterative meethods] *numerical method generating sequence of points* $x^{(0)}, x^{(1)}, \ldots \in S \subset \mathbf{R}^n$ *to make* $f(x^{(k)})$ *approaches to some desired value from some* $f : S \to \mathbf{R}$*, called* iterative method

Definition 69. [iterative meethods with search directions] *iterative method generating search direction* $\Delta x^{(k)} \in \mathbf{R}^n$ *and step length* $t^{(k)} > 0$ *at each step* k *such that*

$$x^{(k+1)} = x^{(k)} + t^{(k)} \Delta x^{(k)}$$

called iterative method with search direction *where* $\Delta x^{(k)}$*, called* search direction*,* $t^{(k)}$*, called* step length *(which actually is not length)*

Definition 70. [descent methods] *for function* $f : S \to \mathbf{R}$*, iterative method reducing function value,* $i.e.$

$$f(x^{(k+1)}) \leq f(x^{(k)})$$

for $k = 0, 1, \ldots$*, called* descent method

Line search methods

Definition 71. [line search method] *for iterating method with search directions, determining search direction* $\Delta x^{(k)}$ *and step length* $t^{(k)}$ *for each step, called* line search method

Algorithm 1. [exact line search] *for descent iterating method with search directions, determine* t *by*

$$t = \underset{s>0}{\operatorname{argmin}} f(x + s\Delta x)$$

Algorithm 2. [backtracking line search] *for descent iterating method with search directions, determine* t *by*

Require: f, $\Delta x^{(k)}$, $\alpha \in (0, 0.5)$, $\beta \in (0, 1)$
  $t := 1$
  while $f(x^{(k)} + t\Delta x^{(k)}) > f(x^{(k)}) + \alpha t \nabla f(x^{(k)})^T \Delta x^{(k)}$ ***do***
    $t := \beta t$
  end while

Gradient descent method

Algorithm 3. [gradient descent method]

Require: *f, initial point $x \in \mathbf{dom}\, f$*

  repeat

    search direction - $\Delta x := -\nabla f(x)$

    do line search to choose $t > 0$

    update - $x := x + t\Delta x$

  until *stopping criterion satisfied*

Summary of gradient descent method

- gradient method often exhibits approximately linear convergence, $i.e.$, error $f(x^{(k)}) - p^*$ converges to zero approximately as geometric series
- choice of backtracking parameters α and β has noticeable but not dramatic effect on convergence
- exact line search sometimes improves convergence of gradient method, but not by large, hence mostly not worth implementation
- converge rate depends greatly on condition number of Hessian or sublevel sets - when condition number if large, gradient method can be useless

Newton's method - motivation

- second-order Taylor expansion of f - $\hat{f}(\Delta x) = f(x+\Delta x) = f(x) + \nabla f(x)^T \Delta x + \frac{1}{2}\Delta x^T \nabla^2 f(x) \Delta x$
- minimum of Taylor expansion achieved when $\nabla \hat{f}(\Delta x) = \nabla f(x) + \nabla^2 f(x) v = 0$
- solution called *Newton step*

$$\Delta x_{\mathrm{nt}}(x) = -\nabla^2 f(x)^{-1} \nabla f(x)$$

  assuming $\nabla^2 f(x) \succ 0$
- thus Newton step minimizes local quadratic approximation of function
- difference of current and quadratic approximation minimum

$$f(x) - \hat{f}(\Delta x_{\mathrm{tn}}(x)) = \frac{1}{2}\Delta x_{\mathrm{nt}}^T \nabla^2 f(x) \Delta x_{\mathrm{nt}} = \frac{1}{2}\lambda(x)^2$$

- *Newton decrement*

$$\lambda(x) = \sqrt{\Delta x_{\mathrm{nt}}(x)^T \nabla^2 f(x) \Delta x_{\mathrm{nt}}(x)} = \sqrt{\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)}$$

Newton's method

Algorithm 4. [Newton's method] *damped descent method using Newton step*

Require: *f, initial point $x \in \mathbf{dom}\, f$, tolerance $\epsilon > 0$*

loop

computer Newton step and descrement

$$\Delta x_{\mathrm{nt}}(x) := -\nabla^2 f(x)^{-1} \nabla f(x) \quad \lambda(x)^2 := \nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)$$

stopping criterion - quit if $\lambda(x)^2/2 < \epsilon$

do line search to choose $t > 0$

update - $x := x + t\Delta x_{\mathrm{nt}}$

end loop

- Newton step is descent direction since

$$\left(\frac{d}{dx} f(x + t\Delta x_{\mathrm{nt}}) \right) \Big|_{t=0} = \nabla f(x)^T \Delta x_{\mathrm{nt}} = -\lambda(x)^2 < 0$$

Assumptions for convergence analysis of Newton's method

- assumptions
  - strong convexity and boundedness of Hessian on sublevel set

$$(\exists\, m, M > 0)\,(\forall x \in S)\left(mI \preceq \nabla^2 f(x) \preceq MI\right)$$

  - Lipschitz continuity of Hessian on sublevel set

$$(\exists L > 0)\,(\forall x, y \in S)\left(\|\nabla^2 f(x) - \nabla^2 f(y)\|_2 \leq L\|x - y\|_2\right)$$

- Lipschitz continuity constant L plays critical role in performance of Newton's method
  - intuition says Newton's method works well for functions whose quadratic approximations do not change fast, $i.e.$, when L is small

Convergence analysis of Newton's method

Theorem 19. [convergence analysis of Newton's method] *for function f satisfying strong convexity, Hessian continuity & Lipschitz continuity with $m, M, L > 0$, exist $0 < \eta < m^2/L$ and $\gamma > 0$ such that for each step k*

- *damped Newton phase - if $\|\nabla f(x^{(k)})\|_2 \geq \eta$,*

$$f(x^{(k+1)}) - f(x^{(k)}) \leq -\gamma$$

- *quadratic convergence phase - if $\|\nabla f(x^{(k)})\|_2 < \eta$, backtracking line search selects step length $t^{(k)} = 1$*

$$\frac{L}{2m^2}\|\nabla f(x^{(k+1)})\|_2 \leq \left(\frac{L}{2m^2}\|\nabla f(x^{(k)})\|_2\right)^2$$

 # iterations of Newton's method required to satisfy stopping criterion $f(x^{(k)}) - p^* \leq \epsilon$ is

$$\frac{f(x^{(0)}) - p^*}{\gamma} + \log_2 \log_2(\epsilon_0/\epsilon) \quad \text{where } \epsilon_0 = 2m^3/L^2$$

Summary of Newton's method

- Newton's method is *affine invariant*, hence *performance is independent of condition number unlike gradient method*
- once entering quadratic convergence phase, Newton's method converges extremely fast
- performance not much dependent on choice of algorithm parameters
- big disadvantage is computational cost for evaluating search direction, $i.e.$, solving linear system

Self-concordance

Definition 72. [self-concordance] *convex function* $f : X \to \mathbf{R}$ *with* $X \subset \mathbf{R}^n$ *such that for all* $x \in X, v \in \mathbf{R}^n$, $g(t) = f(x + tv)$ *with* $\mathbf{dom}\, g = \{t \in \mathbf{R} | x + tv \in X\}$ *satisfies*

$$(\forall t \in \mathbf{dom}\, g) \left(|g'''(t)| \leq 2g''(t)^{3/2} \right)$$

Proposition 11. [self-concordance for logarithms] *if convex function* $g : X \to \mathbf{R}$ *with* $X \subset \mathbf{R}_{++}$ *satisfies*

$$|g'''(x)| \leq 3g''(x)/x$$

function f *with* $\mathbf{dom}\, f = \{x \in \mathbf{R}_{++} | g(x) < 0\}$ *defined by*

$$f(x) = -\log(-g(x)) - \log x$$

and function h *with* $\mathbf{dom}\, h = \{x \in \mathbf{R}_{++} | g(x) + ax^2 + bx + c < 0\}$ *with* $a \geq 0$ *defined by*

$$h(x) = -\log(-g(x) - ax^2 - bx - c) - \log x$$

are self-concordant

Why self-concordance?

