Convex optimization

Why ? What ? How ?

François Glineur

Université catholique de Louvain - EPL/INMA & CORE
Francois.Glineur@uclouvain.be

eVITA Winter School on eScience on Optimization   January 13th 2009

Questions and comments ...

... are more than welcome, at any time !

Slides will be available on the web :

http://www.core.ucl.ac.be/~glineur/

References

This lecture's material relies on several references (see at the end), but most main ideas can be found in:

- Convex Optimization,

  Stephen Boyd and Lieven Vandenberghe,

  Cambridge University Press, 2004 (and on the web)

Motivation

Modelling and decision-making

Help to choose the best decision

$$\left.\begin{array}{rcl} \text{Decision} & \leftrightarrow & \text{vector of variables} \\ \text{Best} & \leftrightarrow & \text{objective function} \\ \text{Constraints} & \leftrightarrow & \text{feasible domain} \end{array}\right\} \Rightarrow \text{Optimization}$$

Use

- Numerous applications in practice
- Resolution methods efficient in practice
- Modelling and solving large-scale problems

Introduction

Applications

- Planning, management and scheduling

  Supply chain, timetables, crew composition, etc.
- Design

  Dimensioning, structural optimization, networks
- Economics and finance

  Portfolio optimization, computation of equilibrium
- Location analysis and transport

  Facility location, circuit boards, vehicle routing
- And lots of others ...

Two facets of optimization

- Modelling

  Translate the problem into mathematical language (sometimes trickier than you might think)

$$\Updownarrow$$

Formulation of an optimization problem

$$\Updownarrow$$

- Solving

  Develop and implement algorithms that are efficient in *theory* and in *practice*

Close relationship

◇ Formulate models that you know how to solve

$$\updownarrow$$

◇ Develop methods applicable to real-world problems

Classical formulation

$$\min_{x \in \mathbb{R}^n} f(x) \text{ such that } x \in X \subseteq \mathbb{R}^n$$

(finite dimension)

Often, we define

$$X = \{x \in \mathbb{R}^n \mid g_i(x) \leq 0 \text{ and } h_j(x) = 0 \text{ for } i \in \mathcal{I}, j \in \mathcal{E}\}$$

Possible situations: optimal value

$$\min_{x \in \mathbb{R}^n} f(x) \text{ such that } x \in X \subseteq \mathbb{R}^n$$

Optimal value $f^* = \inf\{f(x) \mid x \in X\}$

a. $X = \emptyset$: infeasible problem (convention: $f^* = +\infty$)

b. $X \neq \emptyset$: feasible problem ; in this case

   (a) $f^* > -\infty$: bounded problem

   (b) $f^* = -\infty$: unbounded problem

Possible situations: optimal set

$$\min_{x \in \mathbb{R}^n} f(x) \text{ such that } x \in X \subseteq \mathbb{R}^n$$

Optimal value f^* is not always attained

Consider the optimal set $X^* = \{x^* \in X \mid f(x^*) = f^*\}$

a. $X^* \neq \emptyset$: solvable problem
(at least one optimal solution)

b. $X^* = \emptyset$: unsolvable problem.

There exists feasible, bounded unsolvable problems !

$\min \frac{1}{x}$ such that $x \in \mathbb{R}_+$ gives $f^* = 0$ but $X^* = \emptyset$

Convex optimization: plan

Why

a. Nice case: linear optimization

b. Algorithms and guarantees

What

a. Convex problems: definitions and examples

How

a. Algorithms: interior-point methods

b. Guarantees: duality

c. Framework: conic optimization

Convex optimization: plan

Why

a. Nice case: linear optimization

b. Algorithms and guarantees

What

a. Convex problems: definitions and examples

How

a. Algorithms: interior-point methods

b. Guarantees: duality

c. Framework: conic optimization

Linear optimization: three examples

A. Diet problem

Consider a set of different foods for which you know

- ◇ Quantities of calories, proteins, glucids, lipids, vitamins contained per unit of weight
- ◇ Price per unit of weight

Given the nutritional recommendations with respect to daily supply of proteins, glucids, etc, design an optimal, i.e. meeting the constraints with the lowest cost

Formulation

- Index i for the food types $(1 \leq i \leq n)$
- Index j for the nutritional components $(1 \leq j \leq m)$
- Data (per unit of weight) :

  $c_i \longrightarrow$ price of food type i,

  $a_{ji} \longrightarrow$ amount of component j in food type i,

  $b_j \longrightarrow$ daily recommendations for component j
- Unknowns:

  Quantity x_i of food type i in the optimal diet

Formulation (continued)

This is a linear problem:

$$\min \sum_{i=1}^{n} c_i x_i$$

such that

$$x_i \geq 0 \ \forall i \text{ and } \sum_{i=1}^{n} a_{ji} x_i = b_j \ \forall j$$

Using matrix notations

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \geq 0$$

This is a one of the most simple problems, and can be solved for large dimensions
(1947: 9×77 ; today: m and $n \approx 10^7$)

B. Assignment problem

Given

- n workers
- n tasks to accomplish
- the amount of time needed for each worker to execute each of the tasks

Assign (bijectively) the n tasks to the n workers so that the total execution time is minimized

This is a discrete problem with an (a priori) exponential number of potential solutions ($n!$)
$\rightarrow$ explicit enumeration is impossible in practice

Formulation

First idea: x_i denotes the number of the task assigned to person i (n integer variables between 1 and n)

Problem : how to force a bijection ?

Better formulation:

- ◇ Index i for workers ($1 \leq i \leq n$)
- ◇ Index j for tasks ($1 \leq j \leq n$)
- ◇ Data :

  $a_{ij} \longrightarrow$ duration of task j for worker i
- ◇ Unknowns:

  x_{ij} binary variable $\{0, 1\}$ indicating whether worker i executes task j

Formulation (continued)

$$\min \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_{ij}$$

such that

$$\sum_{i=1}^{n} x_{ij} = 1 \ \forall j, \ \sum_{j=1}^{n} x_{ij} = 1 \ \forall i, \quad \text{and } x_{ij} \in \{0,1\} \ \forall i \ \forall j$$

- ◇ Higher number of variables (n^2) $\longrightarrow$ more difficult ?
- ◇ Linear problem with integer (binary) variables
  $\longrightarrow$ requires completely different algorithms
- ◇ But bijection constraint is simplified and linearized

Although its looks more difficult than A., this problem can also be solved very efficiently !

C. Travelling salesman problem

Given

- ◇ a travelling salesman that has to visit n cities going through each city once and only once
- ◇ the distance (or duration of the journey) between each pair of cities

Find an optimal tour that visits each city once with minimal length (or duration)

Also a discrete and exponential problem

Other application : soldering on circuit boards

Formulation

First idea: x_i describes city visited in position i during the tour (n integer variables between 1 and n)
Problem : how to require that each city is visited ?

Better formulation:

◇ Indices i and j for the cities ($1 \leq i, j \leq n$)

◇ Data :

$a_{ij} \rightarrow$ distance (or journey duration) between i and j

◇ Unknowns:

x_{ij} binary variable $\{0, 1\}$ indicating whether the trip from city i to city j is part of the trip

Formulation (continued)

$$\min \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_{ij}$$

such that

$$\sum_{i=1}^{n} x_{ij} = 1 \ \forall j, \ \sum_{j=1}^{n} x_{ij} = 1 \ \forall i, x_{ij} \in \{0,1\} \ \forall i \ \forall j$$

and $\displaystyle\sum_{i \in S, j \notin S} x_{ij} \geq 1 \ \forall S$ with $S \subseteq \{1, \ldots, n\}, 1 < |S| < n$

◇ High (exponential) number of constraints

◇ Problem is a lot harder than A./B. ($\max n \approx 10^4$)

Algorithms and complexity

Why are these three problems different ?

Three linear problems: a priori among the simplest ... ?

- ◇ A. Diet: continuous variables
  $\longrightarrow$ (continuous) linear optimization
- ◇ B. Assignment: discrete variables + expon. # of soln.
  $\longrightarrow$ linear integer optimization
- ◇ C. Salesman: discrete variables + exp. # of cons./soln.
  $\longrightarrow$ linear integer optimization

However, B is not more difficult than A
while C is a lot harder than A and B !

Convex optimization: plan

Why

a. Nice case: linear optimization

b. Algorithms and guarantees

What

a. Convex problems: definitions and examples

How

a. Algorithms: interior-point methods

b. Guarantees: duality

c. Framework: conic optimization

Algorithmic complexity

Difficulty of a problem depends on the

efficiency of methods that can be applied to solve it

$\Rightarrow$ what is a good algorithm ?

