Squared slack variables in nonlinear symmetric cone programming: Optimality conditions and augmented Lagrangians

Bruno Figueira Lourenço
Joint work with Prof. Ellen H. Fukuda (Kyoto U.) and Prof. Masao Fukushima (Nanzan U.)

Seikei University

August 12th, 2016

Our setting

Consider the following program.

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g(x) \in \mathcal{K}, \end{aligned} \tag{NSCP}$$

where:

- $f : \mathbb{R}^n \to \mathbb{R}$ is a twice differentiable function,
- $g : \mathbb{R}^n \to \mathcal{E}$ is also a twice differentiable function,
- $\mathcal{K} \subset \mathcal{E}$ is a symmetric cone.

Our setting

Consider the following program.

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & g(x) \in \mathcal{K}, \end{array} \tag{NSCP}$$

where:

- $f : \mathbb{R}^n \to \mathbb{R}$ is a twice differentiable function,
- $g : \mathbb{R}^n \to \mathcal{E}$ is also a twice differentiable function,
- $\mathcal{K} \subset \mathcal{E}$ is a symmetric cone. (say, the PSD matrices $\mathcal{S}^n_+$, a second-order cone $\mathcal{Q}^n$)

Our setting

Consider the following program.

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & g(x) \in \mathcal{K}, \end{array} \qquad \text{(NSCP)}$$

where:

- $f : \mathbb{R}^n \to \mathbb{R}$ is a twice differentiable function,
- $g : \mathbb{R}^n \to \mathcal{E}$ is also a twice differentiable function,
- $\mathcal{K} \subset \mathcal{E}$ is a symmetric cone. (say, the PSD matrices $\mathcal{S}^n_+$, a second-order cone $\mathcal{Q}^n$)

Objectives

We want to understand optimality conditions for (NSCP).

What is hard about for this problem?

In the classical nonlinear programming (NLP) world...

What is hard about for this problem?

In the classical nonlinear programming (NLP) world...

- Second Order Necessary conditions: $d^T \nabla^2 \mathcal{L}(x, \lambda) d \geq 0$, for $d \in \mathcal{C}$.

What is hard about for this problem?

In the classical nonlinear programming (NLP) world...

- Second Order Necessary conditions: $d^T \nabla^2 \mathcal{L}(x, \lambda) d \geq 0$, for $d \in \mathcal{C}$.
- Second Order Sufficient conditions: $d^T \nabla^2 \mathcal{L}(x, \lambda) d > 0$, for $d \in \mathcal{C}, d \neq 0$.

What is hard about for this problem?

In the classical nonlinear programming (NLP) world...

- Second Order Necessary conditions: $d^T \nabla^2 \mathcal{L}(x, \lambda) d \geq 0$, for $d \in \mathcal{C}$.
- Second Order Sufficient conditions: $d^T \nabla^2 \mathcal{L}(x, \lambda) d > 0$, for $d \in \mathcal{C}, d \neq 0$.

What is hard about for this problem?

In the classical nonlinear programming (NLP) world...

- Second Order Necessary conditions: $d^T \nabla^2 \mathcal{L}(x, \lambda) d \geq 0$, for $d \in \mathcal{C}$.
- Second Order Sufficient conditions: $d^T \nabla^2 \mathcal{L}(x, \lambda) d > 0$, for $d \in \mathcal{C}, d \neq 0$.

In the conic nonlinear programming (NCLP) world...

What is hard about for this problem?

In the classical nonlinear programming (NLP) world...

- Second Order Necessary conditions: $d^T \nabla^2 \mathcal{L}(x, \lambda) d \geq 0$, for $d \in \mathcal{C}$.
- Second Order Sufficient conditions: $d^T \nabla^2 \mathcal{L}(x, \lambda) d > 0$, for $d \in \mathcal{C}, d \neq 0$.

In the conic nonlinear programming (NCLP) world...

- $d^\top \nabla^2 L(x, \lambda) d \geq 0$, for $d \in \mathcal{C}$ is **NOT** a necessary condition for local min.
- $d^\top \nabla^2 L(x, \lambda) d > 0$, for $d \in \mathcal{C}, d \neq 0$ is too strong.

