

Lecture 3: Convex optimization (Part 2)

Optimization for data sciences



Rémy Sun
remy.sun@inria.fr



Course organization

What can we optimize?

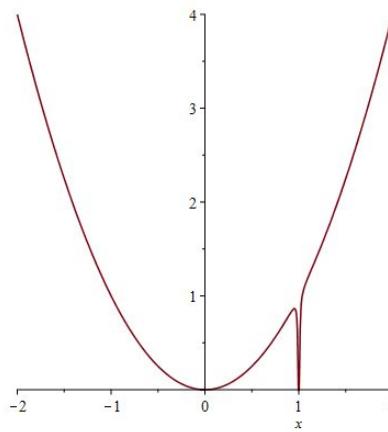
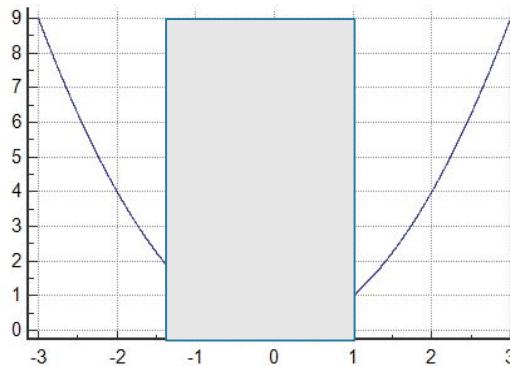
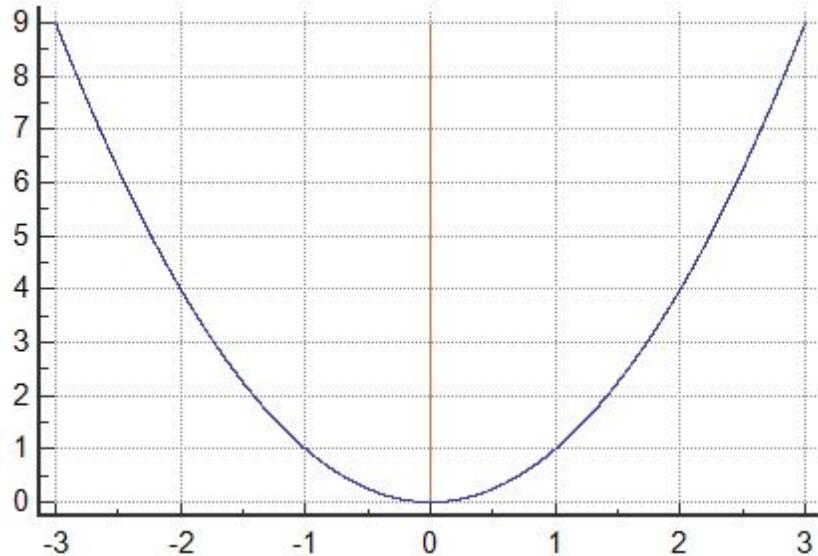
- Reduce the complexity/overhead of a problem
 - E.g. Network quantization
 - E.g. Computational optimization
- **Find the best solution to a problem**
 - **Numerical optimization**
 - **Evaluate solutions according to a criterion**
 - **Look at solutions from some given space of possible solutions to consider**

Defining an optimization problem

- Minimize a quantity $f_0(x)$
 - Under inequality and equality constraints
 - Constraints define a domain D
 - Could have no constraint except $x \in D$

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

Can you formalize these problems?



Course organization

- Introduction to optimization
 - A few problems of interest
 - Quick mathematical refresher
- **Convex problems (Following Stephen Boyd)**
 - **Quick refresher on last week**
 - **Convex sets**
 - **Convex functions**
 - **Convex problems**
 - **Simplex algorithm for Linear Programming**

Course organization

- Duality (for convex problems)
- Newton's Descent and Barrier methods for convex case
- (First order) descent methods for the general case
- Backpropagation
- Some more properties on stochastic gradient descent

- Reports on lab sessions
 - Labs on jupyter notebooks
 - Not every session
 - Explain the code done in the session
 - Summarize what is done in the practical
- Written Exam
 - Theoretical questions
 - We will do exercises in class

Refresher on last week

Convex problem

convex optimization problem:

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

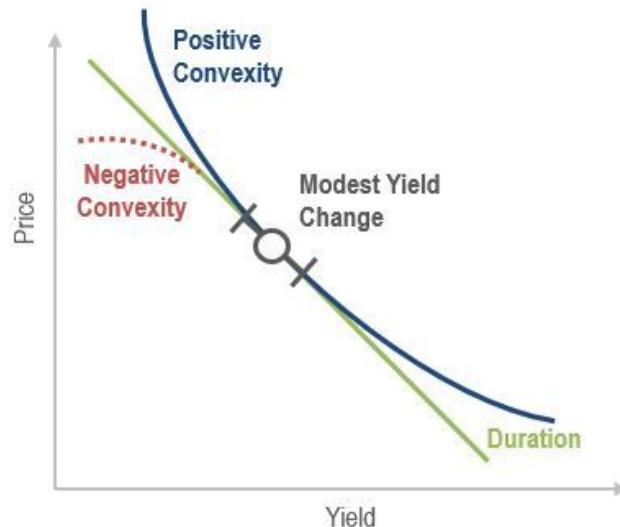
- ▶ variable $x \in \mathbf{R}^n$
- ▶ equality constraints are linear
- ▶ f_0, \dots, f_m are **convex**: for $\theta \in [0, 1]$,

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

i.e., f_i have nonnegative (upward) curvature

Easy problem

- ▶ classical view:
 - linear (zero curvature) is easy
 - nonlinear (nonzero curvature) is hard
- ▶ the classical view is **wrong**
- ▶ the correct view:
 - convex (nonnegative curvature) is easy
 - nonconvex (negative curvature) is hard



Easy to solve!

- ▶ many different algorithms (that run on many platforms)
 - interior-point methods for up to 10000s of variables
 - first-order methods for larger problems
 - do not require initial point, babysitting, or tuning
- ▶ can develop and deploy quickly using modeling languages such as CVXPY
- ▶ solvers are reliable, so can be embedded
- ▶ code generation yields real-time solvers that execute in milliseconds (e.g., on Falcon 9 and Heavy for landing)

- Convex sets
 - Affine sets, norm balls, norm cones
 - Convex combination, convex hull and Convex cones
 - Hyperplanes, halfspaces and polyhedron
 - Positive Semidefinite Cone
- Convexity preserving operations
 - Intersection
 - Affine mapping
 - Perspective and Linear fractional mappings
- Proper cones and generalized inequalities
- Separating and supporting hyperplanes

- **Convex sets (definition!)**
 - **Affine sets, norm balls, norm cones**
 - **Convex combination, convex hull and Convex cones**
 - **Hyperplanes, halfspaces and polyhedron**
 - Positive Semidefinite Cone
- **Showing a set is convex (with operations!)**
 - **Intersection**
 - **Affine mapping**
 - **Perspective and Linear fractional mappings**
- Proper cones and generalized inequalities
- Separating and supporting hyperplanes

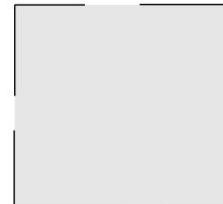
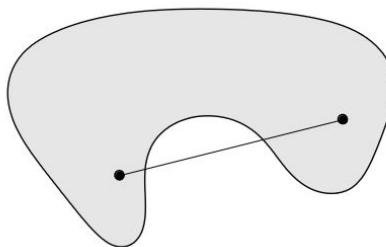
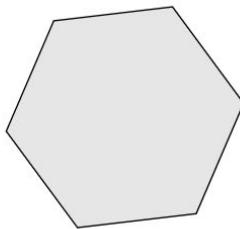
Convex sets

line segment between x_1 and x_2 : all points of form $x = \theta x_1 + (1 - \theta)x_2$, with $0 \leq \theta \leq 1$

convex set: contains line segment between any two points in the set

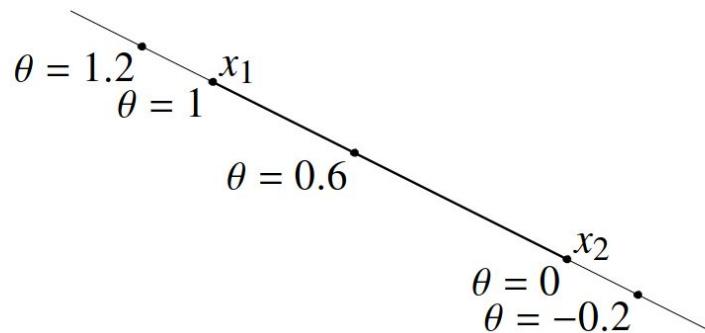
$$x_1, x_2 \in C, \quad 0 \leq \theta \leq 1 \quad \implies \quad \theta x_1 + (1 - \theta)x_2 \in C$$

examples (one convex, two nonconvex sets)