- Solves the problem (approximately)
- Until the middle of the 20th century:

  in finite time (number of elementary operations)
- Now (computers):

  in bounded time (depending on the problem size)

  $\longrightarrow$ algorithmic complexity (worst / average case)

Big distinction: polynomial $\leftrightarrow$ exponential complexity

Algorithms for linear optimization

For linear optimization with continuous variables:
very efficient algorithms ($n \approx 10^7$)

- ◇ Simplex algorithm (Dantzig, 1947)
  Exponential worst-case complexity but ...
  Very efficient in practice (worst-case is rare)
- ◇ Ellipsoid method (analyzed by Khachiyan, 1978)
  Polynomial worst-case complexity but ...
  Poor practical performance (high-degree polynomial)
- ◇ Interior-point methods (Karmarkar, 1985)
  Polynomial worst-case complexity and ...
  Very efficient in practice (large-scale problems)

Algorithms for linear optimization (continued)

For linear optimization with discrete variables:
algorithms a lot less efficient, because problem is intrinsically exponential (cf. class of *NP-complete* problems)

◇ Continuous relaxation (i.e. outer approximation)

◇ Branch and bound

(i.e. explore an exponential solution tree + pruning)

→ Very sophisticated algorithms/heuristics
but still exponential worst-case

→ Middle-scale or even small-scale problems ($n \approx 10^2$) can already be intractable

→ Discrete C. is a lot harder to solve than continuous A.

What about the assignment problem B. ?

Why can it be solved efficiently, despite being discrete ? One can relax variables $x_{ij} \in \{0,1\}$ by $0 \leq x_{ij} \leq 1$ without changing the optimal value and solutions !

$\longrightarrow$ it was a fake discrete problem

$\longrightarrow$ we obtain a continuous linear optimization formulation

$\longrightarrow$ an example of why reformulation is sometimes crucial

In general, if one can replace the binary variables by continuous variables with an additional polynomial number of linear constraints, the resulting problem can be solved in polynomial time

Combinatorial/integer/discrete problems
are not always difficult !

Nonlinear vs. convex optimization

Why nonlinear optimization ?

$$\min_{x \in \mathbb{R}^n} f(x) \text{ such that } x \in X \subseteq \mathbb{R}^n$$

where X is defined (most of the time) by

$$X = \{x \in \mathbb{R}^n \mid g_i(x) \leq 0 \text{ and } h_j(x) = 0 \text{ for } i \in \mathcal{I}, j \in \mathcal{E}\}$$

Linear optimization: any affine functions for f, g_i and h_j but it does not permit satisfactory modelling of all practical problems

$\longrightarrow$ need to consider nonlinear f, g_i and h_j
$\longrightarrow$ nonlinear optimization

A taxonomy

- ◇ Deterministic or stochastic problem
- ◇ Accurate data or inaccurate/fuzzy (robustness)
- ◇ Single or multiple objectives
- ◇ Constrained or unconstrained problem
- ◇ Functions described analytically or using a black box
- ◇ Continuous functions or not, differentiable or not
- ◇ General, polynomial, quadratic, linear functions
- ◇ Continuous or discrete variables

Switch categories: sometimes with *reformulations*

Back to complexity

Discrete sets X can make the problem difficult (with exponential complexity)
but even continuous problems can be difficult!

Consider a simple unconstrained minimization

$$\min f(x_1, x_2, \ldots, x_{10})$$

with smooth f (Lipschitz continuous with $L = 2$):

One can show that for any algorithm there exists some functions where at least 10^{20} iterations (function evaluations) are needed to find a global solution with accuracy better than 1% ! (this is a theorem)

Two paradigms

- ◇ Tackle all problems without any efficiency guarantee
  - Traditional **nonlinear** optimization
  - (Meta)-Heuristic methods
- ◇ Limit the scope to some classes of problems and get in return an efficiency guarantee (complexity)
  - Linear optimization
    - ∗ very fast specialized algorithms
    - ∗ but sometimes too limited in practice
  - **Convex** optimization (this lecture)
    - ∗ (slightly) less efficient but much more general

Compromise: generality $\leftrightarrow$ efficiency

Convex optimization: plan

Why

a. Nice case: linear optimization

b. Algorithms and guarantees

What

a. Convex problems: definitions and examples

How

a. Algorithms: interior-point methods

b. Guarantees: duality

c. Framework: conic optimization

Convex optimization

Introduction

$$\min f(x) \text{ such that } x \in X$$

A feasible solution x^* is a

- global minimum iff $f(x^*) \leq f(x)\ \forall x \in X$
- local minimum iff there exists an open neighborhood $V(x^*)$ such that

$$f(x^*) \leq f(x)\ \forall x \in X \cap V\ .$$

Global minimum $\Rightarrow$ local minimum
Global minima are more interesting but also more difficult to find ... but the notion of convexity can help us !

Convexity definitions

- A set $S \subseteq \mathbb{R}^n$ is convex iff
  $\lambda x + (1-\lambda)y \in S \; \forall x, y \in S, \lambda \in [0\ 1]$
- A function $f : S \mapsto R$ is convex iff
  $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) \; \forall x, y, \lambda \in [0\ 1]$
  (this imposes that the domain S is convex)
- Equivalently, a function $f : S \subseteq \mathbb{R}^n \mapsto \mathbb{R}$ is convex iff its epigraph is convex

$$\text{epi} f = \{(x, t) \in \mathbb{R}^n \times \mathbb{R} \mid x \in S \text{ and } f(x) \leq t\}$$

- An *optimization* problem is *convex* if it deals with the minimization of a convex function on a convex set

Examples

- $\emptyset, \mathbb{R}^n, \mathbb{R}^n_+, \mathbb{R}^n_{++}$
- $\{x \mid \|x-a\| < r\}$ and $\{x \mid \|x-a\| \leq r\}$
- $\{x \mid b^\mathrm{T} x < \beta\}$, $\{x \mid b^\mathrm{T} x \leq \beta\}$ and $\{x \mid b^\mathrm{T} x = \beta\}$
- In $\mathbb{R}$: intervals (open/closed, possibly infinite)
- $x \mapsto c$, $x \mapsto b^\mathrm{T} y + \beta_0$, $x \mapsto \|x\|$ and $x \mapsto \|x\|^2$, $x \mapsto x^\mathrm{T} Q x$ with $Q \in \mathbb{R}^{n \times n}$ positive semidefinite
- In the case $f : \mathbb{R} \mapsto \mathbb{R}$, we mention $x \mapsto e^x$, $x \mapsto -\log x$, $x \mapsto |x|^p$ with $p \geq 1$.
- f is concave iff $-f$ is convex (i.e. reversing inequalities in the definitions) ; there is no notion of concave set!

Fundamental properties of convex optimization

$$\min_{x \in \mathbb{R}^n} f(x) \text{ such that } x \in X \subseteq \mathbb{R}^n$$

When

- ◇ f is a convex function to be minimized
- ◇ X is a convex set

we are dealing with convex optimization problems and

- ◇ Every local minimum is global
- ◇ The optimal set is convex
- ◇ The KKT optimality conditions are sufficient

Basic properties of convex sets

◇ If two sets $S \subseteq \mathbb{R}^n$ and $T \subseteq \mathbb{R}^n$ are convex, so is their intersection $S \cap T \subseteq \mathbb{R}^n$

◇ If two sets $S \subseteq \mathbb{R}^n$ and $T \subseteq \mathbb{R}^m$ are convex, so is their Cartesian product $S \times T \subseteq \mathbb{R}^{n+m}$

◇ For every set $X \subseteq \mathbb{R}^n$, there is a smallest convex set $S \subseteq \mathbb{R}^n$ which includes X, called the convex hull of X

a. all nonlinear problems admit a convex relaxation

b. for a linear objective function (which can be taken w.l.o.g.) this relaxation is exact
(but this does not really help us ...)

A linear objective ?

$$\min_{x \in \mathbb{R}^n} f(x) \text{ such that } x \in X \subseteq \mathbb{R}^n$$

$$\Updownarrow$$

$$\min_{(x,t) \in \mathbb{R}^n \times \mathbb{R}} t \text{ such that } x \in X \text{ and } (x,t) \in \operatorname{epi} f$$

$$\Updownarrow$$

$$\min_{(x,t) \in \mathbb{R}^n \times \mathbb{R}} t \text{ such that } x \in X \text{ and } f(x) \leq t$$

$\Rightarrow$ equivalent convex problem with linear objective

Basic properties of convex functions

- ◇ If two functions $f(x)$ and $g(x)$ are convex
  - Product $af(x)$ is convex for any scalar $a \geq 0$
  - Sum $f(x) + g(x)$ is convex
  - Maximum $\max\{f(x), g(x)\}$ is convex
- ◇ If f is twice differentiable, we have

$$f \text{ convex} \Leftrightarrow \nabla^2 f \succeq 0$$

- ◇ The only functions that are simultaneously convex and concave are the affine functions

Convexity plays nice with linearity

- ◇ If $S \subseteq \mathbb{R}^n$ is convex and $\Phi : \mathbb{R}^n \to \mathbb{R}^m : x \mapsto Ax + b$ a linear function, we have that

$$\Phi S = \{\Phi(x) \mid x \in S\} \text{ is convex}$$

- ◇ This implies that if $f : x \mapsto f(x)$ is a convex function

$$g : x \mapsto g(x) = f(Ax + b) \text{ is convex}$$

(but of course not always true for $af(x) + b$!)

