Sequential Linear Cone Programming (SLCP)

... and Sensitivity of SDP's

Chemnitz, Nov. 9, 2004

R.W. Freund, F. Jarre

Survey

Linear Semidefinite Optimization

- Notation
- Sensitivity Analysis

Survey

Linear Semidefinite Optimization

- Notation
- Sensitivity Analysis

Recent Results on Nonlinear Semidefinite Optimization

- Application (Positive Real Lemma)
- Generalized Sensitivity Result
- An SLCP Method

Notation

$\mathcal{S}^n$: The space of symmetric $n \times n$-matrices

$X \succeq 0, \quad (X \succ 0)$: $X \in \mathcal{S}^n$ is positive semidefinite (positive definite).

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

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

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

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

The adjoint operator $\mathcal{A}^*$ is given by

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

Linear Semidefinite Program in Standard Form:

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

Duality

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

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

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

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

$$\bar{X}\bar{S} = 0.$$

(Converse is true even when Slaters condition is violated.).

Today, linear SDPs are well analyzed and there exists

Numerically efficient

(and polynomial)

public domain software

for linear semidefinite programs,

e.g. SEDUMI by Jos Sturm ($*$1971, †2003).

Motivation:

Plan to solve a nonlinear SDP by a sequence of (approximating) linear SDPs.

Need to understand how the optimal solution of the linear SDP changes when its data is perturbed slightly.

A sensitivity result for linear SDPs

Uniqueness-assumption:

Data $\mathcal{D}$ of a pair (P) and (D) of primal and dual linear semidefinite programs:

$\mathcal{D} = [\mathcal{A}, b, C]$ with $\mathcal{A} : \mathcal{S}^n \to \mathbb{R}^m$, $b \in \mathbb{R}^m$, $C \in \mathcal{S}^n$.

Assume that (P) and (D) satisfy Slater's condition, and that $\bar{X} \in \mathcal{S}^n$ and $\bar{y} \in \mathbb{R}^m$, $\bar{S} \in \mathcal{S}^n$ are unique and strictly complementary solutions of (P) and (D), that is,

$$\begin{aligned} \mathcal{A}(\bar{X}) &= b, \quad \bar{X} \succeq 0, \\ \mathcal{A}^*(\bar{y}) + \bar{S} &= C, \quad \bar{S} \succeq 0, \\ \bar{X}\bar{S} &= 0, \quad \bar{X} + \bar{S} \succ 0. \end{aligned}$$

Theorem (Freund & J. 2003)

If the data of (P) and (D) is changed by sufficiently small perturbations

$$\Delta\mathcal{D} = [\Delta\mathcal{A}, \Delta b, \Delta C],$$

then the optimal solutions $\bar{X}(\mathcal{D}), \bar{y}(\mathcal{D}), \bar{S}(\mathcal{D})$ of the semidefinite programs are differentiable (analytic) functions of the perturbations, i.e.

$$\bar{X}(\mathcal{D} + \Delta\mathcal{D}) = \bar{X}(\mathcal{D}) + D_{\mathcal{D}}\bar{X}[\Delta\mathcal{D}] + O(\|\Delta\mathcal{D}\|^2).$$

Furthermore, the directional derivatives

$$\dot{X} := D_{\mathcal{D}}\bar{X}[\Delta\mathcal{D}], \quad \dot{y} := D_{\mathcal{D}}\bar{y}[\Delta\mathcal{D}], \text{ and } \dot{S} := D_{\mathcal{D}}\bar{S}[\Delta\mathcal{D}],$$

of the solution $\bar{X}(\mathcal{D}), \bar{y}(\mathcal{D}), \bar{S}(\mathcal{D})$ satisfy

$$\mathcal{A}(\dot{X}) = \Delta b - \Delta\mathcal{A}(\bar{X}),$$

$$\mathcal{A}^*(\dot{y}) + \dot{S} = \Delta C - \Delta\mathcal{A}^*(\bar{y}),$$

$$\dot{X}\bar{S} + \bar{X}\dot{S} = 0.$$

The last line can be equivalently rewritten as

$$\dot{S}\bar{X} + \bar{S}\dot{X} + \dot{X}\bar{S} + \bar{X}\dot{S} = 0.$$

More general theory in Bonnans and Shapiro (2000).

