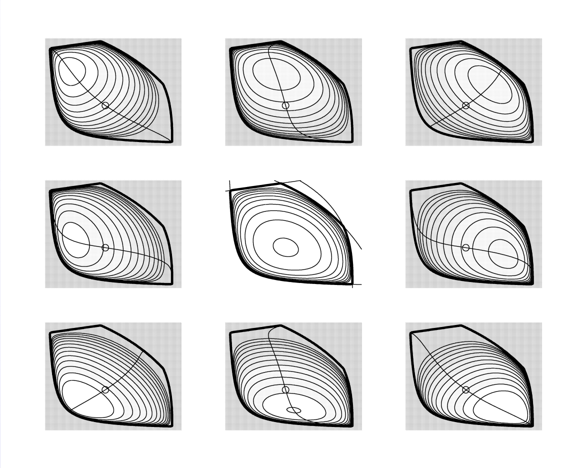
Recent Advances in Structured Convex Optimization

François Glineur
Chargé de Recherches F.N.R.S. - Faculté Polytechnique de Mons
Francois.Glineur@fpms.ac.be



March 8th, 2002

Motivation

Operations research

Model real-life situations to help take the *best* decisions

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

General formulation

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in \mathcal{D} \subseteq \mathbb{R}^n$$

Choice of parameters, optimal design, scheduling ...

Generality

Very general formulation (*too* general ?)
(continuous/discrete, linear or not, smooth or not)
$\Rightarrow$ ability to model a *very large* number of problems

Applicability

To solve problems *in practice*, algorithms are needed
$\Rightarrow$ what is a good algorithm ?

- ◇ Solves (approximately) the problem
- ◇ Until mid 20$^{\text{th}}$ century: in finite time
- ◇ Now (computers): in bounded time
  (depending on size) $\rightarrow$ computational complexity

Two approaches

Bad news: solving all problems *efficiently* is impossible!

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

$\Rightarrow 10^{20}$ operations to be solved with 1% accuracy !

Reaction: two distinct orientations

- General *nonlinear optimization*
  Applicable to all problems but no efficiency guarantee
- *Linear*, *quadratic*, *semidefinite*, ... optimization
  Restrict set of problems to get efficiency guarantee

Tradeoff generality $\leftrightarrow$ efficiency (algorithmic complexity)

Restrict to which class of problems ?

Linear optimization : $+$ specialized, very fast algorithms
$-$ too restricted in practice

$\longrightarrow$ we focus on **Convex optimization**

- Convex objective and convex feasible set
- Many problems are convex or can be convexified
- Efficient *algorithms* (see later)
- Powerful *duality* theory (but weaker than LO)
- But ... establishing convexity *a priori* is difficult

  $\longrightarrow$ work with specific classes of convex constraints: Structured convex optimization (convexity *by design*)

Overview of this talk

Convex Optimization

◇ Conic optimization and duality

An Application to Classification

◇ Pattern separation using ellipsoids

Solving Structured Convex Problems

◇ Interior-point methods

◇ Self-concordant barriers

Separable Convex Optimization

◇ Conic formulation

◇ Duality and self-concordant barriers

Convex optimization

Let $f_0 : \mathbb{R}^n \mapsto \mathbb{R}$ be a convex function, $C \subseteq \mathbb{R}^n$ be a convex set : optimize a vector $x \in \mathbb{R}^n$

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

Properties

- All local optima are global, optimal set is convex
- Lagrange duality $\longrightarrow$ strongly related dual problem
- Objective can be taken linear w.l.o.g. ($f_0(x) = c^{\mathrm{T}} x$)

But we choose a special class of feasible sets ...

Conic formulation

Primal problem

Let $\mathcal{C} \subseteq \mathbb{R}^n$ be a convex cone

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

Formulation is equivalent to convex optimization.

Dual problem

Let $\mathcal{C} \subseteq \mathbb{R}^n$ be a *solid, pointed, closed* convex cone.
The dual cone $\mathcal{C}^* = \left\{ x^* \in \mathbb{R}^n \mid x^{\mathrm{T}} x^* \geq 0 \text{ for all } x \in \mathcal{C} \right\}$ is also convex, solid, pointed and closed $\longrightarrow$ dual problem:

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

Primal-dual pair

Symmetrical pair of primal-dual problems

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

Optimum values p^* and d^* not necessarily attained !

