ESI Summer Institute, Nonlinear Methods in Combinatorial Optimization

Florian Jarre,

Univ. Düsseldorf

Klagenfurt, Aug. 20 – Sept. 3, 2010

Interior-Point Methods for Semidefinite Programs

- Semidefinite Programs, Notation and an Application
- Method of Centers
- Self-Concordance
- Modifications

Notation

$\mathcal{S}^n$: The space of symmetric $n \times n$-matrices
$X \succeq 0$: $X \in \mathcal{S}^n$ is positive semidefinite, $X \in \mathcal{S}^n_+$.
$X \succ 0$: $X \in \mathcal{S}^n$ is positive definite.

Standard scalar product on the space of $n \times n$-matrices

$$\langle C, X \rangle := C \bullet X := \mathrm{trace}(C^T X) = \sum_{i,j} C_{i,j} X_{i,j}$$

inducing the Frobenius norm,

$$X \bullet X = \|X\|_F^2.$$

Notation (continued)

For given symmetric matrices $A^{(i)}$ a *linear* map $\mathcal{A}$ from $\mathcal{S}^n$ to $\mathbb{R}^m$ is given by

$$\mathcal{A}(X) = \begin{bmatrix} A^{(1)} \bullet X \\ \vdots \\ A^{(m)} \bullet X \end{bmatrix}.$$

The adjoint operator $\mathcal{A}^*$ satisfying

$$\langle \mathcal{A}^*(y), X \rangle = y^T \mathcal{A}(X) \qquad \forall\ X \in \mathcal{S}^n,\ y \in \mathbb{R}^m$$

is given by

$$\mathcal{A}^*(y) = \sum_{i=1}^{m} y_i A^{(i)}.$$

Linear semidefinite programs

$$\text{minimize} \quad C \bullet X \quad \text{where} \quad \begin{aligned} \mathcal{A}(X) &= b \\ X &\succeq 0 \end{aligned}$$

Similar structure as linear programs, only
“the condition $x \geq 0$ (componentwise)”

replaced by
“the condition $X \succeq 0$ (semidefinite)”

Can be solved by modified primal-dual methods
or, related, a “method of centers” (next).

Applications

- Relaxations of discrete optimization problems
  Lovasz θ-function
  Max-Cut
  Binary linear programs (Lovasz-Schrijver) ...
- Robust optimization
- Optimal control problems

Example

Consider the differential equation

$$\dot{x}(t) = Ax(t) \text{ with initial value } x(0) = x^{(0)}.$$

The matrix A is called *stable*, if for all initial values $x^{(0)}$ the solutions $x(t)$ converge to zero when $t \to \infty$.

Note that this is the case if, and only if, the real part of all eigenvalues of A is negative,

$$\text{Re}(\lambda_i(A)) < 0 \quad \text{for } 1 \leq i \leq n.$$

(Just for completeness; not needed here.)

Lyapunov's theorem

Stable, if, and only if, $\exists P \succ 0 : \quad A^T P + PA \prec 0.$

Motivation:

$$\begin{aligned}
& \frac{d}{dt} \, \|x(t)\|_P^2 \\
= \;& \frac{d}{dt} \, x(t)^T P x(t) \\
= \;& \dot{x}(t)^T P x(t) + x(t)^T P \dot{x}(t) \\
= \;& (Ax(t))^T P x(t) + x(t)^T P A x(t) \\
= \;& x(t)^T (A^T P + PA) x(t) \\
< \;& 0.
\end{aligned}$$

SDP-formulation

If the matrices $P_i, \quad 1 \le i \le n(n+1)/2$, form a basis of $\mathcal{S}^n$ the determination of a matrix $P = \sum_i y_i P_i$ with $P \succ 0$ and $A^T P + PA \prec 0$ leads to a linear semidefinite program

$$\min \left\{ \lambda \mid \sum y_i P_i \succ 0, \quad \lambda I - \sum y_i (A^T P_i + P_i A) \succ 0 \right\} < 0,$$

a linear semidefinite program in the "dual" form.

(In this form typically unbounded.)

(There are cheaper ways of checking numerically that the real parts of all eigenvalues of A are negative!)

Nonlinear systems

Now consider the system

$$\dot{x}(t) = A(t)x(t) \qquad (*)$$

where the matrix $A(t)$ is not known explicitely.

(For example, when there are small unknown perturbations of the matrix A.)

If there exist matrices $A^{(i)}$, $i = 1, 2, \ldots, K$, with

$$A(t) \in \textbf{conv}\left(\left\{A^{(i)}\right\}_{i \leq i \leq K}\right) \text{ for all } t \geq 0,$$

then the existence of a Lyapunov matrix $P \succ 0$ with

$$(A^{(i)})^T P + P A^{(i)} \prec 0 \;\text{ for }\; 1 \leq i \leq K$$

is a sufficient condition for the stability of $(*)$.

