Duality and algorithms for separable convex optimization: towards separable eigenvalue optimization?

François Glineur
UCL / CORE
glineur@core.ucl.ac.be

Large Scale Nonlinear and Semidefinite Programming Workshop

Waterloo, May 12-15, 2004

Outline

Separable convex optimization

- Problem definition and the separable cone
- Conic formulation and duality
- Example: application to geometric optimization and insight

Interior-point methods

- General principle and self-concordant barriers
- Application to separable convex optimization

Towards separable eigenvalue optimization

- Requirements
- Conic formulation and duality

Conclusions

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 $\{1, \ldots, n\}$

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

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

The separable cone

Definition *[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 $\mathcal{K}^f \subseteq \mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{R}$
- 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)

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

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

Primal

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

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

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

$$\Updownarrow$$

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

(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*** !

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

$$\Updownarrow$$

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

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

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

Duality in separable optimization

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

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

Weak duality (← conic duality)

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

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

Strong duality (← conic duality + *perturbation*)

If the primal and the dual are feasible, their optimum objective values are equal (but not necessarily attained) ⇒ *zero* duality gap *without* Slater condition (not true in general conic case).

Application to geometric optimization

Primal problem

$$\sup_{y \in \mathbb{R}^m} b^{\mathrm{T}} y \quad \text{s.t.} \quad \sum_{i \in I_k} e^{a_i^{\mathrm{T}} y - c_i} \leq 1 \text{ for all } k \in R$$

Convex, many applications (e.g. in engineering design)

$\rightarrow$ separable problem with $f_i : x \mapsto e^{-x}$, $d_k = 1$, $g_k = 0$

Dual problem

Using the fact that $f_i^* : \mathbb{R}_+ \mapsto \mathbb{R} : x^* \mapsto x^*(1 - \log(-x^*))$ for $x^* < 0$, 0 for $x^* = 0$, $+\infty$ for $x^* > 0$

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

(optimal value for z_k variables computed analytically $\rightarrow$ removed)

$\Rightarrow$ zero duality gap property holds without Slater condition

Generalizing geometric optimization

Traditional equivalent (nonconvex) posynomial formulation: change of variables $t_i = e^{y_i} > 0$ and $C_i = e^{-c_i} > 0$ for all i

$$\sup_{y \in \mathbb{R}^m} b^{\mathrm{T}} y \quad \text{s.t.} \quad \sum_{i \in I_k} C_i \prod_{j=1}^{m} t_i^{a_{ij}} \leq 1 \text{ for all } k \in R$$

Allowing now a more general choice for d_k and g_k we obtain

$$\sup_{y \in \mathbb{R}^m} b^{\mathrm{T}} y \quad \text{s.t.} \quad \sum_{i \in I_k} e^{a_i^{\mathrm{T}} y - c_i} \leq d_k - g_k^{\mathrm{T}} y \text{ for all } k \in R$$

and with $D_k = e^{-d_k} > 0$ for all $k \in R$

$$\sup_{y \in \mathbb{R}^m} b^{\mathrm{T}} y \quad \text{s.t.} \quad \sum_{i \in I_k} C_i \prod_{j=1}^{m} t_i^{a_{ij}} \leq -\log D_k \prod_{j=1}^{m} t_i^{g_{kj}} \text{ for all } k \in R$$

More general but *same* cone and conic formulation (and complexity)!

A different conic formulation ?

1. Note that one can choose cones based on a *single* scalar function without loss of generality using a *decomposition* scheme

For example, constraint $\sum_{i\in\{1,2,3\}} f(a_i^{\mathrm{T}} y - c_i) \leq d_k - g_k^{\mathrm{T}} y$ becomes

$f(a_1^{\mathrm{T}} y - c_1) \leq y_1^+$, $f(a_2^{\mathrm{T}} y - c_2) \leq y_2^+$ and $f(a_3^{\mathrm{T}} y - c_3) \leq d_k - g_k^{\mathrm{T}} y - y_1^+ - y_2^+$

using additional variables y_1^+ and y_2^+

2. Geometric case: $e^{-x} \leq y \Leftrightarrow \Leftrightarrow -\log y \leq +x - \frac{1}{2} - \log y \leq -\frac{1}{2} + x$

$\Rightarrow$ introduce $f_g : \mathbb{R}_{++} \mapsto \mathbb{R} : v \mapsto f_g(v) = -\frac{1}{2} - \log v$ with a very nice conjugate $f_g^* : \mathbb{R}_{--} \mapsto \mathbb{R} : v^* \mapsto f_g^*(v^*) = -\frac{1}{2} - \log(-v^*) = f_g(-v^*)$

$\Rightarrow$ corresponding cone $\mathcal{K}^{f_g}$ for geometric optimization is *self-dual* !

