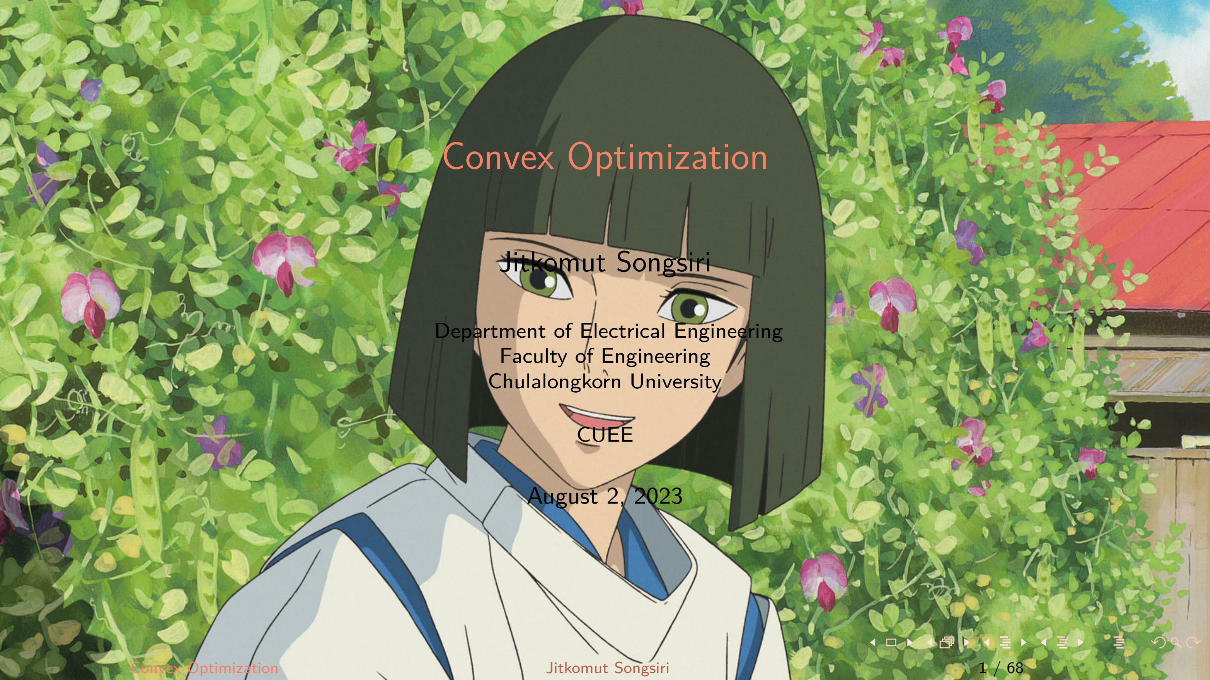


Convex Optimization

Jitkomut Songsiri

Department of Electrical Engineering
Faculty of Engineering
Chulalongkorn University

CUEE

August 2, 2023

Outline

1. General setting
2. Convex optimization
3. Problem Transformation
4. Structured convex problems
5. Optimality conditions
6. Numerical methods

General setting

Problem setting

(mathematical) optimization problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array} \qquad \text{(P1)}$$

- $x = (x_1, \ldots, x_n)$: optimization variable
- $f_0 : \mathbf{R}^n \to \mathbf{R}$: objective function
- $f_i : \mathbf{R}^n \to \mathbf{R}, i = 1, \ldots, m$: inequality constraint functions
- $h_i : \mathbf{R}^n \to \mathbf{R}, i = 1, \ldots, p$: equality constraint functions

constraint set: $\mathcal{C} = \{x \in \mathbf{R}^n \mid f_i(x) \leq 0, i = 1, \ldots, m, h_i(x) = 0, i = 1, \ldots, p\}$

domain of the problem: $\mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom}\, f_i \cap \bigcap_{i=1}^{p} \mathbf{dom}\, h_i$

Optimal value

$$p^\star = \inf \{ f_0(x) \mid f_i(x) \leq 0,\ i = 1, \ldots, m,\ h_i(x) = 0,\ , i = 1, \ldots, p \}$$

- we say x is **feasible** if $x \in \mathbf{dom}\, f_0(x)$ and $x \in \mathcal{C}$
- $p^\star = \infty$ if the problem is **infeasible**
- $p^\star = -\infty$ if the problem is unbounded below
- a feasible x is called **optimal** if $f_0(x) = p^\star$; there can be many
- x is **locally optimal** if $\exists \epsilon > 0$ such that x is optimal for

$$\begin{array}{ll} \text{minimize} & f_0(z) \\ \text{subject to} & z \in \mathcal{C}, \quad \|z - x\|_2 \leq \epsilon \end{array}$$

in other words, a locally optimal point is the best solution in a neighborhood

Terminology

some equivalent definition/setting

- setting: another way of representing (P1)

$$\text{minimize} \quad f_0(x) \quad \text{subject to} \quad x \in \mathcal{C} \qquad \text{(P2)}$$

- optimal point: we can also say $x^\star$ is a **global minimizer** of f_0 over $\mathcal{C}$

$$f_0(x) \geq f_0(x^\star) \quad \forall x \in \mathcal{C}$$

- local optimal point: we can also say $x^\star$ is a **local minimizer** of f_0 over $\mathcal{C}$

$$\exists \epsilon > 0 \quad \text{such that} \quad f_0(x) \geq f_0(x^\star) \quad \forall x \in \mathcal{C} \cap \|x - x^\star\| < \epsilon$$