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

Weak duality

For every feasible x and y $\quad b^{\mathrm{T}}y \leq c^{\mathrm{T}}x \quad \rightarrow$ bounds
with equality iff $x^{\mathrm{T}}s = 0$ (*orthogonality* condition)

$\Delta = p^* - d^*$ is the *duality gap* $\Rightarrow$ always nonnegative

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

Strong duality (with Slater condition $\rightarrow \neq$ LO)

a. Strictly feasible dual point $\Rightarrow p^* = d^*$ (no gap)

b. If in addition primal is bounded
$\Rightarrow$ primal optimum is attained $\Leftrightarrow p^* = \min c^{\mathrm{T}}x$

(dualized result obviously holds)

Overview of this talk

Convex Optimization

◇ Conic optimization and duality

An Application to Classification

◇ Pattern separation using ellipsoids

Solving Structured Convex Problems

◇ Interior-point methods

◇ Self-concordant barriers

Separable Convex Optimization

◇ Conic formulation

◇ Duality and self-concordant barriers

Pattern separation

Problem definition

Let us consider *objects* defined by *patterns*

Object $\equiv$ Pattern $\equiv$ Vector of n attributes

Assume it is possible to group these objects into c classes

Objective

Find a partition of $\mathbb{R}^n$ into c disjoint components such that each component corresponds to one class

Utility: classification

$\Rightarrow$ identify to which class an unknown pattern belongs

Classification

Consider

- ◇ Some well-known objects grouped into classes
- ◇ Some unknown objects

Two-step procedure

a. Separate the patterns of well-known objects
   $\equiv$ *learning* phase

b. Use that partition to classify the unknown objects
   $\equiv$ *generalization* phase

Examples

Medical diagnosis, species identification, credit approval

Our technique

Consider two classes (without loss of generality)

Main idea

Use ellipsoids to separate the patterns *[Glineur 98]*

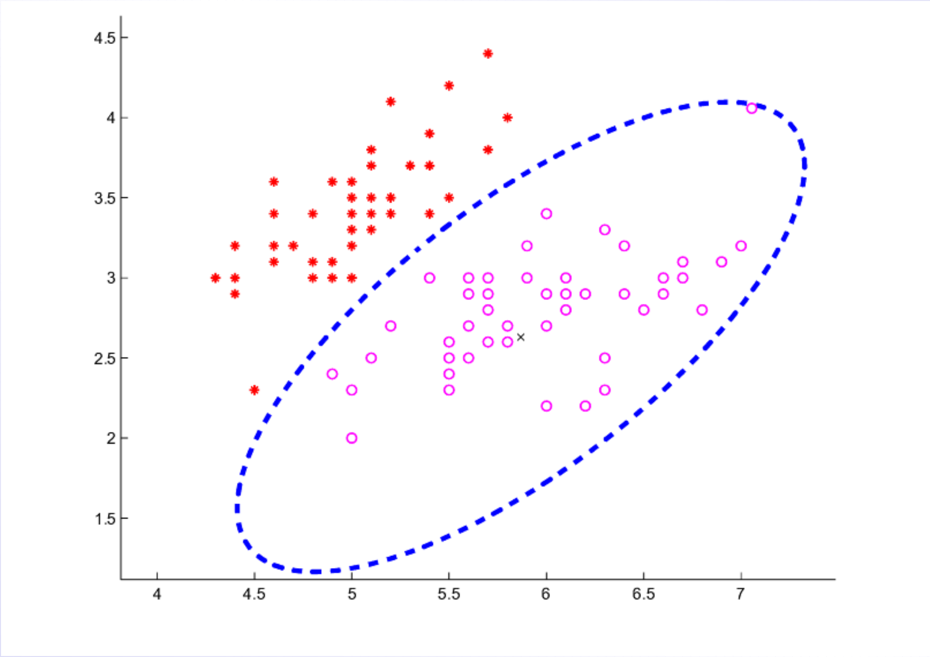
Ellipsoid

An ellipsoid $\mathcal{E} \subseteq \mathbb{R}^n \equiv$ a center $c \in \mathbb{R}^n$ and a positive semidefinite matrix $E \in \mathbb{S}^n_+$

$$\mathcal{E} = \{x \in \mathbb{R}^n \mid (x - c)^{\mathrm{T}} E(x - c) \leq 1\}$$

But which ellipsoid performs the best separation ?