Nonlinear systems

Because then,

$$\begin{aligned} A(t)^T P + PA(t) &\prec 0 \qquad \text{and} \\ \frac{d}{dt}\,\|x(t)\|_P^2 = x(t)^T(A(t)^T P + PA(t))x(t) &< 0 \end{aligned}$$

whenever $x(t) \neq 0$.

Nonlinear systems

Because then,

$$\begin{aligned} A(t)^T P + PA(t) &\prec 0 \qquad \text{and} \\ \frac{d}{dt}\,\|x(t)\|_P^2 = x(t)^T(A(t)^T P + PA(t))x(t) &< 0 \end{aligned}$$

whenever $x(t) \neq 0$.

(The above argument is not complete; it can be made complete by showing that $\frac{d}{dt}\,\|x(t)\|_P^2$ is “sufficiently” negative.)

Note

The condition that the real parts of all eigenvalues of all $A^{(i)}$ are negative is a necessary but not a sufficient condition:

When the real part of all eigenvalues of two matrices A and B are negative, this may not be the case for $\frac{1}{2}(A+B)$, choose e.g.

$$A = \begin{bmatrix} -1 & 4 \\ 0 & -1 \end{bmatrix}, \qquad B = \begin{bmatrix} -1 & 0 \\ 4 & -1 \end{bmatrix}.$$

(All eigenvalues real and equal to -1 but $\frac{1}{2}(A+B)$ has eigenvalue $+1$.)

Method of centers

Consider the following problem:

minimize $f(x)$ s.t. $x \in I\!R^n$: $f_i(x) \leq 0$ for $1 \leq i \leq m$.

Assumptions:

f and f_i three times continuously differentiable
f and $-\sum_{i \geq 1} \ln(-f_i)$ convex

Always satisfied if all f_i are convex

Simplifications

- f is linear, $f(x) \equiv c^T x$
(Without loss of generality!)
- The feasible set

$$\mathcal{S} := \{x \mid f_i(x) \leq 0 \ \text{ for } \ 1 \leq i \leq m\}$$

is bounded and has nonemtpy interior.
(Some restriction of generality)

Logarithmic barrier method

$$\phi(x) := -\sum_{i=1}^{m} \ln(-f_i(x))$$

logarithmic barrier function for $\mathcal{S}$, and for some fixed $\theta \geq 1$

$$\varphi(x,\lambda) := -\theta \ln(\lambda - c^T x) + \phi(x)$$

barrier function for $\mathcal{S}(\lambda) := \mathcal{S} \cap \{x \mid c^T x \leq \lambda\}$.

The minimizer $\bar{x}$ of ϕ is called the analytic center of $\mathcal{S}$.
The minimizer $x(\lambda)$ of $\varphi(\, . \, , \lambda)$ is called analytic center of $\mathcal{S}(\lambda)$.

Method of analytic centers

Let $x^0 \in \mathcal{S}^\circ$ and $\lambda_0 > c^T x^0$ be given. Set $\sigma = \frac{1}{2}$ and $k = 0$.