- the standard form has an **implicit constraint**: $x \in \mathcal{D}$
- the constraint set $\mathcal{C}$ contains **explicit constraints**
- the problem is called **unconstrained** if it has no explicit constraints

Feasibility problem

a feasibility problem

$$\text{find} \quad x \quad \text{subject to} \quad x \in \mathcal{C}$$

can be considered as a special case of the general problem with $f_0(x) = 0$

$$\text{minimize} \quad 0 \quad \text{subject to} \quad x \in \mathcal{C}$$

- $p^\star = 0$ if constraints are feasible; any feasible x is optimal
- $p^\star = \infty$ if constraints are infeasible

examples: $\mathcal{C}_1$ has two-, $\mathcal{C}_2$ has infinitely many feasible points, while $\mathcal{C}_3$ is infeasible

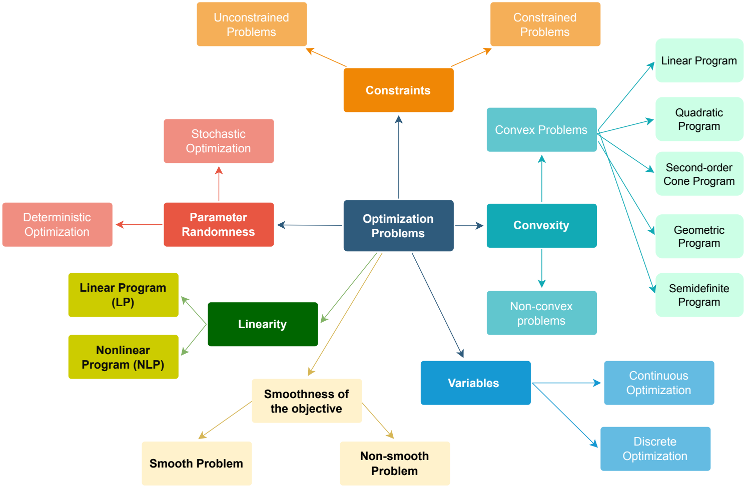
$$\begin{aligned}
\mathcal{C}_1 &= \{x \in \mathbf{R}^2 \mid (x_1 - 1)^2 + x_2^2 = 1, x_1 + x_2 = 1\} \\
\mathcal{C}_2 &= \{x \in \mathbf{R}^2 \mid (x_1 - 1)^2 + x_2^2 \leq 1, x_1 + x_2 = 1\} \\
\mathcal{C}_3 &= \{x \in \mathbf{R}^2 \mid (x_1 - 1)^2 + x_2^2 \leq 1, x_1 + x_2 = -3\}
\end{aligned}$$

Problem types

we can categorize optimization problems by

- **constraints**
  - unconstrained problem
  - constrained problems
- **variable types**
  - continuous optimization
  - discrete optimization
- **linearity of objective and constraints**
  - linear program
  - nonlinear program
- **convexity of objective and constraint set**
  - convex problem
  - non-convex problem
- **smoothness of the objective**
  - smooth problem
  - non-smooth problem
- **parameter randomness**
  - stochastic optimization
  - deterministic optimization

this course focuses on continuous and deterministic optimization



other specific problem types are integer programming, vector optimization.

Optimality of unconstrained problems

assumption: f is twice continuously differentiable (smooth objective)

- **1st-order necessary condition:**

$$\text{if } x^\star \text{ is a local minimizer of } f \text{ then } \nabla f(x^\star) = 0$$

- **2nd-order necessary condition:** if $x^\star$ is a local minimizer of f then $\nabla f(x^\star) = 0$ and $\nabla^2 f(x^\star) \succeq 0$ (positive semidefinite)
- **2nd-order sufficient condition:** if $\nabla f(x^\star) = 0$ and $\nabla^2 f(x^\star) \succ 0$ (pdf)

$$\text{then } x^\star \text{ is a strict local minimizer of } f$$

local minimizers can be distinguished from other stationary points by examining positive definiteness of $\nabla^2 f$

example: $f(x) = x^4$ has $x^\star = 0$ as a local minimizer; $\nabla^2 f(x^\star) = 0$ (hence, 2nd-order sufficient condition fails)

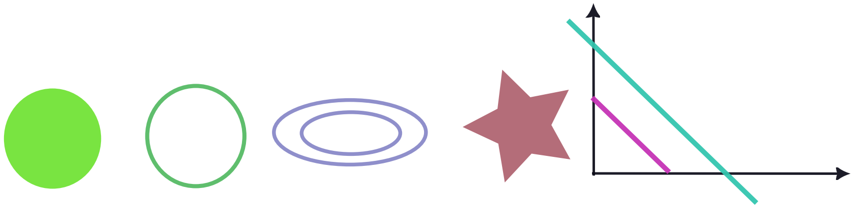
Convex optimization

Convex sets

a set $\mathcal{C}$ is said to be convex if for any $x, y \in \mathcal{C}$ we have

$$\theta x + (1-\theta)y \in \mathcal{C}, \quad \text{for all } 0 \leq \theta \leq 1$$

which of the following sets are convex ?



fact: an intersection of convex sets is convex (even infinitely many number of intersections)

Convex functions

convex function: $f : \mathbf{R}^n \to \mathbf{R}$ is convex if

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$$

for all x, y in the domain of f and $0 \leq \theta \leq 1$

loosely speaking, f is convex if it has an upward shape

examples on $\mathbf{R}$:

- affine: $ax + b$ for any $a, b \in \mathbf{R}$
- exponential: e^{ax} for any $a \in \mathbf{R}$
- powers of absolute value: $|x|^p$ for $p \geq 1$
- negative entropy: $x \log x$ on $\mathbf{R}_{++}$

Examples of convex functions on $\mathbf{R}^n$

- affine: $a^T x + b$
- norm functions: $\|x\|$
- norms of affine: $\|a^T x + b\|$
- quadratic: $x^T P x + q^T x$ when $P \succeq 0$
- negative entropy: $\sum_{i=1}^n x_i \log x_i$ on $\mathbf{R}^n_{++}$

fact: a set of inequality constraints described by convex functions is convex

$$\mathcal{C} = \{x \in \mathbf{R}^n \mid f_i(x) \leq 0, \; i = 1, 2, \ldots, m \}$$

is a convex set if all f_i's are convex functions

First- and second-order conditions of convex functions

suppose f is differentiable; then f is convex if and only if

$$\mathbf{dom}\, f \;\text{ is convex and }\; f(y) \geq f(x) + \nabla f(x)^T(y-x), \quad \forall x, y \in \mathbf{dom}\, f$$

- the first-order Taylor approximation of f is a **global underestimator** of f if and only if f is convex
- if $\nabla f(x) = 0$ then for all $y \in \mathbf{dom}\, f$, $f(y) \geq f(x)$,*i.e.*, x is a global minimizer of f

assume that $\nabla^2 f$ exists at each point in $\mathbf{dom}\, f$; then f is convex if and only if

$$\mathbf{dom}\, f \;\text{ is convex and }\; \nabla^2 f(x) \succeq 0, \;\; \forall x \in \mathbf{dom}\, f$$

f is convex if and only if its Hessian matrix is positive semidefinite

Convex programs

convex optimization problem is one of the form

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\
 & a_i^T x = b_i, \quad i = 1, \ldots, p
\end{array}
$$

where

- objective and constraint functions are **convex**
- equality constraint functions $h_i(x) = a_i^T x - b_i$ must be **affine**

result: an optimal solution of a convex program is a **global** minimizer

Properties of convex problems

convex problems are of interest due to some desirable properties

- many operations preserve convexity of a convex set
  - intersection
  - image (and inverse image) of affine mapping
  - image (and inverse image) of perspective mapping
- many operations preserve convexity of a convex function
  - non-negative weighted sum
  - composition with affine mapping, composition rules
  - pointwise maximum and supremum, minimization over one variable
  - perspective of a function
  - conjugate function (important role in duality theory)
- KKT conditions are *sufficient* and *necessary* for optimality

many optimization problems in engineering are convex programs

Linear program (LP)

a general linear program has the form

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

where $G \in \mathbf{R}^{m \times n}$ and $A \in \mathbf{R}^{p \times n}$

example: minimize the cheapest diet that satisfies the nutritional requiremenets

- $x = (x_1, \ldots, x_n)$ is nonnegative quantity of n different foods
- each food has a cost of c_j; cost objective is $c^T x$
- one unit quantity of food j contains d_{ij} amount of nutrients i
- constraints are $Dx \succeq h$ and $x \succeq 0$

Geometrical interpretation

- hyperplane: solution set of a linear equation with coefficient vector $a \neq 0$

$$\{x \mid a^T x = b\}$$

- halfspace: solution set of a linear inequality with coefficient vector $a \neq 0$

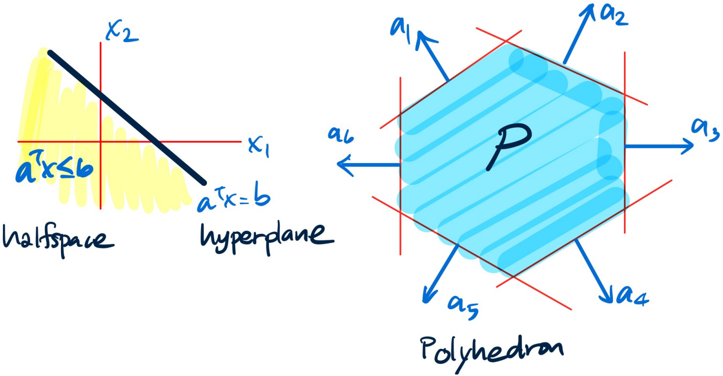
$$\{x \mid a^T x \leq b\}$$

  we say a is the **normal vector**

- polyhedron: solution set of a finite number of linear inequalities

$$\{x \mid a_1^T x \leq b_1, \;\; a_2^T x \leq b_2, \;\; \ldots, \;\; a_m^T x \leq b_m\} = \{x \mid Ax \leq b\}$$

  intersection of a finite number of halfspaces



extreme point of $\mathcal{C}$

a vector $x \in \mathcal{C}$ is an extreme point (or a vertex) if we cannot find $y, z \in \mathcal{C}$ both different from x and a scalar $\alpha \in [0, 1]$ such that $x = \alpha y + (1 - \alpha)z$

Properties of LP

- another standard form: minimize $c^T x$ subject to $Ax = b$, $x \succeq 0$
- an LP may not have a solution (constraints are inconsistent or the feasible set is unbounded)
- we assume A is full row rank; if not, considering $Ax = b$
  - depending on A, the system could be inconsistent (hence, no extreme points), or
  - $Ax = b$ contains redundant equations, which can be removed
- if a standard LP has a finite optimal solution then

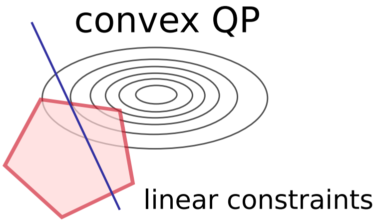
  a solution can always be chosen from among the vertices of the feasible set