- ◇ Similar result holds for $\Theta : \mathbb{R}^m \to \mathbb{R}^n : x \mapsto Ax + b$ and

$$\Theta^{-1} S = \{x \mid \Theta(x) \in S\} \text{ is convex}$$

Feasible set defined with functions

$$X = \{x \in \mathbb{R}^n \mid g_i(x) \leq 0 \text{ and } h_j(x) = 0 \text{ for } i \in \mathcal{I}, j \in \mathcal{E}\}$$

- $X_g = \{x \in \mathbb{R}^n \mid g(x) \leq 0\}$ is convex if g is convex
- When $\mathcal{E} = \emptyset$, X is convex when every g_i is convex
- These two conditions are not necessary
- Allowing now equalities, we note that since $h_j(x) = 0 \Leftrightarrow h_j(x) \leq 0$ and $-h_j(x) \leq 0$, we can guarantee that X is convex when all functions h_j are affine
- To summarize, X is convex as soon as every g_i is convex and every h_j is affine

A few classes of convex problems

General formulation

$$\min_{x\in\mathbb{R}^n} f(x) \text{ s.t. } g_i(x) \leq 0 \ \forall\, i \in \mathcal{I} \text{ and } h_j(x) = 0 \ \forall\, j \in \mathcal{E}$$

where f and g_i for all $i \in \mathcal{I}$ are convex and h_j are affine for all $j \in \mathcal{E}$

$$h_j(x) = a_j^{\mathrm{T}} x - b_j$$

1. Linear optimization (LO):

   f and g_i for all $i \in \mathcal{I}$ are also affine

$$f(x) = c^{\mathrm{T}} x \quad \text{and} \quad g_i(x) = a_i^{\mathrm{T}} x - b_i$$

Linear optimization for data-mining

Given two sets of points in $\mathbb{R}^d$

$$A = \{a_i\}_{1 \leq i \leq n_a} \quad \text{and} \quad B = \{b_i\}_{1 \leq i \leq n_b}$$

find a hyperplane defined by $h \in \mathbb{R}^d$ and $c \in \mathbb{R}$

$$h^T x + c = 0$$

that (strictly) separates them

Applications (medical diagnosis, credit screening, etc.)

a. compute hyperplane with known points (learn)

b. classify new unknown points based on this hyperplane (generalize)

Formulation

$$\min 0 \text{ such that}$$

$$\begin{aligned} h^T a_i + c &\geq +1 \text{ for all } 1 \leq i \leq n_a \\ h^T b_i + c &\leq -1 \text{ for all } 1 \leq i \leq n_b \end{aligned}$$

a. Can add objective function to find the best separator

b. Nonlinear separator can also be found with linear formulation, e.g. $pe^{||x||} + h^T x + c = 0$ leads to

$$pe^{||a_i||} + h^T a_i + c \geq 1 \text{ and } pe^{||b_i||} + h^T b_i + c \leq -1$$

since dependence on decision variables is still linear

c. Ability to solve large-scale problems often needed

Quadratic optimization

$$\min_{x\in\mathbb{R}^n} f(x) \text{ s.t. } g_i(x) \le 0 \;\forall\, i \in \mathcal{I} \text{ and } h_j(x) = 0 \;\forall\, j \in \mathcal{E}$$

where h_j are affine for all $j \in \mathcal{E}$, f is a convex quadratic

$f(x) = x^{\mathrm{T}}Qx + r^{\mathrm{T}}x + s$ with $Q \succeq 0$ (positive semidefinite)

a. $\mathcal{I} = \emptyset$: improper quadratic optimization problem since (necessary and sufficient) optimality conditions consist in a simple linear system of equations

b. $g_i(x)$ are affine: (standard) quadratic optimization (QO), e.g. for Markowitz portfolio selection

c. $g_i(x)$ are also convex quadratic: quadratically constrained quadratic optimization (QCQO)

However remember quadratic equalities are forbidden !

Geometric optimization

A posynomial is a sum of monomials in several positive variables with positive leading coefficients and arbitrary real exponents, such as

$$p(x_1, x_2, x_3) = 3x_1x_3 + \frac{1}{2}\sqrt{x_2x_3} + \frac{x_2}{x_1x_3^2}$$

Geometric optimization (programming) corresponds to

$$\min_{x \in \mathbb{R}^n_{++}} f(x) \text{ s.t. } g_i(x) \leq 1 \; \forall\, i \in \mathcal{I}$$

where f and every g_i are posynomials

These problems are not necessarily convex !

(for example, $\sqrt{x_1}$ is concave)

Geometric optimization in convex form

$$\min_{x \in \mathbb{R}^n_{++}} f(x) \text{ s.t. } g_i(x) \leq 1 \; \forall \, i \in \mathcal{I}$$

fortunately can be convexified by letting $x_i = e^{y_i}$

$$p(x_1, x_2, x_3) = 3x_1x_3 + \frac{1}{2}\sqrt{x_2x_3} + \frac{x_2}{x_1x_3^2}$$

$$\leftrightarrow \; \tilde{p}(y_1, y_2, y_3) = 3e^{y_1+y_3} + \frac{1}{2}e^{\frac{y_2+y_3}{2}} + e^{y_2-y_1-2y_3}$$

$$\min_{y \in \mathbb{R}^n} \tilde{f}(x) \text{ s.t. } \tilde{g}_i(x) \leq 1 \; \forall \, i \in \mathcal{I}$$

(linear equalities correspond here to monomial equalities)

Application example: geometric design, such as wire sizing in circuit optimization

Properties of convex optimization

Why is it interesting to consider (or restrict yourself to) convex optimization problems?

Passive features:

- ◇ every local minimum is a global minimum
- ◇ set of optimal solutions is convex
- ◇ optimality (KKT) conditions are sufficient, in addition to necessary (with regularity assumption)

Any algorithm or solver applied to a convex problem will automatically benefit from those features

but there is more ...

Properties of convex optimization

Active features:

◇ possibility of designing dedicated algorithms with polynomial worst-case algorithmic complexity

(in many situations: an interior-point method based on the theory of self-concordant barriers)

◇ possibility of writing down a dual problem strongly related to original problem

(solutions to the dual problem will provide optimality certificates, i.e. guarantees for the original problem)

Convex optimization: plan

Why

a. Nice case: linear optimization

b. Algorithms and guarantees

What

a. Convex problems: definitions and examples

How

a. Algorithms: interior-point methods

b. Guarantees: duality

c. Framework: conic optimization

Interior-point methods

Convex optimization

Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be a convex function, $C \subseteq \mathbb{R}^n$ be a convex set : optimize a vector $x \in \mathbb{R}^n$

$$\inf_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in C \qquad \text{(P)}$$

Properties

- ◇ All local optima are global, optimal set is convex
- ◇ Lagrange duality $\rightarrow$ strongly related dual problem
- ◇ Objective can be taken linear w.l.o.g. ($f(x) = c^T x$)

Principle

Approximate a constrained problem by

a *family* of unconstrained problems

Use a barrier function F to replace the inclusion $x \in C$

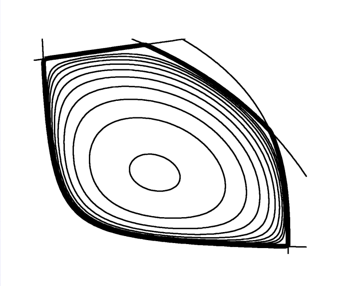
- $\diamond$ F is smooth
- $\diamond$ F is strictly convex on $\operatorname{int} C$
- $\diamond$ $F(x) \rightarrow +\infty$ when $x \rightarrow \partial C$

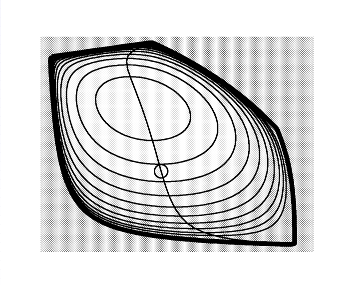
$$\rightarrow \quad C = \operatorname{cl} \operatorname{dom} F = \operatorname{cl} \{x \in \mathbb{R}^n \mid F(x) < +\infty\}$$

Central path

Let $\mu \in \mathbb{R}_{++}$ be a parameter and consider

$$\inf_{x\in\mathbb{R}^n} \frac{c^{\mathrm{T}}x}{\mu} + F(x) \qquad (\mathrm{P}_\mu)$$





$$x^*_\mu \longrightarrow x^* \text{ when } \mu \searrow 0$$

where

- $\diamond$ x^*_μ is the (unique) solution of (P_μ) ($\longrightarrow$ central path)
- $\diamond$ x^* is a solution of the original problem (P)

Ingredients

- ◇ A method for unconstrained optimization
- ◇ A barrier function

Interior-point methods rely on

- ◇ Newton's method to compute x^*_μ
- ◇ When C is defined with convex constraints $g_i(x) \leq 0$, one can introduce the logarithmic barrier function

$$F(x) = -\sum_{i=1}^n \log(-g_i(x))$$

but this is not the only choice

Question: What is a good barrier, i.e. a barrier for which Newton's method is efficient ?