Using ellipsoids to perform separation



Separation ratio

We want the best possible separation
$\Rightarrow$ define and maximize the *separation ratio*

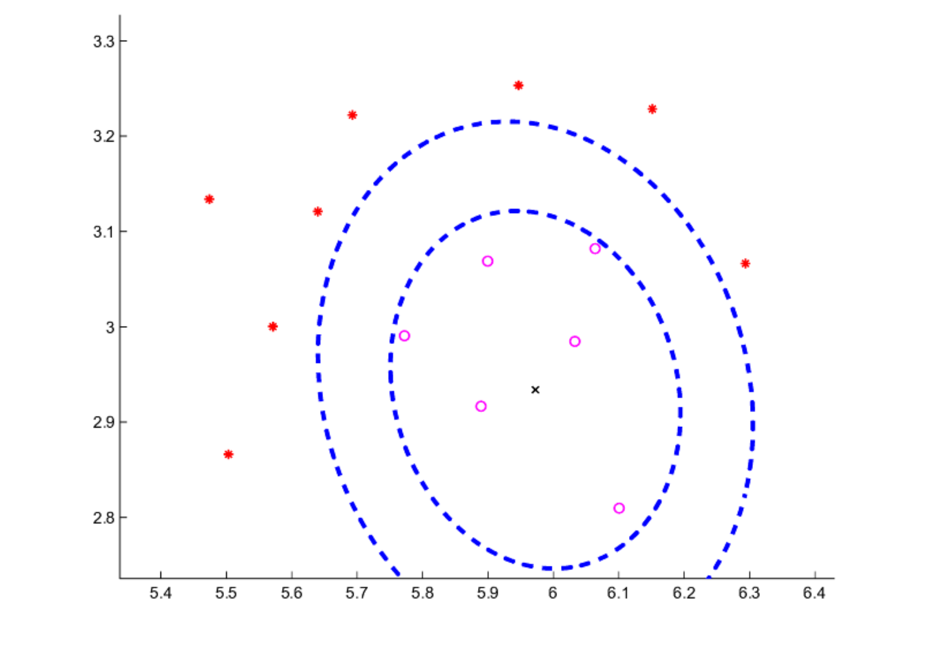
Definition

Pair of homothetic ellipsoids sharing the same center
Separation ratio $\rho \equiv$ ratio of sizes

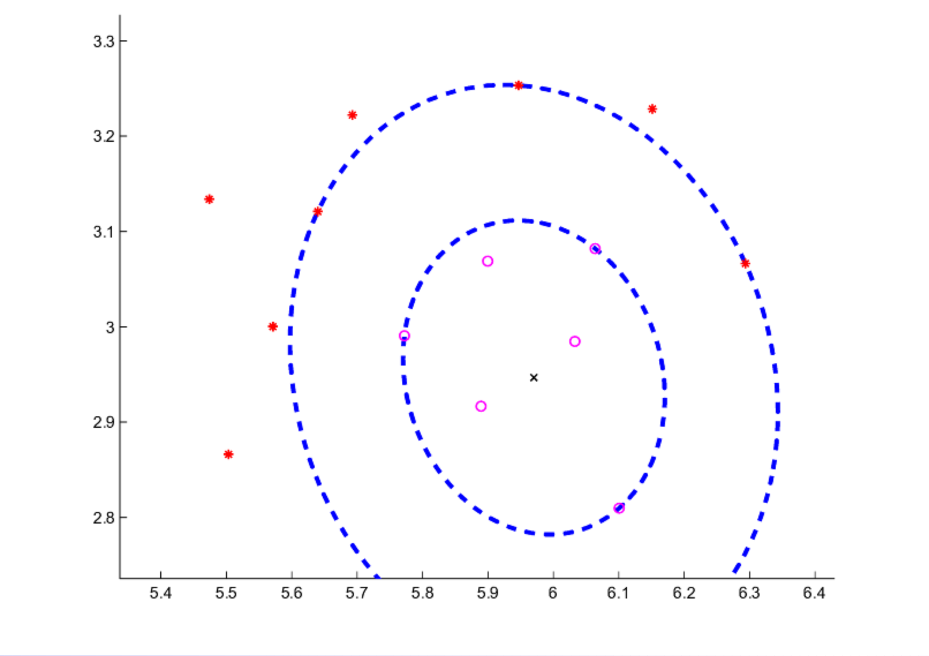
Mathematical formulation

$$\max \rho \quad \text{s.t.} \quad \begin{cases} (a_i - c)^{\mathrm{T}} E (a_i - c) \leq 1 \ \forall i \\ (b_j - c)^{\mathrm{T}} E (b_j - c) \geq \rho^2 \ \forall j \\ E \in \mathbb{S}^n_+ \end{cases}$$

