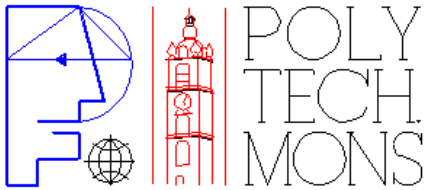
A conic approach for separable convex optimization

François Glineur

F.N.R.S. Postodoctoral Researcher

Francois.Glineur@fpms.ac.be



Faculté Polytechnique de Mons

INFORMS Annual Meeting 2001

Miami Beach, November 5, 2001

Outline

Introduction

- ◇ Convex and conic optimization, duality theorems
- ◇ Geometric optimization: A special case of separable optimization

Geometric optimization

- ◇ The geometric cone
- ◇ Conic formulation and duality

Generalization to separable optimization

- ◇ The separable cone
- ◇ Separable convex optimization
- ◇ Interior-point methods and self-concordant barriers

Concluding remarks

Introduction

Convex optimization

Let $f_0 : \mathbb{R}^n \mapsto \mathbb{R}$ a convex function and $C \subseteq \mathbb{R}^n$ a convex set

$$\inf_{x \in \mathbb{R}^n} f_0(x) \quad \text{s.t.} \quad x \in C$$

Properties

- ◇ Local optima ⇒ global, form a convex optimal set
- ◇ Lagrange duality ⇒ related (asymmetric) dual problem
- ◇ Efficient *interior-point* methods (*self-concordant barriers*)

Conic optimization

Let $\mathcal{C} \subseteq \mathbb{R}^n$ a *solid, pointed, closed convex cone* :

$$\inf_{x \in \mathbb{R}^n} c^T x \quad \text{s.t.} \quad Ax = b \text{ and } x \in \mathcal{C} \quad \Rightarrow \textbf{Equivalent} \text{ setting}$$

Primal-dual pair

Dual cone is also a solid pointed closed convex cone

$$\mathcal{C}^* = \left\{ x^* \in \mathbb{R}^n \mid x^T x^* \geq 0 \text{ for all } x \in \mathcal{C} \right\}$$

$\Rightarrow$ pair of primal-dual problems

$$\inf_{x \in \mathbb{R}^n} \; c^T x \quad \text{s.t.} \quad Ax = b \text{ and } x \in \mathcal{C}$$

$$\sup_{(y,s) \in \mathbb{R}^{m+n}} \; b^T y \quad \text{s.t.} \quad A^T y + s = c \text{ and } s \in \mathcal{C}^*$$

Examples: $\mathcal{C} = \mathbb{R}^n_+ = \mathcal{C}^* \Rightarrow$ linear optimization, $\mathcal{C} = \mathbb{S}^n_+ = \mathcal{C}^* \Rightarrow$ semidefinite optimization (self-duality)

Advantages over classical formulation

- ◇ Remarkable primal-dual symmetry
- ◇ Special handling of (*easy*) linear equality constraints

Duality theorems

Weak duality

Every feasible primal (resp. dual) solution is an upper (resp. lower) bound on all feasible dual (resp. primal) solutions.

(*Conic case* : $b^T y \leq c^T x$ for all feasible x, y)

$\Rightarrow p^* - d^* \geq 0$ is called the *duality gap* (always nonnegative).

This is valid even when no convexity is present.

Strong duality (Slater condition)

The duality gap is guaranteed to be zero ($p^* = d^*$) if

- $\diamond$ Both problems are convex *and*
- $\diamond$ $\exists$ a Slater point, i.e. a strictly feasible (interior) solution

(*Conic case* : x is *strictly feasible* $\Leftrightarrow$ x is feasible and $x \in \operatorname{int} \mathcal{C}$)

But sometimes strong duality holds without Slater condition

Strong duality with no Slater point ...