Answer: A *self-concordant* barrier

Self-concordant barriers

Definition ***[Nesterov & Nemirovski, 1988]***

$F : \operatorname{int} C \mapsto \mathbb{R}$ is called ν-self-concordant on C iff

- F is convex
- F is three times differentiable
- $F(x) \to +\infty$ when $x \to \partial C$
- the following two conditions hold

$$\nabla^3 F(x)[h,h,h] \leq 2 \left(\nabla^2 F(x)[h,h]\right)^{\frac{3}{2}}$$
$$\nabla F(x)^{\mathrm{T}} (\nabla^2 F(x))^{-1} \nabla F(x) \leq \nu$$

for all $x \in \operatorname{int} C$ and $h \in \mathbb{R}^n$

A (simple?) example

For linear optimization, $C = \mathbb{R}^n_+$: take $F(x) = -\sum_{i=1}^n \log x_i$

When $n = 1$, we can choose $\nu = 1$

- $\nabla F(x) = -\frac{1}{x}$ and $\nabla F(x)^{\mathrm{T}} h = -\frac{h}{x}$
- $\nabla^2 F(x) = \frac{1}{x^2}$ and $\nabla^2 F(x)[h, h] = \frac{h^2}{x^2}$
- $\nabla^3 F(x) = -2\frac{1}{x^3}$ and $\nabla^3 F(x)[h, h, h] = -2\frac{h^3}{x^3}$

When $n > 1$, we have

- $\nabla F(x) = (-x_i^{-1})$ and $\nabla F(x)^{\mathrm{T}} h = -\sum h_i x_i^{-1}$
- $\nabla^2 F(x) = \mathrm{diag}(x_i^{-2})$ and $\nabla^2 F(x)[h, h] = \sum h_i^2 x_i^{-2}$
- $\nabla^3 F(x) = \mathrm{diag}_3(-2x_i^{-3})$, $\nabla^3 F(x)[h, h, h] = -2\sum h_i^3 x_i^{-3}$

and one can show that $\nu = n$ is valid

Barrier calculus

Barriers for basic convex sets, for example

- $-\log x$ for $\mathbb{R}_+$; $-\log(1-||x||^2)$ for unit Eucl. ball
- $-\log(\log y - x) - \log y$ for $\{(x,y) \mid e^x \leq y\}$

and convexity-preserving operations to combine them

- Sum:

  F is a ν_1-s.-c. barrier for $\mathcal{C}_1 \subseteq \mathbb{R}^n$

  G is a ν_2-s.-c. barrier for $\mathcal{C}_2 \subseteq \mathbb{R}^n$

  $\Rightarrow (F+G)$ is a $\nu_1+\nu_2$-s.-c. barrier

  for the set $\mathcal{C}_1 \cap \mathcal{C}_2$ (if nonempty)

- Linear transformations preserve self-concordancy

Complexity result

Summary

Self-concordant barrier ⇒ polynomial number of iterations to solve (P) within a given accuracy

Short-step method: follow the central path

- ◇ Measure distance to the central path with $\delta(x, \mu)$
- ◇ Choose a starting iterate with a small $\delta(x_0, \mu_0) < \tau$
- ◇ While accuracy is not attained
  a. Decrease μ geometrically (δ increases above τ)
  b. Take a Newton step to minimize barrier (δ decreases back below the τ threshold)

Geometric interpretation

Two self-concordancy conditions: each has its role

- Second condition bounds the size of the Newton step $\Rightarrow$ controls the increase of the distance to the central path when μ is updated
- First condition bounds the variation of the Hessian $\Rightarrow$ guarantees that the Newton step restores the initial distance to the central path

Summarized complexity result

$$\mathcal{O}\left(\sqrt{\nu}\log\frac{1}{\epsilon}\right)$$

iterations lead a solution with ϵ accuracy on the objective

Complexity result

◇ Let F be a ν-self-concordant barrier for C and let $x_0 \in \operatorname{int} C$ be a (well-chosen) feasible starting point,

a short-step interior-point algorithm can solve problem (P) up to ϵ accuracy within

$$\mathcal{O}\left(\sqrt{\nu}\log\frac{c^T x_0 - p^*}{\epsilon}\right) \text{ iterations,}$$

such that at each iteration the self-concordant barrier and its first and second derivatives have to be evaluated and a linear system has to be solved in $\mathbb{R}^n$

◇ Complexity invariant w.r.t. to scaling of F

◇ Universal bound on complexity parameter: $\nu \geq 1$

Corollary

Assume F, ∇F and $\nabla^2 F$ are polynomially computable
$\Rightarrow$ problem (P) can be solved in polynomial time

Existence

There exists a universal SC barrier with parameters

$$\nu = \mathcal{O}(n)$$

(But it is not necessarily efficiently computable (therefore not a contradiction of the fact that some convex problems are hard to solve)

Other methods

- ◇ Long-step methods: more aggressive reduction of central path parameter but several Newton steps needed to restore proximity
- ◇ Techniques to deal with the lack of an acceptable starting point
- ◇ Non path-following/non interior point techniques, e.g. potential-reduction methods, ellipsoid method, first-order methods (including smoothing techniques), etc.

A few complexity results

- linear optimization with n inequalities: $\nu = n \Rightarrow \mathcal{O}\left(\sqrt{n}\log\frac{1}{\varepsilon}\right)$ (best complexity known so far)
- quadratic optimization with equalities: $\nu = 1$!
- quadratic optimization with m inequalities (linear or quadratic): $\nu = m + 1 \Rightarrow \mathcal{O}\left(\sqrt{m}\log\frac{1}{\varepsilon}\right)$
- geometric optimization with p monomials (objective or constraints): $\nu = p \Rightarrow \mathcal{O}\left(\sqrt{p}\log\frac{1}{\varepsilon}\right)$
- similar results known for (nearly) all practically relevant problems, such as entropy optimization, sum-of-norm minimization, problems with logarithms, etc.

However the main cost of each iteration (i.e. mainly Newton step via a linear system) also grows with # of vars.

Sketch of the proof

Define $n_\mu(x)$ the Newton step taken from x to x^*_μ

$$n_\mu(x) = 0 \text{ if and only if } x = x^*_\mu$$

We take

$$\delta(x, \mu) = \|n_\mu(x)\|_x \quad (\textit{size} \text{ of the } \textit{Newton step})$$

with a well-chosen (*coordinate invariant*) norm $\|\cdot\|_x$
Set $k \leftarrow 0$, perform the following main loop:

a. $\mu_{k+1} \leftarrow \mu_k(1-\theta)$ (*decrease barrier param*)

b. $x_{k+1} \leftarrow x_k + n_{\mu_{k+1}}(x_k)$ (*take Newton step*)

c. $k \leftarrow k + 1$

Sketch of the proof (continued)

Key choice: parameters τ and θ such that

$$\delta(x_k, \mu_k) < \tau \quad \Rightarrow \quad \delta(x_{k+1}, \mu_{k+1}) < \tau$$

To relate $\delta(x_k, \mu_k)$ and $\delta(x_{k+1}, \mu_{k+1})$,
introduce an intermediate quantity

$$\delta(x_k, \mu_{k+1})$$

We will also denote for simplicity

$$x_k \leftrightarrow x$$

$$\mu_k \leftrightarrow \mu$$

Sketch of the proof (end)

Given a ν-self-concordant barrier:

- $\diamond$ $x \in \operatorname{dom} F$ and $\mu^+ = (1-\theta)\mu \Rightarrow$

$$\delta(x, \mu^+) \leq \frac{\delta(x,\mu) + \theta\sqrt{\nu}}{1-\theta}$$

- $\diamond$ $x \in \operatorname{dom} F$ and $\delta(x,\mu) < 1 \Rightarrow$ define $x^+ = x + n_\mu(x)$

$$x^+ \in \operatorname{dom} F \text{ and } \delta(x^+, \mu) \leq 1\left(\frac{\delta(x,\mu)}{1-\delta(x,\mu)}\right)^2$$

with e.g. possible choice for parameters

$$\tau = \frac{1}{4} \text{ and } \theta = \frac{1}{16\sqrt{\nu}}$$

(hence the name short-step)

Convex optimization: plan

Why

a. Nice case: linear optimization

b. Algorithms and guarantees

What

a. Convex problems: definitions and examples

How

a. Algorithms: interior-point methods

b. Guarantees: duality

c. Framework: conic optimization

Duality for linear optimization

Standard formulation

Consider the linear problem(with m variables y_i)

$$\max \sum_{i=1}^{m} b_i y_i \text{ such that } \sum_{i=1}^{m} a_{ij} y_i \leq c_j \ \forall 1 \leq j \leq n$$

(objective and n linear inequalities), or

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y \leq c$$

(matrix notation with $b, y \in \mathbb{R}^m$, $c \in \mathbb{R}^n$ and $A \in \mathbb{R}^{m \times n}$)

All linear problems can be expressed in this format

When is a problem infeasible ?