  (called **basic feasible solutions**)
- the dual of an LP is also an LP
- solutions of some simple LPs can be analytically inspected

Quadratic program (QP)

a **quadratic program (QP)** is in the form

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



where $P \in \mathbf{S}^n, G \in \mathbf{R}^{m \times n}$ and $A \in \mathbf{R}^{p \times n}$

example: constrained least-squares

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|_2^2 \\ \text{subject to} & l \preceq x \preceq u \end{array}$$

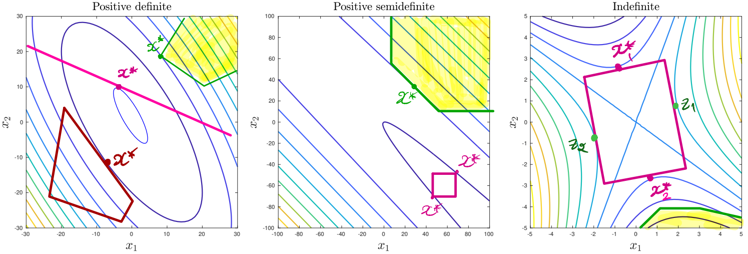
QP has **linear** constraints

Properties of QP

- an unconstrained QP is unbounded below if P is not positive definite
- an unconstrained QP has a unique solution: $x = -P^{-1}q$ when $P \succ 0$
- a QP is a convex problem if P is positive semidifinite definite
  - if $P \succeq 0$ then a local minimizer $x^\star$ is a global minimizer (by convexity)
  - if $P \succ 0$ then $x^\star$ is a *unique* global solution (by strictly convexity)
- the feasible set (polyhedron) may be empty (hence, the problem is infeasible)
- the feasible set can be unbounded (but if $P \succ 0$ it implies boundedness)
- solution of a QP may not be at a vertex
- the dual of a QP is also a QP

Contour of quadratic objective

consider three cases of P and different feasible sets



verify the location of the optimal solution for each constraint set

- left: a bounded set, a line, an unbounded feasible set
- middle: bounded and unbouded feasible sets, while f is unbounded below
- right: a bounded feasible set, while f is unbounded below and above

QCQP

a **quadratically constrained quadratic program (QCQP)** is in the form

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

where P_i's are positive semidefinite, $G \in \mathbf{R}^{m \times n}$ and $A \in \mathbf{R}^{p \times n}$

QCQP has both **linear and quadratic** constraints

Global optimum

consider a convex optimization problem: f is convex and $\mathcal{C}$ is a convex set

$$\underset{x}{\text{minimize}} \quad f(x) \quad \text{subject to} \quad x \in \mathcal{C} \tag{1}$$

Theorem: (Ghaoui book, section 8.3.1)

- any locally optimal solution is also globally optimal
- the set $\mathcal{X}_{\text{opt}}$ of optimal points is convex

Proof

$$f_0(x) = \inf\{ f_0(z) \mid z \text{ is feasible}, \ \ \|z - x\|_2 \leq R \} \quad \text{for some } R > 0$$

Proof of global minimum

proof: let $x^\star$ be a local minimizer and $p^\star = f(x^\star)$

- for any $y \in \mathcal{C}$ then we can write $z \in \mathcal{C}$ as a convex sum: $z = \theta y + (1-\theta)x^\star$
- by the convexity of f

$$f(z) \leq \theta f(y) + (1-\theta) f(x^\star) \quad \Rightarrow \quad f(z) - f(x^\star) \leq \theta(f(y) - f(x^\star))$$

- since $x^\star$ is a local minimizer, LHS is non-negative if θ is small enough, then RHS is also non-negative
- we obtain $f(z) - f(x^\star) \geq 0$ for any $z \in \mathcal{C}$ – $x^\star$ is also global optimal
- the optimal set can be written as the $p^\star$-sublevel set

$$\mathcal{X}_{\rm opt} = \{x \in \mathcal{C} \mid f(x) \leq p^\star \}$$

- since a sublevel set of a convex function is convex, and f is convex, we have $\mathcal{X}_{\rm opt}$ is convex as claimed

Existence of solutions

Weierstrass extreme value theorem:

every continuous function $f : \mathbf{R}^n \to \mathbf{R}$ on a non-empty compact (closed and bounded) set attains its extreme values on that set

Theorem: sufficient condition for the existence

if $\mathcal{C} \subseteq \mathbf{dom}\, f$ is nonempty and compact and f is continuous on $\mathcal{C}$ then the problem (1) attains an optimal solution $x^\star$

note that this theorem is not applicable to an unconstrained convex problem (because $\mathcal{C} = \mathbf{R}^n$ which is not compact)

Coercive function

Definition: coercive functions

a function $f : \mathbf{R}^n \to \mathbf{R}$ is said to be **coercive** if for any sequence $\{x_k\} \subset \int \mathbf{dom}\, f$ tending to the boundary of $\mathbf{dom}\, f$, it holds that the function value sequence $\{f(x_k)\}$ tends to $+\infty$ [1]

Lemma:

a continuous function with open domain is coercive *if and only if* all its sublevel sets $S_\alpha = \{x \mid f(x) \leq \alpha\}, \alpha \in \mathbf{R}$ are **compact**

[1] Ghaoui book, section 8.3.2

Existence of solutions (Coercive function)

Lemma: unconstrained optimization

if $\mathcal{C} = \mathbf{R}^n$ and f is continuous and **coercive**, then the convex optimization (1) attains an optimal solution $x^\star$

proof: take α that S_α is non-empty and follows the Weierstrass theorem

Lemma: constrained optimization

if $\mathcal{C} \subseteq \mathbf{dom}\, f$ is non-empty and closed, and f is continuous on $\mathcal{C}$ and **coercive**, then the convex problem (1) attains an optimal solution $x^\star$

Strictly convex function

a function f is said to be **strictly convex** if

$$f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta)f(y)$$

for all $x \neq y$ in the domain of f and $0 \leq \theta \leq 1$

- a strictly convex f satisfies the convexity condition with strict inequality
- $f(x) = a^T x + b$ is convex but not strictly convex
- intuitively, a convex function that has a 'flat' area is not strictly convex
- what about ϵ-insensitive loss function in SVR ?