Separation ratio equal to $\rho = 1.5$



Optimal separation ratio ($\rho = 1.8$)



Analysis

This problem is not convex but can be convexified (*homogenizing* the description of the ellipsoid)

$\Rightarrow$ we obtain a semidefinite optimization problem
$\equiv$ conic optimization with $\mathcal{C} = \mathbb{S}^n_+$

A general semidefinite optimization problem

$$p^* = \inf_{X \in \mathbb{S}^n} C \bullet X \text{ s.t. } \mathcal{A}X = b \text{ and } X \in \mathbb{S}^n_+$$

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

$\Rightarrow$ efficiently solvable in practice with interior-point method

Numerical experiments

◇ Implementation using MATLAB

◇ Test on sets from the *Repository of Machine Learning Databases and Domain Theories* maintained by the *University of California* at Irvine (widely used)

◇ Cross-validation
divide data set into *learning* and *validation* set

a. Compute best separating ellipsoid on learning set

b. Evaluate accuracy of separating ellipsoid on *validation set* (test generalization capability)

Data sets

Three representative sets

a. **Wisconsin Breast Cancer**.

Predict the benign or malignant nature of a breast tumour (683 patterns, 9 characteristics)

b. **Boston Housing**.

Predict whether a housing value is above or below the median (596 patterns, 12 characteristics)

c. **Pima Indians Diabetes**.

Predict whether a patient is showing signs of diabetes (768 patterns, 8 characteristics)

Results: error rates

	Best ellipsoid		*LAD*	*Best other*
Training %	20 %	50 %	50 %	Variable (% tr.)
Cancer	5.1 %	4.2 %	**3.1** %	3.8 % (80 %)
Housing	15.8 %	**12.4** %	16.0 %	16.8 % (80 %)
Diabetes	28.5 %	28.9 %	28.1 %	**24.1** % (75 %)

- ◇ Competitive error rates ⇒ reliable generalization
- ◇ Best results on the Housing problem (even 20 %)
- ◇ 50 % not always better than 20 % (⇒ overlearning)
- ◇ Results with small learning set already acceptable

Overview of this talk

Convex Optimization

◇ Conic optimization and duality

An Application to Classification

◇ Pattern separation using ellipsoids

Solving Structured Convex Problems

◇ Interior-point methods

◇ Self-concordant barriers

Separable Convex Optimization

◇ Conic formulation

◇ Duality and self-concordant barriers

Interior-point methods

Principle

Approximate a constrained problem by

a *family* of unconstrained problems

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

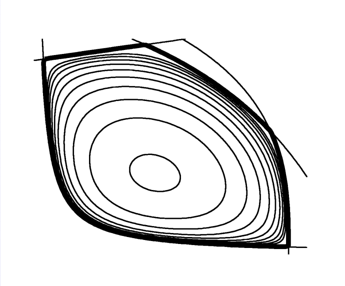
- F is smooth
- F is strictly convex on $\operatorname{int} C$
- $F(x) \rightarrow +\infty$ when $x \rightarrow \partial C$

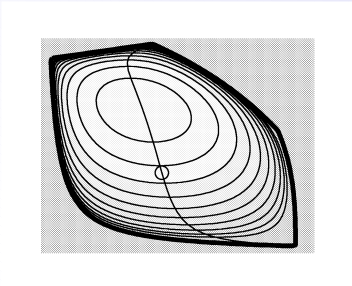
$$\rightarrow \quad C = \operatorname{cl} \operatorname{dom} F = \operatorname{cl} \{x \in \mathbb{R}^n \mid F(x) < +\infty\}$$

Central path

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

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





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

where

- x^*_μ is the (unique) solution of (P_μ) ($\to$ central path)
- 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 $f_i(x) \leq 0$, one can introduce the *logarithmic* barrier function

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

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 (κ, ν)-self-concordant on C iff

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

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

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

Complexity result

Summary

Self-concordant barrier $\Rightarrow$ 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 below τ)