Idea of Proof:

1) Slater and continuity:
   The perturbed problem must have a solution.

2) Subtract optimality conditions and take limit:
   We get precisely the statement of the theorem.

3) It remains to be shown that this system is "nonsingular".
   (It is an overdetermined system just by the number of equations and unknowns.)

Idea of Proof: (continued)

By complementarity, $\bar{X}\bar{S} = 0 = \bar{S}\bar{X}$, and thus the matrices $\bar{X} \succeq 0$ and $\bar{S} \succeq 0$ commute. This guarantees that there exists a unitary matrix U and diagonal matrices

$$\Lambda = \text{Diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_n) \succeq 0 \quad \text{and} \quad \Sigma = \text{Diag}\,(\sigma_1, \sigma_2, \ldots, \sigma_n) \succeq 0$$

such that

$$\bar{X} = U\Lambda U^T \quad \text{and} \quad \bar{S} = U\Sigma U^T.$$

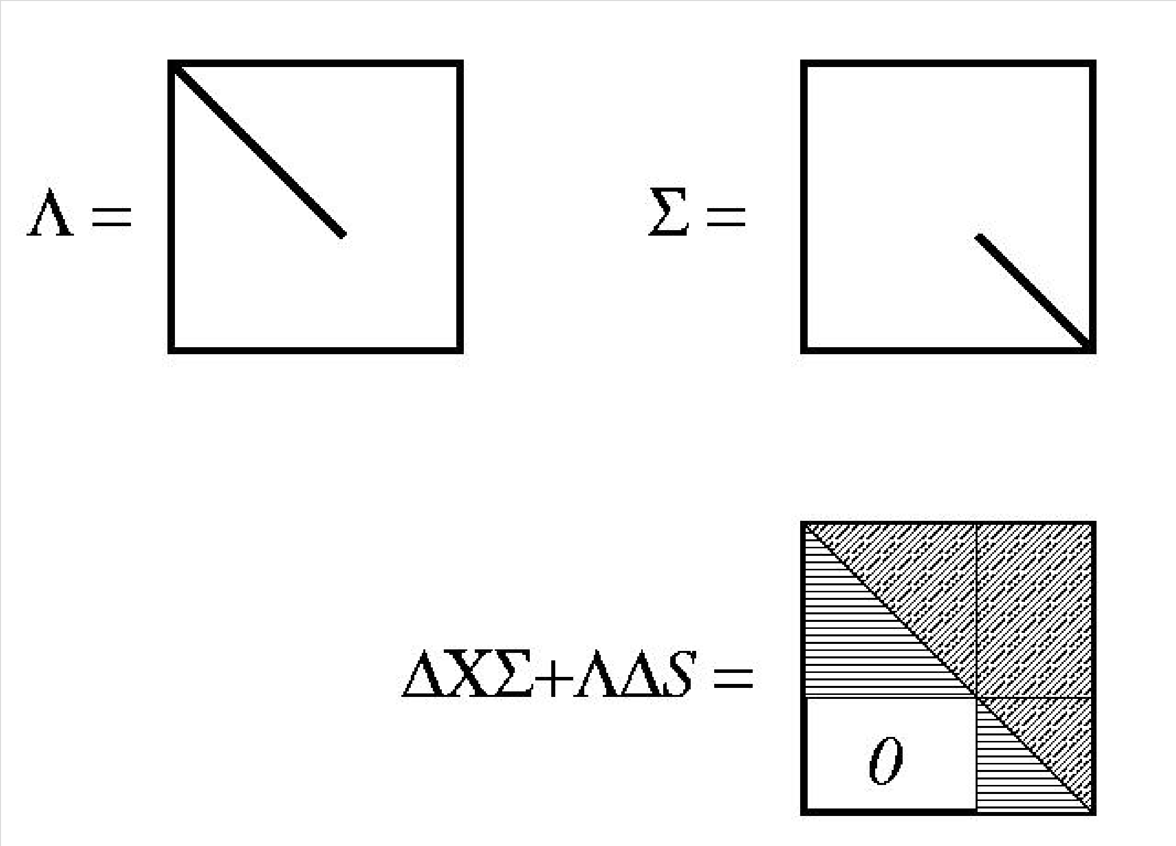
Partition

$$\lambda_1, \lambda_2, \ldots, \lambda_k > 0 \quad \text{and} \quad \sigma_{k+1}, \sigma_{k+2}, \ldots, \sigma_n > 0.$$

Transform so that, without loss of generality, $X = \Lambda, S = \Sigma$ and consider the upper triangular part $\Pi_{up}(\Delta X \Sigma + \Lambda \Delta S) = 0$,
along with $\mathcal{A}(\Delta X) = 0$ and $\mathcal{A}^*(\Delta y) + \Delta S = 0$.