Can **fail**: Using the *second-order cone*
$\mathbb{L}^n = \{(x_1, \ldots, x_{n+1}) \in \mathbb{R}^n \times \mathbb{R}_+ \mid \sum_{i=1}^n x_i^2 \leq x_{n+1}^2\}$

$$\inf x_2 \quad \text{s.t.} \quad (x_1, x_2, x_3) \in \mathbb{L}^2 \text{ and } x_1 - x_3 = 0$$

$$\sup 0\, y \quad \text{s.t.} \quad (x_1^*, x_2^*, x_3^*) \in \mathbb{L}^2 \text{ and } \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} y + \begin{pmatrix} x_1^* \\ x_2^* \\ x_3^* \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

We have $p^* = 0$ and $d^* = -\infty \Rightarrow$ infinite duality gap

Can **hold**: Linear optimization always features a zero duality gap

(*except* when both problems are infeasible: we have then $p^* = +\infty$ and $d^* = -\infty \Rightarrow$ infinite duality gap)

Separable convex optimization

Goal: study a large *well-behaved* subset of convex optimization

---

DEFINITION

Consider n scalar closed proper convex functions $f_i : \mathbb{R} \mapsto \mathbb{R} \cup \{+\infty\}$, $K = \{1, \ldots, r\}$ and a partition $\{I_k\}_{k \in R}$ of $\{1, \ldots, n\}$

$$\sup \; b^T y \;\; \text{s.t.} \; \sum_{i \in I_k} f_i(c_i - a_i^T y) \leq d_k - g_k^T y \quad \forall k \in R$$

- ◇ Linear, quadratic, geometric, entropy, l_p-norm optimization
- ◇ Mix different types of constraints
- ◇ Linear objective without loss of generality
- ◇ Links with Young optimization

Geometric optimization

Optimize positive vector using *posynomials* as objective/constraints

$$\inf_{t \in \mathbb{R}^m_{++}} G_0(t) \quad \text{s.t.} \quad G_k(t) \leq 1 \; \forall k \in R \qquad \text{with posynomials } G_i\text{'s}$$

A *posynomial* is a sum of positive monomials : $2\frac{t_1^2}{t_2} + 3\sqrt{t_2} + \frac{1}{3}\frac{t_2^{2/3}}{t_1 t_3^3}$

Formally: monomials indexed by $i \in I = \{1, \ldots, n\}$, posynomials indexed by $k \in R = \{0, 1, \ldots, r\}$ using a partition $\{I_k\}_{k \in R}$ of I.

$$G_k : \mathbb{R}^m_{++} \mapsto \mathbb{R}_{++} : t \mapsto \sum_{i \in I_k} C_i \prod_{j=1}^{m} t_j^{a_{ij}} \qquad \text{with } a_{ij} \in \mathbb{R} \text{ and } C_i \in \mathbb{R}_{++}$$

Many applications (e.g. in engineering: model or approximate relations), generalizes linear optimization.

Not convex in this form (take $G_0(t) = \sqrt{t_1}$) but can be *convexified* !

Convexification

W.l.o.g. we consider a **linear** objective and let $t_j = e^{y_j}$ for all j

Primal:

$$\sup_{y \in \mathbb{R}^m} b^T y \quad \text{s.t.} \quad g_k(y) \leq 1 \text{ for all } k \in R$$

using $c_i = -\log C_i$ and $g_k(y)$ such that $g_k(y) = G_k(t)$, i.e.

$$g_k : \mathbb{R}^m \mapsto \mathbb{R}_{++} : y \mapsto \sum_{i \in I_k} e^{a_i^T y - c_i}$$

Dual: (Lagrangean)

$$\inf \; c^T x + \sum_{k \in R} \sum_{\substack{i \in I_k \\ x_i > 0}} x_i \log \frac{x_i}{\sum_{i \in I_k} x_i} \quad \text{s.t.} \quad Ax = b \text{ and } x \geq 0$$

Results (Duffin, Peterson and Zener, 1967)

- ◇ Convex problem, weak duality, strong duality with a Slater point
- ◇ Never a duality gap ! ⇒ strong duality without a Slater point

