Zero duality gap for a large class of separable convex problems

François Glineur

F.N.R.S. research fellow
Faculté Polytechnique de MONS
Francois.Glineur@fpms.ac.be

MOPTA'01
August 2, 2001

First McMaster Optimization Conference
MCMASTER university, HAMILTON, CANADA

Outline

Introduction

◇ Convex and conic optimization
◇ Duality properties

Separable convex optimization

◇ Definition
◇ Conic formulation

Duality for separable convex optimization

◇ Weak and strong duality
◇ 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 $\Rightarrow$ global, form a convex optimal set
- Lagrange duality $\Rightarrow$ 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* :

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

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

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

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

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 for convex optimization

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

- ⋄ Both problems are convex and
- ⋄ $\exists$ 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

More on strong duality

Strong duality can fail if there is no Slater point: 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

Linear optimization always features a zero duality gap (no need for a Slater condition) except when both problems are infeasible:

$$\min -x_2 \quad \text{s.t.} \quad x_1 = -1 \text{ and } x_1, x_2 \geq 0$$

$$\max -y \quad \text{s.t.} \quad \begin{pmatrix} 1 \\ 0 \end{pmatrix} y + \begin{pmatrix} x_1^* \\ x_2^* \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

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

Separable convex optimization

Definition

Consider a set of n scalar closed proper convex functions $f_i : \mathbb{R} \mapsto \mathbb{R}$, $K = \{1, \ldots, r\}$ 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

Goal: find the dual problem and study duality properties

Strategy: use conic formulation

The separable cone *[Gl. 00]*

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

- $\mathcal{K}^f$ is a closed convex cone and $\mathcal{K}^f \subseteq \mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{R}$
- The dual can be computed: very symmetric formulation

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

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 and examples

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

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

- 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 $f^*(x^*) = f(-x^*)$)

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

Duality in separable optimization

Weak duality

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

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

Proof. Use weak duality theorem on conic primal-dual pair and extend objective values to the separable optimization problems (easy).

Strong duality

Assume f_i is finite and co-finite for all i, i.e. $\operatorname{dom} f_i = \operatorname{dom} f_i^* = \mathbb{R}$ (e.g. in the case of quadratic optimization, l_p-norm optimization).

Theorem. 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.

This strong duality property is not valid for all convex problems but depends on the *specific scalar structure* of separable optimization.

Proof

Proceed by proving the existence of a strictly feasible point for the dual *conic* program $\Leftrightarrow v_k > z_k \sum_{i \in I_k} f_i^*(-\frac{x_i}{z_k})$ and $z_k > 0$.
But the linear constraints $Ax + Gz = b$ may force $z_k = 0$ for some k for every feasible solution !
$\Rightarrow$ detect these zero z_k components and form a restricted primal-dual pair without these variables $\Rightarrow$ strong duality holds

Detection

Use the following linear problem

$$\min\ 0 \quad \text{s.t.} \quad Ax + Gz = b \quad \text{and} \quad z \geq 0$$

Define $\mathcal{N}$ = set of indices k such that z_k is identically zero on the feasible region and $\mathcal{B}$ the set of the other indices: $(\mathcal{B}, \mathcal{N})$ is the optimal partition of this linear problem (Goldman-Tucker theorem)

Strategy diagram

$$\begin{array}{ccccccc}
(P) & \equiv & (\text{Conic } P) & \xleftrightarrow{\text{Weak duality}} & (\text{Conic } D) & \equiv & (D) \\
 & & \updownarrow^{*\text{p.}} & & \updownarrow & & \\
 & & (\text{Relaxed } P) & \xleftrightarrow{\text{Strong duality}} & (\text{Restricted } D) & & \\
 & & & & \uparrow & & \\
 & & & & (\text{Slater condition}) & &
\end{array}$$

Strategy

Remove variables z_k for all $i \in \mathcal{N}$ from the dual

$\Rightarrow$ restricted dual problem (less variables)

$\Rightarrow$ relaxed primal (less constraints)

$\Rightarrow$ restricted dual has a strictly feasible solution $\Rightarrow$ zero duality gap.

We now have to prove that

- Optimal objective values are equal for restricted and original dual problems (easy)

- Optimal objective values are equal for relaxed and original primal. But the optimal solution of the relaxed primal (with zero duality gap) is not necessarily feasible for the original primal.
  Solution: perturb the relaxed primal optimal solution with a well-chosen vector (existence of a perturbation vector with the correct properties guaranteed by the Goldman-Tucker theorem)

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

- F is convex and three times differentiable
- $F(x) \longrightarrow +\infty$ when $x \longrightarrow \partial\mathcal{C}$
- 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, $1 \leq i \leq n$, 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

a. 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 $\nu_X = \sum_{i=1}^n \nu_i$

b. 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 $\nu_{X'} = \sum_{i=1}^n \nu_i$

c. 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 $\nu_Y = \mathcal{O}\left(\nu_{X'}\right)$

d. 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 $\mathcal{O}\left(\sum_{i=1}^n \nu_i\right)$

Indeed, we have

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

which clearly shows that

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

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]*)