Lecture 4:
Convex optimization
And Duality

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

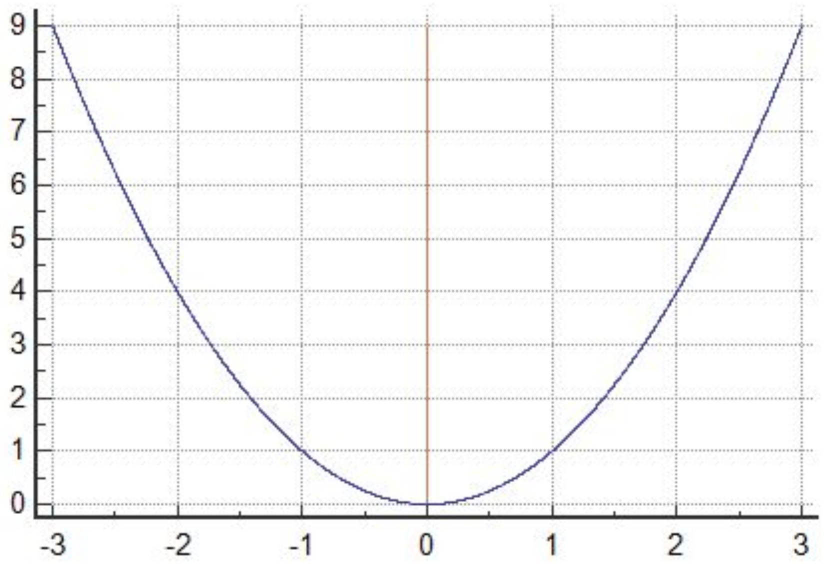
- **Find the best solution to a problem**
  - **Numerical optimization**
  - **Evaluate solutions according to a criterion**
  - **Look at solutions from some given space of possible solutions to consider**

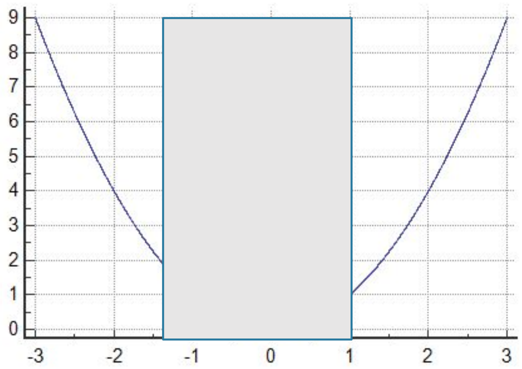
Defining an optimization problem

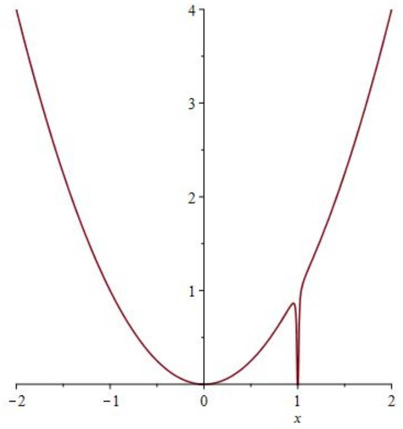
- Minimize a quantity $f_0(x)$
  - Under inequality and equality constraints
  - Constraints define a domain D
  - Could have no constraint except x in D

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

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)**
  - **Lagrangian and dual function**
  - **Dual problem**
  - **Qualification constraints**
  - **Geometric interpretation**
  - **KKT conditions**

- Newton’s Descent and Barrier methods for convex case

Course organization

- (First order) descent methods for the general case
- Backpropagation
- Some more properties on stochastic gradient descent

Evaluation

- 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 weeks

Convex problem

convex optimization problem:

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

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

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

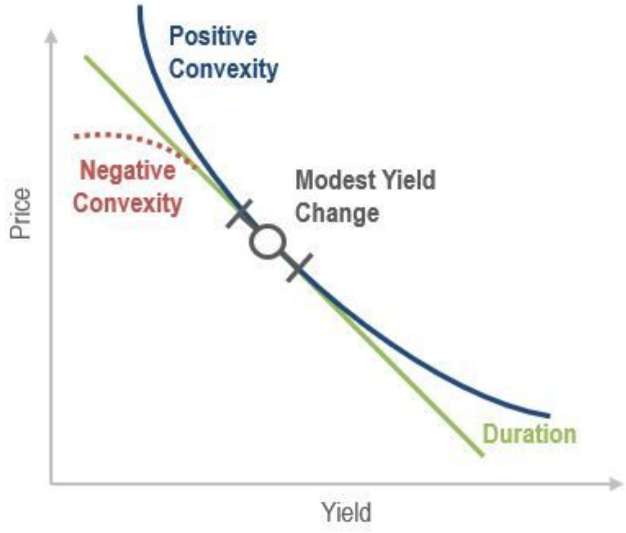
i.e., f_i have nonnegative (upward) curvature