In other terms: when is $A^{\mathrm{T}}y \leq c$ inconsistent ?
And, more importantly: how can we be sure ?

- Feasible $\longrightarrow$ exhibit a feasible solution
- Infeasible $\longrightarrow$??

$$3y_1 + 2y_2 \leq 8, \; -y_2 \leq -3, \; -y_1 \leq -1$$

Add constraints with weights 1, 2 and 3 to obtain $0y_1 + 0y_2 \leq -1 \Leftrightarrow 0 \leq -1 \Leftrightarrow$ a contradiction

In general: consider $A^{\mathrm{T}}y \leq c$ or, equivalently, a set of inequalities $a_i^{\mathrm{T}}y \leq c_i$

Proving infeasibility

Multiply each inequality by $a_i^{\mathrm{T}} y \leq c_i$ by a nonnegative constant x_i and take the sum to obtain a consequence

$$\sum_{i=1}^{n} (a_i^{\mathrm{T}} y) x_i \leq \sum_{i=1}^{n} c_i x_i \text{ with } x_i \geq 0$$

$$\Big(\sum_{i=1}^{n} a_i x_i\Big)^{\mathrm{T}} y \leq c^{\mathrm{T}} x \text{ with } x \geq 0$$

$$(Ax)^{\mathrm{T}} y \leq c^{\mathrm{T}} x \text{ with } x \geq 0$$

Contradiction arises only for $0^{\mathrm{T}} y \leq \alpha$ with $\alpha < 0$

This happens iff $Ax = 0$ et $c^{\mathrm{T}} x < 0 \rightarrow$ sufficient condition for infeasibility but ...

Farkas' Lemma

Theorem: $A^{\mathrm{T}}y \leq c$ is inconsistent if and only if there exists $x \geq 0$ such that $Ax = 0$ et $c^{\mathrm{T}}x < 0$

In other words:
Exactly one of the following two systems is consistent

$$Ax = 0,\ x \geq 0 \text{ and } c^{\mathrm{T}}x < 0$$

$$A^{\mathrm{T}}y \leq c$$

Proof relies on topological notions (separation argument)

There always exists a linear proof for the infeasibility of a system of linear inequalities !

Bounds and optimality

Let $\bar{y}$ a feasible solution (satisfying $A^T y \leq c$)
$\rightarrow b^T \bar{y}$ is a lower bound on the optimal value f^*

But how to

◇ obtain upper bounds on the optimal value ?

◇ prove that a feasible solution y^* is optimal ?

Those questions are linked since

$$\text{proving that } y^* \text{ is optimal}$$
$$\Updownarrow$$
$$\text{proving that } b^T y^* \text{ is an upper bound on the optimal value } f^*$$

Generating upper bounds

Consider

$$\max y_1 + 2y_2 + 3y_3 \text{ such that } \begin{array}{rcll} y_1 + y_2 & \leq & 1 & (a) \\ y_2 + y_3 & \leq & 2 & (b) \\ y_3 & \leq & 3 & (c) \end{array}$$

Solution $y = (1, 0, 2)$ is feasible with objective value 7

$\longrightarrow$ lower bound $f^* \geq 7$

Let us combine constraints: $(a) + (b) + 2(c)$

$$y_1 + y_2 + y_2 + y_3 + 2y_3 \leq 1 + 2 + 2 \times 3 \Leftrightarrow y_1 + 2y_2 + 3y_3 \leq 9$$

$\longrightarrow$ upper bound on the optimal value $f^* \leq 9$

Moreover, considering the feasible solution $y = (2, -1, 3)$ with objective 9 provides a proof that $f^* = 9$ is the optimal value of the problem

The best upper bound

Let us find the best upper bound using this procedure

$$\max \sum_{i=1}^{m} b_i y_i \text{ such that } \sum_{i=1}^{m} a_{ij} y_i \leq c_j \ \forall 1 \leq j \leq n$$

Introducing again n (multiplying) variables $x_i \geq 0$ we get

$$\sum_{j=1}^{n} x_j \sum_{i=1}^{m} a_{ij} y_i \leq \sum_{j=1}^{n} x_j c_j \Leftrightarrow \sum_{i=1}^{m} y_i (\sum_{j=1}^{n} a_{ij} x_j) \leq \sum_{j=1}^{n} c_j x_j$$

The best upper bound (continued)

This provides an upper bound on the objective equal to $\sum_{j=1}^n c_j x_j$, assuming that x satisfies

$$\sum_{j=1}^n a_{ij} x_j = b_i \ \forall 1 \leq i \leq m$$

Minimizing now this upper bound

$$\min \sum_{j=1}^n c_j x_j \text{ s.t. } \sum_{j=1}^n a_{ij} x_j = b_i \ \forall 1 \leq i \leq m \text{ and } x_i \geq 0$$

or

$$\min c^T x \text{ such that } Ax = b \text{ and } x \geq 0$$

We find another linear optimization problem which is dual to our first problem!

Standard denominations

Using a similar reasoning, we could have started with the minimization problem and, looking for the best lower bound, derive the original maximization problem

In fact, it is customary in the literature to call

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \geq 0$$

the primal (P) problem with optimal value p^*
and

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y \leq c$$

the dual (D) problem with optimal value d^*

Duality properties

- Weak duality: any feasible solution for the primal (resp. dual) provides an upper (resp. lower) bound for the dual (resp. primal)

  (immediate consequence of our dualizing procedure)

- Inequality $b^T y \leq c^T x$ holds for any x, y such that $Ax = b$, $x \geq 0$ and $A^T y \leq c$ (corollary)

- If the primal (resp. dual) is unbounded, the dual (resp. primal) must be infeasible

  (but the converse is not true !)

Duality properties (continued)

- Strong duality: If x^* is an optimal solution for the primal, there exists an optimal solution y^* for the dual such that $c^T x^* = b^T y^*$ (in other words: $p^* = d^*$)
- This property (and its dual) is not trivial, and is a generalization of the Farkas Lemma $\longrightarrow$ it is always possible to exhibit a proof that a given solution is optimal !
- However, there are cases where both problems are infeasible: $c = (-1\ 0)^T$, $b = -1$ et $A = (0\ 1)$

Other properties and consequences

	$d^* = -\infty$	d^* finite	$d^* = +\infty$
$p^* = -\infty$	Possible, $p^* = d^*$	Impossible	Impossible
p^* finite	Impossible	Possible, $p^* = d^*$	Impossible
$p^* = +\infty$	Possible, $p^* \neq d^*$	Impossible	Possible, $p^* = d^*$

- ◇ One can also write down the dual to a general linear optimization problem
- ◇ Dual variables can often be interpreted as prices on primal constraints
- ◇ One can indifferently solve the primal or the dual to find the optimal objective value
- ◇ Primal-dual algorithms solve both problems simultaneously

Convex optimization: plan

Why

a. Nice case: linear optimization

b. Algorithms and guarantees

What

a. Convex problems: definitions and examples

How

a. Algorithms: interior-point methods

b. Guarantees: duality

c. Framework: conic optimization

Conic optimization

Motivation

Objective: generalize linear optimization

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y \leq c$$

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \geq 0$$

while trying to preserve the nice duality properties
$\longrightarrow$ change as little as possible

Idea: generalize the inequalities $\leq$ and $\geq$

What are properties of nice inequalities ?

