ESI Summer Institute, Nonlinear Methods in Combinatorial Optimization

Florian Jarre,

Univ. Düsseldorf

Klagenfurt, Aug. 20 – Sept. 3, 2010

Elementary Optimality Conditions for Nonlinear SDPs

- Motivation
- First order conditions
- An example
- Second order conditions
- An example

Motivation

Low-rank approach to solve a large scale linear SDP
(to moderate accuracy):

$\Longrightarrow$ Nonlinear semidefinite program.

Quite different forms of degeneracies compared to “usual” nonlinear programs.

Relation to nonlinear programming

Interior-point methods:
Nonlinear programming methods (Newton and homotopy) to solve linear programs or linear semidefinite programs.

Here:
Look at optimality conditions.

NLSDPs and NLPs

Nonlinear semidefinite program $(NLSDP)$:

$$\text{minimize } f(x) \quad | \quad \begin{aligned} F(x) &= 0, \\ G(x) &\preceq 0. \end{aligned}$$

$f : I\!R^n \to I\!R$, $F : I\!R^n \to I\!R^m$, $G : I\!R^n \to \mathcal{S}^d$ continuously differentiable.

Nonlinear program (NLP):

$$\text{minimize } f(x) \quad | \quad \begin{aligned} F(x) &= 0, \\ G(x) &\leq 0, \end{aligned}$$

with componentwise inequalities $G(x) \leq 0$.

(Redundancies – symmetric multipliers)

Notation

The derivatives of f and F at a point x:
Row vector $Df(x)$ and the $m \times n$ Jacobian matrix $DF(x)$.

Derivative of G at a point x is a linear map
$DG(x) : \mathbb{R}^n \to \mathcal{S}^d$.
Applying the linear map $DG(x)$ to a vector Δx:
$DG(x)[\Delta x] \in \mathcal{S}^d$, or

$$DG(x)[\Delta x] = \sum_{i=1}^{n} \Delta x_i G_i(x) \quad \text{where} \quad G_i(x) := \frac{\partial}{\partial x_i} G(x) \in \mathcal{S}^d.$$

Characteristic equation: $G(x + \Delta x) \approx G(x) + DG(x)[\Delta x]$.

Linearized subproblems

Let some point $\bar{x}$ be given.
"Linearized semidefinite programming problem" $(LSDP)$:

$$\text{minimize} \quad f(\bar{x}) + Df(\bar{x})\Delta x \quad \Big| \quad \begin{aligned} F(\bar{x}) + DF(\bar{x})\Delta x &= 0, \\ G(\bar{x}) + DG(\bar{x})[\Delta x] &\preceq 0. \end{aligned}$$

Likewise (LP), linearized problem for (NLP)

$$\text{minimize} \quad f(\bar{x}) + Df(\bar{x})\Delta x \quad \Big| \quad \begin{aligned} F(\bar{x}) + DF(\bar{x})\Delta x &= 0, \\ G(\bar{x}) + DG(\bar{x})[\Delta x] &\leq 0. \end{aligned}$$

Tangential and linearized cones

Feasible set of $(NLSDP)$: $\mathcal{F}_1$
Tangential cone of $\mathcal{F}_1$ at a point $\bar{x} \in \mathcal{F}_1$:

$$\mathcal{T}_1 := \{\Delta x \mid \exists s^k \to \Delta x, \ \exists \alpha_k > 0, \ \alpha_k \to 0 \ \textbf{s.t.} \ \bar{x} + \alpha_k s^k \in \mathcal{F}_1\}.$$

If $\bar{x}$ is a minimizer of $(NLSDP)$ then $Df(\bar{x})\Delta x \geq 0 \ \forall \Delta x \in \mathcal{T}_1$.

Tangential cone of $(LSDP)$ at $\overline{\Delta x} = 0$:

"linearized cone" $\mathcal{L}_1$

If $\bar{x}$ is a minimizer of $(LSDP)$ then $Df(\bar{x})\Delta x \geq 0 \ \forall \Delta x \in \mathcal{L}_1$.

Back to (NLP)

Feasible set of (NLP): $\mathcal{F}_2$
Tangential cone at a point $\bar{x} \in \mathcal{F}_2$: $\mathcal{T}_2$.
$\mathcal{L}_2$ the “linearized cone” of (NLP) at a point $\bar{x}$.

$$\begin{aligned} \mathcal{L}_2 = \{\Delta x \mid\ & F(\bar{x}) + DF(\bar{x})\Delta x = 0, \\ & G_{k,l}(\bar{x}) + (DG(\bar{x}))_{k,l}\Delta x \leq 0 \\ & \quad \text{for all } k, l \text{ with } G_{k,l}(\bar{x}) = 0\}. \end{aligned}$$

In nonlinear optimization: $\mathcal{T}_2 \subset \mathcal{L}_2$ always holds true.