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 overview

- **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

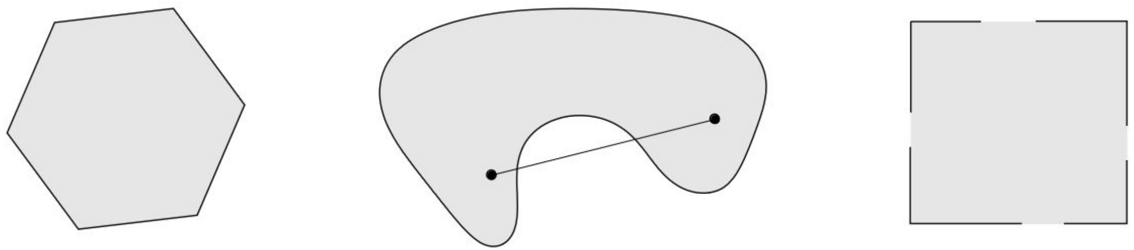
Convex sets

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

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

examples (one convex, two nonconvex sets)



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

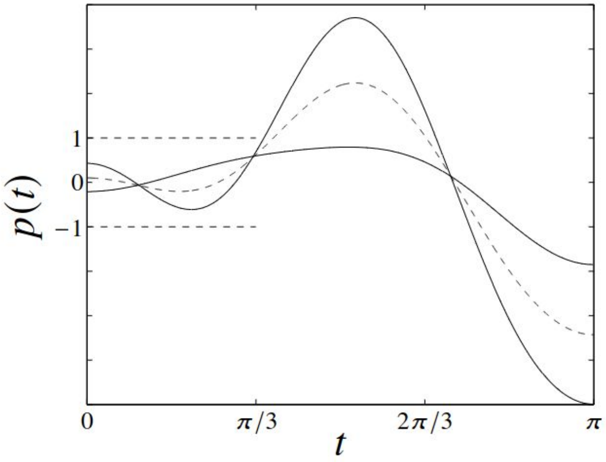
2. use convex functions (next lecture)

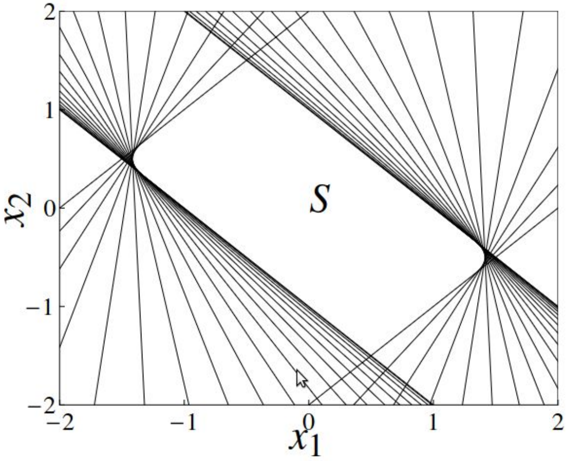
3. show that C is obtained from simple convex sets (hyperplanes, halfspaces, norm balls, ...) by operations that preserve convexity
   - intersection
   - affine mapping
   - perspective mapping
   - linear-fractional mapping

you'll mostly use methods 2 and 3

Showing a set is convex

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





Takeaway

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

- Classic convex sets
  - Affine sets, hyperplanes, cones, balls, polyhedrons

- Convexity preserving operations
  - Intersection
  - Affine mapping
  - Perspective
  - Linear Fractional mapping

Convex function overview

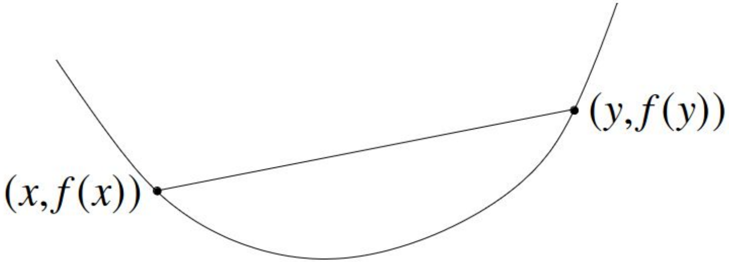
- **Convex functions (with definition!)**
  - **Examples of classic convex functions**
  - *Extended value function*
  - *Line restriction*
  - **First and second order conditions**
  - *Epigraph and sublevel sets*
- **Showing a function is convex with operations**
  - **Non-negative weighted sum and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