- convergence analysis of Newton's method depends on assumptions about function characteristics, $e.g.$, $m, M, L > 0$ for strong convexity, continuity of Hessian, $i.e.$

$$mI \preceq \nabla^2 f(x) \preceq MI \quad \|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L\|x - y\|$$

- *self-concordance* discovered by Nesterov and Nemirovski (who gave name self-concordance) plays important role for reasons such as
  - convergence analysis does not depend any function characterizing paramters
  - many barrier functions which are used for interior-point methods, which are important class of optimization algorithms are self-concordance
  - property of self-concordance is affine invariant

Self-concordance preserving operations

Proposition 12. [self-concordance preserving operations] *self-concordance is preserved by* positive scaling, addition, and affine transformation, *i.e., if* $f, g : X \to \mathbf{R}$ *are self-concordant functions with* $X \subset \mathbf{R}^n$, $h : H \to \mathbf{R}^n$ *with* $H \subset \mathbf{R}^m$ *are affine functions, and* $a > 0$

$$af, \quad f + g, \quad f \circ h$$

are self-concordant where $\mathbf{dom}\, f \circ h = \{x \in H | h(x) \in X\}$

Self-concordant function examples

- negative logarithm - $f : \mathbf{R}_{++} \to \mathbf{R}$ with

$$f(x) = -\log x$$

  is self-concordant since

$$|f'''(x)|/f''(x)^{3/2} = \left(2/x^3\right) / \left((1/x^2)^{3/2}\right) = 2$$

- negative entropy plus negative logarithm - $f : \mathbf{R}_{++} \to \mathbf{R}$ with

$$f(x) = x \log x - \log x$$

  is self-concordant since

$$|f'''(x)|/f''(x)^{3/2} = (x+2)/(x+1)^{3/2} \leq 2$$

- log barrier for linear inequalities - for $A \in \mathbf{R}^{m \times n}$ and $b \in \mathbf{R}^m$

$$f(x) = -\sum \log(b - Ax)$$

with $\mathbf{dom}\, f = \{x \in \mathbf{R}^n | b - Ax \succ 0\}$ is self-concordant by Proposition 12, *i.e.*, f is affine transformation of sum of self-concordant functions

- log-determinant - $f : \mathbf{S}^n_{++} \to \mathbf{R}$ with

$$f(X) = \log\det X^{-1} = -\log\det X$$

is self-concordant since for every $X \in \mathbf{S}^n_{++}$ and $V \in \mathbf{S}^n$ function $g : \mathbf{R} \to \mathbf{R}$ defined by $g(t) = -\log\det(X + tV)$ where $\mathbf{dom}\, f = \{t \in \mathbf{R} | X + tV \succeq 0\}$ is self-concordant since

$$\begin{aligned} g(t) &= -\log\det(X^{1/2}(I + tX^{-1/2}VX^{-1/2})X^{1/2}) \\ &= -\log\det X - \log\det(I + tX^{-1/2}VX^{-1/2}) \\ &= -\log\det X - \sum \log(1 + t\lambda_i(X, V)) \end{aligned}$$

where $\lambda_i(X, V)$ is i-th eigenvalue of $X^{-1/2}VX^{1/2}$ is self-concordant by Proposition 12, *i.e.*, g is affine transformation of sum of self-concordant functions

- log of concave quadratic - $f : X \to \mathbf{R}$ with

$$f(x) = -\log(-x^T P x - q^T x - r)$$

  where $P \in \mathbf{S}_+^n$ and $X = \{x \in \mathbf{R}^n | x^T P x + q^T x + r < 0\}$
- function $f : X \to \mathbf{R}$ with

$$f(x) = -\log(-g(x)) - \log x$$

  where $\mathbf{dom}\, f = \{x \in \mathbf{dom}\, g \cap \mathbf{R}_{++} | g(x) < 0\}$ and function $h : H \to \mathbf{R}$

$$h(x) = -\log(-g(x) - ax^2 - bx - c) - \log x$$

  where $a \geq 0$ and $\mathbf{dom}\, h = \{x \in \mathbf{dom}\, g \cap \mathbf{R}_{++} | g(x) + ax^2 + bx + c < 0\}$ are self-concordant if g is one of below
  - $g(x) = -x^p$ for $0 < p \leq 1$
  - $g(x) = -\log x$
  - $g(x) = x \log x$

  - $g(x) = x^p$ for $-1 \leq p \leq 0$
  - $g(x) = (ax+b)^2/x$ for $a, b \in \mathbf{R}$

  since above g satisfy $|g'''(x)| \leq 3g''(x)/x$ for every $x \in \mathbf{dom}\, g$ (Proposition 11)
- function $f : X \to \mathbf{R}$ with $X = \{(x,y) | \|x\|_2 < y\} \subset \mathbf{R}^n \times \mathbf{R}_{++}$ defined by

$$f(x,y) = -\log(y^2 - x^T x)$$

  is self-concordant - can be proved using Proposition 11
- function $f : X \to \mathbf{R}$ with $X = \{(x,y) | |x|^p < y\} \subset \mathbf{R} \times \mathbf{R}_{++}$ defined by

$$f(x,y) = -2\log y - \log(y^{2/p} - x^2)$$

  where $p \geq 1$ is self-concordant - can be proved using Proposition 11
- function $f : X \to \mathbf{R}$ with $X = \{(x,y) | \exp(x) < y\} \subset \mathbf{R} \times \mathbf{R}_{++}$ defined by

$$f(x,y) = -\log y - \log(\log y - x)$$

  is self-concordant - can be proved using Proposition 11

Properties of self-concordant functions

Definition 73. [Newton decrement] *for convex function* $f : X \to \mathbf{R}$ *with* $X \subset \mathbf{R}^n$, *function* $\lambda : \tilde{X} \to \mathbf{R}_+$ *with* $\tilde{X} = \{x \in X | \nabla^2 f(x) \succ 0\}$ *defined by*

$$\lambda(x) = \left(\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)\right)^{1/2}$$

called Newton decrement

- *note*

$$\lambda(x) = \sup_{v \neq 0} \left(v^T \nabla f(x) / \left(v^T \nabla^2 f(x) v \right)^{1/2} \right)$$

Theorem 20. [optimality certificate for self-concordant functions] *for strictly convex self-concordant function* $f : X \to \mathbf{R}^n$ *with* $X \subset \mathbf{R}^n$, *Hessian is positive definition everywhere (hence Newton decrement is defined everywhere) and for every* $x \in X$

$$p^* \geq f(x) - \lambda(x)^2 \quad \Leftrightarrow \quad f(x) - p^* \leq \lambda(x)^2$$

if $\lambda(x) \leq 0.68$

Stopping criteria for self-concordant objective functions

- recall $\lambda(x)^2$ provides *approximate* optimality certificate, (page 182) $i.e.$, assuming f is well approximated by quadratic function around x

$$f(x) - p^* \lessapprox \lambda(x)^2/2$$

- however, strict convexity together with self-concordance provides proven bound (by Theorem 20)

$$f(x) - p^* \leq \lambda(x)^2$$

  for $\lambda(x) \leq 0.68$

- hence can use following stopping criterion for guaranteed bound

$$\lambda(x)^2 \leq \epsilon \quad \Rightarrow \quad f(x) - p^* \leq \epsilon$$

  for $\epsilon \leq 0.68^2$

Convergence analysis of Newton's method for self-concordant functions

Theorem 21. [convergence analysis of Newton's method for self-concordant functions] *for strictly convex self-concordant function f, exist $0 < \eta \leq 1/4$ and $\gamma > 0$ (which depend only on line search parameters) such that*

- *damped Newton phase - if $\lambda(x^{(k)}) > \eta$*

$$f(x^{(k+1)}) - f(x^{(k)}) \leq -\gamma$$

- *quadratic convergence phase - if $\lambda(x^{(k)}) \leq \eta$ backtracking line search selects step length $t^{(k)} = 1$*

$$2\lambda(x^{(k+1)}) \leq \left(2\lambda(x^{(k)})\right)^2$$

 # iterations required to satisfy stopping criterion $f(x^{(k)}) - p^* \leq \epsilon$ is

$$\left(f(x^{(0)}) - p^*\right)/\gamma + \log_2 \log_2(1/\epsilon)$$

where $\gamma = \alpha\beta(1-2\alpha)^2/(20-8\alpha)$

Equality Constrained Minimization

Equality constrained minimization

- consider equality constrained convex optimization problem, $i.e.$, $m = 0$ in Definition 42

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \end{array}$$

  where $A \in \mathbf{R}^{p \times n}$ and domain of optimization problem is $\mathcal{D} = F \subset \mathbf{R}^n$