Repeat

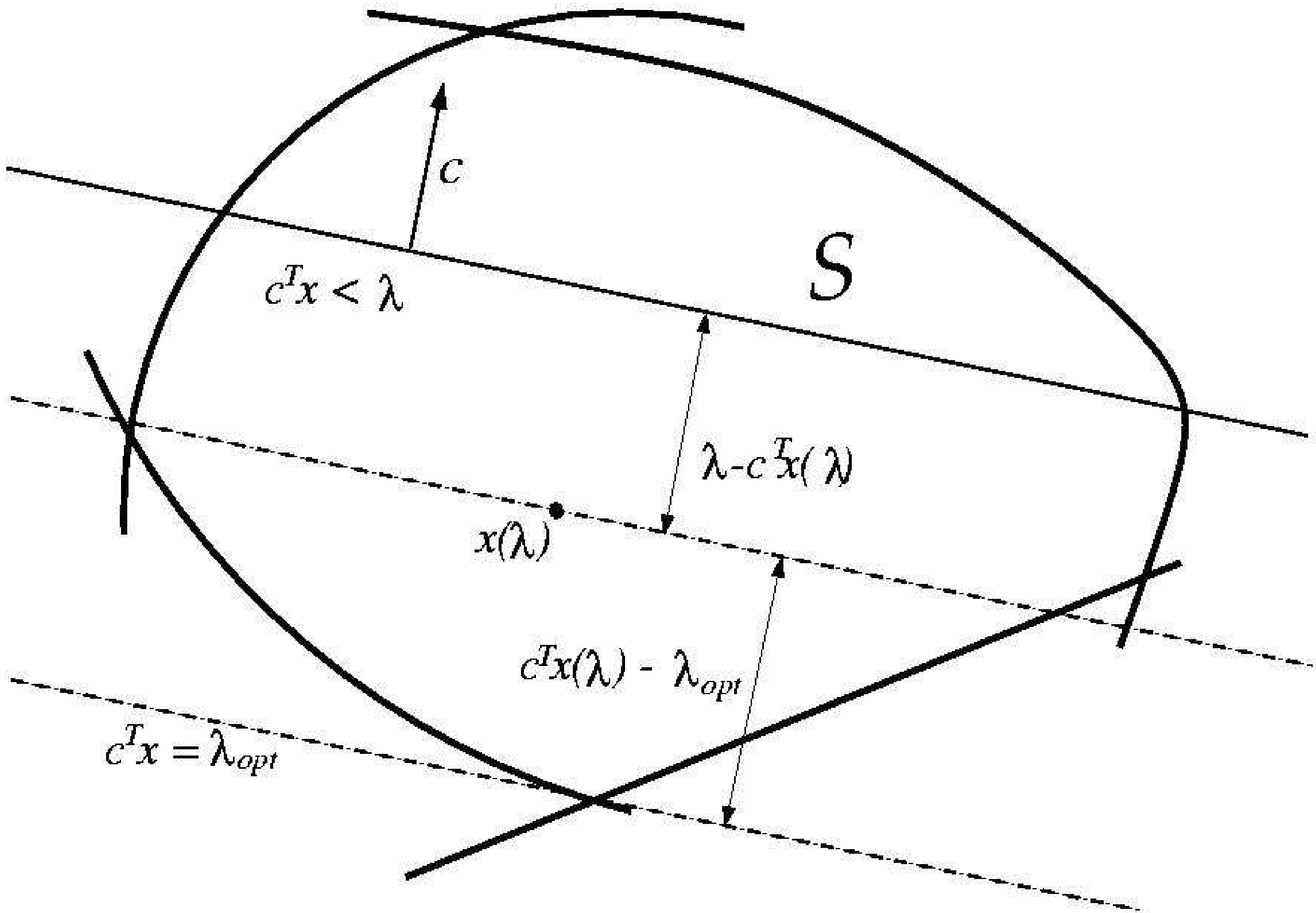
1. Reduce λ_k to $\lambda_{k+1} := \lambda_k - \sigma(\lambda_k - c^T x^k)$.
2. Starting at x^k do few steps of Newton's method for minimizing $\varphi(\ .\ , \lambda_{k+1})$.
3. Set $k := k + 1$.

End.

Crucial questions

1. How fast does Newton's method converge?
2. How big is $\lambda_k - c^T x^k$ compared to the unknown distance $c^T x^k - \lambda_{opt}$?

Distance to optimality



First question

- For simplicity, restrict examination to ϕ
  and apply results for Newton's method later to φ which has the same structure!
- Intuitively, want $\nabla^2\phi$ to be "nearly constant" for Newton's method to work well.
- In fact, want relative change of $\nabla^2\phi$ to be small.
- The absolute change of $\nabla^2\phi$ is given by $\nabla^3\phi$.
- Hence, want $\nabla^3\phi$ to be small compared to $\nabla^2\phi$.
- Look at "Karmarkar's barrier function" for the positive real axis: $\phi(t) := -\ln t$. Here, $\phi'''(t) \le 2(\phi''(t))^{3/2}$.
- Generalize this to n dimensions:

Self-concordance

The barrier function $\phi : \mathcal{S}^\circ \to I\!R$ is called (strongly) self-concordant if for any $x \in \mathcal{S}^\circ$ and any $h \in I\!R^n$ the restriction $l = l_{x,h}$ of ϕ to the line $x + th$,

$$l(t) := \phi(x + th)$$

satisfies $l'''(0) \leq 2(l''(0))^{3/2}$.

- This assumes l is C^3-smooth and $l'' \geq 0$, i.e. l is convex.
- Power $3/2$ guarantees invariance w.r.t. the length of h !

Does this make sense?

I.) Do there exist functions that satisfy this condition?