Convex functions

UNIVERSITÉ CÔTE D'AZUR

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

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



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

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

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 $\nabla^2 f(x) \succeq 0$
   - 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 $f_1, f_2, \ldots$ 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 $\displaystyle\int_{\alpha \in \mathcal{A}} f(x, \alpha)\, d\alpha$ 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

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

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

Showing a function is convex

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

examples

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

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

($x_{[i]}$ is ith 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

UNIVERSITÉ CÔTE D'AZUR

- composition of $g : \mathbf{R}^n \to \mathbf{R}$ and $h : \mathbf{R} \to \mathbf{R}$ 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

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

Standard form optimization problem

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

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

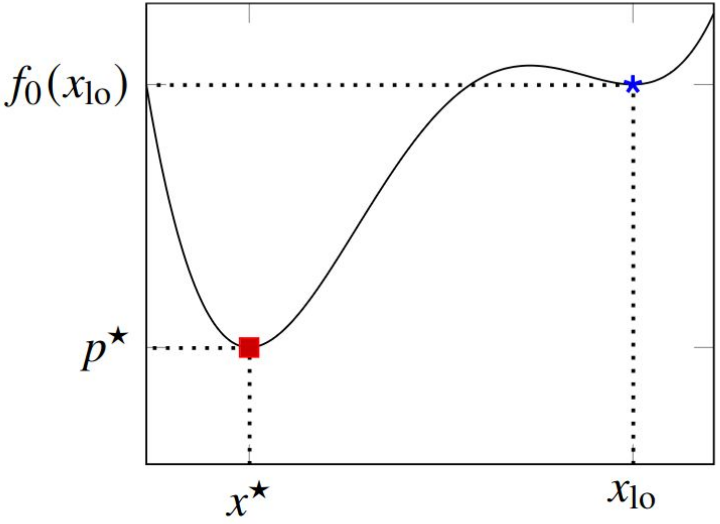
Feasible and optimal points

- $x \in \mathbf{R}^n$ is **feasible** if $x \in \mathbf{dom}\, f_0$ and it satisfies the constraints
- **optimal value** is $p^\star = \inf\{f_0(x) \mid f_i(x) \le 0,\ i = 1, \ldots, m,\ h_i(x) = 0,\ i = 1, \ldots, p\}$
- $p^\star = \infty$ 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

$$
\begin{array}{ll}
\text{minimize (over } z\text{)} & f_0(z) \\
\text{subject to} & f_i(z) \leq 0, \quad i = 1, \ldots, m, \quad h_i(z) = 0, \quad i = 1, \ldots, p \\
& \|z - x\|_2 \leq R
\end{array}
$$



Implicit and explicit constraints

standard form optimization problem has **implicit constraint**

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

- 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:

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

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

Standard form convex problem

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

- objective and inequality constraints $f_0, f_1, \ldots, f_m$ 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, $f_1, \ldots, f_m$ are convex, $h_1, \ldots, h_p$ 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

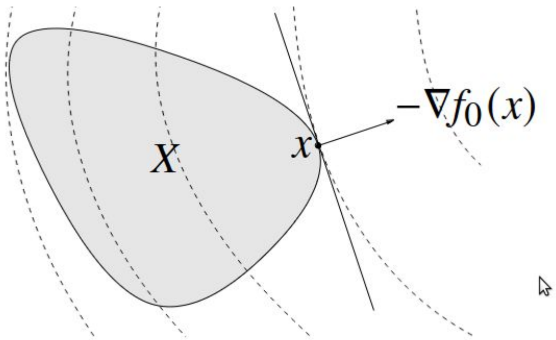
$$z \text{ feasible}, \quad \|z - x\|_2 \leq R \quad \Longrightarrow \quad f_0(z) \geq f_0(x)$$

- 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

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



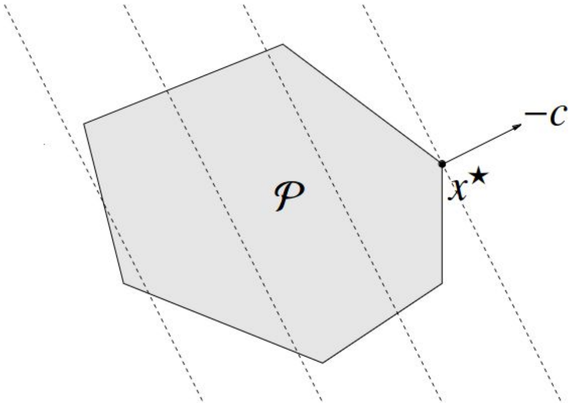
- if nonzero, $\nabla f_0(x)$ defines a supporting hyperplane to feasible set X at x