Strongly convex

a function $f : \mathbf{R}^n \to \mathbf{R}$ is said to be **strongly convex** on S if

$$\exists m > 0 \quad \text{such that} \quad f(x) - \frac{m}{2}\|x\|_2^2$$

is **convex** on S

related definition: if f is twice differentiable and

$$\nabla^2 f(x) \succeq mI, \quad \text{for all } x \in S$$

then f is said to be **strongly convex**

- example: $f(x) = x^T P x$ with $P \succ 0$
- a linear function $f(x) = a^T x + b$ is not strongly convex
- fact: ✎ a sum of convex and strongly convex functions is strongly convex

Strong convexity implies strict convexity

by convexity of $f(x) - \frac{m}{2}\|x\|^2$, that is

$$f(\theta x + (1-\theta)y) - \frac{m}{2}\|\theta x + (1-\theta)y\|_2^2 \le \theta f(x) + (1-\theta)f(y) - \frac{\theta m\|x\|^2}{2} - \frac{(1-\theta)m\|y\|^2}{2}$$

move the squared norm to the RHS and simplify

$$\begin{aligned} f(\theta x + (1-\theta)y) &\le \theta f(x) + (1-\theta)f(y) - \frac{m}{2}\theta(1-\theta)[\|x\|^2 - 2x^T y + \|y\|^2] \\ &\le \theta f(x) + (1-\theta)f(y) - \underbrace{\frac{m}{2}\theta(1-\theta)\|x-y\|_2^2}_{>0 \text{ for all } x \neq y} \end{aligned}$$

clearly, strong convexity implies strict convexity

Uniqueness of the optimal solution

Theorem:

if f is **strictly convex** in the problem (1), and $x^\star$ is an optimal solution, then $x^\star$ is the **unique** optimal solution

proof: let's prove by contradiction; let $x^\star$ be an optimal point and there exists another $y^\star \neq x^\star$ that is also optimal

- both $x^\star$ and $y^\star$ are feasible and $f(x^\star) = f(y^\star) = p^\star$
- let $\theta \in (0,1)$ and let $z = \theta x^\star + (1-\theta)y^\star$
- by convexity of $\mathcal{C}$, z must be also feasible
- by strict convexity of f,

$$f(z) < \theta f(x^\star) + (1-\theta)f(y^\star) = p^\star \quad \Rightarrow \quad z \text{ achieves a lower function value}$$

- this contradicts to the assumption that $x^\star$ is globally optimal

Strict convexity by regularization

let's add a quadratic term to a convex objective function

$$\tilde{f}(x) = f(x) + \gamma \|x - c\|_2^2$$

- clearly, $\|x - c\|_2^2$ is strongly convex
- a sum of convex and strongly convex is strongly convex
- hence, $\tilde{f}$ is strongly convex and also strictly convex
- minimizing $\tilde{f}$ over a convex set attains a unique optimal solution

Implications of strong convexity

obtain a **quadratic lower bound** on f (which is better)

$$f(y) = f(x) + \nabla f(x)^T (y-x) + \frac{1}{2}(y-x)^T \nabla^2 f(z)(y-x), \quad \text{for } z \in [x,y]$$

$$f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{m}{2}\|y-x\|_2^2, \quad \text{for } x, y \in S$$

- when $m = 0$, it reduces to the first-order condition for convexity
- strong convexity provides a higher lower bound than from convexity alone

Implications of strong convexity

obtain a **quadratic upper bound** for f on S

to see this, let $x \in \mathbf{dom}\, f$ and $y \in S = \{y | f(y) \leq f(x)\}$ (x is a fixed point)

$$y \in S \;\Rightarrow\; 0 \geq f(y) - f(x) \geq \underbrace{\nabla f(x)^T(y-x) + \frac{m}{2}\|y-x\|_2^2}_{A}$$

- a set of y such that $A \leq 0$ is the region inside a **bounded** ellipsoid
- then, the sublevel set S is contained in a bounded ellipsoid, so S is bounded
- when $\nabla^2 f$ is assumed to be continuous, it is bounded on a bounded set
- there exists $M > 0$ such that $\nabla^2 f(y) \preceq MI$ for all $y \in S$

$$f(y) \leq f(x) + \nabla f(x)^T(y-x) + \frac{M}{2}\|y-x\|_2^2, \quad \forall x, y \in S$$

Bounds on the optimality gap

for a strongly convex f and twice differentiable, it holds that

$$mI \preceq \nabla^2 f(x) \preceq MI, \quad \forall x \in S$$

and two inequalities for any points $x, y \in S$

$$\begin{aligned} f(y) &\geq f(x) + \nabla f(x)^T (y - x) + \frac{m}{2} \|y - x\|_2^2 \\ f(y) &\leq f(x) + \nabla f(x)^T (y - x) + \frac{M}{2} \|y - x\|_2^2 \end{aligned}$$