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

Complexity result

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

iterations lead a solution with ϵ accuracy on the objective

Optimal choice of parameters *[Glineur 00]*

Two constants define a short-step algorithm

- ◇ Maximum distance τ to the central path
- ◇ Factor of decrease of barrier parameter μ

Optimizing these parameters leads to

$$\left\lceil (8.68\kappa\sqrt{\nu} - 0.5) \log \frac{1.29\mu_0\kappa\sqrt{\nu}}{\epsilon} \right\rceil$$

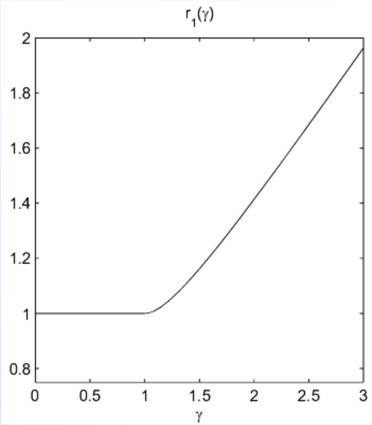
iterations guarantee a solution with ϵ accuracy

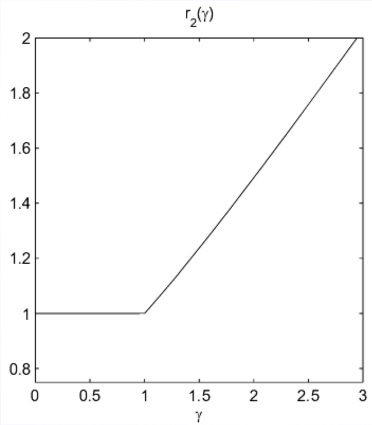
A useful lemma

Proving self-concordancy not always an easy task
$\Rightarrow$ improved version of lemma by *[Den Hertog et al.]*

Auxiliary functions

Define two auxiliary functions r_1 and $r_2 : \mathbb{R}_+ \mapsto \mathbb{R}_+$





Lemma's statement *[Glineur 00]*

Let $F : \mathbb{R}^n \mapsto \mathbb{R}$ be a convex function on $C \subseteq \mathbb{R}^n_{++}$.
If there is a constant $\gamma \in \mathbb{R}_+$ such that

$$\nabla^3 F(x)[h,h,h] \leq 3\gamma \nabla^2 F(x)[h,h] \sqrt{\sum_{i=1}^n \frac{h_i^2}{x_i^2}}$$

then the following barrier functions

$$F_1 : \mathbb{R}^n \mapsto \mathbb{R} : x \mapsto F(x) - \sum_{i=1}^n \log x_i$$

$$F_2 : \mathbb{R}^n \times \mathbb{R} \mapsto \mathbb{R} : (x,u) \mapsto -\log(u - F(x)) - \sum_{i=1}^n \log x_i$$

satisfy the first self-concordancy condition

(respectively) with $\kappa_1 = r_1(\gamma)$ and $\kappa_2 = r_2(\gamma)$
on the sets $\operatorname{dom} F$ and $\operatorname{epi} F = \{(x,u) \mid F(x) \leq u\}$.

Application: extended entropy optimization

$$\inf c^{\mathrm{T}} x + \sum_{i=1}^{n} g_i(x_i) \quad \text{s.t.} \quad Ax = b \text{ and } x \geq 0$$

with scalar functions $g_i : \mathbb{R} \mapsto \mathbb{R}$ such that

$$|g_i'''(x)| \leq \kappa_i \frac{g_i''(x)}{x} \; \forall x \geq 0$$

(which implies convexity)

Special case: classical entropy optimization
when $g_i(x) = x \log x \quad \Rightarrow \quad \kappa_i = 1$

Application of the Lemma

Use Lemma with $F(x_i) = g_i(x_i)$ to prove that

$$-\log\Big(t_i - g_i(x_i)\Big) - \log(x_i) \quad \text{is} \quad \Big(r_2(\frac{\kappa_i}{3}), 2\Big)\text{-SC}$$