Generalizing $\geq$ and $\leq$

Let $K \subseteq \mathbb{R}^n$. Define

$$a \succeq_K 0 \Leftrightarrow a \in K$$

We also have

$$a \succeq_K b \Leftrightarrow a - b \succeq_K 0 \Leftrightarrow a - b \in K$$

as well as

$$a \preceq_K b \Leftrightarrow b \succeq_K a \Leftrightarrow b - a \succeq_K 0 \Leftrightarrow b - a \in K$$

Let us also impose two sensible properties

$$a \succeq_K 0 \Rightarrow \lambda a \succeq_K 0 \; \forall \lambda \geq 0 \text{ (}K\text{ is a cone)}$$

$$a \succeq_K 0 \text{ and } b \succeq_K 0 \Rightarrow a + b \succeq_K 0$$

(K is closed under addition)

Properties of admissible sets K

- K is a convex set!
- In fact, if K is a cone, we have

$$K \text{ is closed under addition } \Leftrightarrow K \text{ is convex}$$

Conic optimization

We can then generalize

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y \leq c$$

to

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y \preceq_K c$$

$\Rightarrow$ This problem is convex
The standard linear cases corresponds to $K = \mathbb{R}^n_+$

More requirements for K

- $x \succeq 0$ and $x \preceq 0 \Rightarrow x = 0$
  which means $K \cap (-K) = \{0\}$ (the cone is pointed)
- We define the strict inequality by $a \succ 0 \Leftrightarrow a \in \operatorname{int} K$ (and $a \succ b$ iff $a - b \in \operatorname{int} K$)
  Hence we require $\operatorname{int} K \neq \emptyset$ (the cone is solid)
- Finally, we would like to be able to take limits:
  If $\{x_i\}_{i\to\infty}$ with $x_i \succeq_K 0 \,\forall i$, then $\lim_{i\to\infty} x_i = \bar{x} \Rightarrow \bar{x} \succeq_K 0$
  which is equivalent to saying that K is closed

Example: second-order (or Lorentz or ice-cream) cone

$$\mathbb{L}^n = \{(x_0, \ldots, x_n) \in \mathbb{R}^{n+1} \mid \sqrt{x_1^2 + \cdots + x_n^2} \leq x_0\}$$

Another example: semidefinite cone $K = \mathbb{S}^n_+$ (symmetric positive semidefinite matrices)

Back to conic optimization

A convex cone $K \subseteq \mathbb{R}^n$ that is solid, pointed and closed will be called a proper cone

In the following, we will always consider proper cones

We obtain

$$\max_{y \in \mathbb{R}^m} b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y \preceq_K c$$

or, equivalently,

$$\max_{y \in \mathbb{R}^m} b^{\mathrm{T}} y \text{ such that } c - A^{\mathrm{T}} y \in K$$

with problem data $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$ and $A \in \mathbb{R}^{m \times n}$

Combining several cones

Considering several conic constraints

$$A_1^T y \preceq_{K_1} c_1 \text{ and } A_2^T y \preceq_{K_2} c_2$$

which are equivalent to

$$c_1 - A_1^T y \in K_1 \text{ and } c_2 - A_2^T y \in K_2$$

one introduces the product cone $K = K_1 \times K_2$ to write

$$(c_1 - A_1^T y, c_2 - A_2^T y) \in K_1 \times K_2$$

$$\Leftrightarrow \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} - \begin{pmatrix} A_1^T \\ A_2^T \end{pmatrix} \in K_1 \times K_2 \Leftrightarrow \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} - \begin{pmatrix} A_1^T \\ A_2^T \end{pmatrix} \succeq_{K_1 \times K_2} 0$$

If K_1 and K_2 are proper, $K_1 \times K_2$ is also proper

Equivalence with convex optimization

Conic optimization is clearly a special case of convex optimization: what about the reverse statement ?

$$\min_{x \in \mathbb{R}^n} f(x) \text{ such that } x \in X \subseteq \mathbb{R}^n$$

- ◇ The objective of a convex problem can be assumed w.l.o.g. to be linear w.l.o.g.: $f(x) = c^{\mathrm{T}} x$
- ◇ The feasible region of a convex problem can be assumed w.l.o.g. to be in the conic standard format:

$$X = \{x \in K \text{ and } Ax = b\}$$

⇒ conic optimization equivalent to convex optimization
Conic format is a standard form for convex optimization

A linear objective ?

$$\min_{x \in \mathbb{R}^n} f(x) \text{ such that } x \in X \subseteq \mathbb{R}^n$$

$$\Updownarrow$$

$$\min_{(x,t) \in \mathbb{R}^n \times \mathbb{R}} t \text{ such that } x \in X \text{ and } (x,t) \in \operatorname{epi} f$$

$$\Updownarrow$$

$$\min_{(x,t) \in \mathbb{R}^n \times \mathbb{R}} t \text{ such that } x \in X \text{ and } f(x) \leq t$$

$\Rightarrow$ equivalent problem with linear objective

Conic constraints ?

$$K_X = \mathrm{cl}\{(x,u) \in \mathbb{R}^n \times \mathbb{R}_{++} \mid \frac{x}{u} \in X\}$$

is called the (closed) conic hull of X

We have that K_X is a closed convex cone and

$$x \in X \Leftrightarrow (x,u) \in K_X \text{ and } u = 1$$

$$\min_{x \in \mathbb{R}^n} c^{\mathrm{T}} x \text{ such that } x \in X \subseteq \mathbb{R}^n$$

$$\Updownarrow$$

$$\min_{(x,u) \in \mathbb{R}^n \times \mathbb{R}} c^{\mathrm{T}} x \text{ such that } (x,u) \succeq_{K_X} 0 \text{ and } u = 1$$

$\Rightarrow$ equivalent problem with a conic constraint

Duality properties

Since we generalized

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y \leq c$$

to

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y \preceq_K c$$

it is tempting to generalize

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \geq 0$$

to

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \succeq_K 0$$

But this is not the right primal-dual pair !

Dualizing a conic problem

Remembering the dualizing procedure for linear optimization, a crucial point lied in the ability to derive consequences by taking nonnegative linear combinations of inequalities

Consider now the following statement

$$\begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix} \succeq_{\mathbb{L}^2} \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

which is true since $(-1)^2 + (-1)^2 \leq 2^2$

Multiplying the first line by $0,1$ and the next two by 1, we get $0.1 \times 2 - 1 \times 1 - 1 \times 1 \geq 0$ or $-1.8 \geq 0$:

$\Rightarrow$ this is a contradiction!

We obtained a contraction although the original system of inequalities was consistent $\Rightarrow$ something is wrong! Some nonnegative linear combinations do not work!

Rescuing duality

Starting with

$$x \in K \subseteq \mathbb{R}^n \Leftrightarrow x \succeq_K 0$$

we identify all vectors (of multipliers) $z \in \mathbb{R}^n$ such that the consequence $z^T x \geq 0$ holds as soon as $x \succeq_K 0$

Hence we define the set

$$K^* = \{z \in \mathbb{R}^n \text{ such that } x^T z \geq 0 \ \forall x \in K\}$$

The dual cone

$$K^* = \{z \in \mathbb{R}^n \text{ such that } x^T z \geq 0 \ \forall x \in K\}$$

- For any $x \in K$ and $z \in K^*$, we have $z^T x \geq 0$
- K^* is a convex cone, called the dual cone of K
- K^* is always closed, and if K is closed, $(K^*)^* = K$
- K is pointed (resp. solid) $\Rightarrow$ K^* is solid (resp. pointed)
- Cartesian products: $(K_1 \times K_2)^* = K_1^* \times K_2^*$
- $(\mathbb{R}^n_+)^* = \mathbb{R}^n_+$, $(\mathbb{L}^n)^* = \mathbb{L}^n$, $(\mathbb{S}^n_+)^* = \mathbb{S}^n_+$:
  these cones are self-dual
- But there exists (many) cones that are not self-dual

Bounds and optimality

Let $\bar{y}$ a feasible solution (satisfying $A^T y \preceq_K c$)
$\rightarrow b^T \bar{y}$ is a lower bound on the optimal value f^*

But how to

- obtain upper bounds on the optimal value ?
- prove that a feasible solution y^* is optimal ?

Those questions are linked since

$$\text{proving that } y^* \text{ is optimal}$$
$$\Updownarrow$$
$$\text{proving that } b^T y^* \text{ is an upper bound on the optimal value } f^*$$

Generating upper bounds

Consider

$$\max 2y_1+3y_2+2y_3 \text{ such that } \begin{pmatrix} y_1+y_2 \\ y_2+y_3 \\ y_3 \end{pmatrix} \preceq_{\mathbb{L}^2} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \quad \begin{matrix} (a) \\ (b) \\ (c) \end{matrix}$$

Solution $y = (-2, 1, 2)$ is feasible with objective value 3
$\rightarrow$ lower bound $f^* \geq 3$ (since $(2, -1, 1) \in \mathbb{L}^2$)

Let us combine constraints: $2(a) + (b) + (c)$
(we have the right to do so since $(2, 1, 1) \in (\mathbb{L}^2)^* = \mathbb{L}^2$)

$$2y_1 + 2y_2 + y_2 + y_3 + y_3 \leq 2 + 2 + 3 \Leftrightarrow 2y_1 + 3y_2 + 2y_3 \leq 7$$

$\rightarrow$ upper bound on the optimal value $f^* \leq 7$

The best upper bound

Let us find the best upper bound using this procedure

$$\max \sum_{i=1}^{m} b_i y_i \text{ such that } \Big(\sum_{i=1}^{m} a_{ij} y_i\Big)_{1\leq j\leq n} \preceq_K \big(c_j\big)_{1\leq j\leq n}$$

Introducing again n (multiplying) variables x_i
we get

$$\sum_{j=1}^{n} x_j \sum_{i=1}^{m} a_{ij} y_i \leq \sum_{j=1}^{n} x_j c_j \Leftrightarrow \sum_{i=1}^{m} y_i \Big(\sum_{j=1}^{n} a_{ij} x_j\Big) \leq \sum_{j=1}^{n} c_j x_j$$

under the assumption that $x \in K^*$

The best upper bound (continued)

This provides an upper bound on the objective equal to $\sum_{j=1}^n c_j x_j$, assuming that x satisfies

$$\sum_{j=1}^n a_{ij} x_j = b_i \; \forall 1 \leq i \leq m$$

Minimizing now this upper bound

$$\min \sum_{j=1}^n c_j x_j \text{ s.t. } \sum_{j=1}^n a_{ij} x_j = b_i \; \forall 1 \leq i \leq m \text{ and } x \in K^*$$

or

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \succeq_{K^*} 0$$

We find another conic optimization problem which is dual to our first problem!

