Nonlinear optimization

Convex optimization - Lecture II

François Glineur

UCL/FSA/INMA & CORE
glineur@core.ucl.ac.be

Graduate School in Systems and Control

March 16 2004

Administrivia

Schedule

- Part I - *Convex optimization*: March 11 & 16
- Part II - *(Traditional) nonlinear optimization*: March 25 and April 1
- Part III - *Some applications in systems and control* (by Michael Overton): April 22 & 29

Questions and comments ...

... are more than welcome, at any time !

Plan of Lecture I

Convex optimization : duality and cones

- ◇ Introduction to nonlinear optimization
- ◇ Motivation : why convex optimization ?
- ◇ Duality
- ◇ Convex optimization
- ◇ Conic optimization

Evaluation

Modelling project (groups of two, using MATLAB)

Plan of Lecture II

Convex optimization : models and algorithms

- *Duality*: from linear to conic optimization
- *Conic* modelling: three very expressive cones
- *Algorithms*: the interior-point revolution
- More *applications*: positive polynomials and max-cut

Corrected slides available on the web :
`http://www.core.ucl.ac.be/~glineur/`

Duality properties

Since we generalized

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y \leq c$$

to

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y \preceq_K c$$

it is tempting to generalize

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \geq 0$$

to

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \succeq_K 0$$

But this is not the right primal-dual pair !

Dualizing a conic problem

Remembering the dualizing procedure for linear optimization, a crucial point lied in the ability to derive consequences by taking nonnegative linear combinations of inequalities

Consider now the following statement

$$\begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix} \succeq_{\mathbb{L}^2} \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

which is true since $(-1)^2 + (-1)^2 \leq 2^2$

Multiplying the first line by $0, 1$ and the next two by 1, we get $0.1 \times 2 - 1 \times 1 - 1 \times 1 \geq 0$ or $-1.8 \geq 0$:

$\Rightarrow$ this is a contradiction!

We obtained a contraction although the original system of inequalities was consistent $\Rightarrow$ something is wrong! Some nonnegative linear combinations do not work!

Rescuing duality

Starting with

$$x \in K \subseteq \mathbb{R}^n \Leftrightarrow x \succeq_K 0$$

we identify all vectors (of multipliers) $z \in \mathbb{R}^n$ such that the consequence $z^{\mathrm{T}} x \geq 0$ holds as soon as $x \succeq_K 0$

Hence we define the set

$$K^* = \{z \in \mathbb{R}^n \text{ such that } x^{\mathrm{T}} z \geq 0 \ \forall x \in K\}$$

The dual cone

$$K^* = \{z \in \mathbb{R}^n \text{ such that } x^{\mathrm{T}} z \geq 0 \ \forall x \in K\}$$

- For any $x \in K$ and $z \in K^*$, we have $z^{\mathrm{T}} x \geq 0$
- K^* is a convex cone, called the dual cone of K
- K^* is always closed, and if K is closed, $(K^*)^* = K$
- K is pointed (resp. solid) $\Rightarrow$ K^* is solid (resp. pointed)
- Cartesian products: $(K_1 \times K_2)^* = K_1^* \times K_2^*$
- $(\mathbb{R}^n_+)^* = \mathbb{R}^n_+$, $(\mathbb{L}^n)^* = \mathbb{L}^n$, $(\mathbb{S}^n_+)^* = \mathbb{S}^n_+$:
  these cones are self-dual
- But there exists (many) cones that are not self-dual

Bounds and optimality

Let $\bar{y}$ a feasible solution (satisfying $A^{\mathrm{T}} y \preceq_K c$)
$\longrightarrow b^{\mathrm{T}} \bar{y}$ is a lower bound on the optimal value f^*

But how to

- $\diamond$ obtain upper bounds on the optimal value ?
- $\diamond$ prove that a feasible solution y^* is optimal ?