Idea of Proof:

$$0 = F(\bar{x} + \alpha_k s^k) = F(\bar{x}) + \alpha_k DF(\bar{x}) s^k + o(\alpha_k).$$

Dividing by $\alpha_k > 0$, using $F(\bar{x}) = 0$ yields

$$0 = DF(\bar{x}) s^k + o(1),$$

Taking the limit as $k \to \infty$ yields $0 = DF(\bar{x}) \Delta x$
(Equality constraints in the definition of $\mathcal{L}_2$)

Same argument for active inequalities.

When the MFCQ-constraint qualification

$DF(\bar{x})$ has linearly independent rows

$\exists \Delta x$ such that $DF(\bar{x})\Delta x = 0$ and

$$(DG(\bar{x}))_{k,l}\Delta x < 0 \quad \text{for all} \quad (k,l) \text{ with } (G(\bar{x}))_{k,l} = 0,$$

is satisfied then, $\mathcal{L}_2 = \mathcal{T}_2$.

In particular,

$\mathcal{T}_2$ and $\mathcal{L}_2$ are polyhedral.

Is $\mathcal{T}_1$ also a polyhedral?

"Should" be so, since $G(x) \preceq 0$ iff all principal submatrices of "$-G(x)$" are nonnegative.

But: NO! (in general).

The determinant does not satisfy MFCQ when zero is an eigenvalue of multiplicity more than one.

Note:

Let some $X \succeq 0$ be given:

$$X = \begin{bmatrix} U^{(1)} & U^{(2)} \end{bmatrix} \begin{bmatrix} D^{(1)} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} (U^{(1)})^T \\ (U^{(2)})^T \end{bmatrix}$$

with $D^{(1)} \succ 0$.

Tangential cone of $\mathcal{S}_+^d$ at X: All matrices of the form

$$W = \begin{bmatrix} U^{(1)} & U^{(2)} \end{bmatrix} \begin{bmatrix} * & * \\ * & \tilde{W}^{(2)} \end{bmatrix} \begin{bmatrix} (U^{(1)})^T \\ (U^{(2)})^T \end{bmatrix}$$

where $\tilde{W}^{(2)} \succeq 0$ and "$*$" can be anything.

Find a representation of $\mathcal{T}_1$

Let $\bar{x} \in \mathcal{F}_1$ and $\alpha_k > 0$ with $\alpha_k \to 0$, and $s^k \to \Delta x$, such that $F(\bar{x} + \alpha_k s^k) = 0$ and $G(\bar{x} + \alpha_k s^k) \preceq 0$ for all k.

As before, $DF(\bar{x})\Delta x = 0$.

Let

$$G(\bar{x}) = U \Lambda U^T = \begin{bmatrix} U^{(1)} & U^{(2)} \end{bmatrix} \begin{bmatrix} \Lambda^{(1)} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} (U^{(1)})^T \\ (U^{(2)})^T \end{bmatrix}$$

where $\Lambda^{(1)} \prec 0$.

Keep U fixed and define $\tilde{G}(x) := U^T G(x) U$.

Partition

$$\tilde{G}(x) := \begin{bmatrix} \tilde{G}^{(1)}(x) & \tilde{G}^{(1,2)}(x) \\ \tilde{G}^{(1,2)}(x)^T & \tilde{G}^{(2)}(x) \end{bmatrix}.$$

So, $\tilde{G}^{(1)}(\bar{x}) = \Lambda^{(1)} \prec 0$.

Since $\tilde{G}(\bar{x} + \alpha_k s^k) \preceq 0$:

$$0 \succeq \tilde{G}^{(2)}(\bar{x} + \alpha_k s^k) - \underbrace{\tilde{G}^{(2)}(\bar{x})}_{=0} \approx \alpha_k D\tilde{G}^{(2)}(\bar{x})[s^k].$$

Taking the limit as $k \to \infty$,

$$D\tilde{G}^{(2)}(\bar{x})[\Delta x] \preceq 0,$$

i.e.,

$$\mathcal{T}_1 \subset \{\Delta x \mid DF(\bar{x})\Delta x = 0, \quad D\tilde{G}^{(2)}(\bar{x})[\Delta x] \preceq 0\}.$$

Example

For $x \in \mathbb{R}$ let

$$G(x) := \begin{bmatrix} -2 & x \\ x & -x^2 \end{bmatrix}.$$

Then, $G(x) \preceq 0$ for all $x \in \mathbb{R}$ and thus, $\mathcal{F}_1 = \mathbb{R}$, and thus, also $\mathcal{T}_1 = \mathbb{R}$, while the feasible set of $(LSDP)$ at $\bar{x} = 0$ is given by

$$\left\{ \Delta x \mid \begin{bmatrix} -2 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & \Delta x \\ \Delta x & 0 \end{bmatrix} \preceq 0 \right\} = \{0\}.$$

Here, $\mathcal{T}_1 \not\subset \mathcal{L}_1$.