Idea of Proof: (continued II)

Using the structure of the equation and uniqueness of the optimal solution shows that this system has only the zero solution.



Idea of Proof: (last)

4) Implicit function theorem ... (Done)

Key was to identify a nonsingular part of the overdetermined system – that can also be used numerically.

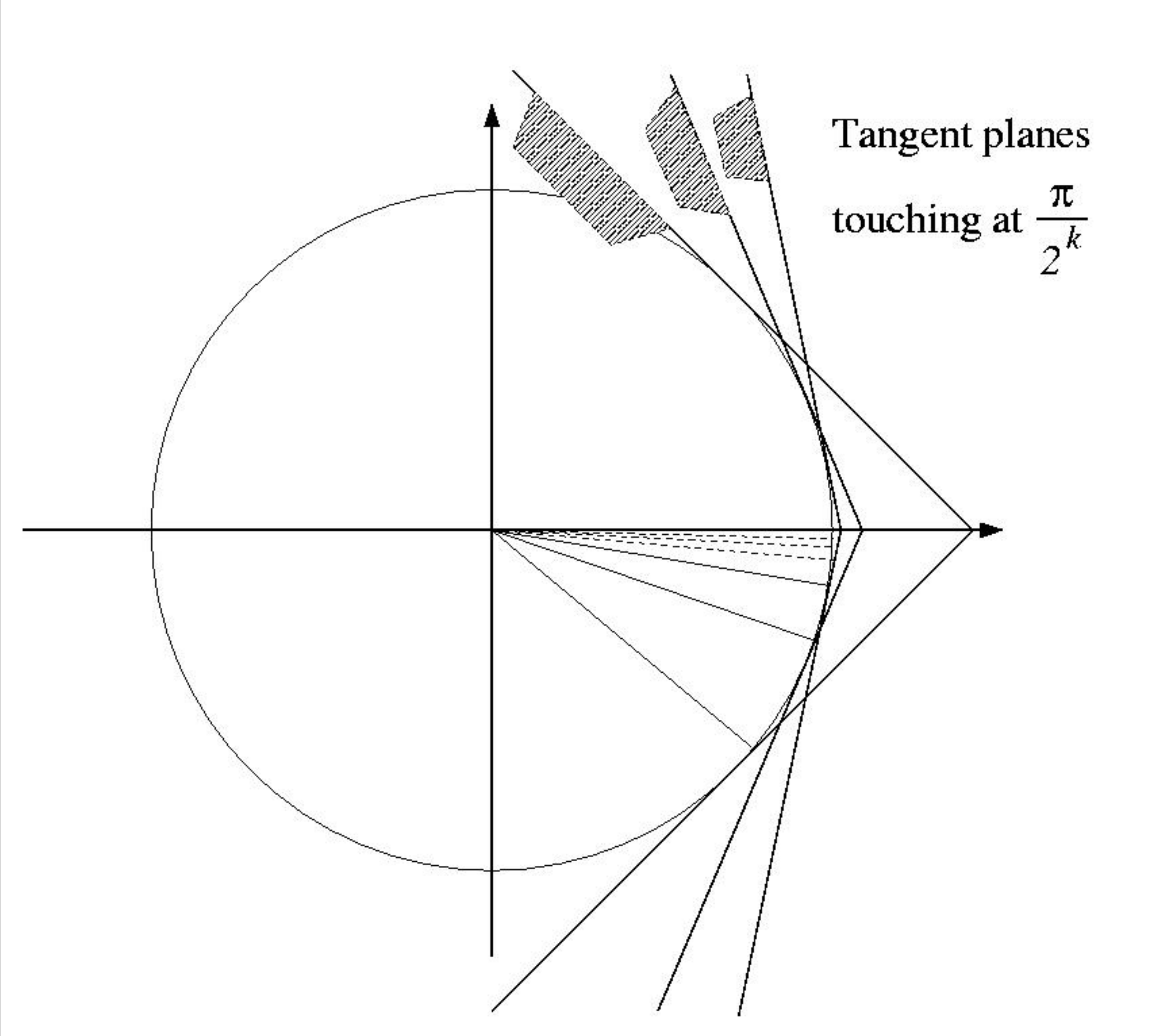
Proof does not use the central path or interior-point techniques.