The geometric cone

Objective: model geometric optimization with cones

DEFINITION *[Gl. 99]*

Let $n \in \mathbb{N}$. We define $\mathcal{G}^n$ as

$$\mathcal{G}^n = \left\{ (x, \theta) \in \mathbb{R}^n_+ \times \mathbb{R}_+ \mid \sum_{i=1}^{n} e^{-\frac{x_i}{\theta}} \leq 1 \right\} \quad (\text{taking } e^{-\frac{x_i}{0}} = 0)$$

PROPERTIES

- $\mathcal{G}^0 = \mathbb{R}_+$, $\mathcal{G}^1 = \mathbb{R}^2_+$, $\mathcal{G}^n$ is a *solid, pointed, closed convex* cone.
- $\operatorname{int} \mathcal{G}^n$ can be computed (replace $\leq$ by $<$ and $\mathbb{R}_+$ by $\mathbb{R}_{++}$).
- Dual cone $(\mathcal{G}^n)^*$ can be computed explicitly (and is an epigraph)

$$(\mathcal{G}^n)^* = \left\{ (x^*, \theta^*) \in \mathbb{R}^n_+ \times \mathbb{R} \mid \theta^* \geq \sum_{x_i^* > 0} x_i^* \log \frac{x_i^*}{\sum_{i=1}^n x_i^*} \right\}$$

A conic formulation for geometric optimization

FORMULATION WITH $\mathcal{G}^n$ CONE

$v_I/M_I \equiv$ restriction of vector v/matrix M to subset of indices I

Primal

$$\sup\ b^T y \quad \text{s.t.} \quad g_k(y) \leq 1 \text{ for all } k \in R$$

Introducing variables $s_i = c_i - a_i^T y$ for all $i \in I$ we get

$$\sup\ b^T y \quad \text{s.t.} \quad s = c - A^T y \text{ and } \sum_{i \in I_k} e^{-s_i} \leq 1 \text{ for all } k \in R$$

$$\Updownarrow$$

$$\sup\ b^T y \quad \text{s.t.} \quad \begin{pmatrix} A^T \\ 0 \end{pmatrix} y + \begin{pmatrix} s \\ v \end{pmatrix} = \begin{pmatrix} c \\ e \end{pmatrix} \text{ and } (s_{I_k}, v_k) \in \mathcal{G}^{n_k}\ \forall k \in R$$

e is the all-one vector, $n_k = \#I_k$ and v_i's are fictitious variables.

Advantages of the conic approach

- ◇ The dual problem can be very easily formulated once the dual cone is computed
- ◇ Duality properties can be derived from the general duality theorems of conic optimization, in particular:
  - Proving weak duality is straightforward
  - Proving strong duality with a Slater point is also simple
  - In the case of geometric optimization, there is *no duality gap*, even in the absence of a Slater point:

    this can also be shown using the conic duality theorem (and gives simpler proofs and more insight than the classical techniques)

Generalizing this conic framework

Comparing cones

$$\mathcal{G}^n = \left\{(x,\theta) \in \mathbb{R}^n_+ \times \mathbb{R}_+ \mid \sum_{i=1}^n e^{-\frac{x_i}{\theta}} \leq 1\right\}$$

$$\mathcal{L}^p = \left\{(x,\theta,\kappa) \in \mathbb{R}^n \times \mathbb{R}^2_+ \mid \sum_{i=1}^n \frac{|x_i|^{p_i}}{p_i \theta^{p_i-1}} \leq \kappa\right\}$$

Variants

$$\mathcal{G}^n_2 = \left\{(x,\theta,\kappa) \in \mathbb{R}^n_+ \times \mathbb{R}^2_+ \mid \theta \sum_{i=1}^n e^{-\frac{x_i}{\theta}} \leq \kappa\right\}$$

$$\mathcal{L}^p = \left\{(x,\theta,\kappa) \in \mathbb{R}^n \times \mathbb{R}^2_+ \mid \theta \sum_{i=1}^n \frac{1}{p_i} \left|\frac{x_i}{\theta}\right|^{p_i} \leq \kappa\right\}$$