What are the right conditions?

In the conic nonlinear programming (NCLP) world...

What are the right conditions?

In the conic nonlinear programming (NCLP) world...

- Second Order Necessary conditions: $d^\top(\nabla^2 L(x,\lambda) + \sigma(x,\lambda))d \geq 0$, for $d \in \mathcal{C}$.
- Second Order Sufficient conditions: $d^\top(\nabla^2 L(x,\lambda)d + \sigma(x,\lambda)) > 0$, for $d \in \mathcal{C}, d \neq 0$,

What are the right conditions?

In the conic nonlinear programming (NCLP) world...

- Second Order Necessary conditions: $d^\top(\nabla^2 L(x,\lambda) + \sigma(x,\lambda))d \geq 0$, for $d \in \mathcal{C}$.
- Second Order Sufficient conditions: $d^\top(\nabla^2 L(x,\lambda)d + \sigma(x,\lambda)) > 0$, for $d \in \mathcal{C}, d \neq 0$,

where $\sigma(x,\lambda)$ is a cumbersome term.

What are the right conditions?

In the conic nonlinear programming (NCLP) world...

- Second Order Necessary conditions: $d^\top(\nabla^2 L(x,\lambda) + \sigma(x,\lambda))d \geq 0$, for $d \in \mathcal{C}$.
- Second Order Sufficient conditions: $d^\top(\nabla^2 L(x,\lambda)d + \sigma(x,\lambda)) > 0$, for $d \in \mathcal{C}, d \neq 0$,

where $\sigma(x,\lambda)$ is a cumbersome term.

- See Kawasaki (1988), Cominetti (1990), Bonnans and Shapiro (2000).

What are the right conditions?

In the conic nonlinear programming (NCLP) world...

- Second Order Necessary conditions: $d^\top(\nabla^2 L(x,\lambda) + \sigma(x,\lambda))d \geq 0$, for $d \in \mathcal{C}$.
- Second Order Sufficient conditions: $d^\top(\nabla^2 L(x,\lambda)d + \sigma(x,\lambda)) > 0$, for $d \in \mathcal{C}, d \neq 0$,

where $\sigma(x,\lambda)$ is a cumbersome term.

- See Kawasaki (1988), Cominetti (1990), Bonnans and Shapiro (2000).
- $\sigma = 0$, if $\mathcal{K}$ is polyhedral.

What are the right conditions?

In the conic nonlinear programming (NCLP) world...

- Second Order Necessary conditions: $d^\top(\nabla^2 L(x,\lambda) + \sigma(x,\lambda))d \geq 0$, for $d \in \mathcal{C}$.
- Second Order Sufficient conditions: $d^\top(\nabla^2 L(x,\lambda)d + \sigma(x,\lambda)) > 0$, for $d \in \mathcal{C}, d \neq 0$,

where $\sigma(x,\lambda)$ is a cumbersome term.

- See Kawasaki (1988), Cominetti (1990), Bonnans and Shapiro (2000).
- $\sigma = 0$, if $\mathcal{K}$ is polyhedral.
- σ is known if $\mathcal{K} = \mathcal{S}^n_+$ (Shapiro, 1997).

What are the right conditions?

In the conic nonlinear programming (NCLP) world...

- Second Order Necessary conditions: $d^\top(\nabla^2 L(x,\lambda) + \sigma(x,\lambda))d \geq 0$, for $d \in \mathcal{C}$.
- Second Order Sufficient conditions: $d^\top(\nabla^2 L(x,\lambda)d + \sigma(x,\lambda)) > 0$, for $d \in \mathcal{C}, d \neq 0$,

where $\sigma(x,\lambda)$ is a cumbersome term.

- See Kawasaki (1988), Cominetti (1990), Bonnans and Shapiro (2000).
- $\sigma = 0$, if $\mathcal{K}$ is polyhedral.
- σ is known if $\mathcal{K} = \mathcal{S}^n_+$ (Shapiro, 1997).
- σ is known if $\mathcal{K} = \mathcal{Q}^n = \{(x_0, \overline{x}) \mid x_0 \geq \|\overline{x}\|_2\}$. (Bonnans and Ramírez, 2005).