1. Easy to verify (binomial formula):
   If ϕ_1 and ϕ_2 are self-concordant, then so is $\phi_1 + \phi_2$.
   (provided the domains of ϕ_1 and ϕ_2 intersect)
2. Very easy to verify:
   If $\mathcal{A}$ is an affine mapping and ϕ is self-concordant, then so is $\phi(\mathcal{A}(\ .\ ))$. (provided the range of $\mathcal{A}$ intersects the domain of ϕ.)
3. This implies that $-\sum \ln(b_i - a_i^T x)$ is a self-concordant barrier function for the polyhedron $\{x \mid a_i^T x \leq b_i\}$.

4. Convex quadratic constraints $q(x) \leq 0$:
   For fixed x with $q(x) < 0$ and fixed h, the term $q(x+th)$ can be factored into the product of two linear terms, so that by 1. and 2., the function $l(t) = -\ln(-q(x+th))$ is self-concordant!

Most important

5. Semidefinite constraints $X \succeq 0$ ($X = X^T \in \mathbb{R}^{n\times n}$):
   The barrier function $\phi(X) := -\ln\det(X)$ is self-concordant:
   For fixed $X \succ 0$ and fixed $H = H^T \in \mathbb{R}^{n\times n}$ it follows

$$\begin{aligned}
l(t) &= -\ln\det(X + tH) \\
&= -\ln\det(X^{1/2}(I + tX^{-1/2}HX^{-1/2})X^{1/2}) \\
&= -2\ln\det(X^{1/2}) - \ln(\prod_{j=1}^{n}(1 + t\lambda_i)) \\
&= -\ln\det(X) - \sum_{j=1}^{n}\ln(1 + t\lambda_i),
\end{aligned}$$

   where λ_i are the eigenvalues of $X^{-1/2}HX^{-1/2}$ independent of t. Again, by 1. and 2. it follows that ϕ is self-concordant.

Note that

$l'(0) = -\operatorname{trace}(X^{-1/2} H X^{-1/2}) = -\operatorname{trace}(H X^{-1}) =: -H \bullet X^{-1}$.

The derivative of $-\log\det(X)$ with respect to the scalar product "$\bullet$" is given by X^{-1}.

(Remember, the derivative is the linear map that best approximates the function locally.)

Moreover, let some symmetric matrices $A^{(0)}, \ldots, A^{(m)}$ be given. Consider the problem

$$\text{minimize } \; b^T y \; \text{ s.t. } \; \mathcal{A}(y) := A^{(0)} + \sum_{1=1}^{m} y_i A^{(i)} \succeq 0.$$

By affine invariance, also the function $\Phi(y) := -\ln\det(\mathcal{A}(y))$ is self-concordant. Moreover, the derivatives of Φ can be explicitely stated:

$$\begin{aligned} \frac{\partial}{\partial y_i}\Phi(y) &= -\mathcal{A}(y)^{-1} \bullet A^{(i)}, \\ \frac{\partial^2}{\partial y_i \partial y_j}\Phi(y) &= \mathcal{A}(y)^{-1} A^{(i)} \mathcal{A}(y)^{-1} \bullet A^{(j)}. \end{aligned}$$

II.) Does this condition really guarantee that Newton's method converges well?

Since $l'''(0) \leq 2l''(0)^{3/2}$ holds for any x (defining l), this implies that $l'''(t) \leq 2l''(t)^{3/2}$ for any t in the domain of l.

Let $u(t) := l''(t)$, then $u'(t) \leq 2u(t)^{3/2}$ is a valid differential inequality.

The extremal solution of $v'(t) = 2v(t)^{3/2}$ with initial value

$$v(0) = u(0) = h^T \nabla^2 \phi(x) h := \delta^2$$

is given by $v(t) = 1/(\delta^{-1} - t)^2$.

Whenever v has finite values, u must be finite, and hence so must be l.