5) Upper semicontinuity of optimal solutions for more general cone programs was established by Robinson (1982).

Here we consider special (linear semidefinite) cone programs.

There is a simple example that this theorem does not hold for strictly complementary solutions of more general cone programs:

Example



Maximize x_1 subject to these (infinitely many) linear constraints. Add the redundant constraint $x_1 \leq 1$. (All other constraints are 'facet defining'.)

Then the optimal solution is unique, strictly complementary, but the only active constraint is the redundant constraint $x_1 \leq 1$.

If the objective gradient $(1, 0)^T$ is changed a bit, the optimal solution jumps between the "vertices" close to $(1, 0)^T$. In particular, it is not differentiable.

From this set form a closed convex cone in $\mathbb{R}^3$ to have a conic program.

Corollary

Any step $\bar{X} + t\dot{X}$ for $t \neq 0$ is (typically) infeasible in the sense that $\bar{X} + t\dot{X} \not\succeq 0$. In some applications the following formula for the second directional derivative $\ddot{X} := \frac{1}{2} D^2_{\mathcal{D}} \bar{X}(\mathcal{D})[\Delta\mathcal{D}, \Delta\mathcal{D}]$ may be useful:

$$\mathcal{A}(\ddot{X}) = -\Delta\mathcal{A}(\dot{X}),$$

$$\mathcal{A}^*(\ddot{y}) + \ddot{S} = -\Delta\mathcal{A}^*(\dot{y}),$$

$$\ddot{X}\bar{S} + \bar{X}\ddot{S} = -\dot{X}\dot{S}.$$

This is the same system matrix as for the first derivative with different right hand side.

Note:

When $\bar{X} = U\Lambda U^T$ where the diagonal matrix Λ has a leading nonzero diagonal block Λ_1 as in the preceeding proof, then $\dot{X}$ has the following structure:

$$\dot{X} = U \begin{pmatrix} A & B \\ B^T & 0 \end{pmatrix} U^T,$$

and $\ddot{X}$ has the structure

$$\ddot{X} = U \begin{pmatrix} * & * \\ * & B^T\Lambda_1^{-1}B \end{pmatrix} U^T.$$

Setting $* = 0$ yields a minimum norm second order correction towards the positive semidefinite cone maintaining the multiplicity of the zero eigenvalue (up to third order terms).

Nonlinear SDP's

Many important applications:

Truss design with buckling constraints

(Ben Tal, J., Nemirovski, Kocvara, Zowe)

Robust optimization

(Ben Tal and Nemirovski)

Passivity of Pade-approximations in circuit design, positive real lemma.

(Freund and J.)

Applications in stochastics, ...

Often obtained but sometimes "nearly unsolvable".

Circuit design (talk of R.W. Freund)

The physical behavior of a (planned) VLSI design is given by a system of the form

$$\begin{aligned} E\dot{x} &= Ax + Bu \\ y &= C^T x + Du \end{aligned}$$

The system is passive (does not generate energy). Generalized version of the positive real lemma for descriptor systems (Freund and J. 2003):

Under mild conditions there exists a matrix P such that

$$\begin{pmatrix} PA + A^T P & PB - C \\ B^T P - C^T & -D - D^T \end{pmatrix} \preceq 0$$

and

$$E^T P = P^T E \succeq 0.$$

In applications, A and E are not known exactly. Instead some Pade-approximation $\tilde{A}, \tilde{E}$ of A, E is given with an optimal order of approximation. However, the system with $\tilde{A}, \tilde{E}$ is typically not passive.