Affine sets

line through x_1, x_2 : all points of form $x = \theta x_1 + (1 - \theta)x_2$, with $\theta \in \mathbf{R}$



affine set: contains the line through any two distinct points in the set

example: solution set of linear equations $\{x \mid Ax = b\}$

(conversely, every affine set can be expressed as solution set of system of linear equations)

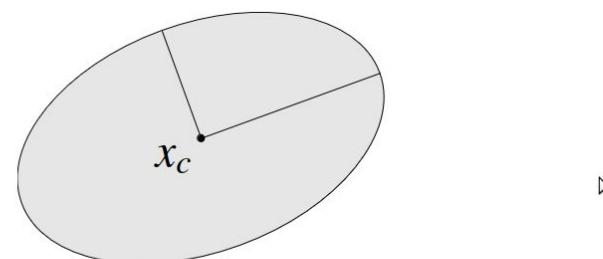
(Euclidean) ball with center x_c and radius r :

$$B(x_c, r) = \{x \mid \|x - x_c\|_2 \leq r\} = \{x_c + ru \mid \|u\|_2 \leq 1\}$$

ellipsoid: set of the form

$$\{x \mid (x - x_c)^T P^{-1} (x - x_c) \leq 1\}$$

with $P \in \mathbf{S}_{++}^n$ (i.e., P symmetric positive definite)



another representation: $\{x_c + Au \mid \|u\|_2 \leq 1\}$ with A square and nonsingular

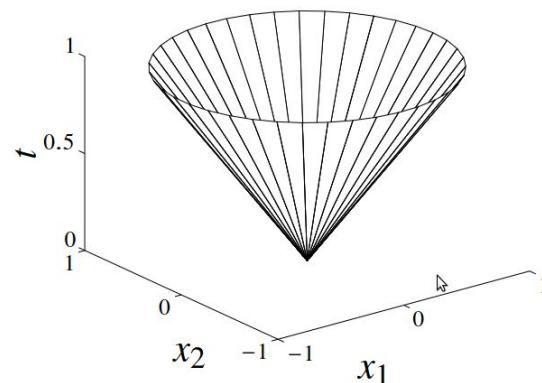
Norm balls and norm cones

- ▶ **norm:** a function $\|\cdot\|$ that satisfies
 - $\|x\| \geq 0$; $\|x\| = 0$ if and only if $x = 0$
 - $\|tx\| = |t| \|x\|$ for $t \in \mathbf{R}$
 - $\|x + y\| \leq \|x\| + \|y\|$
- ▶ notation: $\|\cdot\|$ is general (unspecified) norm; $\|\cdot\|_{\text{symb}}$ is particular norm
- ▶ **norm ball** with center x_c and radius r : $\{x \mid \|x - x_c\| \leq r\}$
- ▶ **norm cone:** $\{(x, t) \mid \|x\| \leq t\}$
- ▶ norm balls and cones are convex

Euclidean norm cone

$$\{(x, t) \mid \|x\|_2 \leq t\} \subset \mathbf{R}^{n+1}$$

is called **second-order cone**



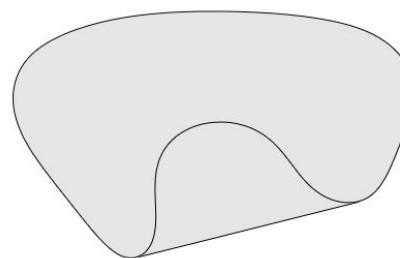
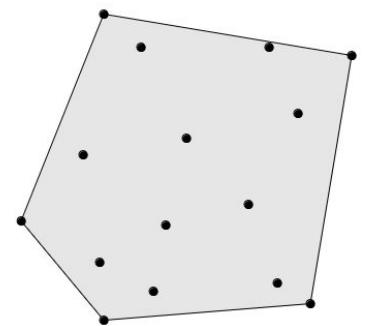
Convex combination and hull

convex combination of x_1, \dots, x_k : any point x of the form

$$x = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k$$

with $\theta_1 + \cdots + \theta_k = 1$, $\theta_i \geq 0$

convex hull $\text{conv } S$: set of all convex combinations of points in S

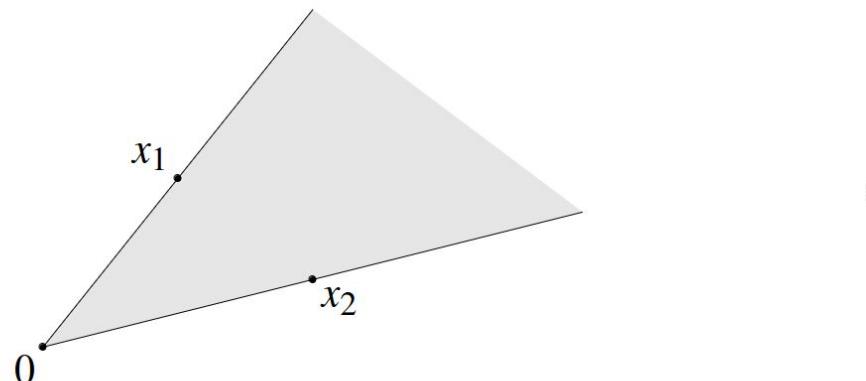


Convex cone

conic (nonnegative) combination of x_1 and x_2 : any point of the form

$$x = \theta_1 x_1 + \theta_2 x_2$$

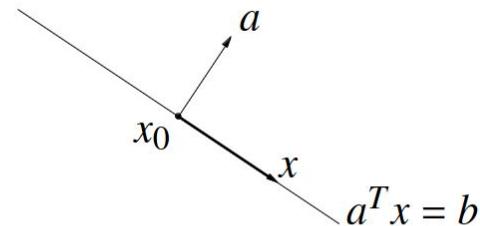
with $\theta_1 \geq 0, \theta_2 \geq 0$



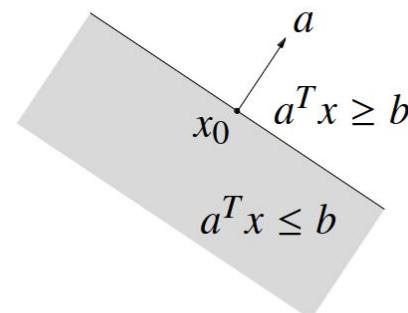
convex cone: set that contains all conic combinations of points in the set

Hyperplanes and halfspaces

hyperplane: set of the form $\{x \mid a^T x = b\}$, with $a \neq 0$



halfspace: set of the form $\{x \mid a^T x \leq b\}$, with $a \neq 0$



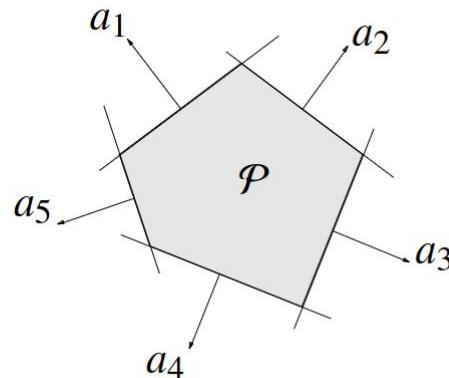
- ▶ a is the normal vector
- ▶ hyperplanes are affine and convex; halfspaces are convex

- ▶ **Polyhedron** is solution set of finitely many linear inequalities and equalities

$$\{x \mid Ax \leq b, Cx = d\}$$

($A \in \mathbf{R}^{m \times n}$, $C \in \mathbf{R}^{p \times n}$, \leq is componentwise inequality)

- ▶ intersection of finite number of halfspaces and hyperplanes
- ▶ example with no equality constraints; a_i^T are rows of A



Showing a set is convex

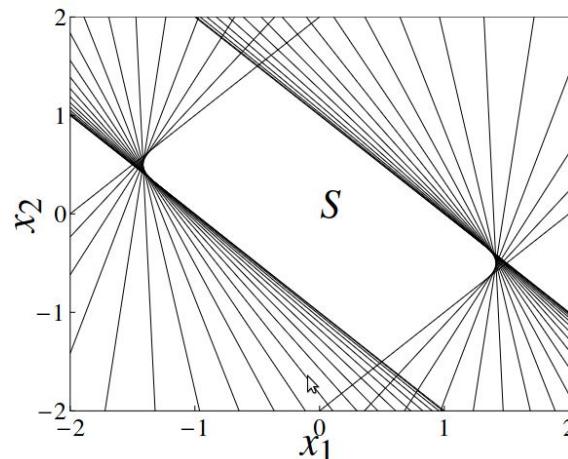
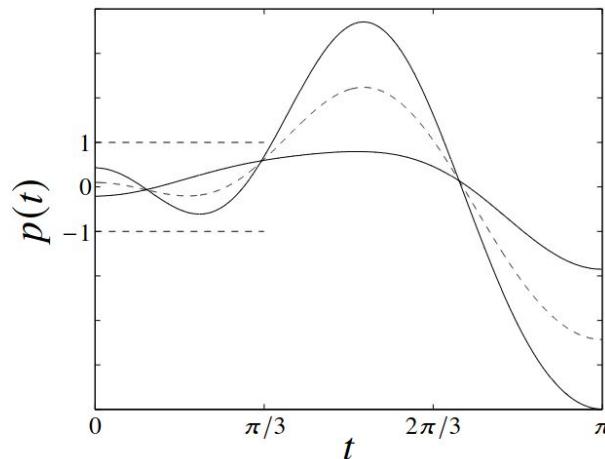
methods for establishing convexity of a set C