- assume
  - $\mathbf{rank}\, A = p < n$, $i.e.$, rows of A are linearly independent
  - f is twice-differentiable (hence by definition F is open)
  - optimal solution x^* exists, $i.e.$, $p^* = \inf_{x \in \mathcal{F}} f(x) = f(x^*)$ and $Ax^* = b$

Solving KKT for equality constrained minimization

- Theorem 10 implies $x^* \in F$ is optimal solution *if and only if* exists $\nu^* \in \mathbf{R}^p$ satisfy KKT optimality conditions, $i.e.$,

$$Ax^* = b \qquad \textit{primal feasibility equations}$$
$$\nabla f(x^*) + A^T\nu^* = 0 \qquad \textit{dual feasibility equations}$$

- solving equality constrained problem is equivalent to solving KKT equations
  - handful types of problems can be solved analytically
- using unconstrained minimization methods
  - can eliminate equality constraints and apply unconstrained minimization methods
  - solving dual problem using unconstrained minimization methods and retrieve primal solution (refer to page 133)
- will discuss Newton's method directly handling equality constraints
  - preserving problem structure such as sparsity

Equality constrained convex quadratic minimization

- equality constrained convex quadratic minimization problem

$$\begin{array}{ll} \text{minimize} & f(x) = (1/2)x^T Px + q^T x \\ \text{subject to} & Ax = b \end{array}$$

  where $P \in \mathbf{S}_+^n$ and $A \in \mathbf{R}^{p \times n}$
- important since basis for extension of Newton's method to equality constrained problems
- *KKT system*

$$Ax^* = b \;\&\; Px^* + q + A^T\nu^* = 0 \;\Leftrightarrow\; \underbrace{\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix}}_{\textit{KKT matrix}} \begin{bmatrix} x^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}$$

- exist primal and dual optimum (x^*, ν^*) *if and only if* KKT system has solution; otherwise, problem is unbounded below

Eliminating equality constraints

- can solve equality constrained convex optimization by
  - eliminating equality constraints and
  - using optimization method for solving unconstrained optimization
- note

$$\mathcal{F} = \{x | Ax = b\} = \{Fz + x_0 | z \in \mathbf{R}^{n-p}\}$$

  for some $F \in \mathbf{R}^{n \times (n-p)}$ where $\mathcal{R}(F) = \mathcal{N}(A)$
- thus original problem equivalent to

$$\text{minimize} \quad f(Fz + x_0)$$

- if z^* is optimal solution, $x^* = Fz^* + x_0$
- optimal dual can be retrieved by

$$\nu^* = -(AA^T)^{-1} A \nabla f(x^*)$$

Solving dual problems

- Lagrange dual function of equality constrained problem

$$\begin{aligned} g(\nu) &= \inf_{x\in\mathcal{D}} \Big(f(x) + \nu^T(Ax-b)\Big) = -b^T\nu - \sup_{x\in\mathcal{D}} \Big((-A^T\nu)^T x - f(x)\Big) \\ &= -b^T\nu - f^*(-A^T\nu) \end{aligned}$$

- dual problem

$$\text{maximize} \quad -b^T\nu - f^*(-A^T\nu)$$

- by assumption, strong duality holds, hence if ν^* is dual optimum

$$g(\nu^*) = p^*$$

- if dual objective is twice-differentiable, can solve dual problem using unconstrained minimization methods
- primal optimum can be retrieved using method on page 133)

Newton's method with equality constraints

- finally discuss Newton's method which directly handles equality constraints
  - similar to Newton's method for unconstrained minimization
  - initial point, however, should be feasible, $i.e.$, $x^{(0)} \in F$ and $Ax^{(0)} = b$
  - Newton step tailored for equality constrained problem

Newton step via second-order approximation

- solve original problem approximately by solving

$$\begin{array}{ll} \text{minimize} & \hat{f}(x+\Delta x) = f(x) + \nabla f(x)^T \Delta x + (1/2)\Delta x^T \nabla^2 f(x) \Delta x \\ \text{subject to} & A(x+\Delta x) = b \end{array}$$

  where $x \in \mathcal{F}$

- *Newton step for equality constrained minimization problem*, defined by solution of KKT system for above convex quadratic minimization problem

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\text{nt}} \\ w \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix}$$

  only when KKT system is nonsingular

Newton step via solving linearized KKT optimality conditions

- recall KKT optimality conditions for equality constrained convex optimization problem

$$Ax^* = b \quad \& \quad \nabla f(x^*) + A^T \nu^* = 0$$

- linearize KKT conditions

$$\begin{aligned} & A(x + \Delta x) = b \quad \& \quad \nabla f(x) + \nabla^2 f(x) \Delta x + A^T w = 0 \\ \Leftrightarrow \quad & A\Delta x = 0 \quad \& \quad \nabla^2 f(x) \Delta x + A^T w = -\nabla f(x) \end{aligned}$$

  where $x \in \mathcal{F}$

- Newton step defined by above equations is equivalent to that obtained by second-order approximation

Newton decrement for equality constrained minimization

- *Newton descrement for equality constrained problem* is defined by

$$\lambda(x) = \left(\Delta x_{\mathrm{nt}} \nabla^2 f(x) \Delta x_{\mathrm{nt}}\right)^{1/2}$$

- same expression as that for unconstrained minimization, but is *different* since Newton step Δx_{nt} is different from that for unconstrained minimization, $i.e.$, $\Delta x_{\mathrm{nt}} \neq -\nabla^2 f(x)^{-1} \nabla f(x)$ (refer to Definition 73)
- however, as before,

$$f(x) - \inf_{\Delta x \in \mathbf{R}^n} \{\hat{f}(x + \Delta x) | A(x + \Delta x) = b\} = \lambda(x)^2/2$$

and

$$\left(\frac{d}{dt} f(x + t\Delta x_{\mathrm{nt}})\right)\Big|_{t=0} = \nabla f(x)^T \Delta x_{\mathrm{nt}} = -\lambda(x)^2 < 0$$

Feasible Newton's method for equality constrained minimization

Algorithm 5. [feasible Newton's method for equality constrained minimization]

Require: *f, initial point $x \in \mathbf{dom}\, f$ with $Ax = b$, tolerance $\epsilon > 0$*

loop

computer Newton step and descrement $\Delta x_{\mathrm{nt}}(x)$ & $\lambda(x)$

stopping criterion - quit if $\lambda(x)^2/2 < \epsilon$

do line search on f to choose $t > 0$

update - $x := x + t\Delta x_{\mathrm{nt}}$

end loop

- Algorithm 5
  - assumes KKT matrix is nonsingular for every step
  - is *feasible descent method* since all iterates are feasible with $f(x^{(k+1)}) < f(x^{(k)})$

Assumptions for convergence analysis of feasible Newton's method for equality constrained minimization

- feasibility of initial point - $x^{(0)} \in \mathbf{dom}\, f \;\&\; Ax^{(0)} = b$
- sublevel set $S = \{x \in \mathbf{dom}\, f | f(x) \leq f(x^{(0)}),\; Ax = b\}$ is closed
- boundedness of Hessian on S

$$(\exists M > 0)\,(\forall x \in S)\left(\nabla^2 f(x) \preceq MI\right)$$

- boundedness of KKT matrix on S - corresponds to strong convexity assumption in unconstrained minimization

$$(\exists K > 0)\,(\forall x \in S)\left(\left\|\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix}^{-1}\right\|_2 \leq K\right)$$

- Lipschitz continuity of Hessian on S

$$(\exists L > 0)\,(\forall x, y \in S)\left(\left\|\nabla^2 f(x) - \nabla^2 f(y)\right\|_2 \leq L\|x - y\|_2\right)$$

Convergence analysis of feasible Newton's method for equality constrained minimization

- convergence analysis of Newton's method for equality constrained minimization can be done by analyzing unconstrained minimization after eliminating equality constraints
- thus, yield *exactly same* results as for unconstrained minimization (Theorem 19) (with different parameter values), $i.e.$,
  - consists of damped Newton phase and quadratic convergence phase
  - # iterations required to achieve $f(x^{(k)}) - p^* \leq \epsilon$ is

$$\left(f(x^{(0)}) - p^*\right)/\gamma + \log_2 \log_2(\epsilon_0/\epsilon)$$

- for # iterations required to achieve $f(x^{(k)}) - p^* \leq \epsilon$ for self-concordant functions is also same as for unconstrained minimization (Theorem 21)

$$\left(f(x^{(0)}) - p^*\right)/\gamma + \log_2 \log_2(1/\epsilon)$$

where $\gamma = \alpha\beta(1-2\alpha)^2/(20-8\alpha)$

Newton step at infeasible points

- only assume that $x \in \mathbf{dom}\, f$ (hence, can be infeasible)