Problem transformation

Equivalent convex problems

two problems are (informally) **equivalent** if the solution of one can be obtained from the solution of the other, and vice versa

examples: P1 and P2 are equivalent (but they are not the same)

minimize	$\|Ax - y\|_2$	(P1)	minimize	$\|Ax - y\|_2^2$	(P2)
maximize	$\frac{1}{\|Ax-y\|_2}$	(P1)	minimize	$\|Ax - y\|_2^2$	(P2)
maximize	$\lvert f(x)\rvert$	(P1)	maximize	$\log \lvert f(x)\rvert$	(P2)

using monotonically increasing property of squared and log functions

Transformation that yield equivalent problems

some transformations are useful for problem re-formulation

- eliminating equality constraints
- introducing slack variables
- epigraph form
- minimizing over some variables
- using indicator function to represent constraints

Eliminating equality constraints

the problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

is equivalent to

$$\begin{array}{ll} \text{minimize} & f_0(Fz + x_0) \\ \text{subject to} & f_i(Fz + x_0) \leq 0, \quad i = 1, \ldots, m \end{array}$$

where F and x_0 are such that

$$Ax = b \quad \Longleftrightarrow \quad x = Fz + x_0 \text{ for some } x_0$$

Example: eliminating equality constraints

equality constraint in the form of $Ax = b$ (non-trivial when A is fat)

$$\begin{array}{lll} \text{minimize} & \|Hx - y\|_2 & \text{(P1)} \\ \text{subject to} & x_1 + x_2 = 0 & \end{array} \qquad \begin{array}{lll} \text{minimize} & \|\tilde{H}x - y\|_2 & \text{(P2)} \\ \text{where} & \tilde{H} = \begin{bmatrix} h_1 - h_2 & h_3 & \cdots & h_n \end{bmatrix} & \end{array}$$

- find the nullspace of A and its basis vectors

$$\dim \mathcal{N}(A) = r \quad \Leftrightarrow \quad \exists F \in \mathbf{R}^{n \times r} \text{ such that } AF = 0 \text{ and } F \text{ is full column rank}$$

- find a particular solution of $Ax = b$, says x_0
- a general solutions to $Ax = b$ is expressed as $x = Fz + x_0$ for any z

Introducing slack variables

the problem

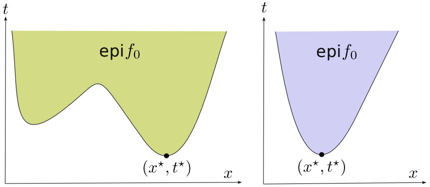
$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \dots, m \end{array}$$

is equivalent to

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) + s_i = 0, \quad i = 1, \dots, m \\ & s_i \geq 0, \quad i = 1, 2, \dots, m \end{array}$$

Epigraph form

the epigraph of a function f_0 is the area above the graph f_0



epi f_0 $= \{(x,t); | \; x \in \mathbf{dom} f_0, \;\; f_0(x) \leq t\,\}$

the standard problem is equivalent to

$$\begin{array}{ll} \text{minimize (over } x, t) & t \\ \text{subject to} & f_0(x) - t \leq 0, \\ & f_i(x) \leq 0, \quad i = 1, \dots, m \\ & Ax = b \end{array}$$

we minimize t over the epigraph of f_0 (objective is now linear of (x, t))

Example: epigraph form

example 1: $\|z\|_\infty \leq t$ if and only if $|z_i| \leq t$ for all i

$$\text{minimize}_x \quad \|Ax - y\|_\infty \quad \text{(P1)}$$

$$\begin{array}{lll} \text{minimize}_{(x,t)} & t & \text{(P2)} \\ \text{subject to} & -t \leq a_i^T x - y_i \leq t & , i = 1, \ldots, m \end{array}$$

example 2: for a symmetric F, $\|F\|_2 \leq t$ if and only if $-tI \preceq F \preceq tI$

given symmetric matrices F_i for $i = 0, 1, \ldots, n$

$$\text{minimize}_x \quad \|F_0 + x_1 F_1 + x_2 F_2 + \cdots + x_n F_n\|_2 \quad \text{(P1)}$$

$$\begin{array}{lll} \text{minimize}_{(x,t)} & t & \text{(P2)} \\ \text{subject to} & -tI \preceq F_0 + \sum_{i=1}^n x_i F_i \preceq tI & \end{array}$$

Minimizing over some variables

the problem

$$\begin{array}{ll} \text{minimize} & f_0(x_1, x_2) \\ \text{subject to} & f_i(x_1) \leq 0, \quad i = 1, \dots, m \end{array}$$

is equivalent to

$$\begin{array}{ll} \text{minimize} & \tilde{f}_0(x_1) \\ \text{subject to} & f_i(x_1) \leq 0, \quad i = 1, \dots, m \end{array}$$

where $\tilde{f}_0(x_1) = \inf_{x_2} f_0(x_1, x_2)$

if the objective can be minimized over one variable easily, we can reduce the problem dimension