1. apply definition: show $x_1, x_2 \in C, 0 \leq \theta \leq 1 \implies \theta x_1 + (1 - \theta)x_2 \in C$
 - recommended only for **very simple** sets
2. use convex functions (next lecture)
3. show that C is obtained from simple convex sets (hyperplanes, halfspaces, norm balls, ...) by operations that preserve convexity
 - intersection
 - affine mapping
 - perspective mapping
 - linear-fractional mapping

you'll mostly use methods 2 and 3

Showing a set is convex

- ▶ the intersection of (any number of) convex sets is convex
- ▶ **example:**
 - $S = \{x \in \mathbf{R}^m \mid |p(t)| \leq 1 \text{ for } |t| \leq \pi/3\}$, with $p(t) = x_1 \cos t + \cdots + x_m \cos mt$
 - write $S = \bigcap_{|t| \leq \pi/3} \{x \mid |p(t)| \leq 1\}$, i.e., an intersection of (convex) slabs
- ▶ picture for $m = 2$:



Showing a set is convex

- ▶ suppose $f : \mathbf{R}^n \rightarrow \mathbf{R}^m$ is affine, i.e., $f(x) = Ax + b$ with $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$
- ▶ the **image** of a convex set under f is convex

$$S \subseteq \mathbf{R}^n \text{ convex} \implies f(S) = \{f(x) \mid x \in S\} \text{ convex}$$

- ▶ the **inverse image** $f^{-1}(C)$ of a convex set under f is convex

$$C \subseteq \mathbf{R}^m \text{ convex} \implies f^{-1}(C) = \{x \in \mathbf{R}^n \mid f(x) \in C\} \text{ convex}$$

Showing a set is convex

- ▶ **perspective function** $P : \mathbf{R}^{n+1} \rightarrow \mathbf{R}^n$:

$$P(x, t) = x/t, \quad \text{dom } P = \{(x, t) \mid t > 0\}$$

- ▶ images and inverse images of convex sets under perspective are convex
 - ▶ **linear-fractional function** $f : \mathbf{R}^n \rightarrow \mathbf{R}^m$:
- $$f(x) = \frac{Ax + b}{c^T x + d}, \quad \text{dom } f = \{x \mid c^T x + d > 0\}$$
- ▶ images and inverse images of convex sets under linear-fractional functions are convex

Takeaway

$$x_1, x_2 \in C, \quad 0 \leq \theta \leq 1 \quad \implies \quad \theta x_1 + (1 - \theta)x_2 \in C$$

- Classic convex sets
 - Affine sets, hyperplanes, cones, balls, polyhedrons
- Convexity preserving operations
 - Intersection
 - Affine mapping
 - Perspective
 - Linear Fractional mapping

Convex function overview

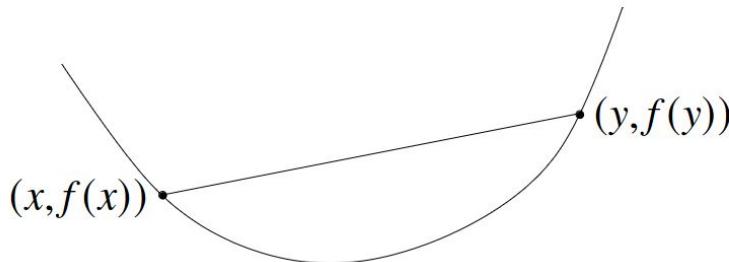
- Convex functions (with definition!)
 - Examples of classic convex functions
 - Extended value function
 - Line restriction
 - First and second order conditions
 - Epigraph and sublevel sets
- Showing a function is convex with operations
 - Non-negative weighted sum and affine composition
 - Pointwise maximum
 - Composition rules
 - Partial minimization and perspective
- Conjugate function

- **Convex functions (with definition!)**
 - **Examples of classic convex functions**
 - *Extended value function*
 - *Line restriction*
 - **First and second order conditions**
 - *Epigraph and sublevel sets*
- **Showing a function is convex with operations**
 - **Non-negative weighted sum and affine composition**
 - **Pointwise maximum**
 - **Composition rules**
 - *Partial minimization and perspective*
- *Conjugate function*

Convex functions

- $f : \mathbf{R}^n \rightarrow \mathbf{R}$ is convex if $\mathbf{dom}f$ is a convex set and for all $x, y \in \mathbf{dom}f$, $0 \leq \theta \leq 1$,

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



- f is concave if $-f$ is convex
- f is strictly convex if $\mathbf{dom}f$ is convex and for $x, y \in \mathbf{dom}f$, $x \neq y$, $0 < \theta < 1$,

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

↳

Convex functions (Examples)

convex functions:

- ▶ affine: $ax + b$ on \mathbf{R} , for any $a, b \in \mathbf{R}$
- ▶ exponential: e^{ax} , for any $a \in \mathbf{R}$
- ▶ powers: x^α on \mathbf{R}_{++} , for $\alpha \geq 1$ or $\alpha \leq 0$
- ▶ powers of absolute value: $|x|^p$ on \mathbf{R} , for $p \geq 1$
- ▶ positive part (relu): $\max\{0, x\}$

concave functions:

- ▶ affine: $ax + b$ on \mathbf{R} , for any $a, b \in \mathbf{R}$
- ▶ powers: x^α on \mathbf{R}_{++} , for $0 \leq \alpha \leq 1$
- ▶ logarithm: $\log x$ on \mathbf{R}_{++}
- ▶ entropy: $-x \log x$ on \mathbf{R}_{++}
- ▶ negative part: $\min\{0, x\}$

Extended value

- ▶ suppose f is convex on \mathbf{R}^n , with domain $\mathbf{dom} f$
- ▶ its extended-value extension \tilde{f} is function $\tilde{f} : \mathbf{R}^n \rightarrow \mathbf{R} \cup \{\infty\}$

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

- ▶ often simplifies notation; for example, the condition

$$0 \leq \theta \leq 1 \implies \tilde{f}(\theta x + (1 - \theta)y) \leq \theta \tilde{f}(x) + (1 - \theta) \tilde{f}(y)$$

(as an inequality in $\mathbf{R} \cup \{\infty\}$), means the same as the two conditions

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

- ▶ $f : \mathbf{R}^n \rightarrow \mathbf{R}$ is convex if and only if the function $g : \mathbf{R} \rightarrow \mathbf{R}$,

$$g(t) = f(x + tv), \quad \text{dom } g = \{t \mid x + tv \in \text{dom } f\}$$



- is convex (in t) for any $x \in \text{dom } f$, $v \in \mathbf{R}^n$
- ▶ can check convexity of f by checking convexity of functions of one variable

First order condition

- f is **differentiable** if $\text{dom } f$ is open and the gradient

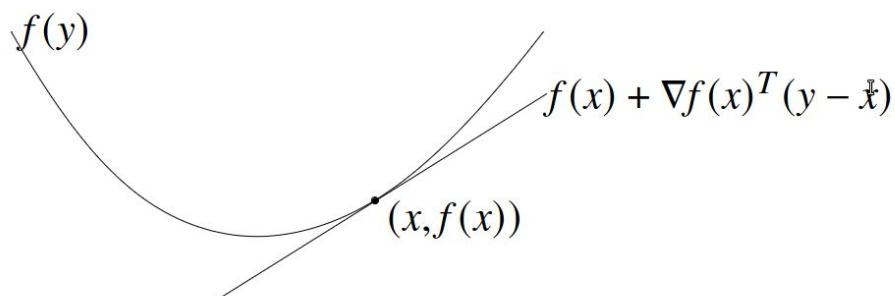
$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \dots, \frac{\partial f(x)}{\partial x_n} \right) \in \mathbf{R}^n$$

exists at each $x \in \text{dom } f$

- **1st-order condition:** differentiable f with convex domain is convex if and only if

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad \text{for all } x, y \in \text{dom } f$$

- first order Taylor approximation of convex f is a **global underestimator** of f



- ▶ f is **twice differentiable** if $\text{dom } f$ is open and the Hessian $\nabla^2 f(x) \in \mathbf{S}^n$,