To maintain a high order of approximation, a structured low-rank perturbation $A(x), E(x)$ of $\tilde{A}$ and $\tilde{E}$ is searched for such that

$$\exists\, P: \qquad Z(x,P) := \begin{pmatrix} PA(x) + A(x)^T P & PB - C \\ B^T P - C^T & -D - D^T \end{pmatrix} \preceq 0$$

and

$$E(x)^T P = P^T E(x) \succeq 0.$$

This leads to the nonlinear SDP:

$$\operatorname*{minimize}_{x,P} \; \{\lambda_{max}(Z(x,P)) \;\mid\; \text{subject to } E(x)^T P = P^T E(x) \succeq 0\}.$$

First Approach:

A Predictor Corrector Barrier Approach

• using ellipsoidal trust regions of variing shape and radii

(considerably more expensive than standard trust region steps)

Make cheaper subproblems once the overall method converges rapidly.

• and adaptive step length control in predictor step (depending on the difficulty for approximately solving the previous corrector step).

This framework had worked reliably for any convex problem tried so far.

Good results for ramdomly generated BMIs (Fukuda and Kojima).

Devastating number of iterations for BMIs from VLSI design.

Nonlinear semidefinite program:

$$\text{minimize}_{x \in \mathbb{R}^n} \quad b^T x \quad \text{subject to} \quad \begin{aligned} \mathcal{B}(x) &\preceq 0, \\ c(x) &\leq 0, \\ d(x) &= 0. \end{aligned}$$

For this talk: Omit $c(x)$ and $d(x)$.

$$\begin{aligned} \mathcal{L}(x,Y) &:= & b^T x + \mathcal{B}(x) \bullet Y, \\ g(x,Y) &:= & \nabla_x \mathcal{L}(x,Y) = b + \nabla_x \left(\mathcal{B}(x) \bullet Y\right), \\ H(x,Y) &:= & \nabla_x^2 \mathcal{L}(x,Y) = \nabla_x^2 \left(\mathcal{B}(x) \bullet Y\right). \end{aligned}$$

Notation

We assume that the nonlinear function $\mathcal{B} : I\!R^n \to \mathcal{S}^m$ is at least $\mathcal{C}^2$-differentiable and denote by

$$B^{(i)}(x) := \frac{\partial}{\partial x_i}\mathcal{B}(x) \quad \text{and} \quad B^{(i,j)}(x) := \frac{\partial^2}{\partial x_i \partial x_j}\mathcal{B}(x), \quad i,j = 1,2,\dots,n,$$

the first and second partial derivatives of $\mathcal{B}$, respectively. For each $x \in I\!R^n$, the derivative $D_x\mathcal{B}$ at x induces a linear function $D_x\mathcal{B}(x) : I\!R^n \to \mathcal{S}^m$, which is given by

$$D_x\mathcal{B}(x)[\Delta x] := \sum_{i=1}^{n}(\Delta x)_i B^{(i)}(x) \in \mathcal{S}^m \quad \text{for all} \quad \Delta x \in I\!R^n.$$

In particular,

$$\mathcal{B}(x + \Delta x) \approx \mathcal{B}(x) + D_x\mathcal{B}(x)[\Delta x], \quad \Delta x \in I\!R^n,$$

is the linearization of $\mathcal{B}$ at the point x. For any linear function $\mathcal{A} : I\!R^n \to \mathcal{S}^m$, we have

$$D_x\mathcal{A}(x)[\Delta x] = \mathcal{A}(\Delta x) \quad \text{for all} \quad x,\, \Delta x \in I\!R^n.$$

Notation (continued)

For any fixed matrix $Y \in \mathcal{S}^m$, the map $x \mapsto \mathcal{B}(x) \bullet Y$ is a scalar-valued function of $x \in I\!R^n$. Its gradient at x is given by

$$\nabla_x \left(\mathcal{B}(x) \bullet Y \right) = \left(D_x \left(\mathcal{B}(x) \bullet Y \right) \right)^T = \begin{pmatrix} B^{(1)}(x) \bullet Y \\ \vdots \\ B^{(n)}(x) \bullet Y \end{pmatrix} \in I\!R^n$$

and its Hessian by

$$\nabla_x^2 \left(\mathcal{B}(x) \bullet Y \right) = \begin{pmatrix} B^{(1,1)}(x) \bullet Y & \cdots & B^{(1,n)}(x) \bullet Y \\ \vdots & & \vdots \\ B^{(n,1)}(x) \bullet Y & \cdots & B^{(n,n)}(x) \bullet Y \end{pmatrix} \in \mathcal{S}^n.$$