Unclear whether ***1.*** or ***2.*** brings any theoretical/practical advantage

Interior-point methods and self-concordant barriers

Principle of an interior-point method

Approximate a constrained problem by a *family* of unconstrained problems relying on a *barrier* to replace the constraints

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

Conic case: barrier $F_{\mathcal{C}}$ for $\mathcal{C}$, keep (easy) linear equality constraints

$$\inf_{x \in \mathbb{R}^n} \frac{c^{\mathrm{T}} x}{\mu} + F_{\mathcal{C}}(x) \quad \text{s.t.} \quad Ax = b \tag{CO_μ}$$

Depends on a parameter $\mu \in \mathbb{R}_{++}$. We have

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

- x^*_μ is the (unique) solution of (CO_μ) ($\to$ central path)
- x^* is a solution of the original problem (CO)

Ingredients

- *Newton's method* to (approximately) compute x^*_μ
- An *efficient* barrier function → *self-concordant* barrier

Definition *[Nesterov & Nemirovsky]*

Function $F : \operatorname{int}\mathcal{C} \mapsto \mathbb{R}$ is a *ν-self-concordant* barrier on $\mathcal{C}$ iff

- F is convex and three times differentiable
- $F(x) \to +\infty$ when $x \to \partial\mathcal{C}$ (barrier property)

$$\nabla^3 F(x)[h,h,h] \leq 2\left(\nabla^2 F(x)[h,h]\right)^{\frac{3}{2}} \quad \forall\, x \in \operatorname{int}\mathcal{C}, \forall\, h \in \mathbb{R}^n$$

$$\nabla F(x)^{\mathrm{T}}(\nabla^2 F(x))^{-1}\nabla F(x) \leq \nu \quad \forall\, x \in \operatorname{int}\mathcal{C}$$

Main result

A conic convex problem endowed with a *computable* self-concordant barrier for $\mathcal{C}$ is solvable *in polynomial time* by a (primal) IPM.

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

Application to separable optimization

Our result

Given a self-concordant barrier F_i with parameter ν_i for each set epi $f_i = \{(x_i, u_i) \in \mathbb{R}^2 \mid f_i(x_i) \le u_i\}$, $1 \le i \le n$ (usually *easy* to find)

One can find a self-concordant barrier F for $\mathcal{K}^f$ with a parameter ν proportional to $\sum_{i=1}^n \nu_i$

$\Rightarrow$ 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)

Example: geometric optimization

We have epi $f_i = \{(x_i, u_i) \in \mathbb{R}^2 \mid e^{-x_i} \le u_i\}$ ($\Leftrightarrow x_i + \log u_i \ge 0$) and

$$F_i = -\log(x_i + \log u_i) - \log u_i$$

is a 2-self-concordant barrier for this set $\Rightarrow$ one can solve geometric optimization up to ϵ accuracy with $\mathcal{O}\left(\sqrt{n} \log \frac{1}{\epsilon}\right)$ iterations

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

Outline of the proof (cont.)

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 $\mathcal{K}^f$

   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

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

$\Rightarrow$ we have a s.c.b. for $\mathcal{K}^f$ with parameter $\mathcal{O}\left(\sum_{i=1}^{n} \nu_i\right)$

Towards a primal-dual interior-point method

$$\mathcal{K}^f = \mathrm{cl}\Big\{(x,\theta,\kappa) \in \mathbb{R}^{n+2} \mid h_f(x,\theta) = \theta \sum_{i=1}^n f_i(\frac{x_i}{\theta}) \le \kappa,\ \theta > 0\Big\}$$

$$(\mathcal{K}^f)^* = \mathrm{cl}\Big\{(x^*,\theta^*,\kappa^*) \in \mathbb{R}^{n+2} \mid h_f^*(x^*,\kappa^*) = \kappa^* \sum_{i=1}^n f_i^*(-\frac{x_i^*}{\kappa^*}) \le \theta^*\Big\}$$

If $v = (x,\theta,\kappa) \in \mathcal{K}^f$ and $(x^*,\theta^*,\kappa^*) \in (\mathcal{K}^f)^*$, we have $v^{\mathrm{T}}v^* = 0$ iff

$$0 = \kappa^*(\kappa - h_f(x,\theta)) = \theta(\theta^* - h_f^*(x^*,\kappa^*)) = f_i(\frac{x_i}{\theta}) + f_i^*(\frac{-x_i^*}{\kappa^*}) + \frac{x_i x_i^*}{\theta\kappa^*}\ \forall\, i$$