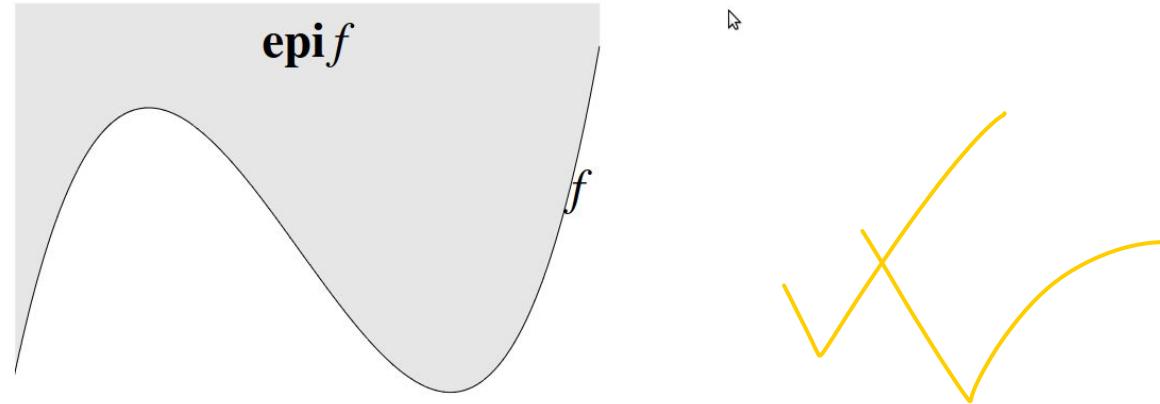
$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \quad i, j = 1, \dots, n,$$

exists at each $x \in \text{dom } f$

- ▶ **2nd-order conditions:** for twice differentiable f with convex domain
 - f is convex if and only if $\nabla^2 f(x) \succeq 0$ for all $x \in \text{dom } f$
 - if $\nabla^2 f(x) > 0$ for all $x \in \text{dom } f$, then f is strictly convex

Epigraph and sublevel set

- ▶ **α -sublevel set** of $f : \mathbf{R}^n \rightarrow \mathbf{R}$ is $C_\alpha = \{x \in \text{dom } f \mid f(x) \leq \alpha\}$
- ▶ sublevel sets of convex functions are convex sets (but converse is false)
- ▶ **epigraph** of $f : \mathbf{R}^n \rightarrow \mathbf{R}$ is $\text{epi } f = \{(x, t) \in \mathbf{R}^{n+1} \mid x \in \text{dom } f, f(x) \leq t\}$



- ▶ f is convex if and only if $\text{epi } f$ is a convex set

Showing a function is convex

methods for establishing convexity of a function f

1. verify definition (often simplified by restricting to a line)
2. for twice differentiable functions, show $\nabla^2 f(x) \geq 0$
 - recommended only for **very simple** functions
3. show that f is obtained from simple convex functions by operations that preserve convexity
 - nonnegative weighted sum
 - composition with affine function
 - pointwise maximum and supremum
 - composition
 - minimization
 - perspective

you'll mostly use methods 2 and 3



Showing a function is convex

- ▶ **nonnegative multiple:** αf is convex if f is convex, $\alpha \geq 0$
 - ▶ **sum:** $f_1 + f_2$ convex if f_1, f_2 convex
 - ▶ **infinite sum:** if f_1, f_2, \dots are convex functions, infinite sum $\sum_{i=1}^{\infty} f_i$ is convex
 - ▶ **integral:** if $f(x, \alpha)$ is convex in x for each $\alpha \in \mathcal{A}$, then $\int_{\alpha \in \mathcal{A}} f(x, \alpha) d\alpha$ is convex
- ↗
- ▶ there are analogous rules for concave functions

Showing a function is convex

(pre-)composition with affine function: $f(Ax + b)$ is convex if f is convex

examples

- ▶ log barrier for linear inequalities

$$f(x) = - \sum_{i=1}^m \log(b_i - a_i^T x), \quad \text{dom } f = \{x \mid a_i^T x < b_i, i = 1, \dots, m\}$$

- ▶ norm approximation error: $f(x) = \|Ax - b\|$ (any norm)

Showing a function is convex

if f_1, \dots, f_m are convex, then $f(x) = \max\{f_1(x), \dots, f_m(x)\}$ is convex

examples

- ▶ piecewise-linear function: $f(x) = \max_{i=1,\dots,m}(a_i^T x + b_i)$
- ▶ sum of r largest components of $x \in \mathbf{R}^n$:

$$f(x) = x_{[1]} + x_{[2]} + \cdots + x_{[r]}$$

($x_{[i]}$ is i th largest component of x)

□

proof: $f(x) = \max\{x_{i_1} + x_{i_2} + \cdots + x_{i_r} \mid 1 \leq i_1 < i_2 < \cdots < i_r \leq n\}$

Showing a function is convex

if $f(x, y)$ is convex in x for each $y \in \mathcal{A}$, then $g(x) = \sup_{y \in \mathcal{A}} f(x, y)$ is convex

examples

- ▶ distance to farthest point in a set C : $f(x) = \sup_{y \in C} \|x - y\|$
- ▶ maximum eigenvalue of symmetric matrix: for $X \in \mathbf{S}^n$, $\lambda_{\max}(X) = \sup_{\|y\|_2=1} y^T X y$ is convex
- ▶ support function of a set C : $S_C(x) = \sup_{y \in C} y^T x$ is convex

Showing a function is convex

- ▶ the function $g(x) = \inf_{y \in C} f(x, y)$ is called the **partial minimization** of f (w.r.t. y)
- ▶ if $f(x, y)$ is convex in (x, y) and C is a convex set, then partial minimization g is convex

examples

- ▶ $f(x, y) = x^T A x + 2x^T B y + y^T C y$ with

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0, \quad C > 0$$



minimizing over y gives $g(x) = \inf_y f(x, y) = x^T (A - BC^{-1}B^T)x$
 g is convex, hence Schur complement $A - BC^{-1}B^T \succeq 0$

- ▶ distance to a set: $\text{dist}(x, S) = \inf_{y \in S} \|x - y\|$ is convex if S is convex

Showing a function is convex

- ▶ composition of $g : \mathbf{R}^n \rightarrow \mathbf{R}$ and $h : \mathbf{R} \rightarrow \mathbf{R}$ is $f(x) = h(g(x))$ (written as $f = h \circ g$)
- ▶ composition f is convex if
 - g convex, h convex, \tilde{h} nondecreasing
 - or g concave, h convex, \tilde{h} nonincreasing(monotonicity must hold for extended-value extension \tilde{h})
- ▶ proof (for $n = 1$, differentiable g, h)

$$f''(x) = h''(g(x))g'(x)^2 + h'(g(x))g''(x)$$



examples

- ▶ $f(x) = \exp g(x)$ is convex if g is convex
- ▶ $f(x) = 1/g(x)$ is convex if g is concave and positive



Showing a function is convex

- ▶ composition of $g : \mathbf{R}^n \rightarrow \mathbf{R}^k$ and $h : \mathbf{R}^k \rightarrow \mathbf{R}$ is $f(x) = h(g(x)) = h(g_1(x), g_2(x), \dots, g_k(x))$
- ▶ f is convex if h is convex and for each i one of the following holds
 - g_i convex, \tilde{h} nondecreasing in its i th argument
 - g_i concave, \tilde{h} nonincreasing in its i th argument
 - g_i affine
- ▶ you will use this composition rule **constantly** throughout this course
- ▶ you need to commit this rule to memory



- ▶ the **perspective** of a function $f : \mathbf{R}^n \rightarrow \mathbf{R}$ is the function $g : \mathbf{R}^n \times \mathbf{R} \rightarrow \mathbf{R}$,

$$g(x, t) = tf(x/t), \quad \text{dom } g = \{(x, t) \mid x/t \in \text{dom } f, t > 0\}$$

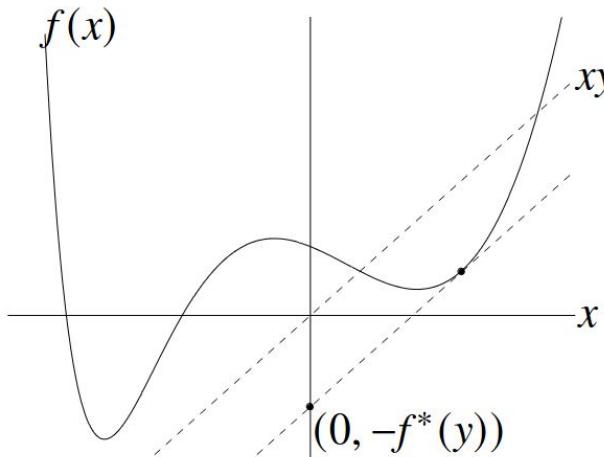
- ▶ g is convex if f is convex

examples

- ▶ $f(x) = x^T x$ is convex; so $g(x, t) = x^T x/t$ is convex for $t > 0$
- ▶ $f(x) = -\log x$ is convex; so relative entropy $g(x, t) = t \log t - t \log x$ is convex on \mathbf{R}_{++}^2