What are the right conditions?

In the conic nonlinear programming (NCLP) world...

- Second Order Necessary conditions: $d^\top(\nabla^2 L(x,\lambda) + \sigma(x,\lambda))d \geq 0$, for $d \in \mathcal{C}$.
- Second Order Sufficient conditions: $d^\top(\nabla^2 L(x,\lambda)d + \sigma(x,\lambda)) > 0$, for $d \in \mathcal{C}, d \neq 0$,

where $\sigma(x,\lambda)$ is a cumbersome term.

- See Kawasaki (1988), Cominetti (1990), Bonnans and Shapiro (2000).
- $\sigma = 0$, if $\mathcal{K}$ is polyhedral.
- σ is known if $\mathcal{K} = \mathcal{S}^n_+$ (Shapiro, 1997).
- σ is known if $\mathcal{K} = \mathcal{Q}^n = \{(x_0, \overline{x}) \mid x_0 \geq \|\overline{x}\|_2\}$. (Bonnans and Ramírez, 2005).
- A concrete formula for σ is not known if $\mathcal{K}$ is an arbitrary symmetric cone.

What are the right conditions?

In the conic nonlinear programming (NCLP) world...

- Second Order Necessary conditions: $d^\top(\nabla^2 L(x,\lambda) + \sigma(x,\lambda))d \geq 0$, for $d \in \mathcal{C}$.
- Second Order Sufficient conditions: $d^\top(\nabla^2 L(x,\lambda)d + \sigma(x,\lambda)) > 0$, for $d \in \mathcal{C}, d \neq 0$,

where $\sigma(x,\lambda)$ is a cumbersome term.

- See Kawasaki (1988), Cominetti (1990), Bonnans and Shapiro (2000).
- $\sigma = 0$, if $\mathcal{K}$ is polyhedral.
- σ is known if $\mathcal{K} = \mathcal{S}^n_+$ (Shapiro, 1997).
- σ is known if $\mathcal{K} = \mathcal{Q}^n = \{(x_0, \overline{x}) \mid x_0 \geq \|\overline{x}\|_2\}$. (Bonnans and Ramírez, 2005).
- A concrete formula for σ is not known if $\mathcal{K}$ is an arbitrary symmetric cone.
- See also (Forsgreen,2010) and (Jarre, 2012).

Slack variables approach

$$\begin{aligned} \underset{x,y}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g(x) - y \circ y = 0. \end{aligned} \qquad \text{(Slack)}$$

where $\circ$ is a Jordan Product:

1. $y \circ z = y \circ z$,
2. $\langle y \circ z, w \rangle = \langle y, z \circ w \rangle$,
3. $y \circ (y^2 \circ z) = y^2 \circ (y \circ z)$,
4. $\mathcal{K} = \{y \circ y \mid y \in \mathcal{E}\}$

- (Slack) is a run-of-the-mill nonlinear program.
- Optimality conditions for (Slack) are much easier.
- There are many solvers for (Slack) but not many for (NSCP).

Objectives 2

- Examine the difference between optimality conditions and regularity conditions for (NSCP) and (Slack).
- Check the computational prospects of the slack variables approach.

KKT conditions

Let $L(x,\lambda) := f(x) - \langle g(x), \lambda \rangle$.
(x,λ) is a KKT pair for (NSCP) if

$$\nabla_x L(x,\lambda) = \nabla f(x) - \nabla g(x)^* \lambda = 0, \qquad \text{(P1.1)}$$
$$\lambda \in \mathcal{K}, \qquad \text{(P1.2)}$$
$$g(x) \in \mathcal{K}, \qquad \text{(P1.3)}$$
$$\lambda \circ g(x) = 0, \qquad \text{(P1.4)}$$