Those questions are linked since

$$\text{proving that } y^* \text{ is optimal}$$
$$\Updownarrow$$
$$\text{proving that } b^{\mathrm{T}} y^* \text{ is an upper bound on the optimal value } f^*$$

Generating upper bounds

Consider

$$\max 2y_1+3y_2+2y_3 \text{ such that } \begin{pmatrix} y_1+y_2 \\ y_2+y_3 \\ y_3 \end{pmatrix} \preceq_{\mathbb{L}^2} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \quad \begin{matrix} (a) \\ (b) \\ (c) \end{matrix}$$

Solution $y = (-2, 1, 2)$ is feasible with objective value 3
$\rightarrow$ lower bound $f^* \geq 3$ (since $(2, -1, 1) \in \mathbb{L}^2$)

Let us combine constraints: $2(a) + (b) + (c)$
(we have the right to do so since $(2, 1, 1) \in (\mathbb{L}^2)^* = \mathbb{L}^2$)

$$2y_1+2y_2+y_2+y_3+y_3 \leq 2+2+3 \Leftrightarrow 2y_1+3y_2+2y_3 \leq 7$$

$\rightarrow$ upper bound on the optimal value $f^* \leq 7$

The best upper bound

Let us find the best upper bound using this procedure

$$\max \sum_{i=1}^{m} b_i y_i \text{ such that } \Big(\sum_{i=1}^{m} a_{ij} y_i\Big)_{1\leq j\leq n} \preceq_K \big(c_j\big)_{1\leq j\leq n}$$

Introducing again n (multiplying) variables x_i
we get

$$\sum_{j=1}^{n} x_j \sum_{i=1}^{m} a_{ij} y_i \leq \sum_{j=1}^{n} x_j c_j \Leftrightarrow \sum_{i=1}^{m} y_i (\sum_{j=1}^{n} a_{ij} x_j) \leq \sum_{j=1}^{n} c_j x_j$$

under the assumption that $x \in K^*$

The best upper bound (continued)

This provides an upper bound on the objective equal to $\sum_{j=1}^n c_j x_j$, assuming that x satisfies

$$\sum_{j=1}^n a_{ij} x_j = b_i \; \forall 1 \leq i \leq m$$

Minimizing now this upper bound

$$\min \sum_{j=1}^n c_j x_j \text{ s.t. } \sum_{j=1}^n a_{ij} x_j = b_i \; \forall 1 \leq i \leq m \text{ and } x \in K^*$$

or

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \succeq_{K^*} 0$$

We find another conic optimization problem which is dual to our first problem!

Duality for conic optimization

We have completely mimicked the dualizing procedure used for linear optimization

The problem of finding the best upper bound

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \geq 0$$

becomes thus

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \succeq_{K^*} 0$$

The correct primal-dual pair is thus

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y \preceq_K c$$

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \succeq_{K^*} 0$$

Primal-dual pair

Again, for historical reasons, the min problem is called the primal. Since our cones are closed, $(K^*)^* = K^*$, which means we can write the primal conic problem

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \succeq_K 0$$

and the dual conic problem

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y \preceq_{K^*} c$$

- Very symmetrical formulation
- Computing the dual essentially amounts to finding K^*
- All nonlinearities are confined to the cones K and K^*

Duality properties

- ◇ Weak duality: any feasible solution for the primal (resp. dual) provides an upper (resp. lower) bound for the dual (resp. primal)

  (immediate consequence of our dualizing procedure)

- ◇ Inequality $b^T y \leq c^T x$ holds for any x, y such that $Ax = b$, $x \succeq_K 0$ and $A^T y \preceq_{K^*} c$ (corollary)

- ◇ If the primal (resp. dual) is unbounded, the dual (resp. primal) must be infeasible

  (but the converse is not true!)

Completely similar to the situation for linear optimization

Duality properties (continued)

What about strong duality ?

If y^* is an optimal solution for the dual, does there exist an optimal solution x^* for the primal such that $c^T x^* = b^T y^*$ (in other words: $p^* = d^*$) ?

Consider $K = \mathbb{L}^2$ with

$$A = \begin{pmatrix} -1 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix}, \ b = \begin{pmatrix} 0 & -1 \end{pmatrix}^T \text{ and } c = \begin{pmatrix} 0 & 0 & 0 \end{pmatrix}^T$$

We can easily check that

- the primal is infeasible
- the dual is bounded and solvable

$\Rightarrow$ strong duality does not hold for conic optimization ...