The separable cone *[Gl. 00]*

Consider a set of n scalar closed proper convex functions $f_i : \mathbb{R} \mapsto \mathbb{R}$

$$\mathcal{K}^f = \mathrm{cl}\,\mathcal{K}^{\circ f} = \mathrm{cl}\Big\{(x, \theta, \kappa) \in \mathbb{R}^n \times \mathbb{R}_{++} \times \mathbb{R} \mid \theta \sum_{i=1}^{n} f_i(\frac{x_i}{\theta}) \leq \kappa\Big\}$$

- $\mathcal{K}^f$ is a closed convex cone and $\mathcal{K}^f \subseteq \mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{R}$
- $\mathcal{K}^f$ generalizes $\mathcal{L}^p$ and $\mathcal{G}_2^n$ (case $\theta = 0$ handled with cl operation)
- The dual can be computed: very symmetric formulation

$$(\mathcal{K}^f)^* = \mathrm{cl}\Big\{(x^*, \theta^*, \kappa^*) \in \mathbb{R}^n \times \mathbb{R}_{++} \times \mathbb{R} \mid \kappa^* \sum_{i=1}^{n} f_i^*(-\frac{x_i^*}{\kappa^*}) \leq \theta^*\Big\}$$

using the conjugate functions (also closed, proper and convex)

$$f_i^* : \mathbb{R} \mapsto \mathbb{R} : x^* \mapsto \sup_{x \in \mathbb{R}^n} \{xx^* - f_i(x)\}$$

Additional properties

- $\operatorname{int} \mathcal{K}^f$ can be identified ($\Rightarrow$ Slater condition): $\operatorname{int} \mathcal{K}^f = \operatorname{int} \mathcal{K}^{\circ f} =$

$$\left\{(x, \theta, \kappa) \in \mathbb{R}^n \times \mathbb{R}_{++} \times \mathbb{R} \mid x_i \in \operatorname{int} \operatorname{dom} f_i \text{ and } \theta \sum_{i=1}^{n} f_i(\frac{x_i}{\theta}) < \kappa\right\}.$$

- If we require in addition that $\operatorname{int} \operatorname{dom} f_i \neq \emptyset$ and $\operatorname{int} \operatorname{dom} f_i^* \neq \emptyset$, we have that $\mathcal{K}^f$ and $(\mathcal{K}^f)^*$ are solid and pointed
- Points with $\theta = 0$ can be identified

$$\mathcal{K}^f \setminus \mathcal{K}^{\circ f} = \left\{(x, 0, \kappa) \mid \sum_{i=1}^{n} \lim_{\theta \to 0} \theta f_i(\frac{x_i}{\theta}) \leq \kappa\right\}.$$

- l_p-norm optimization: $f_i : x \mapsto \frac{1}{p_i} |x|^{p_i}$ and $f_i^* : x^* \mapsto \frac{1}{q_i} |x^*|^{q_i}$
- Geometric optimization: $f_i : x \mapsto e^{-x}$ and $f_i^* : \mathbb{R}_+ \mapsto \mathbb{R}$: $x^* \mapsto x^*(1 - \log(-x^*))$ for $x^* < 0$, 0 for $x^* = 0$, $+\infty$ for $x^* > 0$.

Formulation with $\mathcal{K}^f$ cone

Primal

$$\sup\ b^T y\ \text{ s.t. }\ \sum_{i\in I_k} f_i(c_i - a_i^T y) \le d_k - g_k^T y \quad \forall k \in R$$

Introducing variables $x_i^* = c_i - a_i^T y$ and $z_k^* = d_k - g_k^T y$ we get

$$\sup\ b^T y \quad \text{s.t.} \quad x^* = c - A^T y,\ z^* = d - G^T y,\ \sum_{i\in I_k} f_i(x_i^*) \le z_k^*\ \forall k \in R$$

$$\Updownarrow$$

$$\sup\ b^T y \quad \text{s.t.} \quad \begin{pmatrix} A^T \\ G^T \\ 0 \end{pmatrix} y + \begin{pmatrix} x^* \\ z^* \\ v^* \end{pmatrix} = \begin{pmatrix} c \\ d \\ e \end{pmatrix},\ (x_{I_k}^*, v_k^*, z_k^*) \in \mathcal{K}^{f_{I_k}}\ \forall k \in R$$