Conjugate

- ▶ the **conjugate** of a function f is $f^*(y) = \sup_{x \in \text{dom } f} (y^T x - f(x))$



↳

- ▶ f^* is convex (even if f is not)
- ▶ will be useful in chapter 5

Takeaway

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

- Classic convex functions
 - Affine, exponential, norms, max, ...
- Convexity preserving operations
 - Non negative weighted sum, composition with affine
 - Pointwise maximum and supremum
 - Composition
 - Minimization
 - Perspective

1. Convex problems

Standard form optimization problem

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

- ▶ $x \in \mathbf{R}^n$ is the optimization variable
- ▶ $f_0 : \mathbf{R}^n \rightarrow \mathbf{R}$ is the objective or cost function
- ▶ $f_i : \mathbf{R}^n \rightarrow \mathbf{R}, i = 1, \dots, m$, are the inequality constraint functions
- ▶ $h_i : \mathbf{R}^n \rightarrow \mathbf{R}$ are the equality constraint functions



Feasible and optimal points

- ▶ $x \in \mathbf{R}^n$ is **feasible** if $x \in \text{dom } f_0$ and it satisfies the constraints
- ▶ **optimal value** is $p^\star = \inf\{f_0(x) \mid f_i(x) \leq 0, i = 1, \dots, m, h_i(x) = 0, i = 1, \dots, p\}$
- ▶ $p^\star = \infty$ if problem is infeasible
- ▶ $p^\star = -\infty$ if problem is **unbounded below**
- ▶ a feasible x is **optimal** if $f_0(x) = p^\star$
- ▶ X_{opt} is the set of optimal points



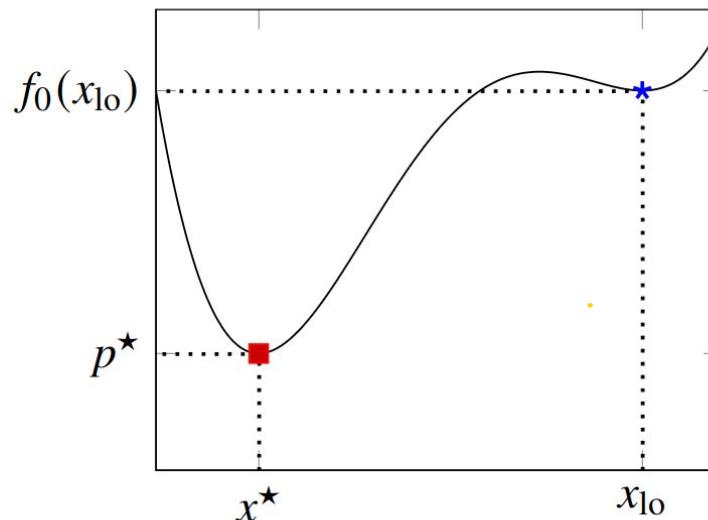
Optimal points

x is **locally optimal** if there is an $R > 0$ such that x is optimal for

minimize (over) $f_0(z)$

subject to $f_i(z) \leq 0, \quad i = 1, \dots, m, \quad h_i(z) = 0, \quad i = 1, \dots, p$

$$\|z - x\|_2 \leq R$$



Implicit and explicit constraints

standard form optimization problem has **implicit constraint**

$$x \in \mathcal{D} = \bigcap_{i=0}^m \text{dom } f_i \cap \bigcap_{i=1}^p \text{dom } h_i,$$

- ▶ we call \mathcal{D} the **domain** of the problem
- ▶ the constraints $f_i(x) \leq 0, h_i(x) = 0$ are the **explicit constraints**
- ▶ a problem is **unconstrained** if it has no explicit constraints ($m = p = 0$)

example:

$$\text{minimize } f_0(x) = -\sum_{i=1}^k \log(b_i - a_i^T x)$$

is an unconstrained problem with implicit constraints $a_i^T x < b_i$

Feasibility problem

$$\begin{aligned} & \text{find} && x \\ & \text{subject to} && f_i(x) \leq 0, \quad i = 1, \dots, m \\ & && h_i(x) = 0, \quad i = 1, \dots, p \end{aligned}$$

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

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

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

Standard form convex problem

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



- ▶ objective and inequality constraints f_0, f_1, \dots, f_m are convex
 - ▶ equality constraints are affine, often written as $Ax = b$
 - ▶ feasible and optimal sets of a convex optimization problem are convex
- ▶ problem is **quasiconvex** if f_0 is quasiconvex, f_1, \dots, f_m are convex, h_1, \dots, h_p are affine



Optimum in a convex set

any locally optimal point of a convex problem is (globally) optimal

proof:

- ▶ suppose x is locally optimal, but there exists a feasible y with $f_0(y) < f_0(x)$
- ▶ x locally optimal means there is an $R > 0$ such that

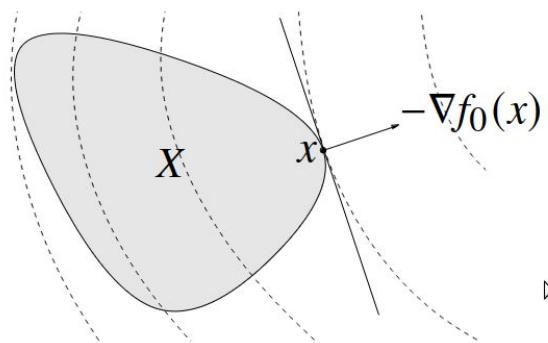
$$z \text{ feasible}, \quad \|z - x\|_2 \leq R \implies f_0(z) \geq f_0(x)$$

- 
- ▶ consider $z = \theta y + (1 - \theta)x$ with $\theta = R/(2\|y - x\|_2)$
 - ▶ $\|y - x\|_2 > R$, so $0 < \theta < 1/2$
 - ▶ z is a convex combination of two feasible points, hence also feasible
 - ▶ $\|z - x\|_2 = R/2$ and $f_0(z) \leq \theta f_0(y) + (1 - \theta)f_0(x) < f_0(x)$, which contradicts our assumption that x is locally optimal

First order criterion

- x is optimal for a convex problem if and only if it is feasible and

$$\nabla f_0(x)^T(y - x) \geq 0 \text{ for all feasible } y$$

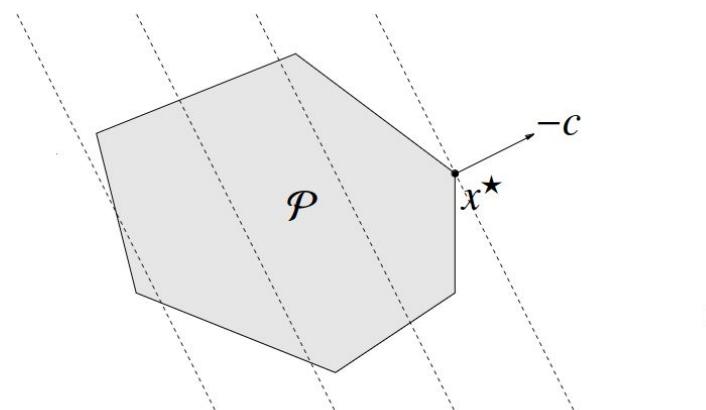


- if nonzero, $\nabla f_0(x)$ defines a supporting hyperplane to feasible set X at x

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

Often written as
a maximization

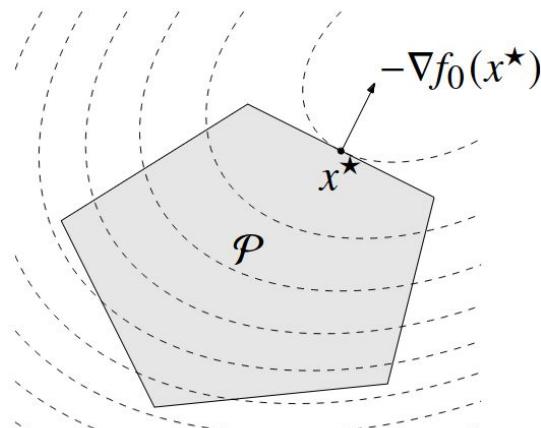
- ▶ convex problem with affine objective and constraint functions
- ▶ feasible set is a polyhedron



Quadratic programming

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

- ▶ $P \in \mathbf{S}_+^n$, so objective is convex quadratic
- ▶ minimize a convex quadratic function over a polyhedron



Quadratically constrained Quadratic programming (QCQP)



$$\begin{aligned} & \text{minimize} && (1/2)x^T P_0 x + q_0^T x + r_0 \\ & \text{subject to} && (1/2)x^T P_i x + q_i^T x + r_i \leq 0, \quad i = 1, \dots, m \\ & && Ax = b \end{aligned}$$

- ▶ $P_i \in \mathbf{S}_+^n$; objective and constraints are convex quadratic
- ▶ if $P_1, \dots, P_m \in \mathbf{S}_{++}^n$, feasible region is intersection of m ellipsoids and an affine set