Other troublesome situations

Let $\lambda \in \mathbb{R}_+$: consider

$$\min \lambda x_3 - 2x_4 \text{ s.t. } \begin{pmatrix} x_1 & x_4 & x_5 \\ x_4 & x_2 & x_6 \\ x_5 & x_6 & x_3 \end{pmatrix} \succeq_{\mathbb{S}^3_+} 0, \begin{pmatrix} x_3 + x_4 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

In this case, $p^* = \lambda$ but $d^* = 2$: duality gap!

$$\min x_1 \text{ such that } x_3 = 1 \text{ and } \begin{pmatrix} x_1 & x_3 \\ x_3 & x_2 \end{pmatrix} \succeq_{\mathbb{S}^2_+} 0$$

In this case, $p^* = 0$ but the problem is unsolvable!

In all cases, one can identify the cause for our troubles: the affine subspace defined by the linear constraints is tangent to the cone (it does not intersect its interior)

Rescuing strong duality

A feasible solution to a conic (primal or dual) problem is strictly feasible iff it belongs to the interior of the cone In other words, we must have $Ax = b$ and $x \succ_K 0$ for the primal and $A^T y \prec_{K^*} c$ for the dual

Strong duality: If the dual problem admits a strictly feasible solution, we have either

- ◇ an unbounded dual, in which case $d^* = +\infty = p^*$ and the primal is infeasible
- ◇ a bounded dual, in which case the primal is solvable with $p^* = d^*$ (hence there exists at least one feasible primal solution x^* such that $c^T x^* = p^* = d^*$)

Strong duality (continued)

- If the primal problem admits a strictly feasible solution, we have either
  - an unbounded primal, in which case $p^* = -\infty = d^*$ and the dual is infeasible
  - a bounded primal, in which case the dual is solvable with $d^* = p^*$ (hence there exists at least one feasible dual solution y^* such that $b^{\mathrm{T}} y^* = d^* = p^*$)
- The first case is a mere consequence of weak duality
- Finally, when both problems admit a strictly feasible solution, both problems are solvable and we have

$$c^{\mathrm{T}} x^* = p^* = d^* = b^{\mathrm{T}} y^*$$

Conic modelling with three cones

A first cone: $\mathbb{R}^n_+$

Standard meaning for inequalities:

$$\succeq_{\mathbb{R}^n_+} \quad \Leftrightarrow \quad \geq$$

$\Rightarrow$ linear optimization

But we can also model some nonlinearities!

$$|x_1 - x_2| \leq 1 \quad \Leftrightarrow \quad -1 \leq x_1 - x_2 \leq 1$$

$$|x_1 - x_2| \leq t \quad \Leftrightarrow \quad \begin{pmatrix} x_1 - x_2 - t \\ x_2 - x_1 - t \end{pmatrix} \leq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Terminology: conic representability

- Set S is K-representable if can be expressed as feasible region of conic problem using cone K
- Closed under intersection and Cartesian product
- Function f is K-representable iff its epigraph is K-representable
- Closed under sum, positive multiplication and max
- What we can do in practice: minimize a K-representable function over a K-representable set where K is a product of cones $\mathbb{R}^n_+$, $\mathbb{L}^n$, $\mathbb{S}^n_+$ and $\mathbb{R}^n$

A simple example

Consider set

$$S = \{x_1^2 + x_2^2 \leq 1\}$$

$\longrightarrow$ can be modelled as

$$(x_0, x_1, x_2) \in \mathbb{L}^2 \text{ and } x_0 = 1$$

$\Rightarrow$ S is $\mathbb{L}^2$-representable

but an additional variable x_0 was needed

$\Rightarrow$ *formally*, $S \subseteq \mathbb{R}^n$ is K-representable

iff there *exists* a set $T \subseteq \mathbb{R}^{n+m}$ such that

a. T is K-representable

b. $x \in S$ iff there exists $t \in \mathbb{R}^m$ such that $(x, t) \in T$

(i.e. S is the projection of T on $\mathbb{R}^n$)

Back to $\mathbb{R}^n_+$

- ◇ Polyhedrons and polytopes are $\mathbb{R}^n_+$-representable
- ◇ Hyperplanes and half-planes are $\mathbb{R}^n_+$-representable
- ◇ Affine functions $x \mapsto a^{\mathrm{T}}x + b$ are $\mathbb{R}^n_+$-representable
- ◇ Absolute values $x \mapsto \left|a^{\mathrm{T}}x + b\right|$ are $\mathbb{R}^n_+$-representable
- ◇ *Convex* piecewise linear function are $\mathbb{R}^n_+$-representable

Two potential issues with $\mathbb{R}^n_+$:

a. free variables in the primal $\rightarrow x = x^+ - x^-$

b. equalities in the dual $\rightarrow a^{\mathrm{T}}x \leq c$ and $a^{\mathrm{T}}x \geq c$

But these are **wrong** solutions !

