A conic approach for separable convex optimization

François Glineur

Aspirant F.N.R.S.

`Francois.Glineur@fpms.ac.be`

Faculté Polytechnique de Mons

Mathematical Programming Seminars

CORE, UCL, April 24, 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

- $\diamond$ Local optima $\Rightarrow$ global, form a convex optimal set
- $\diamond$ Lagrange duality $\Rightarrow$ related (asymmetric) dual problem
- $\diamond$ 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

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

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

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 for conic optimization

Weak duality.

$b^T y \leq c^T x$ for every feasible x and y, with equality iff $x^T s = 0$

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

Strong duality (Slater condition)

Definition: x is *strictly feasible* $\Leftrightarrow$ x is feasible and $x \in \operatorname{int} \mathcal{C}$

$\exists$ strictly feasible point for primal $\Rightarrow$ zero duality gap, i.e. $p^* = d^*$.

Moreover, if dual is feasible, d^* is attained, i.e. $d^* = \max b^T y$

Corollary

If both the primal and the dual have a strictly feasible point, the duality gap is zero and both optimal values are attained.

Our approach: Use these general theorems to prove duality properties for specific classes of problems

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.

But **not** convex (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)

- $\diamond$ Convex problem $\Rightarrow$ weak duality
- $\diamond$ No duality gap !

The geometric cone

Objective: model geometric optimization with cones

Definition *[Gl. 99]*

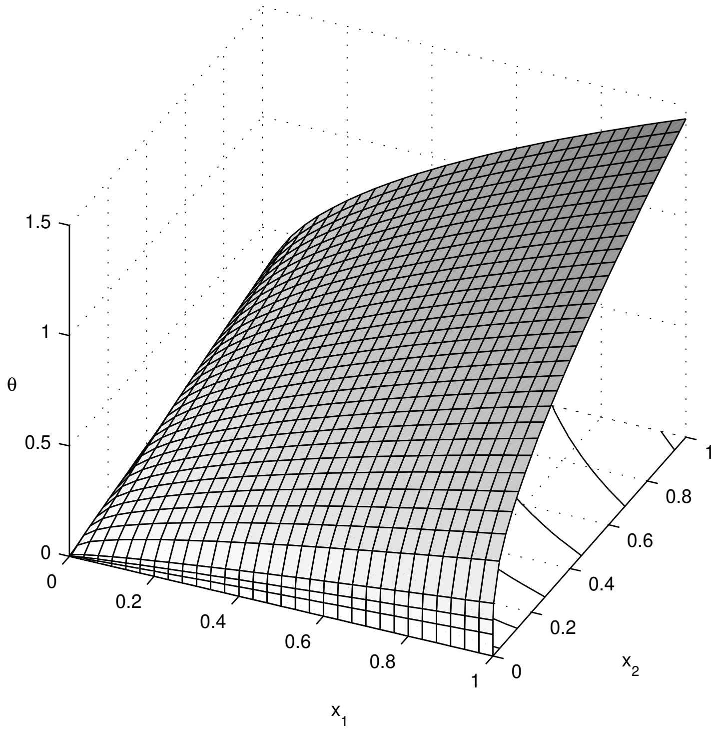
Let $n \in \mathbb{N}$. We define $\mathcal{G}^n$ as

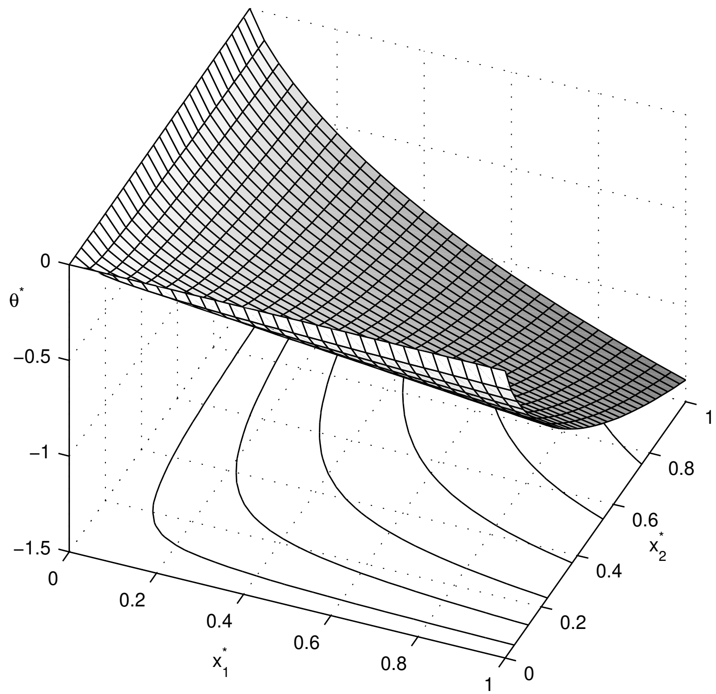
$$\mathcal{G}^n = \left\{ (x, \theta) \in \mathbb{R}^n_+ \times \mathbb{R}_+ \mid \sum_{i=1}^n e^{-\frac{x_i}{\theta}} \leq 1 \right\} \quad \text{(taking } e^{-\frac{x_i}{0}} = 0\text{)}$$

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)

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


1.5
1
θ
0.5
0
0
0.2
0.4
0.6
0.8
1
x1
0
0.2
0.4
0.6
0.8
1
x2


0
θ*
−0.5
−1
−1.5
0
0.2
0.4
0.6
0.8
1
x1*
0
0.2
0.4
0.6
0.8
1
x2*

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.