Change of variable

- ▶ $\phi : \mathbf{R}^n \rightarrow \mathbf{R}^n$ is one-to-one with $\phi(\mathbf{dom} \phi) \supseteq \mathcal{D}$
- ▶ consider (possibly non-convex) problem

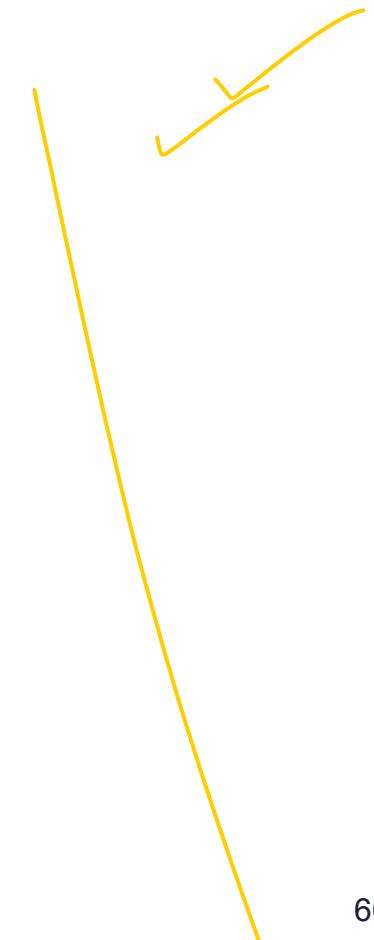
$$\begin{aligned} & \text{minimize} && f_0(x) \\ & \text{subject to} && f_i(x) \leq 0, \quad i = 1, \dots, m \\ & && h_i(x) = 0, \quad i = 1, \dots, p \end{aligned}$$

- ▶ change variables to z with $x = \phi(z)$
- ▶ can solve equivalent problem

$$\begin{aligned} & \text{minimize} && \tilde{f}_0(z) \\ & \text{subject to} && \tilde{f}_i(z) \leq 0, \quad i = 1, \dots, \overset{\mathbb{I}}{m} \\ & && \tilde{h}_i(z) = 0, \quad i = 1, \dots, p \end{aligned}$$

where $\tilde{f}_i(z) = f_i(\phi(z))$ and $\tilde{h}_i(z) = h_i(\phi(z))$

- ▶ recover original optimal point as $x^\star = \phi(z^\star)$



Transformation

suppose

- ▶ ϕ_0 is monotone increasing
- ▶ $\psi_i(u) \leq 0$ if and only if $u \leq 0$, $i = 1, \dots, m$
- ▶ $\varphi_i(u) = 0$ if and only if $u = 0$, $i = 1, \dots, p$

standard form optimization problem is equivalent to

$$\begin{aligned} & \text{minimize} && \phi_0(f_0(x)) \\ & \text{subject to} && \psi_i(f_i(x)) \leq 0, \quad i = 1, \dots, m \\ & && \varphi_i(h_i(x)) = 0, \quad i = 1, \dots, p \end{aligned}$$



example: minimizing $\|Ax - b\|$ is equivalent to minimizing $\|Ax - b\|^2$

Maximization and minimization

- ▶ suppose ϕ_0 is monotone decreasing
- ▶ the maximization problem

$$\begin{aligned} & \text{maximize} && f_0(x) \\ & \text{subject to} && f_i(x) \leq 0, \quad i = 1, \dots, m \\ & && h_i(x) = 0, \quad i = 1, \dots, p \end{aligned}$$

.

is equivalent to the minimization problem

$$\begin{aligned} & \text{minimize} && \phi_0(f_0(x)) \\ & \text{subject to} && f_i(x) \leq 0, \quad i = 1, \dots, m \\ & && h_i(x) = 0, \quad i = 1, \dots, p \end{aligned}$$

- ▶ **examples:**

- $\phi_0(u) = -u$ transforms maximizing a concave function to minimizing a convex function
- $\phi_0(u) = 1/u$ transforms maximizing a concave positive function to minimizing a convex function



Eliminating equality constraints



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

is equivalent to

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

where F and x_0 are such that $Ax = b \iff x = Fz + x_0$ for some z

Introducing equality constraints

is equivalent to

$$\begin{aligned} & \text{minimize} && f_0(A_0x + b_0) \\ & \text{subject to} && f_i(A_i x + b_i) \leq 0, \quad i = 1, \dots, m \\ \\ & \text{minimize (over } x, y_i) && f_0(y_0) \\ & \text{subject to} && f_i(y_i) \leq 0, \quad i = 1, \dots, m \\ & && y_i = A_i x + b_i, \quad i = 0, 1, \dots, m \end{aligned}$$

Slack variables for linear equalities

is equivalent to

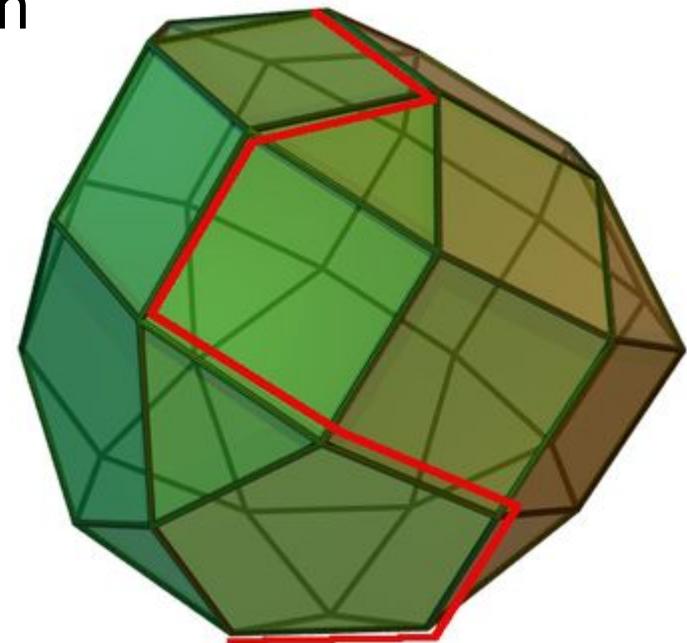


$$\begin{aligned} & \text{minimize} && f_0(x) \\ & \text{subject to} && a_i^T x \leq b_i, \quad i = 1, \dots, m \end{aligned}$$

$$\begin{aligned} & \text{minimize (over } x, s) && f_0(x) \\ & \text{subject to} && a_i^T x + s_i = b_i, \quad i = 1, \dots, m \\ & && s_i \geq 0, \quad i = 1, \dots, m \end{aligned}$$

2. Simplex algorithm (for LP)

- Historical algorithm of optimization
 - Proposed by Dantzig
- Widely used even nowadays
 - “Usually” efficient
 - Bad worst case
- Move along constraint edges



Standard form LP

Maximize $5x_1 + 4x_2 + 3x_3$

Subject to :

$$2x_1 + 3x_2 - x_3 \leq 5$$

$$4x_1 + x_2 - 2x_3 \leq 11$$

$$3x_1 + 4x_2 - 2x_3 \leq 8$$

$$x_1, x_2, x_3 \geq 0.$$

Maximize $5x_1 + 4x_2 + 3x_3$

Subject to :

$$2x_1 + 3x_2 - x_3 \leq 5$$

$$4x_1 + x_2 - 2x_3 \leq 11$$

$$3x_1 + 4x_2 - 2x_3 \leq 8$$

$$x_1, x_2, x_3 \geq 0.$$

Add slack variables

$$2x_1 + 3x_2 + x_3 \leq 5$$

Add slack variables

$$2x_1 + 3x_2 + x_3 \leq 5$$

$$x_4 = 5 - 2x_1 - 3x_2 - x_3.$$

$$x_4 \geq 0.$$

Add slack variables

$$2x_1 + 3x_2 + x_3 \leq 5$$

$$\begin{array}{rclcl} 4x_1 & + & x_2 & - & 2x_3 \leq 11 \\ 3x_1 & + & 4x_2 & - & 2x_3 \leq 8 \end{array}$$

$$x_4 = 5 - 2x_1 - 3x_2 - x_3.$$

$$x_4 \geq 0.$$

Add slack variables

$$2x_1 + 3x_2 + x_3 \leq 5$$

$$\begin{array}{rclclclcl} 4x_1 & + & x_2 & - & 2x_3 & \leq & 11 \\ 3x_1 & + & 4x_2 & - & 2x_3 & \leq & 8 \end{array}$$

$$x_4 = 5 - 2x_1 - 3x_2 - x_3.$$

$$\begin{array}{rclclclcl} x_5 & = & 11 & - & 4x_1 & - & x_2 & - & 2x_3 \\ x_6 & = & 8 & - & 3x_1 & - & 4x_2 & - & 2x_3 \end{array}$$

$$x_4 \geq 0.$$

$$x_5 \geq 0, x_6 \geq 0.$$

Maximize z subject to $x_1, x_2, x_3, x_4, x_5, x_6 \geq 0$.