A constraint qualification (MFCQ)

$DF(\bar{x})$ has linearly independent rows, and $\exists\ d$ s.t.

$DF(\bar{x})d = 0$ and

$G(\bar{x}) + DG(\bar{x})[d] \prec 0.$

Lemma: If MFCQ is satisfied, then

$$\mathcal{L}_1 = \mathcal{T}_1 = \{\Delta x \mid DF(\bar{x})\Delta x = 0, \quad D\tilde{G}^{(2)}(\bar{x})[\Delta x] \preceq 0\}.$$

Thus, the tangential cone of $(NLSPD)$ is not polyhedral, in general.

KKT conditions

Lemma: Let $\bar{x}$ be a local minimizer of $(NLSDP)$ and let $(NLSDP)$ be regular at $\bar{x}$ in the sense of MFCQ. Then there exists a matrix $\bar{Y} \in \mathcal{S}^d$ and a vector $\bar{y} \in \mathbb{R}^m$ such that

$$Df(\bar{x})^T + DF(\bar{x})^T \bar{y} + \begin{bmatrix} G_1(\bar{x}) \bullet \bar{Y} \\ \vdots \\ G_m(\bar{x}) \bullet \bar{Y} \end{bmatrix} = 0 \quad \text{and} \quad G(\bar{x}) \bullet \bar{Y} = 0.$$

Lagrangian function

$$L(x, y, Y) := f(x) + F(x)^T y + G(x) \bullet Y$$

with $Y \in \mathcal{S}^d_+$.
Lagrangian for (NLP) and $(NLSDP)$ is identical.

Second order conditions

D^2G at a point $x \in I\!R^n$:

$$D^2G(x)[\Delta x, \Delta x] = \sum_{i,j=1}^{n} \Delta x_i \Delta x_j G_{i,j}(x) \in \mathcal{S}^d$$

with $G_{i,j}(x) := \frac{\partial^2}{\partial x_i \partial x_j} G(x) \in \mathcal{S}^d$.
For $Y \in \mathcal{S}^d$:

$$Y \bullet D^2G(x)[\Delta x, \Delta x] = \sum_{i,j=1}^{n} \Delta x_i \Delta x_j (Y \bullet G_{i,j}(x)) = \Delta x^T H \Delta x,$$

where $H = H(x, Y)$ is the symmetric $n \times n$-matrix with matrix entries $Y \bullet G_{i,j}(x)$.

Cone of critical directions

Let $\bar{x}$ be a local minimizer of $(NLSDP)$ and let the MFCQ be satisfied at $\bar{x}$.
Then, the KKT conditions state that $Df(\bar{x})\Delta x \geq 0$ for all $\Delta x \in \mathcal{T}_1$.

The cone of critical directions is given by

$$\mathcal{C}_1 := \{\Delta x \in \mathcal{T}_1 \mid Df(\bar{x})\Delta x = 0\}.$$

For (NLP) the “critical cone” depends on three conditions:

$$DF(\bar{x})\Delta x = 0,$$
$$(DG(\bar{x}))_{k,l}\Delta x = 0 \quad \text{for all} \quad (k,l) \text{ with } \bar{Y}_{k,l} > 0$$
$$(DG(\bar{x}))_{k,l}\Delta x \leq 0 \quad \text{for all} \quad (k,l) \text{ with } (G(\bar{x}))_{k,l} = 0,\ \bar{Y}_{k,l} = 0.$$

If we assume strict complementarity, then

$$\mathcal{C}_2 = \{\Delta x \mid \text{the first two conditions hold}\}.$$

For $(NLSDP)$ we also assume uniqueness of the multipliers and strict complementarity, $G(\bar{x}) - \bar{Y} \prec 0$.
The cone of critical directions at $\bar{x}$ is then

$$\mathcal{C}_1 = \{\Delta x \mid DF(\bar{x})\Delta x = 0, \quad \underbrace{(U^{(2)})^T DG(\bar{x})[\Delta x] U^{(2)}}_{D\tilde{G}^{(2)}(\bar{x})[\Delta x]} = 0\}.$$

Let $U^{(2)}$ have q columns, i.e. $U^{(2)}$ is a $d \times q$ matrix.
$\mathcal{C}_1$ is the tangent cone of the following boundary manifold of the feasible set

$$\mathcal{F}_1^{bd} := \{x \mid F(x) = 0, \quad \text{rank}(G(x)) = d - q\}$$

at $\bar{x}$.