- (as before) linearize KKT conditions

$$\begin{aligned} & A(x+\Delta x_{\mathrm{nt}}) = b \quad \& \quad \nabla f(x) + \nabla^2 f(x)\Delta x_{\mathrm{nt}} + A^T w = 0 \\ \Leftrightarrow \quad & A\Delta x_{\mathrm{nt}} = b - Ax \quad \& \quad \nabla^2 f(x)\Delta x_{\mathrm{nt}} + A^T w = -\nabla f(x) \\ \Leftrightarrow \quad & \begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\mathrm{nt}} \\ w \end{bmatrix} = - \begin{bmatrix} \nabla f(x) \\ Ax - b \end{bmatrix} \end{aligned}$$

- same as feasible Newton step *except second component on RHS of KKT system*

Interpretation as primal-dual Newton step

- update both primal and dual variables x and ν

- define $r : \mathbf{R}^n \to \mathbf{R}^p \to \mathbf{R}^n \times \mathbf{R}^p$ by

$$r(x, \nu) = (r_{\text{dual}}(x, \nu), r_{\text{pri}}(x, \nu))$$

where

$$\begin{aligned} \textit{dual residual} \quad &- \quad r_{\text{dual}}(x, \nu) = \nabla f(x) + A^T \nu \\ \textit{primal residual} \quad &- \quad r_{\text{pri}}(x, \nu) = Ax - b \end{aligned}$$

Equivalence of infeasible Newton step to primal-dual Newton step

- linearize r to obtain primal-dual Newton step, $i.e.$

$$r(x,\nu) + D_{x,\nu} r(x,\nu) \begin{bmatrix} \Delta x_{\text{pd}} \\ \Delta \nu_{\text{pd}} \end{bmatrix} = 0$$

$$\Leftrightarrow \quad \begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\text{pd}} \\ \Delta \nu_{\text{pd}} \end{bmatrix} = - \begin{bmatrix} \nabla f(x) + A^T \nu \\ Ax - b \end{bmatrix}$$

- letting $\nu^+ = \nu + \Delta \nu_{\text{pd}}$ gives

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\text{pd}} \\ \nu^+ \end{bmatrix} = - \begin{bmatrix} \nabla f(x) \\ Ax - b \end{bmatrix}$$

  - equivalent to infeasible Newton step
  - reveals that current value of dual variable not needed

Residual norm reduction property

- infeasible Newton step is *not* descent direction (unlike feasible Newton step) since

$$\left(\frac{d}{dt}f(x+t\Delta x_{\mathrm{pd}})\right)\Big|_{t=0} = \nabla f(x)^T\Delta x_{\mathrm{pd}}$$
$$= \quad -\Delta x_{\mathrm{pd}}^T\left(\nabla^2 f(x)\Delta x_{\mathrm{pd}} + A^T w\right) = -\Delta x_{\mathrm{pd}}^T\nabla^2 f(x)\Delta x_{\mathrm{pd}} + (Ax-b)^T w$$

  which is not necessarily negative

- however, norm of residual decreases in infeasible Newton direction

$$\left(\frac{d}{dx}\|r(y+t\Delta y_{\mathrm{pd}})\|_2^2\right)\Big|_{t=0} = -2r(y)^T r(y) = -2\|r(y)\|_2^2$$
$$\Leftrightarrow \quad \left(\frac{d}{dx}\|r(y+t\Delta y_{\mathrm{pd}})\|_2\right)\Big|_{t=0} = \frac{-2\|r(y)\|_2^2}{2\|r(y)\|_2} = -\|r(y)\|_2$$

  where $y=(x,\nu)$ and $\Delta y_{\mathrm{pd}} = (\Delta x_{\mathrm{pd}}, \Delta\nu_{\mathrm{pd}})$

- can use $r(x^{(k)},\nu^{(k)})$ to measure optimization progress for infeasible Newton's method

Full and damped step feasibility property

- assume step length is t at some iteration, then

$$r_{\text{pri}}(x^+, \nu^+) = Ax^+ - b = A(x + t\Delta x_{\text{pd}}) - b = (1-t)r_{\text{pri}}(x, \nu)$$

- hence $l > k$

$$r^{(l)} = \left(\prod_{i=k}^{l-1} (1 - t^{(i)}) \right) r^{(k)}$$

  - primal residual reduced by $1 - t^{(k)}$ at step k
  - Newton step becomes feasible step once full step length ($t = 1$) taken

Infeasible Newton's method for equality constrained minimization

Algorithm 6. [infeasible Newton's method for equality constrained minimization]

Require: *f, initial point $x \in \mathbf{dom}\, f$ & ν, tolerance $\epsilon_{\text{pri}} > 0$ & $\epsilon_{\text{dual}} > 0$*

repeat

computer Newton step and descrement $\Delta x_{\text{pd}}(x)$ & $\Delta \nu_{\text{pd}}(x)$,

do line search on $r(x, \nu)$ to choose $t > 0$

update - $x := x + t\Delta x_{\text{pd}}$ & $\nu := \nu + t\Delta \nu_{\text{pd}}$

until $\|r_{\text{dual}}(x, \nu)\| \leq \epsilon_{\text{dual}}$ & $\|Ax - b\| \leq \epsilon_{\text{pri}}$

- note similarity and difference of Algorithm 6 & Algorithm 5
  - line search done not on f, but on primal-dual residuals $r(x, \nu)$
  - stopping criteria depends on $r(x, \nu)$, not on Newton decrementa $\lambda(x)^2$
  - primal and dual feasibility checked separately - here norm in $\|Ax - b\|$ can be any norm, *e.g.*, $\|\cdot\|_0$, $\|\cdot\|_1$, $\|\cdot\|_2$, $\|\cdot\|_\infty$, depending on specific application

Line search methods for infeasible Newton's method

- line search methods for infeasible Newton's method, $i.e.$, Algorithm 1 & Algorithm 2 same with f replaced by $\|r(x,\nu)\|_2$,
- but they have special forms (of course) - refer to below special case descriptions

Algorithm 7. [exact line search for infeasible Newton's method]

$$t = \underset{s>0}{\operatorname{argmin}} \|r(x + s\Delta x_{\mathrm{pd}}, \nu + s\Delta\nu_{\mathrm{pd}})\|_2$$

Algorithm 8. [backtracking line search for infeasible Newton's method]

Require: Δx, $\Delta\nu$, $\alpha \in (0, 0.5)$, $\beta \in (0, 1)$

$t := 1$

while $\|r(x + t\Delta x_{\mathrm{pd}}, \nu + t\Delta\nu_{\mathrm{pd}})\|_2 > (1 - \alpha t)\|r(x,\nu)\|_2$ ***do***

$t := \beta t$

end while

Pros and cons of infeasible Newton's method

- pros
  - do not need to find feasible point separately, $e.g.$
    - "minimize $-\log(Ax) + b^T x$"

    can be solved by converting to
    - "minimize $-\log(y) + b^T x$ s.t. $y = Ax$"

    and solved by infeasible Newton's method
  - if step length is one at any iteration, following steps coincides with feasible Newton's method - could switch to feasible Newton's method
- cons
  - exists no clear way to detect feasibility - primal residual decreases slowly (phase I method in interior point method resolves this problem)
  - convergence of infeasible Newton's method can be very slow (until feasibility is achieved0

Assumptions for convergence analysis of infeasible Newton's method for equality constrained minimization

- sublevel set $S = \left\{(x,\nu) \in \mathbf{dom}\, f \times \mathbf{R}^m \,\middle|\, \|r(x,\nu)\|_2 \leq \|r(x^{(0)},\nu^{(0)})\|_2\right\}$ is closed, which always holds because $\|r\|_2$ is closed
- boundedness of KKT matrix on S

$$(\exists K > 0)\,(\forall x \in S)\left(\left\|Dr(x,\nu)^{-1}\right\|_2 = \left\|\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix}^{-1}\right\|_2 \leq K\right)$$