Example: minimizing over one variable

given $g_i : \mathbf{R}^n \to \mathbf{R}, y_i \in \mathbf{R}$ for $i = 1, \ldots, N$, consider the problem

$$\underset{x,d}{\text{minimize}} \quad -N \log \left[\frac{1}{d}\right] + \frac{1}{d} \sum_{i=1}^{N} (g_i(x) - y_i)^2$$

first, we can minimize over d by setting the gradient w.r.t. $1/d$ to zero

$$d = \frac{1}{N} \sum_{i=1}^{N} (g_i(x) - y_i)^2$$

the reduced problem is

$$\underset{x}{\text{minimize}} \quad \log \left[\frac{1}{N} \sum_{i=1}^{N} (g_i(x) - y_i)^2\right] \quad \Longleftrightarrow \quad \underset{x}{\text{minimize}} \quad \sum_{i=1}^{N} (g_i(x) - y_i)^2$$

Constraints expressed as indicator functions

introduce the **indicator function** associated with a set $\mathcal{C}$

$$I_{\mathcal{C}}(x) = \begin{cases} 0, & x \in \mathcal{C} \\ +\infty, & x \notin \mathcal{C} \end{cases}$$

the minimization of $f_0(x)$ subject to $x \in \mathcal{C}$ is equivalent to

$$\underset{x}{\text{minimize}} \quad f_0(x) + I_{\mathcal{C}}(x)$$

note that $I_{\mathcal{C}} : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$ is extended-value function

we express the original constrained problem as an unconstrained problem usign $I_{\mathcal{C}}$

Structured convex problems

some structures that are amenable for parallel and distributed algorithms

- **separable sum**

$$\underset{x_1,\ldots,x_m}{\text{minimize}} \quad f(x) := \sum_{i=1}^{m} f_i(x_i)$$

  it is obvious that we can minimize over x_i independently

- **global consensus**

$$\underset{x}{\text{minimize}} \quad f(x) := \sum_{i=1}^{m} f_i(x)$$

  f_i is a local objective; x is the global variable

  consensus form: add a consensus constraint that makes all local x_i's agree

$$\underset{x_1,\ldots,x_m}{\text{minimize}} \quad \sum_{i=1}^{m} f_i(x) \quad \text{subject to} \quad x_1 = x_2 = \cdots = x_m$$

Structured convex problems

Structured convex problems

- **global exchange**

$$\underset{x}{\text{minimize}} \quad \sum_{i=1}^{m} f_i(x_i) \quad \text{subject to} \quad \sum_{i=1}^{m} x_i = 0$$

interpretation: x_i's are quantities of commodities exchanged among m agents

goal: minimize total social cost subject to the market clearing

- **allocation**

$$\underset{x}{\text{minimize}} \quad \sum_{i=1}^{m} f_i(x_i) \quad \text{subject to} \quad x_i \geq 0, \ \sum_{i=1}^{m} x_i = b$$

interpretation: x_i's are non-negative resources allocated to m activities

goal: minimize each activity cost while the total resource is limited to a budget

Distributed model fitting

a problem of fitting y using a linear model Ax using a loss function l

$$\underset{x}{\text{minimize}} \quad l(Ax - y) + r(x)$$

$l(Ax - y) = \sum_{i=1}^{N} l_i(a_i^T x - y_i)$ represents the model cost due to error $Ax - y$

r is a separable function representing **regularization**, *e.g.*, $\|\cdot\|_1, \|\cdot\|_2^2$

this is an example of global consensus

a common model parameter x that makes the model fits with *all* data samples

Nonsmooth optimization

a function is smooth if it is differentiable and the derivatives are continuous

- example: $f(x) = |x|$ is not smooth at $x = 0$
- example: $f(x) = \|x\|$ is not smooth at $x = 0$

a problem is called **nonsmooth** if the objective or constraints are nonsmooth functions

example: lasso problems

$$\text{minimize} \quad \|Ax - b\|_2 + \gamma\|x\|_1$$

then the methods relying on the gradient should be carefully revisited

Scalarized multi-objective optimization

a common form of multi-objective problem: for a given $\gamma > 0$,

$$\text{minimize} \quad f(x) + \gamma g(x)$$

- we desire both f and g to be small but they are weighed in by a given weight, γ (or often called penalty parameter)
- as γ is higher, we penalize more on g, then the minimized g is smaller; in this case, we care less about f
- appear in model performance evaluation where two diffferent metrics are desired to be small
- example 1: minimize model error + model complexity
- example 2: minimize system tracking error + input power

Multi-objective optimization

setting: minimizing $f_0 : \mathbf{R}^n \to \mathbf{R}^m$ (vector-valued function) over a feasible set

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & x \in \mathcal{C} \end{array}$$

a vector optimization has a **vector-valued** objective function

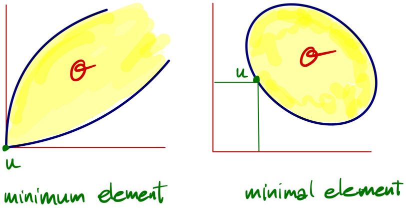
- example: $f_0(x) = (\text{fuel,time})$ the energy used and time spent of a vehicle parameter x
- require a generalized inequality definition for comparing any two vectors of $f_0(x)$

$$\begin{bmatrix} 5 \\ 2 \end{bmatrix} \preceq \begin{bmatrix} 10 \\ 3 \end{bmatrix} \quad \text{but} \quad \begin{bmatrix} 5 \\ 2 \end{bmatrix} \not\preceq \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

here, for $f_0(x) \in \mathbf{R}^n$, we typically use the **non-negative orthant** to define $\preceq$

Achievable objective values

define $\mathcal{O} = \{ f_0(x) \mid x \in \mathcal{C} \}$ the set of objective values of feasible points