In particular, for any linear function $\mathcal{A} : I\!R^n \to \mathcal{S}^m$, we have

$$\nabla_x \left(\mathcal{A}(x) \bullet Y \right) = \mathcal{A}^*(Y).$$

Optimality condition:

$$\mathcal{B}(x) + S = 0,$$

$$g(x, Y) = 0,$$

$$YS = 0, \leftarrow \mu I$$

$$Y,\ S \succeq 0.$$

Linearization of the optimality condition (which is a "nonsingular" condition):

Find S, Δx and ΔY such that

$$\mathcal{B}(x^k) + D_x\mathcal{B}(x^k)[\Delta x] + S = 0,$$

$$b + H^k \Delta x + \nabla_x(\mathcal{B}(x^k) \bullet (Y^k + \Delta Y)) = 0,$$

$$(Y^k + \Delta Y^k)S = 0, \qquad \text{bilinear}$$

$$Y^k + \Delta Y,\ S \succeq 0. \qquad \text{nonlinear inequalities}$$

Equivalent quadratic semidefinite program

$$\text{minimize} \quad b^T \Delta x + \frac{1}{2}(\Delta x)^T H^k \Delta x$$

$$\text{subject to} \quad \Delta x \in I\!R^n : \quad \mathcal{B}(x^k) + D_x \mathcal{B}(x^k)[\Delta x] \preceq 0.$$

If H is positive semidefinite

the quadratic objective function can be rewritten as

- a convex quadratic constraint in plain primal methods,
- a second order constraint (SEDUMI, Jos Sturm).

Let $H^k = LL^T$ (not necessarily triangular or square L).

Then, $x^T H^k x \leq t \Longleftrightarrow y^T y \leq t \cdot 1, \quad y = L^T x$

(rotated second order cone constraint in SeDuMi-notation, linear constraints).

Introducing slack variables, we obtain a linear conic subproblem:

$$\begin{array}{rl} \text{minimize} & \tilde{c}^T \Delta \tilde{x} \\ \text{subject to} & \Delta \tilde{x} \in \mathbb{R}^{2n+k+2} : \quad \mathcal{A}[\Delta \tilde{x}] = \tilde{b}, \qquad \tilde{x} \in \mathcal{K} \end{array}$$

where $\mathcal{K}$ is the cartesian product of free variables and self-dual cones.

• Need positive semidefinite approximation H^k of the Hessian of the Lagrangian (in contrast to standard nonlinear SQP-methods, where *nonconvex* QPs can be solved for a suitable local solution.)

"SLCP-method $\approx$ SQP-method with conic self-dual constraints and the additional restriction for a semidefinite approximation of the Hessian."

Local quadratic convergence

for unique, strictly complementary local solutions that satisfy a second order growth condition using exact Hessians.

Outline of Proof:

1) Straightforward (but tedious) generalization of sensitivity analysis for linear SDP to nonconvex quadratic SDP and to general nonlinear SDP.

2) Implicit function theorem. #

Sensitivity analysis also in Bonnans and Shapiro, 2000, with more general concept, transversality condition, inf sup compactness condition.

Local analysis and comparison with Augmented Lagrangian:

Fares, Noll, Apkarian (2000)

In Step 1): Nonconvex quadratic SDP

$$\text{minimize} \quad \left\{ b^T x + \frac{1}{2} x^T H x \mid \mathcal{A}(x) + C \preceq 0 \right\}.$$

Here, $\mathcal{A} : \mathbb{R}^n \to \mathcal{S}^m$ is a linear operator;

and the data is $\mathcal{D} := [\mathcal{A}, b, C, H]$.

Assumptions:

There exists x with $\mathcal{A}(x) + C \prec 0$.