- Lipschitz continuity of Hessian on S

$$(\exists L > 0)\,(\forall (x,\nu),(y,\mu) \in S)\,(\|Dr(x,\nu) - Dr(y,\mu)\|_2 \leq L\|(x,\nu) - (y,\mu)\|_2)$$

- above assumptions imply $\{x \in \mathbf{dom}\, f | Ax = b\} \neq \emptyset$ and exist optimal point (x^*,ν^*)

Convergence analysis of infeasible Newton's method for equality constrained minimization

- very simliar to that for Newton's method for unconstrained minimization
- consist of two phases - like unconstrained minimization or infeasible Newton's method (refer to Theorem 19 or page 210)
  - damped Newton phase - if $\|r(x^{(k)}, \nu^{(k)})\|_2 > 1/(K^2L)$

$$\|r(x^{(k+1)}, \nu^{(k+1)})\|_2 \leq \|r(x^{(k)}, \nu^{(k)})\|_2 - \alpha\beta/K^2L$$

  - quadratic convergence damped Newton phase - if $\|r(x^{(k)}, \nu^{(k)})\|_2 \leq 1/(K^2L)$

$$\left(K^2L\|r(x^{(k)}, \nu^{(k)})\|_2/2\right) \leq \left(K^2L\|r(x^{(k-1)}, \nu^{(k-1)})\|_2/2\right)^2 \leq \cdots \leq (1/2)^{2^k}$$

- $\#$ iterations of infeasible Newton's method required to satisfy $\|r(x^{(k)}, \nu^{(k)})\|_2 \leq \epsilon$

$$\|r(x^{(0)}, \nu^{(0)})\|/(\alpha\beta/K^2L) + \log_2\log_2(\epsilon_0/\epsilon) \quad \text{where } \epsilon_0 = 2/(K^2L)$$

- $(x^{(k)}, \nu^{(k)})$ converges to (x^*, ν^*)

Barrier Interior-point Methods

Interior-point methods

- want to solve inequality constrained minimization problem
- interior-point methods solve convex optimization problem (Definition 42) or KKT optimality conditions (Definition 62) by
  - applying Newton's method to sequence of
    - equality constrained problems or
    - modified versions of KKT optimality conditions
- discuss interior-point *barrier method* & interior-point *primal-dual method*
- hierarchy of convex optimization algorithms
  - simplest - linear equality constrained quadratic program - can solve analytically
  - Newton's method - solve linear equality constrained convex optimization problem by solving sequence of linear equality constrained quadratic programs
  - interior-point methods - solve linear equality & convex inequality constrained problem by solving sequence of lienar equality constrained convex optimization problem

Indicator function barriers

- approxmiate general convex inequality constrained problem as linear equality constrained problem

- make inequality constraints implicit in objective function

$$\begin{array}{ll} \text{minimize} & f(x) + \sum I_-(q(x)) \\ \text{subject to} & Ax = b \end{array}$$

where $I_- : \mathbf{R} \to \mathbf{R}$ is indicator function for nonpositive real numbers, $i.e.$

$$I_-(u) = \begin{cases} 0 & u \leq 0 \\ \infty & u > 0 \end{cases}$$

Logarithmic barriers

- approximate indicator function by logarithmic function

$$\hat{I}_- = -(1/t)\log(-u) \quad \mathbf{dom}\, \hat{I}_- = -\mathbf{R}_{++}$$

  for $t > 0$ to obtain

$$\begin{array}{ll} \text{minimize} & f(x) + \sum -(1/t)\log(-q(x)) \\ \text{subject to} & Ax = b \end{array}$$

- objective function is convex due to composition rule for convexity preservation (page 41), and differentiable
- hence, can use Newton's method to solve it
- function ϕ defined by

$$\phi(x) = -\sum \log(-q(x))$$

  with $\mathbf{dom}\, \phi\{x \in X | q(x) \prec 0\}$ called *logarithmic barrier* or *log barrier*
- solve sequence of log barrier problems as we increase t

Central path

- optimization problem

$$\begin{array}{ll} \text{minimize} & tf(x) + \phi(x) \\ \text{subject to} & Ax = b \end{array}$$

  with $t > 0$ where

$$\phi(x) = -\sum \log(-q(x))$$

- solution of above problem, called *central point*, denoted by $x^*(t)$, set of central points, called *central path*

- intuition says $x^*(t)$ will converge to x^* as $t \to \infty$

- KKT conditions imply

$$Ax^*(t) = b \quad q(x^*(t)) \prec 0$$

and exists $\nu^*(t)$ such that

$$\begin{aligned} 0 &= t\nabla f(x^*(t)) + \nabla\phi(x^*(t)) + tA^T\nu^*(t) \\ &= t\nabla f(x^*(t)) - \sum \frac{1}{q_i(x^*(t))}\nabla q_i(x^*(t)) + tA^T\nu^*(t) \end{aligned}$$

- thus if we let $\lambda^*(t) = -1/tq_i(x^*(t))$, $x^*(t)$ minimizes

$$L(x, \lambda^*(t), \nu^*(t)) = f(x) + \lambda^*(t)^T q(x) + \nu^*(t)^T(Ax - b)$$

where L is Lagrangian of original problem in Definition 42

- hence, dual function $g(\lambda^*(t), \nu^*(t))$ is finite and

$$g(\lambda^*(t), \nu^*(t)) = \inf_{x \in X} L(x, \lambda^*(t), \nu^*(t)) = L(x^*(t), \lambda^*(t), \nu^*(t))$$
$$= f(x^*(t)) + \lambda^*(t)^T q(x^*(t)) + \nu^*(t)^T (Ax^*(t) - b) = f(x^*(t)) - m/t$$

and

$$f(x^*(t)) - p^* \leq f(x^*(t)) - g(\lambda^*(t), \nu^*(t)) = m/t$$

that is,

$x^*(t)$ is no more than m/t-suboptimal

which confirms out intuition that $x^*(t) \to x^*$ as $t \to \infty$

Central path interpretation via KKT conditions

- previous arguments imply that x is central point, *i.e.*, $x = x^*(t)$ for some $t > 0$ *if and only if* exist λ and ν such that

$$\begin{aligned} Ax = b \quad q(x) &\preceq 0 \quad \text{- primal feasibility} \\ \lambda &\succeq 0 \quad \text{- dual feasibility} \\ -{\lambda_i}^T q_i(x) &= 1/t \quad \text{- complementary } 1/t\text{-slackness} \\ \nabla_x L(x, \lambda, \nu) &= 0 \quad \text{- vanishing gradient of Lagrangian} \end{aligned}$$

  called *centrality conditions*
- only difference between centrality conditions and KKT conditions in Definition 62 is *complementary $1/t$-slackness*
  - note that I've just made up term "complementary $1/t$-slackness" - you won't be able to find terminology in any literature
- for large t, $\lambda^*(t)$ & $\nu^*(t)$ *very closely* satisfy (true) complementary slackness

Central path interpretation via force field

- assume exist no equality constraints
- interpret ϕ as potential energy by some force field, $e.g.$, electrical field and tf as potential energy by some other force field, $e.g.$, gravity
- then
  - force by first force field (in n-dimensional space), which we call *barrier force*, is

$$-\nabla\phi(x) = \sum \frac{1}{q_i(x)} \nabla q_i(x)$$

  - force by second force field, which we call *objective force*, is

$$-\nabla(tf(x)) = -t\nabla f(x)$$

- $x^*(t)$ is point where two forces exactly balance each other
  - as x approach boundary, barrier force pushes x harder from barriers,
  - as t increases, objective force pushes x harder to point where objective potential energy is minimized

Equality constrained problem using log barrier

- central point $x^*(t)$ is m/t-suboptimal point guaranteed by optimality certificate $g(\lambda^*(t), \nu^*(t))$

- hence solving below problem provides solution with ϵ-suboptimality

$$\begin{array}{ll} \text{minimize} & (m/\epsilon)f(x) + \phi(x) \\ \text{subject to} & Ax = b \end{array}$$

- but works only for small problems since for large m/ϵ, objective function ill behaves

Barrier methods

Algorithm 9. [barrier method]

Require: *strictly feasible* x, $t > 0$, $\mu > 1$, *tolerance* $\epsilon > 0$

repeat

centering step - find $x^*(t)$ *by minimizing* $tf + \phi$ *subject to* $Ax = b$ *starting at* x