"*complementarity*" conditions → primal-dual optimality conditions

Perturbation → *central path* ? An idea: use the fact that

$$f_i(\frac{x_i}{\theta}) + f_i^*(\frac{-x_i^*}{\kappa^*}) + \frac{x_i x_i^*}{\theta\kappa^*} \ge 0 \ \rightarrow\ f_i(\frac{x_i}{\theta}) + f_i^*(\frac{-x_i^*}{\kappa^*}) + \frac{x_i x_i^*}{\theta\kappa^*} = \mu > 0$$

Separable eigenvalue optimization

Requirements

$X \in \mathbb{S}^n$ and $\lambda : \mathbb{S}^n \mapsto \mathbb{R}^n : X \mapsto \lambda(X)$ is the *ordered* eigenvalue map

Let $f : \mathbb{R}^n \mapsto \mathbb{R}$; for $h = f \circ \lambda$ to be convex on its domain, we need

- f *convex* (on $\mathbb{R}^n$)
- f *symmetric* (w.r.t. permutations of its arguments)

Here in addition we require f to be *separable*, i.e.

$$f : \mathbb{R}^n \mapsto \mathbb{R} : x \mapsto \sum_{i=1}^{n} f_i(x)$$

with scalar functions $f_i : \mathbb{R} \mapsto \mathbb{R}$

Separable eigenvalue functions

Combining both requirements we consider functions f of the form

$$f_\phi = \phi \times \phi \times \cdots \times \phi : \mathbb{R}^n \mapsto \mathbb{R} : x \mapsto \sum_{i=1}^{n} \phi(x_i)$$

using the *same* scalar *convex* function ϕ which leads to

$$h_\phi = f_\phi \circ \lambda : \mathbb{S}^n \mapsto \mathbb{R} : X \mapsto \sum_{i=1}^{n} \phi(\lambda_i(X))$$

(hence we have ϕ convex $\Rightarrow h_\phi$ convex)

Examples

- Taking $\phi(x) = -\log(x)$ leads to $-\log\det X$ ($\rightarrow$ MAXDET)
- Taking $\phi(x) = x^2$ leads to the Frobenius norm (not that useful)
- Taking $\phi(x) = e^x$ leads to trace e^X (trace of matrix exponential)

Duality results

$$\mathcal{K}^{h_\phi} = \mathrm{cl}\,\mathcal{K}^{\circ h_\phi} = \mathrm{cl}\Big\{(X,\theta,\kappa) \in \mathbb{S}^n \times \mathbb{R}_{++} \times \mathbb{R} \mid \theta \sum_{i=1}^n \phi(\frac{\lambda_i(X)}{\theta}) \leq \kappa\Big\}$$

admits as a *dual* (using the fact that $(h_\phi)^* = h_{\phi^*}$)

$$(\mathcal{K}^{h_\phi})^* = \mathrm{cl}\Big\{(X^*,\theta^*,\kappa^*) \in \mathbb{S}^n \times \mathbb{R}_{++} \times \mathbb{R} \mid \kappa^* \sum_{i=1}^n \phi^*(-\frac{\lambda_i(X^*)}{\kappa^*}) \leq \theta^*\Big\}$$

For example, one can use this to show that

$$\inf A_0 \bullet X + \sum_{i=1}^n \phi(\lambda_i(X)) \text{ s.t. } A_i \bullet X = b_i \ (1 \leq i \leq m)$$

admits the following dual problem

$$\sup b^{\mathrm{T}} y - \sum_{i=1}^n \phi^*(\lambda_i(X^*)) \text{ s.t. } X^* = \sum_{i=1}^m A_i y_i - A_0$$

(adding *constraints* based on different functions h_{ϕ_k} is also possible)

Concluding remarks

- A very wide class of *separable* convex problems can be formulated (including linear, quadratic, entropy, l_p-norm and geometric optimization) in the *conic* setting
- Self-concordant barriers on 2-dimensional epigraphs of scalar functions $f_i \to$ *self-concordant* barrier for the $\mathcal{K}^f$ cone
  $\to$ primal algorithm with *polynomial* complexity
- *However* this does not use primal-dual symmetry
- There are other separable convex eigenvalue functions (e.g. λ_{max}) $\Rightarrow$ *more general* setting?
- Generalization of self-concordant construction for $\mathcal{K}^f$ to $\mathcal{K}^{h\phi}$?