There exists a locally unique strictly complementary solution $\bar{x}(\mathcal{D}), \bar{Y}(\mathcal{D})\bar{S}(\mathcal{D})$,

$$\begin{aligned} \mathcal{A}(\bar{x}) + C + \bar{S} &= 0, \quad \bar{S} \succeq 0, \\ b + H\bar{x} + \mathcal{A}^*(\bar{Y}) &= 0, \quad \bar{Y} \succeq 0, \\ \bar{Y}\bar{S} &= 0, \quad \bar{X} + \bar{S} \succ 0. \end{aligned}$$

$\bar{x}(\mathcal{D}), \bar{Y}(\mathcal{D})\bar{S}(\mathcal{D})$ satisfies a second order growth condition: There exists some $\mu > 0$ such that for any feasible direction h with $h^T(b + H\bar{x}) = 0$ (i.e. where the objective function is approximately constant) the inequality

$$h^T H h \geq \mu \|h\|^2$$

is true. (Note that the Hessian of the Lagrangian $\mathcal{L}^q(x, Y) := b^T x + \frac{1}{2} x^T H x + (\mathcal{A}(x) + C) \bullet Y$ coincides with the Hessian H of the objective function since the constraints are linear.)

Theorem

If the data $\mathcal{D}$ is changed by sufficiently small perturbations

$$\Delta\mathcal{D} = [\Delta\mathcal{A}, \Delta b, \Delta C, \Delta H],$$

then the optimal solutions of the perturbed semidefinite programs are differentiable functions of the perturbations. Furthermore, the derivatives

$$\dot{x} := D_{\mathcal{D}}\bar{x}[\Delta\mathcal{D}], \quad \dot{Y} := D_{\mathcal{D}}\bar{Y}[\Delta\mathcal{D}], \text{ and } \dot{S} := D_{\mathcal{D}}\bar{S}[\Delta\mathcal{D}],$$

of the solution x, Y, S at $\bar{x}$, $\bar{Y}$, $\bar{S}$ satisfy

$$\mathcal{A}(\dot{x}) + \dot{S} = -\Delta C - \Delta\mathcal{A}(\bar{x}),$$

$$H\dot{x} + \mathcal{A}^*(\dot{Y}) = -\Delta b - \Delta H\bar{x} - \Delta\mathcal{A}^*(\bar{Y}),$$

$$\dot{Y}\bar{S} + \bar{Y}\dot{S} = 0.$$

Consequence

Since the necessary and sufficient optimality conditions of a general nonlinear SDP, are precisely the same as those for the associated quadratic SDP, (with H given by the Hessian of the Lagrangian) this sensitivity result generalizes.

Under uniqueness, strict complementarity, and second order growth condition it can be used to show

• that the size of the "residual" in the SSP sub-problem is of the same order as the distance to optimality,

• that the SSP step is also of this size,

• and – as the residual after the SSP step is just the linearization error – that the residual after the SSP step therefore must be "squared".

Practical refinements (C. Vogelbusch, in progress)

- BFGS updates for H^k based on augmented Lagrangian

$$\Lambda(x,Y;r) = b^T x + \frac{r}{2}\left(\left(B(x)+\frac{Y}{r}\right)^+\right)^2 \bullet I - \frac{Y \bullet Y}{2r}.$$

(When $M := B(x) + \frac{Y}{r}$ is invertible, the first derivative of $((M^+)^2)$ is given by $(L_{M_+})^2 L_{|M|}^{-1}$ where $L_M(H) = MH + HM$ is the Lyapunov operator.)

- Second order correction (to avoid Maratos effect).

- Filter approach.

Open Questions

• Theoretical rate of convergence when using the exact (or an approximate) Hessian of the augmented Lagrangian.

• Choice of the penalty parameter.

• Sparsity (so far dense Hessian) — may need to include limited memory BFGS for Hessian of Lagrangian and different SDP-solver.

(This is completely open, so far work with 40×40 systems is possible, but still in debugging stage.)

• Reformulation of nonlinear SDP to less ill-conditioned form?

Concluding Remark

• Sensitivity result for nonlinear semidefinite programs.

• This result can be used to derive an elementary and self-contained proof of local quadratic convergence of the sequential linear conic programming (SLCP) method (a generalization of the well-known SQP method).

• For interior methods that are applied directly to nonlinear semidefinite programs, the choice of the symmetrization procedure is considerably more complicated than in the linear case since the system matrix is no longer positive semidefinite.

• In the SLCP method, the choice of the symmetrization scheme is shifted to the subproblems and is thus separated from the linearization.