(optionally) compute $\lambda^*(t)$ *&* $\nu^*(t)$

stopping criterion - quit if $m/t < \epsilon$

increase t *-* $t := \mu t$

update x *-* $x := x^*(t)$

until

- *barrier method*, also called *path-following method*, solves sequence of equality constrained optimization problem with log barrier
  - when first proposed by Fiacco and McCormick in 1960s, it was called *sequential unconstrained minimization technique (SUMT)*
- *centering step* also called *outer iteration*
- each iteration of algorithm used for equality constrained problem, called *inner iteration*

Accuracy in centering in barrier method

- accuracy of centering
  - only goal of centering is getting close to x^*, hence exact calculation of $x^*(t)$ not critical as long as approximates of $x^*(t)$ go to x^*
  - while cannot calculate $g(\lambda, \nu)$ for this case, below provides dual feasible point when Newton step Δx_{nt} for optimization problem on page 225 is small, $i.e.$, for nearly centered

$$\tilde{\lambda}_i = -\frac{1}{tq_i(x)}\left(1 - \frac{\nabla q_i(x)^T \Delta x_{\mathrm{nt}}}{q_i(x)}\right)$$

Choices of parameters of barrier method

- choice of μ
  - μ determines aggressiveness of t-update
    - larger μ, less outer iterations, but more inner iterations
    - smaller μ, less outer iterations, but more inner iterations
  - values from 10 to 20 for μ seem to work well
- candidates for choice of initial t - choose $t^{(0)}$ such that

$$m/t^{(0)} \approx f(x^{(0)}) - p^*$$

or make central path condition on page 225 maximally satisfied

$$t^{(0)} = \operatorname*{arginf}_t \inf_{\tilde{\nu}} \left\| t\nabla f(x^{(0)}) + \nabla\phi(x^{(0)}) + A^T\tilde{\nu} \right\|$$

Convergence analysis of barrier method

- assuming $tf + \phi$ can be minimized by Newton's method for $t^{(0)}$, $\mu t^{(0)}$, $\mu^2 t^{(0)}$, . . .
- at k'th step, duality gap achieved is $m/\mu^k t^{(0)}$
- # centering steps required to achieve accuracy of ϵ is

$$\left\lceil \frac{\log\left(m/\epsilon t^{(0)}\right)}{\log \mu} \right\rceil$$

  plus one (initial centering step)
- for convergence of centering
  - for feasible centering problem, $tf + \phi$ should satisfy conditions on page 209, *i.e.*, initial sublevel set is closed, associated inverse KKT matrix is bounded & Hessian satisfies Lipschitz condition
  - for infeasible centering problem, $tf + \phi$ should satisfy conditions on page 219

Primal-dual Interior-point Methods

Primal-dual & barrier interior-point methods

- in primal-dual interior-point methods
  - both primal and dual variables are updated at each iteration
  - search directions are obtained from Newton's method, applied to modified KKT equations, $i.e.$, optimality conditions for logarithmic barrier centering problem
  - primal-dual search directions are similar to, but not quite the same as, search directions arising in barrier methods
  - primal and dual iterates are not necessarily feasible
- primal-dual interior-point methods
  - often more efficient than barrier methods especially when high accuracy is required - can exhibit better than linear convergence
  - (customized versions) outperform barrier method for several basic problems, such as, LP, QP, SOCP, GP, SDP
  - can work for feasible, but *not* strictly feasible problems
  - still active research topic, but show great promise

Modified KKT conditions and central points

- modified KKT conditions (for convex optimization in Definition 42) expressed as

$$r_t(x, \lambda, \nu) = \begin{bmatrix} \nabla f(x) + Dq(x)^T \lambda + A^T \nu \\ -\mathbf{diag}(\lambda) f(x) - (1/t)\mathbf{1} \\ Ax - b \end{bmatrix}$$

  where

$$\begin{aligned} \textit{dual residual} \quad &- \quad r_{\text{dual}}(x, \lambda, \nu) = \nabla f(x) + Dq(x)^T \lambda + A^T \nu \\ \textit{centrality residual} \quad &- \quad r_{\text{cent}}(x, \lambda, \nu) = -\mathbf{diag}(\lambda) f(x) - (1/t)\mathbf{1} \\ \textit{primal residual} \quad &- \quad r_{\text{pri}}(x, \lambda, \nu) = Ax - b \end{aligned}$$

- if x, λ, ν satisfy $r_t(x, \lambda, \nu) = 0$ (and $q(x) \prec 0$), then
  - $x = x^*(t)$, $\lambda = \lambda^*(t)$, $\nu = \nu^*(t)$
  - x is primal feasible and λ & ν are dual feasible with duality gap m/t

Primal-dual search direction

- assume current (primal-dual) point $y = (x, \lambda, \nu)$ and Newton step $\Delta y = (\Delta x, \Delta\nu, \Delta\lambda)$
- as before, linearize equation to obtain Newton step, *i.e.*,

$$r_t(y + \Delta y) \approx r_t(y) + Dr_t(y)\Delta y = 0 \quad \Leftrightarrow \quad \Delta y = -Dr_t(y)^{-1} r_t(y)$$

  hence

$$\begin{bmatrix} \nabla^2 f(x) + \sum \lambda_i \nabla^2 q_i(x) & Dq(x)^T & A^T \\ -\mathbf{diag}(\lambda) Df(x) & -\mathbf{diag}(f(x)) & 0 \\ A & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta\lambda \\ \Delta\nu \end{bmatrix} = - \begin{bmatrix} r_{\text{dual}} \\ r_{\text{cent}} \\ r_{\text{pri}} \end{bmatrix}$$

- above equation determines *primal-dual search direction* $\Delta y_{\text{pd}} = (\Delta x_{\text{pd}}, \Delta\lambda_{\text{pd}}, \Delta\nu_{\text{pd}})$

Surrogate duality gap

- iterates $x^{(k)}$, $\lambda^{(k)}$, and $\nu^{(k)}$ of primal-dual interior-point method are *not* necessarily feasible
- hence, cannot easily evaluate duality gap $\eta^{(k)}$ as for barrier method
- define *surrogate duality gap* for $q(x) \prec 0$ and $\lambda \succeq 0$ as

$$\hat{\eta}(x, \lambda) = -q(x)^T \lambda$$

- $\hat{\eta}$ would be duality gap if x were primal feasible and λ & ν were dual feasible
- value t corresponding to surrogate duality gap $\hat{\eta}$ is $m/\hat{\eta}$

Primal-dual interior-point method

Algorithm 10. [primal-dual interior-point method]

Require: *initial point* x *with* $q(x) \prec 0$, $\lambda \succ 0$, $\mu > 1$, $\epsilon_{\text{pri}} > 0$, $\epsilon_{\text{dual}} > 0$, $\epsilon > 0$

repeat

set $t := \mu m/\hat{\eta}$

computer primal-dual search direction $\Delta y_{\text{pd}} = (\Delta x_{\text{pd}}, \Delta\lambda_{\text{pd}}, \Delta\nu_{\text{pd}})$

do line search to choose $s > 0$

update - $x := x + s\Delta x_{\text{pd}}$, $\lambda := \lambda + s\Delta\nu_{\text{pd}}$, $\nu := \nu + s\Delta\nu_{\text{pd}}$

until $\|r_{\text{pri}}(x, \lambda, \nu)\|_2 \leq \epsilon_{\text{pri}}$, $\|r_{\text{dual}}(x, \lambda, \nu)\|_2 \leq \epsilon_{\text{dual}}$, $\hat{\eta} \leq \epsilon$

- common to choose small ϵ_{pri}, ϵ_{dual}, & ϵ since primal-dual method often shows faster than linear convergence

Line search for primal-dual interior-point method

- liner search is standard backtracking line search on $r(x, \lambda, \nu)$ similar to that in Algorithm 7 except making sure that $q(x) \prec 0$ and $\lambda \succ 0$
- note initial s in Algorithm 11 is largest s that makes $\lambda + s\Delta\lambda_{\text{pd}}$ positive

Algorithm 11. [backtracking line search for primal-dual interior-point method]

Require: Δx_{pd}, $\Delta\lambda_{\text{pd}}$, $\Delta\nu_{\text{pd}}$, $\alpha \in (0.01, 0.1)$, $\beta \in (0.3, 0.8)$