KKT conditions

Let $L(x, \lambda) := f(x) - \langle g(x), \lambda \rangle$.
(x, λ) is a KKT pair for (NSCP) if

$$\nabla_x L(x, \lambda) = \nabla f(x) - \nabla g(x)^* \lambda = 0, \tag{P1.1}$$
$$\lambda \in \mathcal{K}, \tag{P1.2}$$
$$g(x) \in \mathcal{K}, \tag{P1.3}$$
$$\lambda \circ g(x) = 0, \tag{P1.4}$$

(x, y, λ) is a KKT triple for (Slack) if:

$$\nabla f(x) - \nabla g(x)^* \lambda = 0, \tag{P2.1}$$
$$\lambda \circ y = 0, \tag{P2.2}$$
$$g(x) - y \circ y = 0. \tag{P2.3}$$

λ is arbitrary.

Regularity Conditions

For (NSCP) we have:

- Mangasarian-Fromovitz: if there exists some d such that

$$g(x) + \nabla g(x)d \in \operatorname{int} \mathcal{K},$$

- Nondegeneracy:

$$\mathcal{E} = T_{\mathcal{K}}(g(x)) + \operatorname{Im} \nabla g(x), \qquad \text{(Nondegeneracy)}$$

- Suppose (x, λ) is KKT for (NSCP) and

$$\operatorname{rank} g(x) + \operatorname{rank} \lambda = m, \qquad (1)$$

  then (x, λ) is said to satisfy the *strict complementarity condition*.

For (Slack) we have the "linear independence constraint qualification" (LICQ).

Second order sufficient condition for (Slack)

The second order sufficient condition for (SOSC-NLP) holds if

$$\langle \nabla_x^2 \mathcal{L}(x, y, \lambda)(v, w), (v, w) \rangle > 0,$$

for every $(v, w) \in \mathbb{R}^n \times \mathcal{E}$ such that $\nabla g(x)v - 2y \circ w = 0$, where $\mathcal{L}$ is the Lagrangian of (Slack).

Proposition

Let $(x, y, \lambda) \in \mathbb{R}^n \times \mathcal{E} \times \mathcal{E}$ be KKT for (Slack)*. The second-order sufficient condition (SOSC-NLP) holds if*

$$\langle \nabla_x^2 L(x, \lambda)v, v \rangle + 2\langle w \circ w, \lambda \rangle > 0 \qquad (2)$$

for every non-zero $(v, w) \in \mathbb{R}^n \times \mathcal{E}$ such that $\nabla g(x)v - 2y \circ w = 0$.

L is the Lagrangian of (NSCP).

Second order necessary condition for (Slack)

Proposition

Let (x, y) *be local min for* (Slack) *and* $(x, y, \lambda) \in \mathbb{R}^n \times \mathcal{E} \times \mathcal{E}$ *be KKT such that LICQ holds. Then the second order necessary condition holds (SONC-NLP):*

$$\langle \nabla_x^2 L(x, \lambda) v, v \rangle + 2\langle w \circ w, \lambda \rangle \geq 0 \quad (3)$$

for every $(v, w) \in \mathbb{R}^n \times \mathcal{E}$ *such that* $\nabla g(x) v - 2y \circ w = 0$.

Relation between KKT points and CQs

(NSCP)		(Slack)

Relation between KKT points and CQs

(NSCP)		(Slack)
(x, λ) is a KKT pair	$\Rightarrow$	$\exists y$ such that (x, y, λ) is a KKT triple

Relation between KKT points and CQs

(NSCP)		(Slack)
(x, λ) is a KKT pair	$\Rightarrow$	$\exists y$ such that (x, y, λ) is a KKT triple
(x, λ) is a KKT pair	$\not\Leftarrow$	(x, y, λ) is a KKT triple

Relation between KKT points and CQs

(NSCP)		(Slack)
(x, λ) is a KKT pair	$\Rightarrow$	$\exists y$ such that (x, y, λ) is a KKT triple
(x, λ) is a KKT pair	$\not\Leftarrow$	(x, y, λ) is a KKT triple
(x, λ) is a KKT pair	$\Longleftarrow$	(x, y, λ) is a KKT triple $+ \lambda \in \mathcal{K}$