Total complexity of EEO is *[Glineur 00]*

$$O\Big(\sqrt{2\sum_{i=1}^{n} r_2(\frac{\kappa_i}{3})^2}\Big) \text{ iterations}$$

or

$$O(\sqrt{2n}) \text{ iterations for entropy optimization}$$

Possible application: *polynomial* g_i's

Specializing interior-point methods

Consider class of cones that are *[Nesterov & Todd 97]*

self-dual and *homogeneous*

- ◇ The nonnegative orthant $\mathbb{R}^n_+$ ($\Rightarrow$ LO)
- ◇ The second-order cone $\mathbb{L}^n_+$ ($\Rightarrow$ QO and SOCO)

$$\mathbb{L}^n_+ = \{(r, x) \in \mathbb{R} \times \mathbb{R}^n \mid r \geq \|x\|\}$$

- ◇ The cone of positive semidefinite matrices $\mathbb{S}^n_+$ ($\Rightarrow$ SDO)

Applications

- ◇ $\mathbb{L}^n_+ \quad \longrightarrow$ truss topology, limit analysis, etc.
- ◇ $\mathbb{S}^n_+ \quad \longrightarrow$ control, combinatorial optimization, etc.

Modelling

- ◇ Linear, quadratic and semidefinite optimization
- ◇ Nonlinear (convex) objectives such as $\frac{x^2}{y}$, $\|x\|$, $\lambda_{max}(E)$
- ◇ Constraints such as $xy \geq 1$, $E \in \mathbb{S}^n_+$, $\lambda_{max}(E) \leq 1$
- ◇ Handle free variables in a very natural way (using $\mathbb{L}^n_+$)
- ◇ But only *convex* programs

Cones $\mathbb{R}^n_+$, $\mathbb{L}^n_+$ and $\mathbb{S}^n_+$ allow the design of very efficient primal-dual algorithms (that also work well in practice)

$\longrightarrow$ What about other classes of convex problems ?
(in particular with non-symmetric duality)

Overview of this talk

Convex Optimization

◇ Conic optimization and duality

An Application to Classification

◇ Pattern separation using ellipsoids

Solving Structured Convex Problems

◇ Interior-point methods

◇ Self-concordant barriers

Separable Convex Optimization

◇ Conic formulation

◇ Duality and self-concordant barriers

Separable convex optimization

Definition

Set of n scalar closed proper convex functions $f_i : \mathbb{R} \mapsto \mathbb{R}$
$R = \{1, \ldots, r\}$ 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 objective without loss of generality
- ◇ Linear, quadratic, geometric, entropy, l_p-norm opt.
- ◇ Mix different types of constraints

Goal: study duality and algorithms

Examples - primal separable problems

◇ Geometric optimization (after convexification)

$$\sup b^{\mathrm{T}} y \quad \text{s.t.} \quad \sum_{i \in I_k} e^{a_i^{\mathrm{T}} y - c_i} \leq 1 \quad \forall k \in R$$

using $f_i : x \mapsto e^{-x}$.

◇ l_p-norm optimization

$$\sup\ b^T y \ \text{ s.t. } \sum_{i \in I_k} \frac{1}{p_i} \left| c_i - a_i^T y \right|^{p_i} \leq d_k - g_k^T y \quad \forall k \in R$$

using $f_i : x \mapsto \frac{1}{p_i} |x|^{p_i}$

Strategy: use conic formulation

The separable cone *[Glineur 00]*

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

- $\mathcal{K}^f$ is a closed convex cone and $\mathcal{K}^f \subseteq \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}_+$
- Dual cone $(\mathcal{K}^f)^*$ is computable and very symmetric

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

using the conjugate functions

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

(also closed, proper and convex)

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 $z_k^* = d_k - g_k^T y$

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

$$\Updownarrow$$

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

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

$\Rightarrow$ dual 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** !

$$\inf \begin{pmatrix} c \\ d \\ e \end{pmatrix}^T \begin{pmatrix} x \\ z \\ v \end{pmatrix}$$

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

$$\Updownarrow$$

$$\inf c^T x + d^T z + e^T v$$

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

$$\Updownarrow$$

Dual

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

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

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

$\Rightarrow$ find e.g. the dual quadratic, geometric and l_p-norm optimization problems *in a completely seamless way*

Examples - dual separable problems

◇ Geometric optimization, using $f_i^* : x \mapsto x - x \log(-x)$

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

◇ l_p-norm optimization, $f_i^*(x) = \frac{1}{q_i} |x|^{q_i} \quad (\frac{1}{p_i} + \frac{1}{q_i} = 1)$

$$\inf\ \psi(x, z) \ = \ c^{\mathrm{T}} x + d^{\mathrm{T}} z + \sum_{k=1}^{r} z_k \sum_{i \in I_k} \frac{1}{q_i} \left| \frac{x_i}{z_k} \right|^{q_i}$$

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

$\longrightarrow$ standard dual problems found in the literature

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.