What use is $K = \mathbb{R}^n$?

- $K = \mathbb{R}^n$ and $K^* = \{0\}$
- Can be used to introduce free variables in the primal $Ax = b, x \succeq_K 0$

$$x \succeq_{\mathbb{R}^n} 0 \quad \Leftrightarrow \quad x \text{ is free}$$

- or equalities in the dual $A^{\mathrm{T}} y \preceq_{K^*} c$

$$A^{\mathrm{T}} y \preceq_{\{0\}} c \quad \Leftrightarrow \quad A^{\mathrm{T}} y = c$$

  in combination with other cones
- $\mathbb{R}^n$ in dual or $\{0\}$ is primal is useless!

What use is $\mathbb{L}^n$?

- $f : x \mapsto \|x\|$, $f : x \mapsto \|x\|^2$ and $f : (x, z) \mapsto \frac{\|x\|^2}{z}$
- $B_r = \{x \in \mathbb{R}^n \mid \|x\| \leq r\}$
- $\{(x, y) \in \mathbb{R}^2_+ \mid xy \geq 1\}$
- $\{(x, y, z) \in \mathbb{R}^2_+ \times \mathbb{R} \mid xy \geq z^2\}$
- $\{(a, b, c, d) \in \mathbb{R}^4_+ \mid abcd \geq 1\}$
- $\{(x, t) \in \mathbb{R}^n \times \mathbb{R}\times \mid x^T Q x \leq t\}$ with $Q \in \mathbb{S}^n_+$

$\Rightarrow$ second-order cone optimization

Very useful trick: $xy \geq z^2 \Leftrightarrow (x + y, x - y, 2z) \in \mathbb{L}^2$

Unfortunately, $(x, y) \mapsto \frac{x}{y}$ is not convex!

What use is $\mathbb{S}^n_+$?

Preliminary remark: for the purpose of conic optimization, members of $\mathbb{S}^n$ are viewed as vectors in $\mathbb{R}^{n\times n}$

What about constraint $Ax = b$?

$$Ax = b \Leftrightarrow a_i^{\mathrm{T}} x = b_i \ \forall i$$

$a_i^{\mathrm{T}} x$ can be views as the inner product between a_i and x

Let $X, Y \in \mathbb{S}^n$: their inner product is

$$X \bullet Y = \sum_{1 \leq i,j \leq n} X_{i,j} Y_{i,j} = \operatorname{trace}(XY)$$

$\rightarrow$ replace $a_i^{\mathrm{T}} x$ by $A_i \bullet X$ with $A_i, X \in \mathbb{S}^n$

Standard format for semidefinite optimization

The primal becomes

$\min C \bullet X$ such that $A_i \bullet X = b_i \; \forall 1 \leq i \leq m$ and $X \succeq 0$

In the conic dual, we have

$$A^{\mathrm{T}} y = \sum a_i y_i, \text{ an application from } \mathbb{R}^m \mapsto \mathbb{R}^n$$

$\Rightarrow$ with the $\mathbb{S}^n_+$ cone, we have

$$\mathcal{A}(y) = \sum A_i y_i, \text{ an application from } \mathbb{R}^m \mapsto \mathbb{S}^n$$

which gives for the dual

$$\max b^{\mathrm{T}} y \text{ such that } \sum_{i=1}^{m} A_i y_i \preceq C$$

What use is $\mathbb{S}^n_+$ (continued) ?

- $\mathbb{S}^n_+$ *generalizes* both $\mathbb{R}^n_+$ and $\mathbb{L}^n$ (arrow matrices) (however, using $\mathbb{R}^n_+$ and $\mathbb{L}^n$ is more efficient)
- $f : X \mapsto \lambda_{max}(X)$ and $f : X \mapsto -\lambda_{min}(X)$
- $f : X \mapsto \max_i |\lambda_i|(X)$ (spectral norm)
- Describing ellipsoids $\{x \in \mathbb{R}^n \mid (x-c)^{\mathrm{T}} E(x-c) \leq 1\}$ with $E \succeq 0$
- Matrix constraint $XX^{\mathrm{T}} \preceq Y$ using the Schur Complement lemma

  When $A \succ 0$: $\begin{pmatrix} A & B \\ B^{\mathrm{T}} & C \end{pmatrix} \succeq 0 \Leftrightarrow C - B^{\mathrm{T}} A^{-1} B \succeq 0$
- And more ...