$$x_4 = 5 - 2x_1 - 3x_2 - x_3$$

$$x_5 = 11 - 4x_1 - x_2 - 2x_3$$

$$x_6 = 8 - 3x_1 - 4x_2 - 2x_3$$

$$z = 5x_1 + 4x_2 + 3x_3.$$

Maximize z subject to $x_1, x_2, x_3, x_4, x_5, x_6 \geq 0$.

$$x_4 = 5 - 2x_1 - 3x_2 - x_3$$

$$x_5 = 11 - 4x_1 - x_2 - 2x_3$$

$$x_6 = 8 - 3x_1 - 4x_2 - 2x_3$$

$$z = 5x_1 + 4x_2 + 3x_3.$$

Find initial feasible solution

$$x_1 = 0, x_2 = 0, x_3 = 0$$

Find initial feasible solution

$$x_1 = 0, x_2 = 0, x_3 = 0$$

$$x_4 = 5 - 2x_1 - 3x_2 - x_3 = 5$$

Find initial feasible solution

$$x_1 = 0, x_2 = 0, x_3 = 0$$

$$x_4 = 5 - 2x_1 - 3x_2 - x_3 = 5$$

$$x_1 = 0, x_2 = 0, x_3 = 0, x_4 = 5, x_5 = 11, x_6 = 8$$

$$z = 0$$

Solving the simplex dictionary

Let's borrow an unconventional presentation from F. Giroire!

$$\begin{array}{rcl} x_4 & = & 5 - 2x_1 - 3x_2 - x_3 \\ x_5 & = & 11 - 4x_1 - x_2 - 2x_3 \\ x_6 & = & 8 - 3x_1 - 4x_2 - 2x_3 \\ \hline z & = & 5x_1 + 4x_2 + 3x_3. \end{array}$$

Basic variables: x_4, x_5, x_6 , variables on the left.

Non-basic variable: x_1, x_2, x_3 , variables on the right.

A dictionary is **feasible** if a feasible solution is obtained by setting all non-basic variables to 0.

Find the most influential variable

Simplex strategy: find an optimal solution by successive improvements.

Rule: we increase the value of the variable of **largest positive coefficient** in z .

$$\begin{array}{rcl} x_4 & = & 5 - 2x_1 - 3x_2 - x_3 \\ x_5 & = & 11 - 4x_1 - x_2 - 2x_3 \\ x_6 & = & 8 - 3x_1 - 4x_2 - 2x_3 \\ \hline z & = & \boxed{5}x_1 + 4x_2 + 3x_3. \end{array}$$

Here, we try to increase x_1 .

How far can we go?

How much can we increase x_1 ?

$$\begin{array}{rcl} x_4 & = & 5 - 2x_1 - 3x_2 - x_3 \\ x_5 & = & 11 - 4x_1 - x_2 - 2x_3 \\ x_6 & = & 8 - 3x_1 - 4x_2 - 2x_3 \\ \hline z & = & 5x_1 + 4x_2 + 3x_3. \end{array}$$

We have $x_4 \geq 0$.

It implies $5 - 2x_1 \geq 0$,

that is $x_1 \leq \frac{5}{2}$.

How far can we go?

How much can we increase x_1 ?

$$\begin{array}{rcl} x_4 & = & 5 - 2x_1 - 3x_2 - x_3 \\ x_5 & = & 11 - 4x_1 - x_2 - 2x_3 \\ x_6 & = & 8 - 3x_1 - 4x_2 - 2x_3 \\ \hline z & = & 5x_1 + 4x_2 + 3x_3. \end{array}$$

We have $x_4 \geq 0$.

It implies $5 - 2x_1 \geq 0$, that is $x_1 \leq 5/2$.

Similarly,

$x_5 \geq 0$ gives $x_1 \leq 11/4$.

$x_6 \geq 0$ gives $x_1 \leq 8/3$.

How far can we go?

How much can we increase x_1 ?

$$\begin{array}{rcl} x_4 & = & 5 - 2x_1 - 3x_2 - x_3 \\ x_5 & = & 11 - 4x_1 - x_2 - 2x_3 \\ x_6 & = & 8 - 3x_1 - 4x_2 - 2x_3 \\ \hline z & = & 5x_1 + 4x_2 + 3x_3. \end{array}$$

We have $x_4 \geq 0$.

It implies $5 - 2x_1 \geq 0$,

that is $x_1 \leq 5/2$

Strongest constraint

Similarly,

$x_5 \geq 0$ gives $x_1 \leq 11/4$.

$x_6 \geq 0$ gives $x_1 \leq 8/3$.

How far can we go?

How much can we increase x_1 ?

$$\begin{array}{rcl} x_4 & = & 5 - 2x_1 - 3x_2 - x_3 \\ x_5 & = & 11 - 4x_1 - x_2 - 2x_3 \\ x_6 & = & 8 - 3x_1 - 4x_2 - 2x_3 \\ \hline z & = & 5x_1 + 4x_2 + 3x_3. \end{array}$$

We have $x_4 \geq 0$.

It implies $5 - 2x_1 \geq 0$, that is $x_1 \leq 5/2$ Strongest constraint

We get a new solution: $x_1 = 5/2$, $x_4 = 0$

with better value $z = 5 \cdot 5/2 = 25/2$.

We still have $x_2 = x_3 = 0$ and now $x_5 = 11 - 4 \cdot 5/2 = 1$,
 $x_6 = 8 - 3 \cdot 5/2 = 1/2$

Pivot around the chosen variable

We build a new feasible dictionary.

$$\begin{array}{rcl} x_4 & = & 5 - 2x_1 - 3x_2 - x_3 \\ x_5 & = & 11 - 4x_1 - x_2 - 2x_3 \\ x_6 & = & 8 - 3x_1 - 4x_2 - 2x_3 \\ \hline z & = & 5x_1 + 4x_2 + 3x_3. \end{array}$$

x_1 enters the bases and x_4 leaves it:

$$x_1 = 5/2 - 3/2x_2 - 1/2x_3 - 1/2x_4$$

Pivot around the chosen variable

We replace x_1 by its expression in function of x_2, x_3, x_4 .

$$\begin{array}{rcl} x_1 & = & 5/2 - 1/2x_4 - 3/2x_2 - 1/2x_3 \\ x_5 & = & 11 - 4(5/2 - 3/2x_2 - 1/2x_3 - 1/2x_4) - x_2 - 2x_3 \\ x_6 & = & 8 - 3(5/2 - 3/2x_2 - 1/2x_3 - 1/2x_4) - 4x_2 - 2x_3 \\ \hline z & = & 5(5/2 - 3/2x_2 - 1/2x_3 - 1/2x_4) + 4x_2 + 3x_3. \end{array}$$

Pivot around the chosen variable

Finally, we get the new dictionary:

$$\begin{array}{rcl} x_1 & = & \frac{5}{2} - \frac{3}{2}x_2 - \frac{1}{2}x_3 - \frac{1}{2}x_4 \\ x_5 & = & 1 + 5x_2 \\ x_6 & = & \frac{1}{2} + \frac{1}{2}x_2 - \frac{1}{2}x_3 + \frac{3}{2}x_4 \\ \hline z & = & \frac{25}{2} - \frac{7}{2}x_2 + \frac{1}{2}x_3 - \frac{5}{2}x_4. \end{array}$$

New system

Finally, we get the new dictionary:

$$\begin{array}{rcl} x_1 & = & \boxed{\frac{5}{2}} - \frac{3}{2}x_2 - \frac{1}{2}x_3 - \frac{1}{2}x_4 \\ x_5 & = & \boxed{1} + 5x_2 \quad \quad \quad + 2x_4 \\ x_6 & = & \boxed{\frac{1}{2}} + \frac{1}{2}x_2 - \frac{1}{2}x_3 + \frac{3}{2}x_4 \\ \hline z & = & \boxed{\frac{25}{2}} - \frac{7}{2}x_2 + \frac{1}{2}x_3 - \frac{5}{2}x_4. \end{array}$$

We can read the solution directly from the dictionary:

Non basic variables: $x_2 = x_3 = x_4 = 0$.

Basic variables: $x_1 = 5/2$, $x_5 = 1$, $x_6 = 1/2$.

Value of the solution: $z = 25/2$.

$$\begin{array}{rcl}
 x_1 & = & \frac{5}{2} - \frac{3}{2}x_2 - \frac{1}{2}x_3 - \frac{1}{2}x_4 \\
 x_5 & = & 1 + 5x_2 + 2x_4 \\
 x_6 & = & \frac{1}{2} + \frac{1}{2}x_2 - \frac{1}{2}x_3 + \frac{3}{2}x_4 \\
 \hline
 z & = & \frac{25}{2} - \frac{7}{2}x_2 + \frac{1}{2}x_3 - \frac{5}{2}x_4.
 \end{array}$$

New step of the simplex:

- x_3 enters the basis (variable with largest positive coefficient).
- 3^d equation is the strictest constraint $x_3 \leq 1$.
- x_6 leaves the basis.