Relation between KKT points and CQs

(NSCP)		(Slack)
(x, λ) is a KKT pair	$\Rightarrow$	$\exists y$ such that (x, y, λ) is a KKT triple
(x, λ) is a KKT pair	$\not\Leftarrow$	(x, y, λ) is a KKT triple
(x, λ) is a KKT pair	$\Longleftarrow$	(x, y, λ) is a KKT triple $+ \lambda \in \mathcal{K}$
(x, λ) is a KKT pair + strict complementarity	$\Longleftarrow$	(x, y, λ) is a KKT triple + SOSC-NLP

Relation between KKT points and CQs

(NSCP)		(Slack)
(x, λ) is a KKT pair	$\Rightarrow$	$\exists y$ such that (x, y, λ) is a KKT triple
(x, λ) is a KKT pair	$\not\Leftarrow$	(x, y, λ) is a KKT triple
(x, λ) is a KKT pair	$\Longleftarrow$	(x, y, λ) is a KKT triple $+ \lambda \in \mathcal{K}$
(x, λ) is a KKT pair + strict complementarity	$\Longleftarrow$	(x, y, λ) is a KKT triple + SOSC-NLP

(NSCP)		(Slack)

Relation between KKT points and CQs

(NSCP)		(Slack)
(x, λ) is a KKT pair	$\Rightarrow$	$\exists y$ such that (x, y, λ) is a KKT triple
(x, λ) is a KKT pair	$\not\Leftarrow$	(x, y, λ) is a KKT triple
(x, λ) is a KKT pair	$\Longleftarrow$	(x, y, λ) is a KKT triple $+ \lambda \in \mathcal{K}$
(x, λ) is a KKT pair + strict complementarity	$\Longleftarrow$	(x, y, λ) is a KKT triple + SOSC-NLP

(NSCP)		(Slack)
x satisfies nondegeneracy	$\Longleftrightarrow$	$\exists y$ such that (x, y) satisfies LICQ

Relations between second order sufficient conditions

Suppose $\mathcal{K} = \mathcal{S}^n_+$ or a direct product of second order cones $\mathcal{Q}^n$.

(NSCP)		(Slack)

Relations between second order sufficient conditions

Suppose $\mathcal{K} = \mathcal{S}^n_+$ or a direct product of second order cones $\mathcal{Q}^n$.

(NSCP)		(Slack)
(x, λ) is a KKT pair + strict complementarity + SOSC-NSCP	$\iff$	$\exists y$ such that (x, y, λ) a KKT triple + SOSC-NLP

Relations between second order sufficient conditions

Suppose $\mathcal{K} = \mathcal{S}^n_+$ or a direct product of second order cones $\mathcal{Q}^n$.

(NSCP)		(Slack)
(x, λ) is a KKT pair + strict complementarity + SOSC-NSCP	$\iff$	$\exists y$ such that (x, y, λ) a KKT triple + SOSC-NLP

Proposition (A Sufficient Condition via Slack Variables)

Suppose

1. (x, λ) *is KKT pair for* (NSCP).
2. *Strict complementarity.*
3. *The following inequality is satisfied:*

$$\langle \nabla^2_x L(x, \lambda) v, v \rangle + 2\langle w \circ w, \lambda \rangle > 0 \quad (4)$$

for every non-zero $(v, w) \in \mathbb{R}^n \times \mathcal{E}$ *such that* $\nabla g(x) v - 2\sqrt{g(x)} \circ w = 0$.

Then, x *is a local minimum for* (NSCP).

Relations between second order necessary conditions

Suppose $\mathcal{K} = \mathcal{S}^n_+$ or a direct product of second order cones $\mathcal{Q}^n$.

(NSCP)		(Slack)

Relations between second order necessary conditions

Suppose $\mathcal{K} = \mathcal{S}^n_+$ or a direct product of second order cones $\mathcal{Q}^n$.

(NSCP)		(Slack)
(x, λ) is a KKT pair + strict complementarity + nondegeneracy + SOSC-NSCP	$\Longleftrightarrow$	$\exists y$ such that (x, y, λ) a KKT triple + LICQ + SOSC-NLP

Relations between second order necessary conditions

Suppose $\mathcal{K} = \mathcal{S}^n_+$ or a direct product of second order cones $\mathcal{Q}^n$.