$\Rightarrow$ dual conic problem with variables $(\tilde{y}, \tilde{s})$, data $(\tilde{A}, \tilde{b}, \tilde{c})$ and cone $\mathcal{C}^*$

$$\tilde{y} = y,\ \tilde{s} = \begin{pmatrix} s \\ v \end{pmatrix},\ \tilde{A} = \begin{pmatrix} A & 0 \end{pmatrix},\ \tilde{b} = b,\ \tilde{c} = \begin{pmatrix} c \\ e \end{pmatrix},\ \mathcal{C}^* = \mathcal{G}^{n_1} \times \cdots \times \mathcal{G}^{n_r}\ .$$

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

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

$$\Updownarrow$$

$$\inf c^T x + e^T z \quad \text{s.t.} \quad Ax = b,\ x_{I_k} \geq 0 \text{ and } z_k \geq \sum_{\substack{i \in I_k \\ x_i > 0}} x_i \log \frac{x_i}{\sum_{i \in I_k} x_i}$$

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

Duality in geometric optimization

WEAK DUALITY

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

$$b^T y \leq c^T x + \sum_{k \in R} \sum_{\substack{i \in I_k \\ x_i > 0}} x_i \log \frac{x_i}{\sum_{i \in I_k} x_i} \ .$$

Equality occurs if and only if

$$\Big(\sum_{i \in I_k} x_i \Big) e^{a_i^T y - c_i} = x_i \text{ for all } i \in I_k, k \in R \ .$$

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

◇ This proof and the corresponding classical proof are not difficult.

◇ This property is a merely a consequence of convexity

Strong duality

If the primal and the dual are feasible, their optimum objective values are equal (but not necessarily attained).

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

Proof.

Proceed by proving the existence of a strictly feasible point for the dual *conic* program $\Leftrightarrow z_k > \sum_{i \in I_k} x_i \log \frac{x_i}{\sum_{i \in I_k} x_i}$ **and** $x_i > 0 \ \forall i$.

But the linear constraints $Ax = b$ may force $x_i = 0$ for some i for every feasible solution !

$\Rightarrow$ detect these zero x_i components and form a restricted primal-dual pair without these variables $\Rightarrow$ strong duality holds

Strategy diagram

$$\begin{array}{ccccccc}
(P) & \equiv & (ConeP) & \overset{\text{Weak duality}}{\longleftrightarrow} & (ConeD) & \equiv & (D) \\
 & & \updownarrow^{*} & & \updownarrow & & \\
 & & (RestrP) & \overset{\text{Strong duality}}{\longleftrightarrow} & (RestrD) & & \\
 & & & & \uparrow & & \\
 & & & & \text{(Slater condition)} & &
\end{array}$$

Detection. Use the following linear problem

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

Define $\mathcal{N}$ = set of indices i such that x_i 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: Remove variables x_i for all $i \in \mathcal{N}$
$\Rightarrow$ restricted primal-dual conic pair
$\Rightarrow$ the dual has a strictly feasible solution $\Rightarrow$ zero duality gap.

We still have to prove that

- $\diamond$ Optimal objective values are equal for restricted and original dual problems (easy)
- $\diamond$ Optimal values are equal for restricted and original primal (more difficult: need to perturb optimal solution $\Rightarrow$ attainment is lost)

Generalizing our 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 $f_i : \mathbb{R} \mapsto \mathbb{R}$

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

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

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

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

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

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

$$\mathcal{K}^f = \mathcal{K}^{\circ f} \cup \left\{ (x, 0, \kappa) \mid x^T x^* \leq \kappa \text{ for all } x_i^* \in \operatorname{dom} f_i^* \right\} .$$

- 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 $x^* = 0$.

Separable convex optimization

Use $\mathcal{K}^f$ and $(\mathcal{K}^f)^*$ in a primal-dual conic pair of problems $\Rightarrow$

Primal

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

Dual

$$\inf \; c^T x + d^T z + \sum_{k \in K \,|\, z_k > 0} z_k \sum_{i \in I_k} f_i^*(-\frac{x_i}{z_k}) - \sum_{k \in K \,|\, z_k = 0} \inf_{x^*_{I_k} \in \mathrm{dom}\, f_{I_k}} x_{I_k}^T x_{I_k}^*$$

s.t. $Ax + Fz = b$ and $z \ge 0$.

- ◇ Gen. linear, quadratic, geometric, entropy, l_p-norm optimization
- ◇ Mix different types of constraints
- ◇ Hints towards a primal-dual setting
- ◇ Links (?) with extended geometric opt. (Peterson), Young opt.

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), $0 < p < 1, q < 0, 1/p + 1/q = 1$

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

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

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

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 $h \in \mathbb{R}^n$

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

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

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

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

$\Rightarrow F_X(x, y) = \sum_{i=1}^n F_i(x_i, y_i)$ 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 $\nu_Y = \mathcal{O}\left(\nu_{X'}\right)$

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

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 problems can be formulated (links with extended geometric optimization *[Peterson 80]*)

◇ Using this setting, duality properties can be obtained easily (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]*)