Linear programming

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

<u>Often written as a maximization</u>

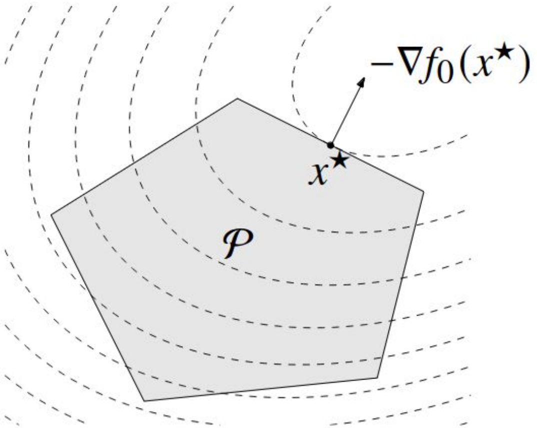
- convex problem with affine objective and constraint functions
- feasible set is a polyhedron



Quadratic programming

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

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



Quadratically constrained Quadratic programming (QCQP)

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

- $P_i \in \mathbf{S}^n_+$; objective and constraints are convex quadratic
- if $P_1, \ldots, P_m \in \mathbf{S}^n_{++}$, 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

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

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

$$\begin{array}{lll} \text{minimize} & \tilde{f}_0(z) & \\ \text{subject to} & \tilde{f}_i(z) \leq 0, & i = 1, \ldots, m \\ & \tilde{h}_i(z) = 0, & i = 1, \ldots, p \end{array}$$

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 $u \leq 0$, $i = 1, \ldots, m$
- $\varphi_i(u) = 0$ if and only if $u = 0$, $i = 1, \ldots, p$

standard form optimization problem is equivalent to

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

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

Maximization and minimization

- suppose ϕ_0 is monotone decreasing
- the maximization problem

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

  is equivalent to the minimization problem

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

- **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

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

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

Introducing equality constraints

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

is equivalent to

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

Slack variables for linear equalities

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

is equivalent to

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

Slack variables for linear equalities

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

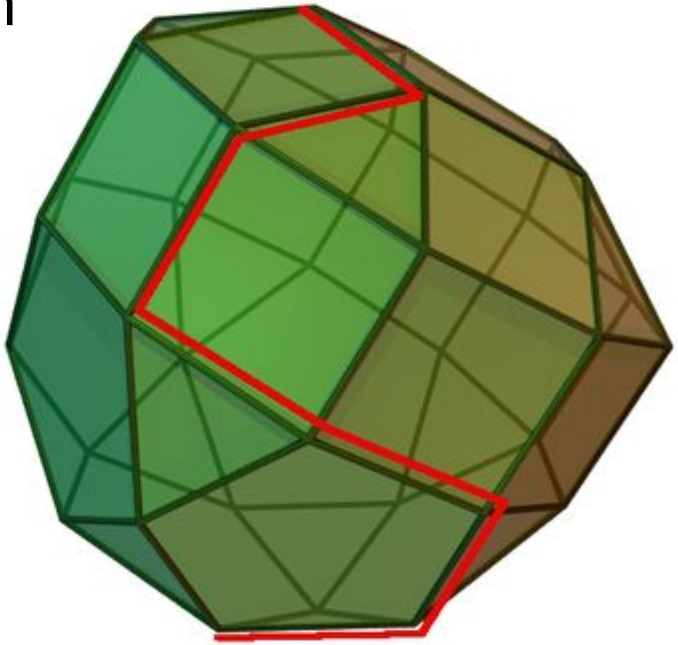
- Convex f and linear h
    - X feasible: satisfies implicit and explicit constraints

- Quite a few classical convex problems(linear, quadratic, …)

- Easy to change variables between equivalent problems

UNIVERSITÉ CÔTE D'AZUR

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



Standard form LP

$$
\begin{array}{lrcrcrcr}
\text{Maximize} & 5x_1 & + & 4x_2 & + & 3x_3 & & \\
\text{Subject to :} & & & & & & & \\
& 2x_1 & + & 3x_2 & - & x_3 & \leq & 5 \\
& 4x_1 & + & x_2 & - & 2x_3 & \leq & 11 \\
& 3x_1 & + & 4x_2 & - & 2x_3 & \leq & 8 \\
& & & & & x_1, x_2, x_3 & \geq & 0.
\end{array}
$$