$s := 0.99 \sup\{s \in [0,1] | \lambda + s\Delta\lambda \succeq 0\} = 0.99 \min\{1, \min\{-\lambda_i/\Delta\lambda_i | \Delta\lambda_i < 0\}\}$

while $q(x + s\Delta x_{\text{pd}}) \not\prec 0$ ***do***

$t := \beta t$

end while

while $\|r(x + s\Delta x_{\text{pd}}, \lambda + s\Delta\lambda_{\text{pd}}, \nu + s\Delta\nu_{\text{pd}})\|_2 > (1 - \alpha s)\|r(x, \lambda, \nu)\|_2$ ***do***

$t := \beta t$

end while

Selected Proofs

Selected proofs

- **Proof 1.** *(Proof for "intersection of convex sets is convex set" on page 20)*
  Suppose $\mathcal{C}$ is a collection of convex sets. Suppose $x, y \in \bigcap_{C\in\mathcal{C}} C$ and $0 < \theta < 1$. Then for each $C \in \mathcal{C}$ and $\theta x + (1-\theta)y \in C$, hence, $\theta x + (1-\theta)y \in \bigcap_{C\in\mathcal{C}} C$, $\bigcap_{C\in\mathcal{C}} C$ is a convex set. ∎

- **Proof 2.** *(Proof for "theorem of alternative for linear strict generalized inequalities" on page 30)*
  Suppose $Ax \prec_K b$ is infeasible. Then $\{b - Ax | x \in \mathbf{R}^n\} \cap K^\circ = \emptyset$. Theorem 1 implies there exist nonzero $\lambda \in \mathbf{R}^n$ and $c \in \mathbf{R}$ such that

$$(\forall x \in \mathbf{R}^n)\left(\lambda^T(b - Ax) \leq c\right) \tag{1}$$

  and

$$\left(\forall y \in K^\circ\right)\left(\lambda^T y \geq c\right). \tag{2}$$

  The former equation (1) implies $\lambda^T A = 0$ and $\lambda^T b \leq c$. and the latter $a \succeq_{K^*} 0$. If $c > 0$, there exists $y \in K^\circ$ such that $\lambda^T y \geq c > 0$. Then $\lambda^T((c/2\lambda^T y)y) =$

$c/2 < c$, but $(c/2\lambda^T y)y \in K^\circ$, hence contradiction. Thus, $c \leq 0$. If $\lambda^T y < 0$ for some $y \in K^\circ$, then $\alpha y \in K^\circ$ for any $\alpha > 0$, thus there exists $z \in K^\circ$ which makes $\lambda^T z$ arbitrarily large toward $-\infty$. Therefore $\lambda^T y$ is nonnegative for every $y \in K^\circ$. Then the latter equation (2) implies $(\forall y \in K^\circ)\left(\lambda^T y \geq 0\right)$, hence $\lambda \in K^*$ (by Definition 26). Therefore we have

$$\lambda \neq 0,\ \lambda \succeq_{K^*} 0,\ A^T\lambda = 0,\ \lambda^T b \leq 0.$$

Conversely, assume that all of above are satisfied. Then for every $x \in \mathbf{R}^n$, there exists nonzero $\lambda \succeq_{K^*} 0$ such that

$$\lambda^T(Ax) \geq \lambda^T b,$$

thus Proposition 3 implies $Ax \not\prec_K b$. ∎

- **Proof 3.** *(Proof for "convexity of infimum of convex function" on page 42)*

Note

$$\begin{aligned}\mathbf{epi}\inf_{y\in C} f(x,y) &= \{(x,t)|(\forall\epsilon>0)(\exists y\in C)(f(x,y)\le t+\epsilon)\}\\ &= \bigcap_{n\in\mathbf{N}}\{(x,t)\,|(\exists y\in C)(f(x,y,t+1/n)\in\mathbf{epi}\,f)\}\\ &= \bigcap_{n\in\mathbf{N}}(\{(x,t)\,|(\exists y\in C)(f(x,y,t)\in\mathbf{epi}\,f)\}-(0,1/n))\end{aligned}$$

where $\{(x,t)\,|(\exists y\in C)(f(x,y,t)\in\mathbf{epi}\,f)\}-(0,1/n)$ for each n since $\mathbf{epi}\,f$ is convex and projection of a convex set is convex. Since the intersection of any collection of convex sets is convex, $\mathbf{epi}\inf_{y\in C} f(x,y)$ is convex, thus $\inf_{y\in C} f(x,y)$ is convex function. ∎

- **Proof 4.** *(Proof for "Lagrange dual is lower bound for optimal value" on page 81)* For every $\lambda\succeq 0$ and $y\in\mathcal{F}$

$$g(\lambda,\nu)\le f(y)+\lambda^T q(y)+\nu^T h(y)\le f(y)\le\inf_{x\in\mathcal{F}} f(x)=p^*.$$

∎

- **Proof 5.** *(Proof for "max-min inequality" on page 121)*
  For every $x \in X, y \in Y$

$$f(x,y) \le \sup_{x' \in X} f(x',y)$$

  hence for every $x \in X$

$$\inf_{y'' \in Y} f(x,y'') \le \inf_{y' \in Y} \sup_{x' \in X} f(x',y')$$

  i.e., $\inf_{y' \in Y} \sup_{x' \in X} f(x',y')$ is upper bound of $\inf_{y'' \in Y} f(x,y'')$, hence

$$\sup_{x \in X} \inf_{y'' \in Y} f(x,y'') \le \inf_{y' \in Y} \sup_{x' \in X} f(x',y')$$

∎

- **Proof 6.** *(Proof for "epigraph of convex optimization is convex" on page 134)*
  Assume $(u_1,v_1,t_1),(u_2,v_2,t_2) \in H$. Then there exist $x_1, x_2 \in \mathcal{D}$ such that $q(x_1) \preceq u_1$, $h(x_1) = v_1$, $f(x_1) \le t_1$, $q(x_2) \preceq u_2$, $h(x_2) = v_2$, and $f(x_2) \le t_2$.

Then for every $0 \leq \theta \leq 1$

$$q(\theta x_1 + (1-\theta)x_2) \preceq \theta q(x_1) + (1-\theta)q(x_2) = \theta u_1 + (1-\theta)u_2$$
$$h(\theta x_1 + (1-\theta)x_2) = \theta h(x_1) + (1-\theta)h(x_2) = \theta v_1 + (1-\theta)v_2$$
$$f(\theta x_1 + (1-\theta)x_2) \preceq \theta f(x_1) + (1-\theta)f(x_2) = \theta t_1 + (1-\theta)t_2$$

thus $\theta(u_1, v_1, t_1) + (1-\theta)(u_2, v_2, t_2) \in H$, hence H is a convex set. ∎

References

References

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.

Index

K-convex functions, 50

a.e.
almost everywhere, 7

a.s.
almost surely, 7

affine dimension, 12

affine hulls, 12

affine sets, 12

algorithms
backtracking line search, 179
backtracking line search for infeasible Newton's method, 217
backtracking line search for primal-dual interior-point method, 241
barrier method, 231
exact line search, 179
exact line search for infeasible Newton's method, 217
feasible Newton's method for equality constrained minimization, 208
gradient descent method, 180
infeasible Newton's method for equality constrained minimization, 216
Newton's method, 183
primal-dual interior-point method, 240

almost everywhere, 7

almost everywhere - a.e., 7

almost surely, 7

backtracking line search, 179

backtracking line search for infeasible Newton's method, 217

backtracking line search for primal-dual interior-point method, 241

barrier method, 231

boundary
    set, 4

closure
    set, 4

complement
    set, 3

complementary slackness, 128

complex number, 3

concave functions, 32

cones, 15

conic programming, 74

conjugate functions, 46

conjugate of conjugate, 46

convergence analysis of Newton's method, 185

convergence analysis of Newton's method for self-concordant functions, 197

convex cone, 15

convex functions, 32

convex hulls, 14

convex optimization, 57

convex optimization with generalized inequality constraints, 73

convex sets, 14

convexity of level sets, 38

convexity preserving function operations, 40

convexity preserving set operations, 20

definitions

- K-convex functions, 50
- affine dimension, 12
- affine hulls, 12
- affine sets, 12
- almost everywhere - a.e., 7
- concave functions, 32
- cones, 15
- conic programming, 74
- conjugate functions, 46
- convex cone, 15
- convex functions, 32
- convex hulls, 14