Interior-point methods

Back to convex optimization

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

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

Properties

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

Principle

Approximate a constrained problem by

a *family* of unconstrained problems

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

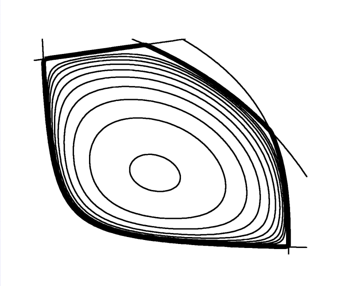
- F is smooth
- F is strictly convex on $\operatorname{int} C$
- $F(x) \longrightarrow +\infty$ when $x \longrightarrow \partial C$

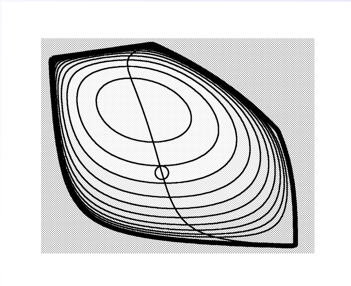
$$\longrightarrow \quad C = \operatorname{cl} \operatorname{dom} F = \operatorname{cl} \{x \in \mathbb{R}^n \mid F(x) < +\infty\}$$

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

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

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

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
- $F(x) \to +\infty$ when $x \to \partial C$
- 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 $h \in \mathbb{R}^n$

A (simple?) example

For linear optimization, $C = \mathbb{R}^n_+$: take $F(x) = -\sum_{i=1}^n \log x_i$

When $n = 1$, we can choose $(\kappa, \nu) = (1, 1)$

- ◇ $\nabla F(x) = -\frac{1}{x}$ and $\nabla F(x)^{\mathrm{T}} h = -\frac{h}{x}$
- ◇ $\nabla^2 F(x) = \frac{1}{x^2}$ and $\nabla^2 F(x)[h, h] = \frac{h^2}{x^2}$
- ◇ $\nabla^3 F(x) = -2\frac{1}{x^3}$ and $\nabla^3 F(x)[h, h, h] = -2\frac{h^3}{x^3}$

When $n > 1$, we have

- ◇ $\nabla F(x) = (-x_i^{-1})$ and $\nabla F(x)^{\mathrm{T}} h = -\sum h_i x_i^{-1}$
- ◇ $\nabla^2 F(x) = \mathrm{diag}(x_i^{-2})$ and $\nabla^2 F(x)[h, h] = \sum h_i^2 x_i^{-2}$
- ◇ $\nabla^3 F(x) = \mathrm{diag}_3(-2x_i^{-3})$, $\nabla^3 F(x)[h, h, h] = -2\sum h_i^3 x_i^{-3}$

and one can show that $(\kappa, \nu) = (1, n)$ is valid

Barrier calculus

Two elementary results:

◇ Scaling:

F is a (κ, ν)-s.-c. barrier for $\mathcal{C} \subseteq \mathbb{R}^n$ and $\lambda \in \mathbb{R}_{++}$

$\Rightarrow (\lambda F)$ is a $(\frac{\kappa}{\sqrt{\lambda}}, \lambda\nu)$-s.-c. barrier for $\mathcal{C}$

◇ Sum:

F is a (κ_1, ν_1)-s.-c. barrier for $\mathcal{C}_1 \subseteq \mathbb{R}^n$

G is a (κ_2, ν_2)-s.-c. barrier for $\mathcal{C}_2 \subseteq \mathbb{R}^n$

$\Rightarrow (F + G)$ is a $(\max\{\kappa_1, \kappa_2\}, \nu_1 + \nu_2)$-s.-c. barrier for the set $\mathcal{C}_1 \cap \mathcal{C}_2$ (if nonempty)

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

Summarized complexity result

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

iterations lead a solution with ϵ accuracy on the objective

Complexity result

- Let F be a (κ, ν)-self-concordant barrier for C and let $x_0 \in \operatorname{int} C$ be a starting point,

  a short-step interior-point algorithm can solve problem (P) up to ϵ accuracy within

  $$\mathcal{O}\left(\kappa\sqrt{\nu}\log\frac{c^T x_0 - p^*}{\epsilon}\right) \text{ iterations,}$$

  such that at each iteration the self-concordant barrier and its first and second derivatives have to be evaluated and a linear system has to be solved in $\mathbb{R}^n$