Standard form LP

$$
\begin{array}{lrcrcrcr}
\boxed{\text{Maximize}} & 5x_1 & + & 4x_2 & + & 3x_3 & & \\
\text{Subject to :} & & & & & & & \\
& 2x_1 & + & 3x_2 & - & x_3 & \leq & 5 \\
& 4x_1 & + & x_2 & - & 2x_3 & \leq & 11 \\
& 3x_1 & + & 4x_2 & - & 2x_3 & \leq & 8 \\
& & & & & x_1, x_2, x_3 & \geq & 0.
\end{array}
$$

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

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

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

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

$$x_4 \geq 0.$$

Add slack variables

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

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

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

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

$$x_4 \geq 0.$$

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

Canonical form LP

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

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

Canonical form LP

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

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

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

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

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

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

$$z = 0$$

Solving the simplex dictionary

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

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

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.

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

Here, we try to increase x_1.

How far can we go?

How much can we increase x_1?

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

We have $\boxed{x_4 \geq 0}$.

It implies $5 - 2x_1 \geq 0$, that is $\boxed{x_1 \leq \frac{5}{2}}$.

How far can we go?

How much can we increase x_1?

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

We have $\boxed{x_4 \geq 0}$.

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

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?

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

We have $\boxed{x_4 \geq 0}$.

It implies $5 - 2x_1 \geq 0$, that is $\boxed{x_1 \leq 5/2}$ 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?

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

We have $\boxed{x_4 \geq 0}$.

It implies $5 - 2x_1 \geq 0$, that is $\color{red}{\boxed{x_1 \leq 5/2}}$ 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.

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

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.

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

Pivot around the chosen variable

Finally, we get the new dictionary:

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

New system

UNIVERSITÉ CÔTE D'AZUR

Finally, we get the new dictionary:

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

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: $z = 25/2$.

New step!

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

New step of the simplex:

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

And we're done!

New feasible dictionary:

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

With new solution:

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

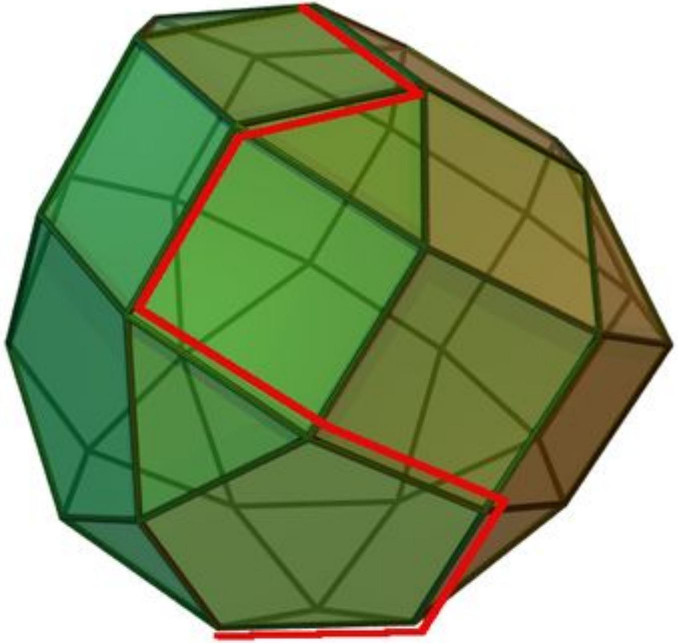
of value $z = 13$.