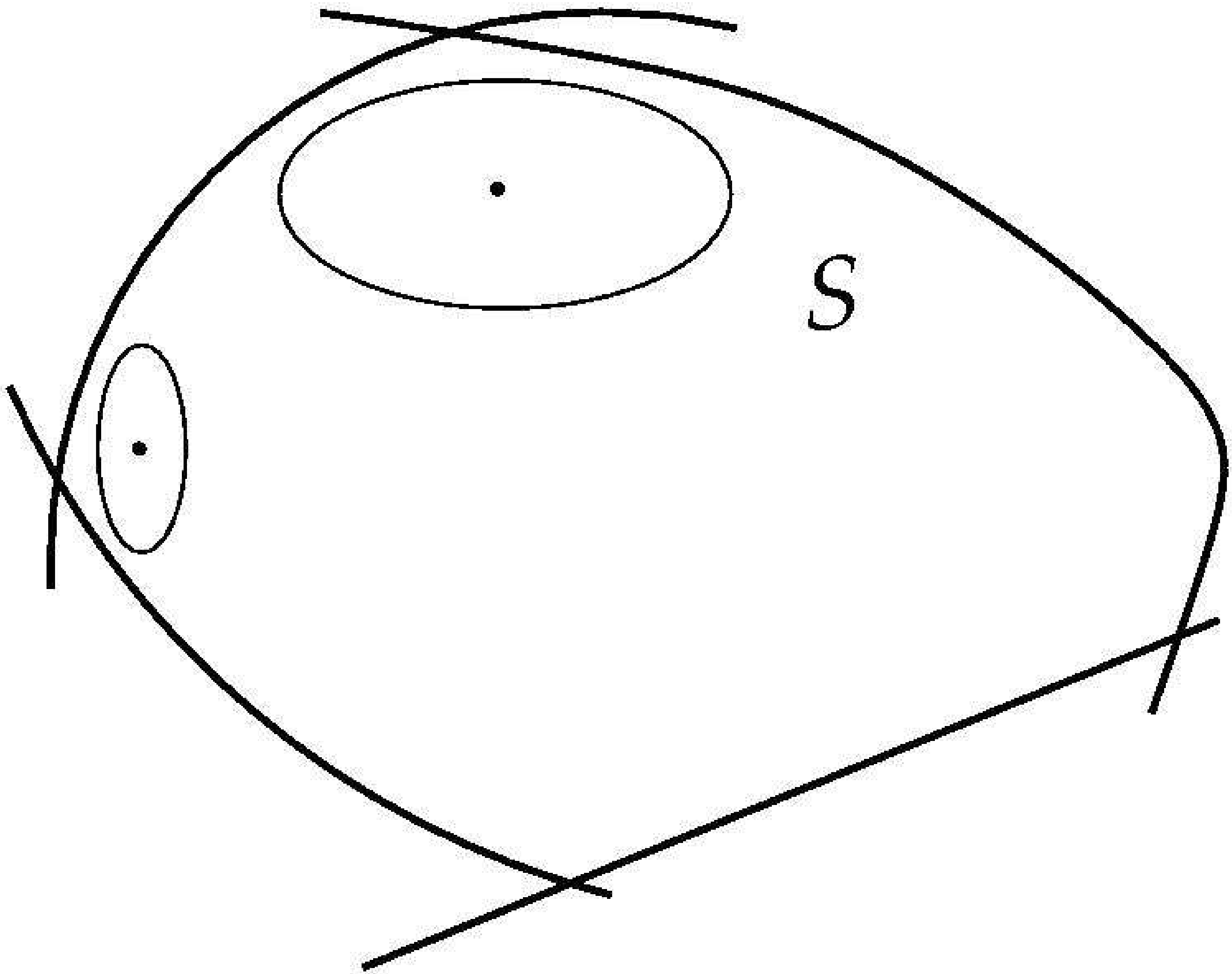
Inner Ellipsoid

Define a (semi-) norm as $\|h\|_{H_x} := \sqrt{h^T H_x h}$ based on the Hessian $H_x := \nabla^2 \phi(x)$ of ϕ at x

Let $E(x) := \{h \mid h^T \nabla^2 \phi(x) h \leq 1\}$ be the unit ball of the (semi-) norm.

Then, for any $x \in \mathcal{S}^\circ$ the inclusion $x + E(x) \subset \mathcal{S}$ holds true.

Inner ellipsoid



Further Results

Equivalent Relative Lipschitz condition:

$$|h^T(\nabla^2\phi(x+\Delta x) - \nabla^2\phi(x))h| \leq \delta M(\delta) h^T \nabla^2\phi(x) h,$$

where

$$\delta := \|\Delta x\|_{H_x} \quad \text{and} \quad M(\delta) := \frac{2}{1-\delta} + \frac{\delta}{(1-\delta)^2} = 2 + O(\delta).$$

(Somewhat more difficult to show.)

Newton's method

Let $x \in \mathcal{S}^\circ$ be given. Denote $H_x := \nabla^2 \phi(x)$.
Let $\Delta x := -H_x^{-1} \nabla \phi(x)$ be the Newton step for minimizing ϕ starting at $x \in \mathcal{S}^\circ$. Assume that $\delta := \|\Delta x\|_{H_x} < 1$.
Then, by the inner ellipsoid $x + \Delta x \in \mathcal{S}^\circ$.

Let $\tilde{x} := x + \Delta x$ and $\Delta \tilde{x} := -H_{\tilde{x}}^{-1} \nabla \phi(\tilde{x})$ be the "next" Newton step. Then

$$\|\Delta \tilde{x}\|_{H_{\tilde{x}}} \leq \frac{\delta^2}{(1-\delta)^2}.$$

This implies quadratic convergence in at least one fifth of the inner ellipsoid about the center.