- Complexity invariant w.r.t. to scaling of F

- Universal bound on complexity parameter: $\kappa\sqrt{\nu} \geq 1$

Corollary

Assume F, ∇F and $\nabla^2 F$ are polynomially computable $\Rightarrow$ problem (P) can be solved in polynomial time

Existence

There exists a universal SC barrier with parameters

$$\kappa = 1 \text{ and } \nu = \mathcal{O}(n)$$

(but not necessarily efficiently computable)

Examples

- linear optimization: $(\kappa, \nu) = (1, n) \Rightarrow \mathcal{O}\left(\sqrt{n} \log \frac{1}{\varepsilon}\right)$
- entropy optimization: $\kappa = 1$ and $\nu = 2n \Rightarrow \mathcal{O}\left(\sqrt{n} \log \frac{1}{\varepsilon}\right)$ ($\inf c^{\mathrm{T}} x + \sum_i x_i \log x_i$ such that $Ax = b$ and $x \geq 0$)

Sketch of the proof

Define $n_\mu(x)$ the Newton step taken from x to x^*_μ

$$n_\mu(x) = 0 \text{ if and only if } x = x^*_\mu$$

We take

$$\delta(x,\mu) = \|n_\mu(x)\|_x \quad (\textit{size of the Newton step})$$

with a well-chosen (*coordinate invariant*) norm $\|\cdot\|_x$
Set $k \leftarrow 0$, perform the following main loop:

a. $\mu_{k+1} \leftarrow \mu_k(1-\theta)$ (*decrease barrier param*)

b. $x_{k+1} \leftarrow x_k + n_{\mu_{k+1}}(x_k)$ (*take Newton step*)

c. $k \leftarrow k+1$

Sketch of the proof (continued)

Key choice: parameters τ and θ such that

$$\delta(x_k, \mu_k) < \tau \quad \Rightarrow \quad \delta(x_{k+1}, \mu_{k+1}) < \tau$$

To relate $\delta(x_k, \mu_k)$ and $\delta(x_{k+1}, \mu_{k+1})$, introduce an intermediate quantity

$$\delta(x_k, \mu_{k+1})$$

We will also denote for simplicity

$$x_k \leftrightarrow x$$

$$\mu_k \leftrightarrow \mu$$

Sketch of the proof (end)

Given a (κ, ν)-self-concordant barrier:

- $\diamond$ $x \in \operatorname{dom} F$ and $\mu^+ = (1-\theta)\mu \Rightarrow$

$$\delta(x, \mu^+) \leq \frac{\delta(x,\mu) + \theta\sqrt{\nu}}{1-\theta}$$

- $\diamond$ $x \in \operatorname{dom} F$ and $\delta(x,\mu) < \frac{1}{\kappa} \Rightarrow$ define $x^+ = x + n_\mu(x)$

$$x^+ \in \operatorname{dom} F \text{ and } \delta(x^+, \mu) \leq \kappa\left(\frac{\delta(x,\mu)}{1-\kappa\delta(x,\mu)}\right)^2$$

with e.g. possible choice for parameters

$$\tau = \frac{1}{4\kappa} \text{ and } \theta = \frac{1}{16\kappa\sqrt{\nu}}$$

(hence the name short-step)

Primal-dual algorithms

Advantage of conic optimization over standard convex optimization is (symmetric) duality

However previous approach does not seem to use it !

$\Rightarrow$ a better approach that uses duality is needed

The linear case (again)

Introduce additional vector of variables $s \in \mathbb{R}^n$

$$\min c^T x \text{ such that } Ax = b \text{ and } x \geq 0$$

and

$$\max b^T y \text{ such that } A^T y + s = c \text{ and } s \geq 0$$

Primal-dual optimality conditions

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \geq 0$$

and

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y + s = c \text{ and } s \geq 0$$

Duality tells us x^* and y^* are optimal **iff** they satisfy

$$Ax = ,\ x \geq 0,\ A^{\mathrm{T}} y + s = c,\ s \geq 0 \text{ and } c^{\mathrm{T}} x = b^{\mathrm{T}} y$$

or

$$Ax = b,\ x \geq 0,\ A^{\mathrm{T}} y + s = c,\ s \geq 0 \text{ and } x_i s_i = 0\ \forall i$$