$\boxed{\text{This solution is optimal.}}$

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

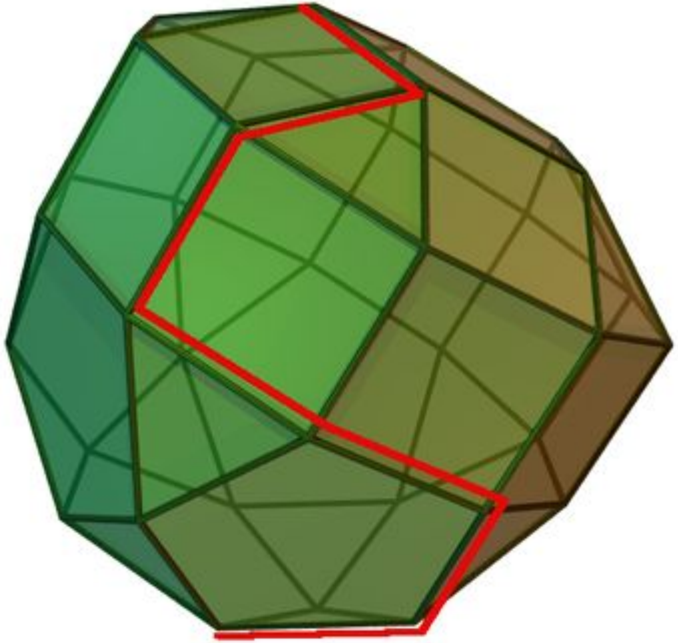
Let's recap

- 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

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

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

UNIVERSITÉ CÔTE D'AZUR

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

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

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

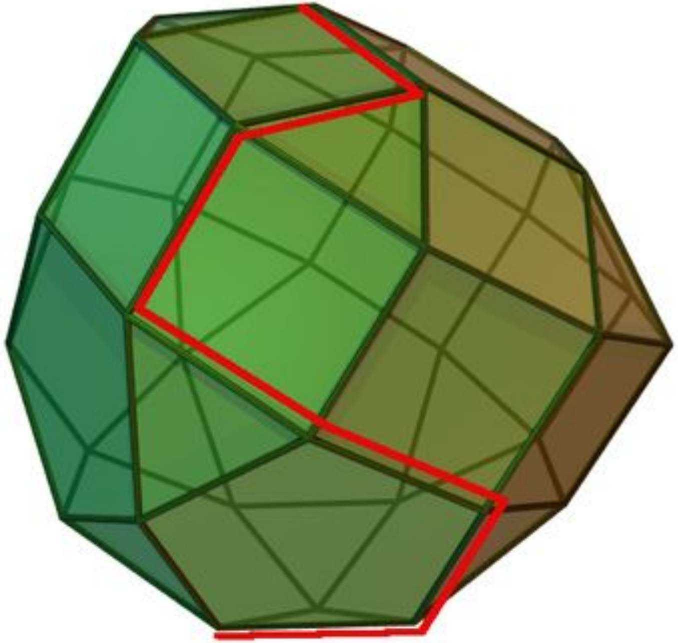
Simplex tableau

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

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)



1. Lagrangian and dual function

Unconstrained problem version

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

Unconstrained problem version

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

$$I_-(u) = \begin{cases} 0 & u \leq 0 \\ \infty & u > 0, \end{cases}$$

$$\text{minimize} \quad f_0(x) + \sum_{i=1}^{m} I_-(f_i(x)) + \sum_{i=1}^{p} I_0(h_i(x)),$$

Relaxing the indicator function

UNIVERSITÉ CÔTE D'AZUR

$$I_-(u) = \begin{cases} 0 & u \leq 0 \\ \infty & u > 0, \end{cases}$$

$$\text{minimize} \quad f_0(x) + \textstyle\sum_{i=1}^{m} I_-(f_i(x)) + \sum_{i=1}^{p} I_0(h_i(x)),$$

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

Lagrangian

UNIVERSITÉ CÔTE D'AZUR

- **standard form problem** (not necessarily convex)

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

variable $x \in \mathbf{R}^n$, domain $\mathcal{D}$, optimal value $p^\star$

- **Lagrangian:** $L : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$, with $\mathbf{dom}\, L = \mathcal{D} \times \mathbf{R}^m \times \mathbf{R}^p$,

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

  - weighted sum of objective and constraint functions
  - λ_i is **Lagrange multiplier** associated with $f_i(x) \le 0$
  - ν_i is Lagrange multiplier associated with $h_i(x) = 0$

Lagrange dual

UNIVERSITÉ CÔTE D'AZUR

- **Lagrange dual function:** $g : \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$,

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)$$

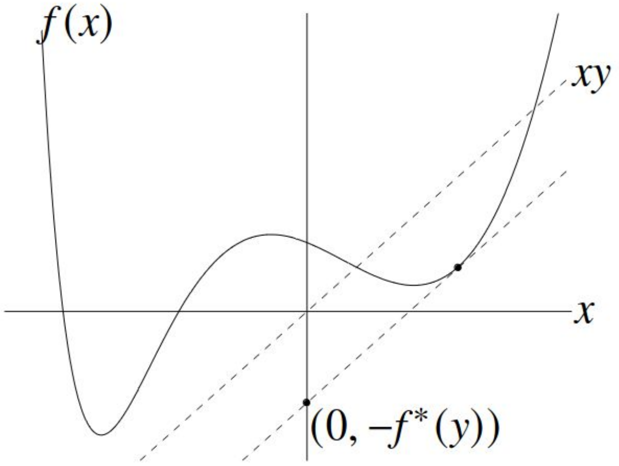
- g is concave, can be $-\infty$ for some λ, ν
- **lower bound property:** if $\lambda \succeq 0$, then $g(\lambda, \nu) \le p^\star$
- proof: if $\tilde{x}$ is feasible and $\lambda \succeq 0$, then

$$f_0(\tilde{x}) \ge L(\tilde{x}, \lambda, \nu) \ge \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = g(\lambda, \nu)$$

minimizing over all feasible $\tilde{x}$ gives $p^\star \ge g(\lambda, \nu)$

Conjugate

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



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

Lagrange dual and conjugate

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & Ax \preceq b, \quad Cx = d \end{array}$$