Duality for conic optimization

We have completely mimicked the dualizing procedure used for linear optimization

The problem of finding the best upper bound

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \geq 0$$

becomes thus

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \succeq_{K^*} 0$$

The correct primal-dual pair is thus

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y \preceq_K c$$

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \succeq_{K^*} 0$$

Primal-dual pair

Again, for historical reasons, the min problem is called the primal. Since our cones are closed, $(K^*)^* = K^*$, which means we can write the primal conic problem

$$\min c^T x \text{ such that } Ax = b \text{ and } x \succeq_K 0$$

and the dual conic problem

$$\max b^T y \text{ such that } A^T y \preceq_{K^*} c$$

- Very symmetrical formulation
- Computing the dual essentially amounts to finding K^*
- All nonlinearities are confined to the cones K and K^*

Duality properties

- Weak duality: any feasible solution for the primal (resp. dual) provides an upper (resp. lower) bound for the dual (resp. primal)

  (immediate consequence of our dualizing procedure)

- Inequality $b^T y \leq c^T x$ holds for any x, y such that $Ax = b$, $x \succeq_K 0$ and $A^T y \preceq_{K^*} c$ (corollary)

- If the primal (resp. dual) is unbounded, the dual (resp. primal) must be infeasible

  (but the converse is not true!)

Completely similar to the situation for linear optimization

Duality properties (continued)

What about strong duality ?

If y^* is an optimal solution for the dual, does there exist an optimal solution x^* for the primal such that $c^T x^* = b^T y^*$ (in other words: $p^* = d^*$) ?

Consider $K = \mathbb{L}^2$ with

$$A = \begin{pmatrix} -1 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix}, \; b = \begin{pmatrix} 0 & -1 \end{pmatrix}^T \text{ and } c = \begin{pmatrix} 0 & 0 & 0 \end{pmatrix}^T$$

We can easily check that

- the primal is infeasible
- the dual is bounded and solvable

$\Rightarrow$ strong duality does not hold for conic optimization ...

Other troublesome situations

Let $\lambda \in \mathbb{R}_+$: consider

$$\min \lambda x_3 - 2x_4 \text{ s.t. } \begin{pmatrix} x_1 & x_4 & x_5 \\ x_4 & x_2 & x_6 \\ x_5 & x_6 & x_3 \end{pmatrix} \succeq_{\mathbb{S}^3_+} 0, \quad \begin{pmatrix} x_3 + x_4 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

In this case, $p^* = \lambda$ but $d^* = 2$: duality gap!

$$\min x_1 \text{ such that } x_3 = 1 \text{ and } \begin{pmatrix} x_1 & x_3 \\ x_3 & x_2 \end{pmatrix} \succeq_{\mathbb{S}^2_+} 0$$

In this case, $p^* = 0$ but the problem is unsolvable!

In all cases, one can identify the cause for our troubles: the affine subspace defined by the linear constraints is tangent to the cone (it does not intersect its interior)

Rescuing strong duality

A feasible solution to a conic (primal or dual) problem is strictly feasible iff it belongs to the interior of the cone

In other words, we must have $Ax = b$ and $x \succ_K 0$ for the primal and $A^T y \prec_{K^*} c$ for the dual

Strong duality: If the dual problem admits a strictly feasible solution, we have either

- ◇ an unbounded dual, in which case $d^* = +\infty = p^*$ and the primal is infeasible
- ◇ a bounded dual, in which case the primal is solvable with $p^* = d^*$ (hence there exists at least one feasible primal solution x^* such that $c^T x^* = p^* = d^*$)

Strong duality (continued)

- If the primal problem admits a strictly feasible solution, we have either
  - an unbounded primal, in which case $p^* = -\infty = d^*$ and the dual is infeasible
  - a bounded primal, in which case the dual is solvable with $d^* = p^*$ (hence there exists at least one feasible dual solution y^* such that $b^T y^* = d^* = p^*$)
- The first case is a mere consequence of weak duality
- Finally, when both problems admit a strictly feasible solution, both problems are solvable and we have

$$c^T x^* = p^* = d^* = b^T y^*$$

Conic modelling with three cones

A first cone: $\mathbb{R}^n_+$

Standard meaning for inequalities:

$$\succeq_{\mathbb{R}^n_+} \quad \Leftrightarrow \quad \geq$$

$\Rightarrow$ linear optimization
But we can also model some nonlinearities!

$$|x_1 - x_2| \leq 1 \quad \Leftrightarrow \quad -1 \leq x_1 - x_2 \leq 1$$

$$|x_1 - x_2| \leq t \quad \Leftrightarrow \quad \begin{pmatrix} x_1 - x_2 - t \\ x_2 - x_1 - t \end{pmatrix} \leq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Terminology: conic representability

- Set S is K-representable if can be expressed as feasible region of conic problem using cone K
- Closed under intersection and Cartesian product
- Function f is K-representable iff its epigraph is K-representable
- Closed under sum, positive multiplication and max
- What we can do in practice: minimize a K-representable function over a K-representable set where K is a product of cones $\mathbb{R}^n_+$, $\mathbb{L}^n$, $\mathbb{S}^n_+$ and $\mathbb{R}^n$

A simple example

Consider set

$$S = \{x_1^2 + x_2^2 \leq 1\}$$

$\longrightarrow$ can be modelled as

$$(x_0, x_1, x_2) \in \mathbb{L}^2 \text{ and } x_0 = 1$$

$\Rightarrow$ S is $\mathbb{L}^2$-representable

but an additional variable x_0 was needed

$\Rightarrow$ *formally*, $S \subseteq \mathbb{R}^n$ is K-representable

iff there *exists* a set $T \subseteq \mathbb{R}^{n+m}$ such that

a. T is K-representable

b. $x \in S$ iff there exists $t \in \mathbb{R}^m$ such that $(x, t) \in T$

(i.e. S is the projection of T on $\mathbb{R}^n$)

Back to $\mathbb{R}^n_+$

- ◇ Polyhedrons and polytopes are $\mathbb{R}^n_+$-representable
- ◇ Hyperplanes and half-planes are $\mathbb{R}^n_+$-representable
- ◇ Affine functions $x \mapsto a^T x + b$ are $\mathbb{R}^n_+$-representable
- ◇ Absolute values $x \mapsto \left|a^T x + b\right|$ are $\mathbb{R}^n_+$-representable
- ◇ *Convex* piecewise linear function are $\mathbb{R}^n_+$-representable

Two potential issues with $\mathbb{R}^n_+$:

a. free variables in the primal $\rightarrow x = x^+ - x^-$

b. equalities in the dual $\rightarrow a^T x \leq c$ and $a^T x \geq c$

But these are **wrong** solutions !

What use is $K = \mathbb{R}^n$?

- $K = \mathbb{R}^n$ and $K^* = \{0\}$
- Can be used to introduce free variables in the primal $Ax = b, x \succeq_K 0$

$$x \succeq_{\mathbb{R}^n} 0 \quad \Leftrightarrow \quad x \text{ is free}$$

- or equalities in the dual $A^T y \preceq_{K^*} c$

$$A^T y \preceq_{\{0\}} c \quad \Leftrightarrow \quad A^T y = c$$

  in combination with other cones
- $\mathbb{R}^n$ in dual or $\{0\}$ is primal is useless!

What use is $\mathbb{L}^n$?

everything both convex and quadratic ...

- $f : x \mapsto \|x\|$, $f : x \mapsto \|x\|^2$ and $f : (x, z) \mapsto \frac{\|x\|^2}{z}$
- $B_r = \{x \in \mathbb{R}^n \mid \|x\| \leq r\}$
- $\{(x, y) \in \mathbb{R}^2_+ \mid xy \geq 1\}$
- $\{(x, y, z) \in \mathbb{R}^2_+ \times \mathbb{R} \mid xy \geq z^2\}$
- $\{(a, b, c, d) \in \mathbb{R}^4_+ \mid abcd \geq 1\}$
- $\{(x, t) \in \mathbb{R}^n \times \mathbb{R} \times \mid x^{\mathrm{T}} Q x \leq t\}$ with $Q \in \mathbb{S}^n_+$

$\Rightarrow$ second-order cone optimization

Very useful trick: $xy \geq z^2 \Leftrightarrow (x + y, x - y, 2z) \in \mathbb{L}^2$

Unfortunately, $(x, y) \mapsto \frac{x}{y}$ is not convex!

What use is $\mathbb{S}^n_+$?