And we're done!

New feasible dictionary:

$$\begin{array}{rcl} x_3 & = & 1 + x_2 + 3x_4 - 2x_6 \\ x_1 & = & 2 - 2x_2 - 2x_4 + x_6 \\ \hline x_5 & = & 1 + 5x_2 + 2x_4 \\ \hline z & = & 13 - 3x_2 - x_4 - x_6. \end{array}$$

With new solution:

$$x_1 = 2, x_2 = 0, x_3 = 1, x_4 = 0, x_5 = 1, x_6 = 0$$

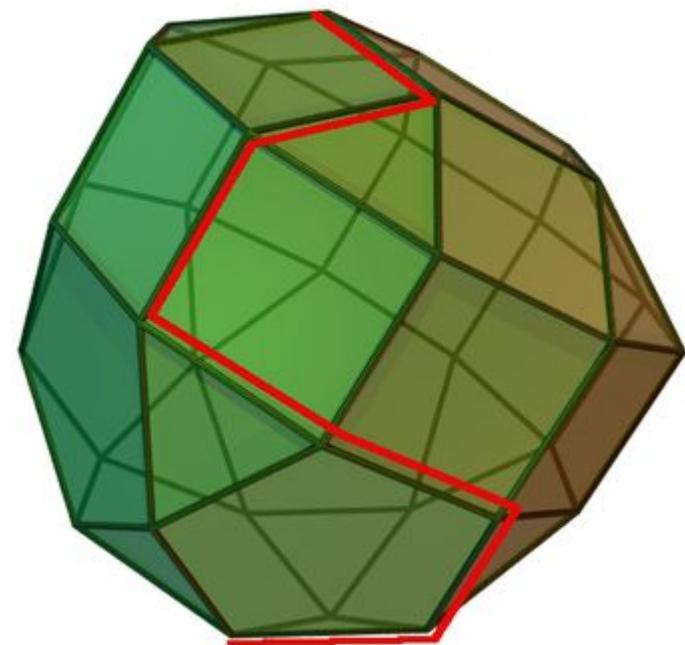
of value $z = 13$.

This solution is optimal.

All coefficients in z are negative and $x_2 \geq 0, x_4 \geq 0, x_6 \geq 0$, so $z \leq 13$.

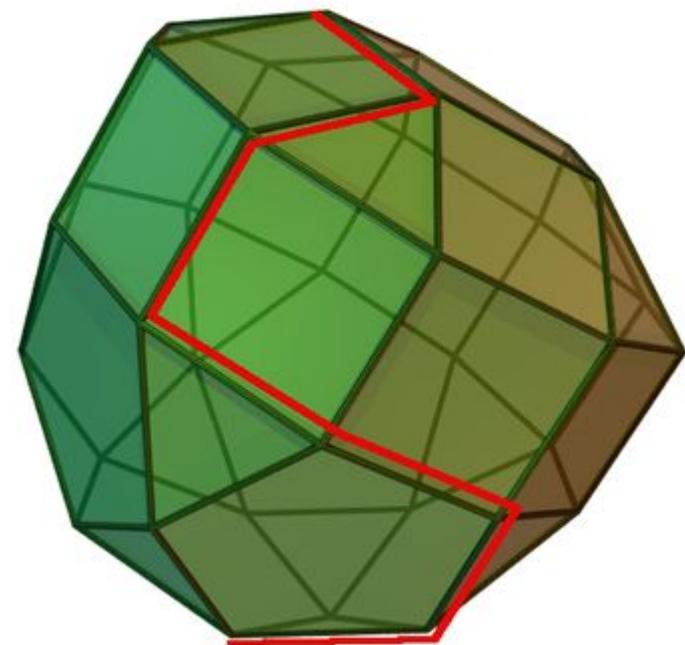
Let's recap

- We convert to canonical form to work along equalities
- We start along some corner
- We move as far as we can on the best edge
 - Until we are done



Let's recap

- We convert to canonical form to work along equalities
- We start along some corner
- We move as far as we can on the best edge
 - Until we are done
- **Usually solved with a simplex tableau**



Simplex tableau

$$\begin{array}{rcl}
 x_4 & = & 5 - 2x_1 - 3x_2 - x_3 \\
 x_5 & = & 11 - 4x_1 - x_2 - 2x_3 \\
 x_6 & = & 8 - 3x_1 - 4x_2 - 2x_3 \\
 \hline
 z & = & 5x_1 + 4x_2 + 3x_3.
 \end{array}$$

x1	x2	x3	x4	x5	x6	z	c
2	3	1	1	0	0	0	5
4	1	2	0	1	0	0	11
3	4	2	0	0	1	0	8
-5	-4	-3	0	0	0	1	0

Simplex tableau

x1	x2	x3	x4	x5	x6	Z	C
2	3	1	1	0	0	0	5 (5/2)
4	1	2	0	1	0	0	11 (11/4)
3	4	2	0	0	1	0	8 (8/3)
-5	-4	-3	0	0	0	1	0

x1	x2	x3	x4	x5	x6	Z	C
1	1.5	0.5	0.5	0	0	0	2.5
4	1	2	0	1	0	0	11
3	4	2	0	0	1	0	8
-5	-4	-3	0	0	0	1	0

Simplex tableau

x1	x2	x3	x4	x5	x6	Z	C
2	3	1	1	0	0	0	5 (5/2)
4	1	2	0	1	0	0	11 (11/4)
3	4	2	0	0	1	0	8 (8/3)
-5	-4	-3	0	0	0	1	0

x1	x2	x3	x4	x5	x6	Z	C
1	1.5	0.5	0.5	0	0	0	2.5
0	-5	0	-2	1	0	0	1
0	-0.5	0.5	-1.5	0	1	0	0.5
0	3.5	-0.5	2.5	0	0	1	12.5

Simplex tableau

$$\begin{array}{rcl}
 x_1 & = & \frac{5}{2} - \frac{3}{2}x_2 - \frac{1}{2}x_3 - \frac{1}{2}x_4 \\
 x_5 & = & 1 + 5x_2 + 2x_4 \\
 x_6 & = & \frac{1}{2} + \frac{1}{2}x_2 - \frac{1}{2}x_3 + \frac{3}{2}x_4 \\
 \hline
 z & = & \frac{25}{2} - \frac{7}{2}x_2 + \frac{1}{2}x_3 - \frac{5}{2}x_4.
 \end{array}$$

x1	x2	x3	x4	x5	x6	Z	C
1	1.5	0.5	0.5	0	0	0	2.5
0	-5	0	-2	1	0	0	1
0	-0.5	0.5	-1.5	0	1	0	0.5
0	3.5	-0.5	2.5	0	0	1	12.5

Simplex tableau

x1	x2	x3	x4	x5	x6	Z	C
1	1.5	0.5	0.5	0	0	0	2.5 (5)
0	-5	0	-2	1	0	0	1 (inf)
0	-0.5	0.5	-1.5	0	1	0	0.5 (1)
0	3.5	-0.5	2.5	0	0	1	12.5

x1	x2	x3	x4	x5	x6	Z	C
1	1.5	0.5	0.5	0	0	0	2.5
0	-5	0	-2	1	0	0	1
0	-1	1	-3	0	2	0	1
0	3.5	-0.5	2.5	0	0	1	12.5

Simplex tableau

x1	x2	x3	x4	x5	x6	Z	C
1	1.5	0.5	0.5	0	0	0	2.5 (5)
0	-5	0	-2	1	0	0	1 (inf)
0	-0.5	0.5	-1.5	0	1	0	0.5 (1)
0	3.5	-0.5	2.5	0	0	1	12.5

x1	x2	x3	x4	x5	x6	Z	C
1	2	0	2	0	-1	0	2
0	-5	0	-2	1	0	0	1
0	-1	1	-3	0	2	0	1
0	3	0	1	0	1	1	13

Simplex tableau

$$\begin{array}{rcl}
 x_3 & = & 1 + x_2 + 3x_4 - 2x_6 \\
 x_1 & = & 2 - 2x_2 - 2x_4 + x_6 \\
 x_5 & = & 1 + 5x_2 + 2x_4 \\
 \hline
 z & = & 13 - 3x_2 - x_4 - x_6.
 \end{array}$$

x1	x2	x3	x4	x5	x6	Z	C
1	2	0	2	0	-1	0	2
0	-5	0	-2	1	0	0	1
0	-1	1	-3	0	2	0	1
0	3	0	1	0	1	1	13

Algorithm

- Build tableau from canonical
- Check we have feasible solution
- Do a pivot step if negative coef
 - Pick column c w/ most negative coefficient
 - Pick row r w/ smallest ratio
 - Pivot! (set c to 1 in r and 0 in other rows)