- dual function

$$\begin{aligned} g(\lambda, \nu) &= \inf_{x \in \mathbf{dom} f_0} \left(f_0(x) + (A^T\lambda + C^T\nu)^T x - b^T\lambda - d^T\nu \right) \\ &= -f_0^*(-A^T\lambda - C^T\nu) - b^T\lambda - d^T\nu \end{aligned}$$

  where $f^*(y) = \sup_{x \in \mathbf{dom} f}(y^Tx - f(x))$ is conjugate of f_0
- simplifies derivation of dual if conjugate of f_0 is known
- **example: entropy maximization**

$$f_0(x) = \sum_{i=1}^n x_i \log x_i, \qquad f_0^*(y) = \sum_{i=1}^n e^{y_i - 1}$$

2. Dual Problem

Dual problem

(Lagrange) **dual problem**

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

- finds best lower bound on $p^\star$, obtained from Lagrange dual function
- a convex optimization problem, even if original **primal** problem is not
- dual optimal value denoted $d^\star$
- λ, ν are dual feasible if $\lambda \succeq 0$, $(\lambda, \nu) \in \mathbf{dom}\, g$
- often simplified by making implicit constraint $(\lambda, \nu) \in \mathbf{dom}\, g$ explicit

Weak and strong duality

weak duality: $d^\star \leq p^\star$

- always holds (for convex and nonconvex problems)
- can be used to find nontrivial lower bounds for difficult problems, *e.g.*, solving the SDP

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

  gives a lower bound for the two-way partitioning problem on page 5.7

strong duality: $d^\star = p^\star$

- does not hold in general
- (usually) holds for convex problems
- conditions that guarantee strong duality in convex problems are called **constraint qualifications**

Slater's constraint qualification

strong duality holds for a convex problem

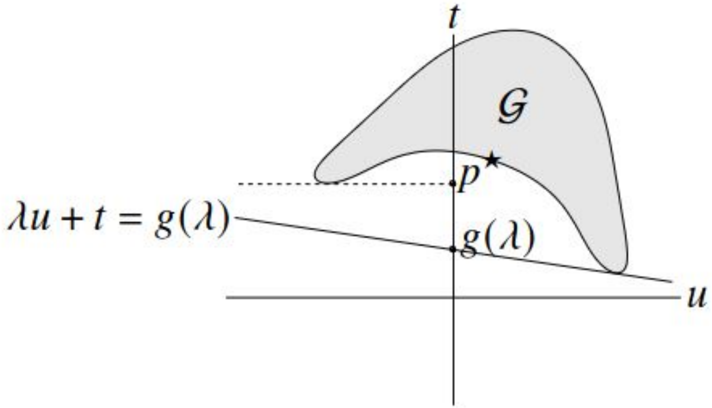
$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \dots, m \\ & Ax = b \end{array}$$

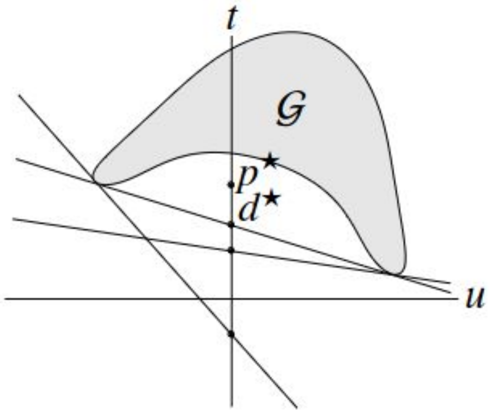
if it is **strictly feasible**, *i.e.*, there is an $x \in \mathbf{int}\,\mathcal{D}$ with $f_i(x) < 0$, $i = 1, \dots, m$, $Ax = b$

- also guarantees that the dual optimum is attained (if $p^\star > -\infty$)
- can be sharpened: *e.g.*,
  - can replace $\mathbf{int}\,\mathcal{D}$ with $\mathbf{relint}\,\mathcal{D}$ (interior relative to affine hull)
  - affine inequalities do not need to hold with strict inequality
- there are many other types of constraint qualifications