(e is the all-one vector and v_i's are fictitious variables)

$\Rightarrow$ standard dual conic problem based on data $(\tilde{A}, \tilde{b}, \tilde{c})$ and cone $\mathcal{C}^*$

$$\tilde{A} = \begin{pmatrix} A & G & 0 \end{pmatrix},\ \tilde{b} = b,\ \tilde{c} = \begin{pmatrix} c \\ d \\ e \end{pmatrix},\ \mathcal{C}^* = \mathcal{K}^{f_{I_1}} \times \cdots \times \mathcal{K}^{f_{I_r}}\ .$$

$\Rightarrow$ we can mechanically derive the ***dual*** !

$$\inf \begin{pmatrix} c \\ d \\ e \end{pmatrix}^T \begin{pmatrix} x \\ z \\ v \end{pmatrix} \quad \text{s.t.} \quad \begin{pmatrix} A & G & 0 \end{pmatrix} \begin{pmatrix} x \\ z \\ v \end{pmatrix} = b \text{ and } (x_{I_k}, v_k, z_k) \in (\mathcal{K}^{f_{I_k}})^*$$

$$\Updownarrow$$

$$\inf c^T x + d^T z + e^T v \quad \text{s.t.} \quad Ax + Gz = b, z \geq 0 \text{ and } v_k \geq z_k \sum_{i \in I_k} f_i^*(-\frac{x_i}{z_k})$$

$$\Leftrightarrow \quad \inf c^T x + d^T z + \sum_{k \in R} z_k \sum_{i \in I_k} f_i^*(-\frac{x_i}{z_k}) \quad \text{s.t.} \quad Ax + Gz = b \text{ and } z \geq 0$$

(taking the limit if necessary when $z_k = 0$)

Some other types of constraints

◇ Model circle/ellipses in $\mathbb{R}^2$

$$f_1 : x \mapsto \begin{cases} -\sqrt{a^2 - x^2} & \text{if } |x| \leq a \\ +\infty & \text{if } |x| > a \end{cases} \qquad f_1^* : x^* \mapsto a\sqrt{1 + x^{*2}}$$

◇ CES functions (consumer theory), $0 < p < 1, q < 0, \frac{1}{p} + \frac{1}{q} = 1$

$$f_2 : x \mapsto \begin{cases} -\frac{1}{p}x^p & \text{if } x \geq 0 \\ +\infty & \text{if } x < 0 \end{cases} \qquad f_2^* : x^* \mapsto \begin{cases} -\frac{1}{q}(-x^*)^q & \text{if } x^* < 0 \\ +\infty & \text{if } x^* \geq 0 \end{cases}$$

◇ Logarithms (with property that $f^*(x^*) = f(-x^*)$)

$$f_3 : x \mapsto \begin{cases} -\frac{1}{2} - \log x & x > 0 \\ +\infty & x \leq 0 \end{cases} \qquad f_3^* : x^* \mapsto \begin{cases} -\frac{1}{2} - \log(-x^*) & x^* < 0 \\ +\infty & x^* \geq 0 \end{cases}$$

Duality in separable optimization

Weak duality

If y is feasible for the primal and (x, z) is feasible for the dual, we have

$$b^T y \leq c^T x + d^T z + \sum_{k \in R} z_k \sum_{i \in I_k} f_i^*(-\frac{x_i}{z_k}) \ .$$

Strong duality

If the primal and the dual are feasible, their optimum objective values are equal (but not necessarily attained) $\Rightarrow$ *zero* duality gap *without* Slater condition.

Interior-point methods and self-concordant barriers

According to *[Nesterov & Nemirovsky]*, a convex problem in conic format is solvable by a primal short-step interior-point algorithm in polynomial time if $\mathcal{C}$ admits a computable *self-concordant* barrier.

A solution of accuracy ϵ can be reached in $\mathcal{O}\left(\sqrt{\nu}\log\frac{1}{\epsilon}\right)$ iterations.