(Related idea of proof as for inner ellipsoid.)

Second question, distance to optimality

Self-concordance is a (relative) Lipschitz condition on the Hessian of ϕ.

- By adding a linear perturbation to ϕ, the self-concordance condition obviously does not change.
- But by adding a linear perturbation to ϕ we can make any point $\hat{x} \in \mathcal{S}^\circ$ a minimizer of the perturbed function.
- For $\hat{x}$ close to the boundary, the perturbation will have to be large, and the perturbed function will have a "large gradient" at the minimizer of ϕ.
- If we want to avoid points close to the boundary to be a minimizer of "our" barrier function, we may limit the norm of its gradient – of course with respect to the canonical norm $\| \, . \, \|_{H_{\hat{x}}}$.

With the notation used in the self-concordance condition, we require for some fixed $\theta \geq 1$:

$$l'(0) \leq \sqrt{\theta} l''(0)^{1/2}.$$

If ϕ is self-concordant and satisfies the above condition, we say ϕ is θ**-self-concordant.**

This condition is also affine invariant.

It is "additive" w.r.t. θ, in the sense that if ϕ_1 and ϕ_2 satisfy the condition with values $\theta_1 \geq 1$ and $\theta_2 \geq 1$, then so does $\phi_1 + \phi_2$ with value $\theta_1 + \theta_2$.

The previous examples satisfy the condition with $\theta = 1$ for a linear or convex quadratic constraint, and $\theta = l$ for an $l \times l$ semidefinite constraint.

Results

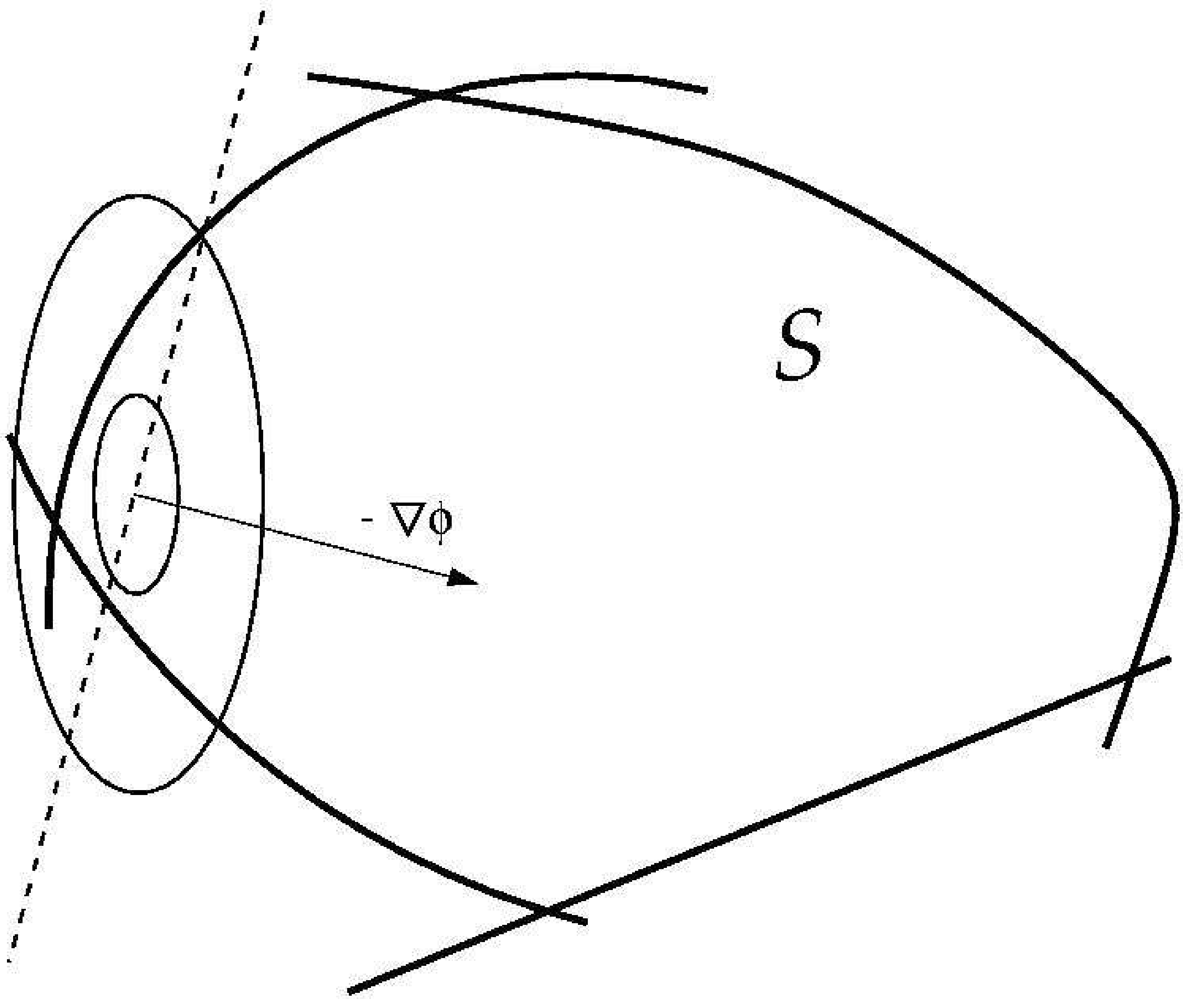
1. We have $\lambda - c^T x(\lambda) > c^T x(\lambda) - \lambda_{opt}$ when θ in the definition of φ is chosen as least as large as the self-concordance parameter of ϕ. "Identical" proof as for inner ellipsoid, just the other way round.