Geometric interpretation

- for simplicity, consider problem with one constraint $f_1(x) \leq 0$
- $\mathcal{G} = \{(f_1(x), f_0(x)) \mid x \in \mathcal{D}\}$ is set of achievable (constraint, objective) values
- **interpretation of dual function:** $g(\lambda) = \inf_{(u,t)\in\mathcal{G}}(t + \lambda u)$

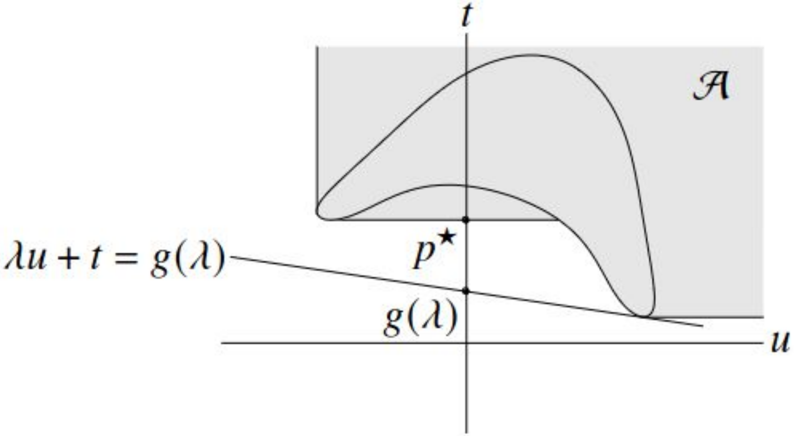




- $\lambda u + t = g(\lambda)$ is (non-vertical) supporting hyperplane to $\mathcal{G}$
- hyperplane intersects t-axis at $t = g(\lambda)$

Geometric interpretation

- same with $\mathcal{G}$ replaced with $\mathcal{A} = \{(u,t) \mid f_1(x) \le u,\ f_0(x) \le t \text{ for some } x \in \mathcal{D}\}$



- strong duality holds if there is a non-vertical supporting hyperplane to $\mathcal{A}$ at $(0, p^\star)$
- for convex problem, $\mathcal{A}$ is convex, hence has supporting hyperplane at $(0, p^\star)$
- Slater's condition: if there exist $(\tilde{u}, \tilde{t}) \in \mathcal{A}$ with $\tilde{u} < 0$, then supporting hyperplane at $(0, p^\star)$ must be non-vertical

3. KKT Conditions

Complementary slackness

- assume strong duality holds, $x^\star$ is primal optimal, $(\lambda^\star, \nu^\star)$ is dual optimal

$$\begin{aligned} f_0(x^\star) = g(\lambda^\star, \nu^\star) &= \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p \nu_i^\star h_i(x) \right) \\ &\leq f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star f_i(x^\star) + \sum_{i=1}^p \nu_i^\star h_i(x^\star) \\ &\leq f_0(x^\star) \end{aligned}$$

- hence, the two inequalities hold with equality
- $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$
- $\lambda_i^\star f_i(x^\star) = 0$ for $i = 1, \dots, m$ (known as **complementary slackness**):

$$\lambda_i^\star > 0 \implies f_i(x^\star) = 0, \qquad f_i(x^\star) < 0 \implies \lambda_i^\star = 0$$

Karush-Kuhn-Tucker (KKT) conditions

UNIVERSITÉ CÔTE D'AZUR

the **KKT conditions** (for a problem with differentiable f_i, h_i) are

1. primal constraints: $f_i(x) \leq 0$, $i = 1, \ldots, m$, $h_i(x) = 0$, $i = 1, \ldots, p$
2. dual constraints: $\lambda \succeq 0$
3. complementary slackness: $\lambda_i f_i(x) = 0$, $i = 1, \ldots, m$
4. gradient of Lagrangian with respect to x vanishes:

$$\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0$$

if strong duality holds and x, λ, ν are optimal, they satisfy the KKT conditions

KKT for convex problems

if $\tilde{x}$, $\tilde{\lambda}$, $\tilde{\nu}$ satisfy KKT for a convex problem, then they are optimal:

- from complementary slackness: $f_0(\tilde{x}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$
- from 4th condition (and convexity): $g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$

hence, $f_0(\tilde{x}) = g(\tilde{\lambda}, \tilde{\nu})$

if Slater's condition is satisfied, then

x is optimal if and only if there exist λ, ν that satisfy KKT conditions

- recall that Slater implies strong duality, and dual optimum is attained
- generalizes optimality condition $\nabla f_0(x) = 0$ for unconstrained problem