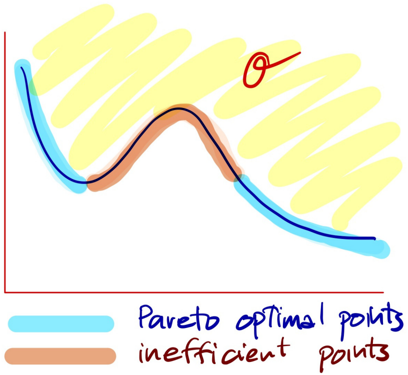
- u is said to be the **minimum** element of $\mathcal{O}$ if $u \preceq v$, for every $v \in \mathcal{O}$
- u is said to be a **minimal** element of $\mathcal{O}$ if $v \in \mathcal{O}$, $v \preceq u$ only if $v = u$
- if $\mathcal{O}$ has a minimum point (then it is unique) and

$$\exists \text{ feasible } x \quad \text{such that} \quad f_0(x) \preceq f_0(y), \quad \text{for all feasible } y$$

then we say x is **optimal**

Pareto optimal points

consider when $\mathcal{O}$ does not have a minimum element



- x is called **Pareto optimal** (or efficient) if $f_0(x)$ is a minimal element of $\mathcal{O}$
- a technique to extract pareto optimal points: scalarization (more on this later)

Optimality conditions

Unconstrained optimality

assumption: f is twice continuously differentiable (smooth objective)

- **necessary condition:** if $x^\star$ is a local minimizer of f then
  1. $\nabla f(x^\star) = 0$
  2. $\nabla^2 f(x^\star) \succeq 0$ (positive semidefinite)
- **sufficient condition:** if $\nabla f(x^\star) = 0$ and $\nabla^2 f(x^\star) \succ 0$ (positive definite), then $x^\star$ is a strict local minimizer of f
- when f is convex and differentiable, any stationary point $x^\star$ is a global minimizer of f

example: the Rosenbrock function:

$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

verify that $x^\star = (1, 1)$ is the only local minimizer of f

Constrained optimality

first, define the Lagrangian function

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

where λ, ν are called the **Lagrange multipliers** for inequality and equality constraints

the KKT conditions are **necessary conditions** for optimality

1. zero-gradient condition of L: $\nabla_x L(x^\star, \lambda^\star, \nu^\star) = 0$
2. primal and dual feasibility

$$f_i(x^\star) \leq 0, i = 1, \ldots, m, \quad h_i(x^\star) = 0, i = 1, \ldots, p, \quad \lambda^\star \succeq 0$$

3. complementary slackness condition: $\lambda_i f_i(x) = 0$ for $i = 1, 2, \ldots, m$

fact: for convex problems, KKT conditions are sufficient and necessary for optimality

Optimality of contrained LS

derive KKT conditions for

$$\underset{x}{\text{minimize}} \quad (1/2)\|Ax - y\|_2^2 \quad \text{subject to} \quad l \preceq x \preceq u$$

the Lagrangian is $L(x, \lambda_1, \lambda_2) = (1/2)\|Ax - y\|_2^2 + \lambda_1^T(l - x) + \lambda_2^T x - u)$

KKT conditions are

1. zero-gradient of L: $A^T(Ax - y) - \lambda_1 + \lambda_2 = 0$
2. primal feasibility: $l \preceq x \preceq u$
3. dual feasibility: $\lambda_1, \lambda_2 \succeq 0$
4. complementary slackness condition:

$$\lambda_{1i}(l_i - x_i) = 0, \quad \lambda_{2i}(x_i - u_i) = 0, \quad i = 1, 2, \ldots, n$$

Intro to duality theory

some quick facts

- define the dual function as the infimum of the Lagrangian over primal variables

$$g(\lambda, \nu) = \inf_{x \in \mathbf{dom}\, \mathcal{D}} L(x, \lambda, \nu)$$

- for any $\lambda \succeq 0$, the dual function provides a lower bound for $p^\star$, *i.e.*, $g(\lambda, \nu) \leq p^\star$
- any optimization problem (called a primal problem) has its dual problem

$$\underset{\lambda, \nu}{\text{maximize}} \quad g(\lambda, \nu) \quad \text{subject to} \quad \lambda \succeq 0$$

  which is the problem of finding the *best* lower bound, denoted as $d^\star$, for $p^\star$
- more theoretical results about relations between primal and dual problems – when $d^\star = p^\star$, we say we have strong duality
- solving the dual can be more beneficial in some cases

Numerical methods

Overview of available methods

- unconstrained problems: gradient descent, Newton, quasi Newton, trust-region
- convex programs: interior point, gradient projection, ellipsoid method
- convex programs of certain structures: proximal methods
- linear programming: simplex, interior point
- quadratic programming: interior point, active set, conjugate gradient, augmented Lagrangian

Softwares

MATLAB: cvx

- CVX is a MATLAB-based modeling system for convex optimization
- `http://cvxr.com/cvx/`

Python

- **CVXPY**: Python-embedded modeling language for convex optimization problems available at `https://www.cvxpy.org/` by Stephen Boyd group
- **CVXOPT**: Python-based package for convex optimization available at `http://cvxopt.org/` by M. Andersen, J. Dahl and L. Vandenberghe

References

1. S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge, 2004
2. G. Calafiore and L. El Ghaoui, *Optimization Models*, Cambridge University Press, 2014
3. Y. Nesterov, *Introductory lectures on convex optimization: A basic course*, KAP, 2004
4. R.T. Rockfellar, *Convex Analysis*, Princeton University Press, 1970
5. S.Boyd, N. Parikh, E. Chu, B.Peleato, and J. Eckstein, *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*, Foundations and Trends in Machine Learning, 2011