2. Let $x \in \mathcal{S}^\circ$ be arbitrary and let ϕ be a θ-self-concordant barrier function for $\mathcal{S}$. Let $\mathcal{H} := \{y \mid (y-x)^T \nabla\phi(x) \geq 0\}$ be a half space cutting through x and $E(x)$ be the inner ellipsoid. Then

$$\mathcal{S} \cap \mathcal{H} \subset x + (\theta + 2\sqrt{\theta})E(x).$$

($\mathcal{H} = \mathbb{R}^n$ when $\nabla\phi(x) = 0$.) Again, similar proof as for inner ellipsoid.

Inner and outer ellipsoid



- Now we have **all essential tools** to show that the method of centers converges at a fixed rate.
- If λ_k is changed only little at each iteration (namely $\sigma = 1/(8\sqrt{\theta})$ rather than $\sigma = 1/2$) then only **one** step of Newton's method suffices at each iteration, and after $12\sqrt{\theta}$ iterations the unknown distance $\lambda_k - \lambda_{opt}$ is reduced by a factor at least $1/2$.

Discussion

The given rate of convergence (for a problem with 10000 convex quadratic constraints, 1200 iterations are needed to reduce the error bound $\lambda_k - \lambda_{opt}$ by a factor $1/2$) is **too slow** for practical implementations.
BUT it guarantees a very weak dependence on the data of the problem — the rate only depends on a weighted number of constraints, where complicated conditions like semidefiniteness constraints are counted with a somewhat higher weight.

- No dependence on the number of unknowns.

(Generalization to Hilbert space by Renegar)

- Assuming exact artihmetic – unlike the conjugate gradient or steepest descent methods, no dependence on any condition numbers of the problem.
- Assuming exact artihmetic – unlike the simplex method, no dependence on degeneracy.

Hence, the CONCEPT is very robust.
Find an acceleration based on this concept.

Modifications

- Infeasible starting points, empty interior, unbounded set of optimal solutions.
- Predictor corrector strategy: Under “mild” conditions, the central path $x(\lambda)$ forms a smooth curve leading to an optimal solution.
  Through any given point $\hat{x} \in \mathcal{S}^\circ$ one can define a perturbed central path $\hat{x}(\lambda)$ leading to an optimal solution as well, and the tangent to this curve is “easily” computable. (Same system as used for Newton’s method.)
  Do some extrapolation along this tangent and start the Newton corrections from the extrapolated point.

Conic formulations

- Each convex program that posesses a self-concordant barrier function can be expressed in conic form with a self-concordant barrier function of the same order of magnitude. (Nesterov and Nemirovskii 1994)
- Conic formulations allow for primal-dual methods that have turned out to be more efficient in practical implementations. (Their theoretical complexity is the same as the one of the method of centers.)
- Many programs (like semidefinite programs) are naturally given in conic form and thus allow for direct application of primal dual methods.

Duality for SDPs

If there exists $X \succ 0$ with $\mathcal{A}(X) = b$ (strict feasibility), then

$$(P) \qquad \text{inf } C \bullet X \text{ s.t. } \mathcal{A}(X) = b,\ X \succeq 0$$

$$(D) \qquad = \text{sup } b^T y \text{ s.t. } \mathcal{A}^*(y) + S = C,\ S \succeq 0.$$

If (P) and (D) have strictly feasible solutions, then the optimal solutions X^{opt} and y^{opt}, S^{opt} of both problems exist and satisfy the equation

$$X^{opt} S^{opt} = 0.$$

Basic theory (continued)

Conversely, any pair X and y, S of feasible points for (P) and (D) satisfying

$$\begin{aligned} \mathcal{A}(X) &= b, \quad X \succeq 0 \\ \mathcal{A}^*(y) + S &= C, \quad S \succeq 0 \\ XS &= 0 \qquad (\text{or } SX = 0) \end{aligned}$$

is optimal for both problems.