Strong duality

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

Strong duality (cont.)

This theorem guarantees a zero duality gap

without any Slater condition

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

Proof

$\exists$ strictly feasible point for the dual *conic* program

$$\Leftrightarrow v_k > z_k \sum_{i \in I_k} f_i^*(-\frac{x_i}{z_k}) \text{ and } z_k > 0$$

$\Rightarrow$ easily prove *strong duality properties* of e.g. quadratic, geometric and l_p-norm optimization problems

Self-concordant barriers for 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

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

$\Rightarrow$ separable convex problems can be solved in

$$\mathcal{O}\left(\sqrt{\sum_{i=1}^{n} \nu_i} \log \frac{1}{\epsilon}\right) \text{ iterations}$$

$\Rightarrow$ polynomial-time if F_i's are polynomial-time computable (unified proof of polynomiality) *[Glineur 00]*

Summary and conclusions

Structured Convex Optimization

- ◇ Models a very large class of problems
- ◇ Powerful duality theory
- ◇ Efficient interior-point methods
- ◇ Symmetric conic formulation

Interior-point methods

- ◇ Self-concordancy theory
- ◇ Optimal complexity of short-step method
- ◇ Improvement of useful Lemma

Application to Classification

- ◇ Pattern separation using SDO and ellipsoids

Separable Convex Optimization

- ◇ Generalizes quadratic optimization, geometric optimization, l_p-norm optimization, etc.
- ◇ Using a conic formulation ≡ unified framework to
  a. Formulate the dual problem,
  b. Prove weak/strong duality,
  c. Find self-concordant barriers
  $\longrightarrow$ polynomial algorithms.

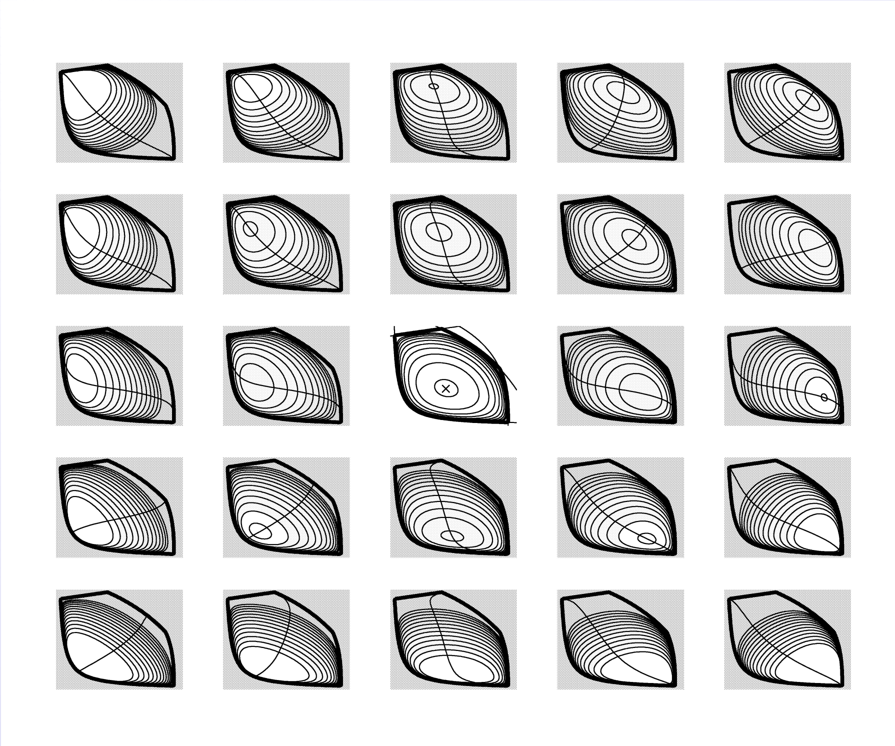
Research directions

Modelling

- Investigate problems that can be modelled as convex separable optimization problems
- Networks (modelled as graphs): objective and constraints naturally separable (*scalar* quantities defined at *arcs* and *nodes*)

Solving Separable Convex Problems

- Develop symmetric primal-dual algorithms
- Implementation $\longrightarrow$ solve large-scale problems



Thank you for your attention