convex optimization, 57

convex optimization with generalized inequality constraints, 73

convex sets, 14

descent methods, 178

determinant maximization problems, 77

dual cones, 26

dual norms, 27

ellipsoids, 17

epigraphs, 39

equivalent optimization problems, 55

Euclidean ball, 17

extended real-value extension of convex functions, 33

generalized inequalities, 23

geometric programming, 71

global optimality, 54

half spaces, 16

hyperplanes, 16
hypographs, 39
infinitely often - i.o., 7
iterative meethods, 178
iterative meethods with search directions, 178
KKT optimality conditions, 129
KKT optimality conditions for generalized inequalities, 163
Lagrange dual functions, 81
Lagrange dual functions for generalized inequalities, 158
Lagrange dual problems, 89
Lagrange dual problems for generalized inequalities, 159
Lagrangian, 80
Lagrangian for generalized inequalities, 157
line search method, 179
line segmenets, 11
linear programming, 62
lines, 11
local optimality, 54
matrix convexity, 51
monomial functions, 71
Newton decrement, 195
norm ball, 18
norm cone, 18
optimal duality gap, 99
optimization problems, 53
optimization problems with generalized inequalities, 156
polyhedra, 19
posynomial functions, 71
proper cones, 23
quadratic programming, 66
quadratically constrained quadratic programming, 68

relative boundaries of sets, 13
relative interiors of sets, 13
saddle-points, 122
second-order cone, 18
second-order cone programming, 69
self-concordance, 187
semidefinite programming, 75
separating hyperplanes, 25
strong duality, 100
strong max-min property, 121
sublevel sets, 38
superlevel sets, 38
supporting hyperplanes, 25
weak duality, 98

descent methods, 178

determinant maximization problems, 77

difference
set, 4

dual characterization of K-convexity, 50

dual cones, 26

dual norms, 27

ellipsoids, 17

epigraphs, 39

equivalent optimization problems, 55

Euclidean ball, 17

exact line search, 179

exact line search for infeasible Newton's method, 217

extended real-value extension of convex functions, 33

Farkas' lemma, 154

Farkas, Julius
    Farkas' lemma, 154

feasible Newton's method for equality constrained minimization, 208

Fenchel's inequality, 46

Fenchel, Moritz Werner
    Fenchel's inequality, 46

finite sequence, 3

first-order condition for convexity, 34

generalized inequalities, 23

geometric programming, 71

geometric programming in convex form, 72

global optimality, 54

gradient descent method, 180

graphs and convexity, 39

half spaces, 16

hyperplanes, 16

hypographs, 39

i.o.
    infinitely often, 7

inequalities
    Fenchel's inequality, 46

max-min inequality, 121

infeasible Newton's method for equality constrained minimization, 216

infinite sequence, 3

infinitely often, 7

infinitely often - i.o., 7

integer, 3

interior

set, 4

iterative meethods, 178

iterative meethods with search directions, 178

KKT and convexity sufficient for optimality with strong duality, 132

KKT necessary for optimality with strong duality, 130

KKT optimality conditions, 129

KKT optimality conditions for generalized inequalities, 163

Lagrange dual functions, 81

Lagrange dual functions for generalized inequalities, 158

Lagrange dual problems, 89

Lagrange dual problems for generalized inequalities, 159

Lagrange, Joseph-Louis

Lagrange dual functions, 81

Lagrange dual functions for generalized inequalities, 158

Lagrange dual problems, 89
Lagrange dual problems for generalized inequalities, 159
Lagrangian, 80
Lagrangian for generalized inequalities, 157
Lagrangian, 80
Lagrangian for generalized inequalities, 157
Lagrangian for generalized inequalities, 157
line search method, 179
line segmenets, 11
linear programming, 62
lines, 11
local optimality, 54
local optimality implies global optimality, 58
matrix
positive definite, 5
positive semi-definite, 5
symmetric, 5
trace, 4
matrix convexity, 51
max-min inequality, 121
monomial functions, 71
natural number, 3
Newton decrement, 182, 195
for equality constrained problem, 207
Newton's method, 183
Newton, Isaac
Newton decrement, 182, 195

for equality constrained problem, 207
Newton's method, 183

norm
vector, 4

norm ball, 18

norm cone, 18

number
complex number, 3
integer, 3
natural number, 3
rational number, 3
real number, 3

optimal duality gap, 99

optimality certificate for self-concordant functions, 195

optimality conditions for convex optimality problems, 58

optimization problems, 53

optimization problems with generalized inequalities, 156

polyhedra, 19

positive definite matrix, 5

positive semi-definite matrix, 5

posynomial functions, 71

primal-dual interior-point method, 240

proper cones, 23

properties of dual cones, 28

propositions

complementary slackness, 128
conjugate of conjugate, 46
convexity of level sets, 38
convexity preserving function operations, 40
convexity preserving set operations, 20
dual characterization of K-convexity, 50
geometric programming in convex form, 72
graphs and convexity, 39
properties of dual cones, 28
self-concordance for logarithms, 187
self-concordance preserving operations, 189

quadratic programming, 66

quadratically constrained quadratic programming, 68

rational number, 3

real number, 3

relative boundaries of sets, 13

relative interior
set, 4

relative interiors of sets, 13

saddle-points, 122

second-order condition for convexity, 35

second-order cone, 18

second-order cone programming, 69

self-concordance, 187

self-concordance for logarithms, 187

self-concordance preserving operations, 189

semidefinite programming, 75

separating hyperplane theorem, 25

separating hyperplanes, 25

sequence, 3
- finite sequence, 3
- infinite sequence, 3

set
- boundary, 4
- closure, 4
- complement, 3
- difference, 4
- interior, 4
- relative interior, 4

Slater's theorem, 101

Slater's theorem for generalized inequalities, 160

smallest σ-algebra containing subsets, 4

strong alternatives for generalized inequalities, 168

strong alternatives of two systems, 150

strong alternatives of two systems with strict inequalities, 151

strong duality, 100

strong max-min property, 121

sublevel sets, 38

superlevel sets, 38

supporting hyperplane theorem, 25

supporting hyperplanes, 25

symmetric matrix, 5

theorem of alternative for linear strict generalized inequalities, 30

theorems
- convergence analysis of Newton's method, 185
- convergence analysis of Newton's method for self-concordant functions, 197
- Farkas' lemma, 154
- first-order condition for convexity, 34
- KKT and convexity sufficient for optimality with strong duality, 132
- KKT necessary for optimality with strong duality, 130
- local optimality implies global optimality, 58
- optimality certificate for self-concordant functions, 195
- optimality conditions for convex optimality problems, 58
- second-order condition for convexity, 35
- separating hyperplane theorem, 25
- Slater's theorem, 101
- Slater's theorem for generalized inequalities, 160

strong alternatives for generalized inequalities, 168

strong alternatives of two systems, 150

strong alternatives of two systems with strict inequalities, 151

supporting hyperplane theorem, 25

theorem of alternative for linear strict generalized inequalities, 30

weak alternatives for generalized inequalities, 167

weak alternatives of two systems, 147

weak alternatives of two systems with strict inequalities, 149

trace
matrix, 4

vector
norm, 4

weak alternatives for generalized inequalities, 167

weak alternatives of two systems, 147

weak alternatives of two systems with strict inequalities, 149

weak duality, 98

ZZ-figures
diagrams for containment of convex optimization problems, 78
dual cone, 26
geometric interpretation of duality - 1, 113
geometric interpretation of duality - 2, 114
geometric interpretation of duality - 3, 116
geometric interpretation of duality - 4, 119
sensitivity analysis of optimal value, 135

ZZ-todo
0 - apply new comma conventions, 0
1 - convert bullet points to proper theorem, definition, lemma, corollary, proposition, etc., 0
CANCELED - < 2024 0421 - python script extracting important list, 0
DONE - 2024 0324 - change tocpageref and funpageref to hyperlink, 0
DONE - 2024 0324 - python script extracting figure list → using "list of figures" functionality on doc, 0
DONE - 2024 0324 - python script extracting theorem-like list → using "list of theorem" functionality on doc, 0

DONE - 2024 0324 - python script for converting slides to doc, 0

DONE - 2025 0414 - 1 - change mathematicians' names, 0