Preliminary remark: for the purpose of conic optimization, members of $\mathbb{S}^n$ are viewed as vectors in $\mathbb{R}^{n \times n}$

What about constraint $Ax = b$?

$$Ax = b \Leftrightarrow a_i^T x = b_i \ \forall i$$

$a_i^T x$ can be views as the inner product between a_i and x

Let $X, Y \in \mathbb{S}^n$: their inner product is

$$X \bullet Y = \sum_{1 \leq i,j \leq n} X_{i,j} Y_{i,j} = \mathrm{trace}(XY)$$

$\rightarrow$ replace $a_i^T x$ by $A_i \bullet X$ with $A_i, X \in \mathbb{S}^n$

Standard format for semidefinite optimization

The primal becomes

$\min C \bullet X$ such that $A_i \bullet X = b_i \; \forall 1 \leq i \leq m$ and $X \succeq 0$

In the conic dual, we have

$$A^{\mathrm{T}} y = \sum a_i y_i, \text{ an application from } \mathbb{R}^m \mapsto \mathbb{R}^n$$

$\Rightarrow$ with the $\mathbb{S}^n_+$ cone, we have

$$\mathcal{A}(y) = \sum A_i y_i, \text{ an application from } \mathbb{R}^m \mapsto \mathbb{S}^n$$

which gives for the dual

$$\max b^{\mathrm{T}} y \text{ such that } \sum_{i=1}^{m} A_i y_i \preceq C$$

What use is $\mathbb{S}^n_+$ (continued) ?

- $\mathbb{S}^n_+$ *generalizes* both $\mathbb{R}^n_+$ and $\mathbb{L}^n$ (arrow matrices) (however, using $\mathbb{R}^n_+$ and $\mathbb{L}^n$ is more efficient)
- $f : X \mapsto \lambda_{max}(X)$ and $f : X \mapsto -\lambda_{min}(X)$
- $f : X \mapsto \max_i |\lambda_i| (X)$ (spectral norm)
- Describing ellipsoids $\{x \in \mathbb{R}^n \mid (x-c)^{\mathrm{T}} E(x-c) \leq 1\}$ with $E \succeq 0$
- Matrix constraint $XX^{\mathrm{T}} \preceq Y$ using the Schur Complement lemma

  When $A \succ 0$: $\begin{pmatrix} A & B \\ B^{\mathrm{T}} & C \end{pmatrix} \succeq 0 \Leftrightarrow C - B^{\mathrm{T}} A^{-1} B \succeq 0$
- And more ...

Primal-dual algorithms

Advantage of conic optimization over standard convex optimization is (symmetric) duality

However previous approach does not seem to use it !

$\Rightarrow$ a better approach that uses duality is needed

The linear case (again)

Introduce additional vector of variables $s \in \mathbb{R}^n$

$$\min c^T x \text{ such that } Ax = b \text{ and } x \geq 0$$

and

$$\max b^T y \text{ such that } A^T y + s = c \text{ and } s \geq 0$$

Primal-dual optimality conditions

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \geq 0$$

and

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y + s = c \text{ and } s \geq 0$$

Duality tells us x^* and y^* are optimal **iff** they satisfy

$$Ax = ,\ x \geq 0,\ A^{\mathrm{T}} y + s = c,\ s \geq 0 \text{ and } c^{\mathrm{T}} x = b^{\mathrm{T}} y$$

or

$$Ax = b,\ x \geq 0,\ A^{\mathrm{T}} y + s = c,\ s \geq 0 \text{ and } x_i s_i = 0\ \forall i$$

Both problems are handled simultaneously

Perturbed optimality conditions

Introducing a logarithmic barrier term in both problems

$$\min c^{\mathrm{T}}x - \mu \sum_i \log x_i \text{ such that } Ax = b \text{ and } x > 0$$

$$\max b^{\mathrm{T}}y + \mu \sum_i \log s_i \text{ such that } A^{\mathrm{T}}y + s = c \text{ and } s > 0$$

one can derive new perturbed optimality conditions

$$Ax = b,\ x \geq 0,\ A^{\mathrm{T}}y + s = c,\ s \geq 0 \text{ and } x_i s_i = \mu \ \forall i$$

Again, both problems are handled simultaneously

Primal-dual path following algorithm

Same principle as in the general case:

◇ Follow the central path

◇ Not wandering too far from it

◇ Until (primal-dual) optimality

◇ Using a polynomial number of iterations

Complexity is also the same:

$$\mathcal{O}\left(\sqrt{n}\log\frac{1}{\varepsilon}\right) \text{ iterations to get } \varepsilon \text{ accuracy}$$

But this scheme is very efficient in practice (long steps)
(all practical implementations use it nowadays)

What about other convex/conic problems ?

This primal-dual scheme is only generalizable to cones that are

a. self-dual ($K = K^*$)

b. homogeneous

(linear automorphism group acts transitively on int K)

(*[Nesterov & Todd 97]*)

There exists a complete classification of these cones : in the real case, they are ...

$$\mathbb{R}^n_+ \quad , \quad \mathbb{L}^n \quad \text{and} \quad \mathbb{S}^n_+$$

and their Cartesian products!

Complexity

Complexity for a product of $\mathbb{R}^n_+, \mathbb{L}^n, \mathbb{S}^n_+$

$$\mathcal{O}\left(\sqrt{\nu}\log\frac{1}{\varepsilon}\right) \text{ iterations to get } \varepsilon \text{ accuracy}$$

where ν is the sum of

- ◇ n for $\mathbb{R}^n_+$ (see above) (barrier term is $-\sum \log x_i$)
- ◇ n for $\mathbb{S}^n_+$ (although there are $n(n+1)/2$ variables) (barrier term is $-\log\det X = -\sum \log \lambda_i$)
- ◇ 2 for $\mathbb{L}^n$ (independently of n !) (barrier term is $-\log(x_0^2 - \sum x_i^2)$; no $-\log x_0$ term!)

$\longrightarrow$ these problems are solved very efficiently in practice

More applications

Using semidefinite optimization:

Positive polynomials

- ◇ Single variable case: exact formulation
- ◇ Test positivity and minimize on an interval
- ◇ Multiple variable case: relaxation only

The MAX-CUT relaxation

- ◇ Relaxation of a difficult discrete problem
- ◇ With a quality guarantee (0.878)

References

Convex analysis, optimization and modelling

◇ Convex Analysis, Rockafellar,
Princeton University Press, 1980

◇ Convex optimization, Boyd and Vandenberghe,
Cambridge University Press, 2004 (on the web)

Conic convex modelling

◇ Lectures on Modern Convex Optimization, Analysis, Algorithms, and Engineering Applications, Ben-Tal and Nemirovski,
MPS/SIAM Series on Optimization, 2001

Interior-point methods (linear)

- ◇ Primal-Dual Interior-Point Methods, Wright
  SIAM, 1997
- ◇ Theory and Algorithms for Linear Optimization, Roos, Terlaky, Vial, John Wiley & Sons, 1997

Interior-point methods (convex)

- ◇ Interior-point polynomial algorithms in convex programming, Nesterov & Nemirovski, SIAM, 1994
- ◇ A Mathematical View of Interior-Point Methods in Convex Optimization, Renegar,
  MPS/SIAM Series on Optimization, 2001

Semidefinite optimization applications

- ◇ Handbook of Semidefinite Programming, WOLKOWICZ, SAIGAL, VANDENBERGHE (eds.) Kluwer, 2000
- ◇ Semidefinite programming, BOYD, VANDENBERGHE, SIAM Review 38 (1), 1996

Software: a few choices among many others

- ◇ Linear & second-order cone: MOSEK (commercial)
- ◇ Linear, sec.-ord. & semidefinite: SeDuMi (free)
- ◇ Modeling languages: AMPL, YALMIP

Thank you for your attention

Does linear optimization **exist** at all ?

Let us only mention the following *not so well-known* theorem, due to Dr. Addock PRILFIRST

Theorem

The objective function of any linear program is constant on its feasible region

Proof

$$
\begin{aligned}
& \{\min c^{\mathrm{T}} x \mid Ax = b, x \geq 0\} = \{\max b^{\mathrm{T}} y \mid A^{\mathrm{T}} y \leq c\} \\
\geq\ & \{\min b^{\mathrm{T}} y \mid A^{\mathrm{T}} y \leq c\} = \{\max c^{\mathrm{T}} x \mid Ax = b, x \leq 0\} \\
\geq\ & \{\min c^{\mathrm{T}} x \mid Ax = bx \leq 0\} = \{\max b^{\mathrm{T}} y \mid A^{\mathrm{T}} y \geq c\} \\
\geq\ & \{\min b^{\mathrm{T}} y \mid A^{\mathrm{T}} y \geq c\} = \{\max c^{\mathrm{T}} x \mid Ax = b, x \geq 0\} \\
\geq\ & \{\min c^{\mathrm{T}} x \mid Ax = b, x \geq 0\}
\end{aligned}
$$