Both problems are handled simultaneously

Perturbed optimality conditions

Introducing a logarithmic barrier term in both problems

$$\min c^{\mathrm{T}} x - \mu \sum_i \log x_i \text{ such that } Ax = b \text{ and } x > 0$$

$$\max b^{\mathrm{T}} y + \mu \sum_i \log s_i \text{ such that } A^{\mathrm{T}} y + s = c \text{ and } s > 0$$

one can derive new perturbed optimality conditions

$$Ax = b,\ x \geq 0,\ A^{\mathrm{T}} y + s = c,\ s \geq 0 \text{ and } x_i s_i = \mu\ \forall i$$

Again, both problems are handled simultaneously

Primal-dual path following algorithm

Same principle as in the general case:

- ◇ Follow the central path
- ◇ Not wandering too far from it
- ◇ Until (primal-dual) optimality
- ◇ Using a polynomial number of iterations

Complexity is also the same:

$$\mathcal{O}\left(\sqrt{n}\log\frac{1}{\varepsilon}\right) \text{ iterations to get } \varepsilon \text{ accuracy}$$

But this scheme is very efficient in practice (long steps) (all practical implementations use it nowadays)

What about other convex/conic problems ?

This primal-dual scheme is only generalizable
to cones that are

a. self-dual ($K = K^*$)

b. homogeneous

(linear automorphism group acts transitively on int K)

(*[Nesterov & Todd 97]*)

There exists a complete classification of these cones :
in the real case, they are ...

$$\mathbb{R}^n_+ \quad , \quad \mathbb{L}^n \quad \text{and} \quad \mathbb{S}^n_+$$

and their Cartesian products!

Complexity

Complexity for a product of $\mathbb{R}^n_+, \mathbb{L}^n, \mathbb{S}^n_+$

$$\mathcal{O}\left(\sqrt{\nu}\log\frac{1}{\varepsilon}\right) \text{ iterations to get } \varepsilon \text{ accuracy}$$

where ν is the sum of

- n for $\mathbb{R}^n_+$ (see above) (barrier term is $-\sum \log x_i$)
- n for $\mathbb{S}^n_+$ (although there are $n(n+1)/2$ variables)
  (barrier term is $-\log\det X = -\sum \log \lambda_i$)
- 2 for $\mathbb{L}^n$ (independently of n !)
  (barrier term is $-\log(x_0^2 - \sum x_i^2)$; no $-\log x_0$ term!)

$\longrightarrow$ these problems are solved very efficiently in practice

More applications

Using semidefinite optimization:

Positive polynomials

- ◇ Single variable case: exact formulation
- ◇ Test positivity and minimize on an interval
- ◇ Multiple variable case: relaxation only

The MAX-CUT relaxation

- ◇ Relaxation of a difficult discrete problem
- ◇ With a quality guarantee (0.878)

References

Convex optimization

◇ Convex Analysis, Rockafellar,
Princeton University Press, 1980

◇ Convex optimization, Boyd and Vandenberghe,
Cambridge University Press, 2004 (on the web)

Convex modelling

◇ Lectures on Modern Convex Optimization, Analysis, Algorithms, and Engineering Applications, Ben-Tal and Nemirovski,
MPS/SIAM Series on Optimization, 2001

Interior-point methods (linear)

◇ Primal-Dual Interior-Point Methods, Wright
SIAM, 1997

◇ Theory and Algorithms for Linear Optimization, Roos, Terlaky, Vial, John Wiley & Sons, 1997

Interior-point methods (convex)

◇ Interior-point polynomial algorithms in convex programming, Nesterov & Nemirovski, SIAM, 1994

◇ A Mathematical View of Interior-Point Methods in Convex Optimization, Renegar,
MPS/SIAM Series on Optimization, 2001

Semidefinite optimization applications

- ◇ Handbook of Semidefinite Programming, Wolkowicz, Saigal, Vandenberghe (eds.) Kluwer, 2000
- ◇ Semidefinite programming, Boyd, Vandenberghe, SIAM Review 38 (1), 1996

Software: two choices among many others

- ◇ Linear & second-order cone: MOSEK (commercial)
- ◇ Linear, sec.-ord. & semidefinite: SeDuMi (free)

Thanks for you attention