Definition: A function $F : \operatorname{int}\mathcal{C} \mapsto \mathbb{R}$ is called a *self-concordant* barrier with parameter ν on $\mathcal{C}$ iff

- $\diamond$ F is convex and three times differentiable
- $\diamond$ $F(x) \to +\infty$ when $x \to \partial\mathcal{C}$
- $\diamond$ the following two conditions hold for all $x \in \operatorname{int}\mathcal{C}$ and $h \in \mathbb{R}^n$

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

$$\nabla F(x)^T(\nabla^2 F(x))^{-1}\nabla F(x) \leq \nu$$

Application to separable optimization

Given a self-concordant barrier F_i with parameter ν_i for each two-dimensional epigraph epi f_i, $1 \leq i \leq n$ (*easy* to find), there exists a self-concordant barrier F for $\mathcal{K}^f$ with parameter $\mathcal{O}\left(\sum_{i=1}^n \nu_i\right)$

Outline of the proof

1. Form the Cartesian product $X = \text{epi}\, f_1 \times \cdots \times \text{epi}\, f_n$

$$X = \{(x, y) \in \mathbb{R}^{2n} \mid f_i(x_i) \leq y_i \ \forall i\}$$

$\Rightarrow F_X(x, y) = \sum_{i=1}^n F_i(x_i, y_i)$ is s.c. for X with $\nu_X = \sum_{i=1}^n \nu_i$

2. Extend X with a linear variable to

$$X' = \{(x, y, \kappa) \in \mathbb{R}^{2n+1} \mid f_i(x_i) \leq y_i \ \forall i \text{ and } \kappa = \sum_{i=1}^n y_i\}$$

This additional variable κ *links* the epigraphs,
$F'_X(x, y, \kappa) = F(x, y)$ is still s.c. for X' with $\nu_{X'} = \sum_{i=1}^n \nu_i$

3. Consider the closed conic hull of X' (introducing a homogenizing variable θ)

$$Y = \mathrm{cl}\{(x, y, \kappa, \theta) \in \mathbb{R}^{2n+2} \mid (x/\theta, y/\theta, \kappa/\theta) \in X'\}$$

   There exists a s.c. barrier F_Y for Y with $\nu_Y = \mathcal{O}(\nu_{X'})$

4. The projection of Y on the space (x, κ, θ) is exactly equal to $\mathcal{K}^f$
   $\Rightarrow$ we have a s.c.b. for $\mathcal{K}^f$ with parameter $\mathcal{O}(\sum_{i=1}^n \nu_i)$

Indeed, we have

$$(x/\theta, y/\theta, \kappa/\theta) \in X' \Leftrightarrow f_i(\frac{x_i}{\theta}) \leq \frac{y_i}{\theta} \ \forall i \text{ and } \frac{\kappa}{\theta} = \sum_{i=1}^n \frac{y_i}{\theta}$$

which clearly shows that

$$(x, \kappa, \theta) \in \mathcal{K}^f \Leftrightarrow \exists y \in \mathbb{R}^n \mid (x, y, \kappa, \theta) \in Y$$

$\Rightarrow$ separable problems can be solved in $\mathcal{O}\left(\sqrt{\sum_{i=1}^n \nu_i} \log \frac{1}{\epsilon}\right)$ iterations
($\rightarrow$ polynomial time if each F_i is computable in polynomial time)

Concluding remarks

- ◇ A very wide class of *separable* convex problems can be formulated (including linear, quadratic, entropy, l_p-norm and geometric optimization)
- ◇ Using this setting, interesting *duality* properties can be obtained in a seamless way (weak and strong duality *without* Slater condition)
- ◇ Finding a computable *self-concordant* barrier for the $\mathcal{K}^f$ cone (which can be done using self-concordant barriers on 2-dimensional epigraphs of scalar functions f_i) provides a primal algorithm with *polynomial* complexity (short-step path-following interior-point method, *[Nesterov & Nemirovsky 83]*)
- ◇ Primal-dual formulation $\Rightarrow$ first step towards *true primal-dual* algorithms (*self-regular* barriers, *[Peng et al. 00]*)