The condition $G(x)$ has rank $d-q$ translates to the condition that the Schur complement

$$\tilde{G}^{(2)}(x) - \tilde{G}^{(1,2)}(x)^T(\tilde{G}^{(1)}(x))^{-1}\tilde{G}^{(1,2)}(x) \equiv 0$$

equals zero (for small $\|x-\bar{x}\|$).

Regularity of $\mathcal{F}_1^{bd}$ is equivalent to linear independence of the $q(q+1)/2$ gradients of $\tilde{G}^{(2)}$ and of the m gradients of $F_\nu(x)$ at $\bar{x}$, ($1 \leq \nu \leq m$).

A second example

For (NLP), if $\bar{x}$ is a local minimizer satisfying LICQ, then the Hessian of the Lagrangian is positive semidefinite on $\mathcal{C}_2$.
Example by Diehl et.al. (2006):

$$\text{minimize} \quad -x_1^2 - (1-x_2)^2 \quad | \quad \begin{bmatrix} -1 & x_1 & x_2 \\ x_1 & -1 & 0 \\ x_2 & 0 & -1 \end{bmatrix} \preceq 0.$$

The semidefiniteness constraint is satisfied if, and only if, $x_1^2 + x_2^2 \leq 1$.
Global optimal solution: $\bar{x} = (0, -1)^T$.

Semidefiniteness constraint is linear; hence,
$D^2_{xx}L(\bar{x}, \bar{y}) = D^2 f(\bar{x}) \prec 0$.
Cone of critical directions $(x_1, 0)^T$ with $x_1 \in I\!R$.

Implications:
Slow convergence for sequential semidefinite programming algorithms that use subproblems with a *convex* quadratic objective function and linearized semidefiniteness constraints.

Contrast to standard SQP methods!

Second order conditions

Lemma: Let $\bar{x}$ be a local minimizer of $(NLSDP)$ and let $\bar{x}, \bar{y}, \bar{Y}$ be a strictly complementary KKT-point. Assume that the set $\mathcal{F}_1^{bd}$ satisfies MFCQ. Then

$$h^T(D_x^2 L(\bar{x}, \bar{y}, \bar{Y}) + \mathcal{H}(\bar{x}, \bar{Y}))h \geq 0 \qquad \forall\ h \in \mathcal{C}_1,$$

where $\mathcal{H}(\bar{x}, \bar{Y}) \succeq 0$ is a matrix depending on the curvature of the semidefinite cone at $G(\bar{x})$ and the directional derivatives of G at $\bar{x}$, and is given by its matrix entries

$$(\mathcal{H}(\bar{x}, \bar{Y}))_{i,j} := -2\bar{Y} \bullet G_i(\bar{x}) G(\bar{x})^{\dagger} G_j(\bar{x})^T.$$

The converse direction also holds under the additional assumption of regularity of $\mathcal{F}_1^{bd}$:

Lemma: Let $\bar{x}$ be a strictly complementary KKT-point of $(NLSDP)$ and assume that

$$h^T(D_x^2 L(\bar{x}, \bar{y}, \bar{Y}) + \mathcal{H}(\bar{x}, \bar{Y}))h > 0 \qquad \forall\, h \in \mathcal{C}_1.$$

Then, $\bar{x}$ is a strict local minimizer of $(NLSDP)$ that satisfies the second order growth condition, $\exists \epsilon > 0,\ \delta > 0:$

$$f(x+s) \geq f(x) + \epsilon \|s\|^2 \quad \forall\, \|s\| \leq \delta \text{ with } x + s \in \mathcal{F}_1.$$

For (NLP) the term $D_x^2 L(\bar{x}, \bar{y}, \bar{Y}) + \mathcal{H}(\bar{x}, \bar{Y})$ is replaced with $D_x^2 L(\bar{x}, \bar{y}, \bar{Y})$.

This yields a stronger second order sufficient condition: $D_x^2 L(\bar{x}, \bar{y}, \bar{Y})$ be positive definite on $\mathcal{C}_1$, see Robinson (1982).

The weaker form above due to Shapiro (1997) explains the “negative example” that complicates the local convergence of sequential semidefinite programming algorithms.

References

F. Bonnans and A. Shapiro, Perturbation Analysis of Optimization Problems, Springer-Verlag, New York, (2000).

Jarre, F., Elementary Optimality Conditions for Nonlinear SDPs, download at: Optimization Online (2010).

Shapiro, A.: First and second order analysis of nonlinear semidefinite programs. Math. Prog. **77**, 301–320 (1997).