Basic theory (continued)

Conversely, any pair X and y, S of feasible points for (P) and (D) satisfying

$$\begin{array}{rcl} \mathcal{A}(X) &=& b, \quad X \succeq 0 \\ \mathcal{A}^*(y) + S &=& C, \quad S \succeq 0 \\ XS &=& 0 \qquad (\text{or } \; SX = 0) \end{array}$$

is optimal for both problems.

For Newton's method symmetrize (Monteiro et. al.) and perturb the last equation for some small $\mu > 0$:
E.g. $\mathcal{M}(X, S) = \frac{1}{2}(XS + SX)$ (AHO)

Basic theory (continued)

Conversely, any pair X and y, S of feasible points for (P) and (D) satisfying

$$\begin{array}{rcl} \mathcal{A}(X) & = & b, \quad X \succeq 0 \\ \mathcal{A}^*(y) + S & = & C, \quad S \succeq 0 \\ \mathcal{M}(X,S) & = & 0 \quad \longleftarrow \mu I \end{array}$$

is optimal for both problems.

For Newton's method symmetrize (Monteiro et. al.) and perturb the last equation for some small $\mu > 0$:
E.g. $\mathcal{M}(X,S) = \frac{1}{2}(XS + SX)$ (AHO)

Solutions coincide with the "analytic centers" of the self-concordance approach (but not so the Newton steps).

Basic Theory (continued)

For example (AHO), the linearization of

$$\begin{aligned} \mathcal{A}(X) &= b, \quad X \succeq 0 \\ \mathcal{A}^*(y) + S &= C, \quad S \succeq 0 \\ XS + SX &= 2\mu I \end{aligned}$$

Basic theory (continued)

For example (AHO), the linearization of

$$\begin{aligned} \mathcal{A}(X) &= b, \quad X \succeq 0 \\ \mathcal{A}^*(y) + S &= C, \quad S \succeq 0 \\ XS + SX &= 2\mu I \end{aligned}$$

yields the linear system for $\Delta X, \Delta y, \Delta S$:

$$\begin{aligned} \mathcal{A}(\Delta X) &= b - \mathcal{A}(X), \quad X \succeq 0 \\ \mathcal{A}^*(\Delta y) + \Delta S &= C - \mathcal{A}^*(y) - S, \quad S \succeq 0 \\ X\Delta S + \Delta X S + S\Delta X + \Delta S X &= 2\mu I - XS - SX. \end{aligned}$$

Interior-point methods

Same principle as in case of linear programs

(linear systems somewhat more complicated,
step length needs to be adjusted more carefully ...)

Repeat

- Update $(X, y, S) \mapsto (X + \Delta X, y + \Delta y, S + \Delta S)$.
- Reduce $\mu > 0$

Until convergence

Sedumi

Again, a very comfortable free software package for this is SEDUMI.

- The nonnegative cone $\{x \mid x \geq 0\}$ can be combined with
- the semidefinite cone $\{X \mid X \succeq 0\}$ and with
- the ice-cream cone, $\{x \mid x_0 \geq \|(x_1, \ldots, x_n)\|_2\}$, also called second order cone or Loren(t)z cone

(and with some rotated second-order cone).

To be specified by an input argument “K”.

References

Nesterov, J.E., Nemirovsky A.S. (1994): Interior Point Polynomial Methods in Convex Programming: Theory and Applications. SIAM, Philadelphia

(Excellent, non-trivial)

First Chapters of:
Jarre, F. (1994): Interior-point methods via self-concordance or relative Lipschitz condition, Habilitiationsschrift. Universität Würzburg
http://www.opt.uni-duesseldorf.de/de/forschung-fs.html

(Much less complete, slightly simpler)

References

Nesterov, J.E., Nemirovsky A.S. (1994): Interior Point Polynomial Methods in Convex Programming: Theory and Applications. SIAM, Philadelphia

(Excellent, non-trivial)

First Chapters of:
Jarre, F. (1994): Interior-point methods via self-concordance or relative Lipschitz condition, Habilitiationsschrift. Universität Würzburg
http://www.opt.uni-duesseldorf.de/de/forschung-fs.html

(Much less complete, slightly simpler)

If you like to get a .pdf-file of the talk write to:
jarre@opt.uni-duesseldorf.de