(NSCP)		(Slack)
(x, λ) is a KKT pair + strict complementarity + nondegeneracy + SOSC-NSCP	$\Longleftrightarrow$	$\exists y$ such that (x, y, λ) a KKT triple + LICQ + SOSC-NLP

For any symmetric cone $\mathcal{K}$ we have.

Proposition (A Necessary Condition via Slack Variables)

Suppose x is a local min and

1. (x, λ) *is KKT for* (NSCP)
2. *Strict complementarity.*
3. *Nondegeneracy.*

Then:

$$\langle \nabla^2_x L(x, \lambda)v, v\rangle + 2\langle w \circ w, \lambda\rangle \geq 0 \tag{5}$$

for every (v, w) *such that* $\nabla g(x)v - 2\sqrt{g(x)} \circ w = 0$.

Sharp characterization of positive semidefiniteness

- $\mathcal{S}^m$: $m \times m$ symmetric matrices.
- $y \circ \lambda = \frac{y\lambda + \lambda y}{2}$.

Lemma (L.,Fukuda,Fukushima,2016)

Let $\lambda \in \mathcal{S}^m$. The following statements are equivalent:

i. $\lambda \succeq 0$,

ii. *there exists $y \in \mathcal{S}^m$ such that $y \circ \lambda = 0$ and*

$$\langle w \circ w, \lambda \rangle > 0, \tag{6}$$

for every nonzero $w \in \mathcal{S}^m$ which satisfies $y \circ w = 0$.

For any y satisfying (6) we have $\operatorname{rank} \lambda = m - \operatorname{rank} y$. Moreover, if σ and σ' are nonzero eigenvalues of y, then $\sigma + \sigma' \neq 0$.

Extension to symmetric cones

Suppose $\mathcal{E}$ is a Jordan algebra and $\mathcal{K} = \mathcal{K}^1 \times \ldots \times \mathcal{K}^s$.

Lemma

Let $\lambda \in \mathcal{E}$. The following statements are equivalent:

i. $\lambda \in \mathcal{K}$,

ii. *there exists $y \in \mathcal{E}$ such that*

$$y \circ \lambda = 0 \text{ and } \langle w \circ w, \lambda \rangle > 0, \tag{7}$$

for every nonzero $w \in \mathcal{E}$ which satisfies $y \circ w = 0$.

Suppose y satisfies Any y satisfying (7) we have that $\operatorname{rank} \lambda = m - \operatorname{rank} y$. *Moreover, if σ and σ' are nonzero eigenvalues of* **the same block** *of y, then $\sigma + \sigma' \neq 0$.*

Is it feasible to use (Slack)?

- There are previous results for linear SDPs by Burer and Monteiro in 2003.
- we used PENLAB to solve the same problems via both (NSCP) and (Slack).

Hock-Schittkowski problem 71

$$
\begin{array}{ll}
\underset{x\in\mathbb{R}^6}{\text{minimize}} & x_1x_4(x_1+x_2+x_3)+x_3 \\
\text{subject to} & x_1x_2x_3x_4 - x_5 - 25 = 0, \\
& x_1^2+x_2^2+x_3^2+x_4^4 - x_6 - 40 = 0, \\
& \begin{pmatrix} x_1 & x_2 & 0 & 0 \\ x_2 & x_4 & x_2+x_3 & 0 \\ 0 & x_2+x_3 & x_4 & x_3 \\ 0 & 0 & x_3 & x_1 \end{pmatrix} \in \mathcal{S}^4_+, \\
& 1 \le x_i \le 5,\ i=1,2,3,4; \quad x_i \ge 0,\ i=5,6.
\end{array} \tag{HS}
$$

Table: Slack vs "native" for (HS)

	functions	gradients	Hessians	iterations	time (s)	opt. value
slack	110	57	44	13	0.54	87.7105
native	123	71	58	13	0.57	87.7105

The closest correlation matrix problem - simple version

Let H be a $m \times m$ symmetric matrix H with diagonal entries equal to one.

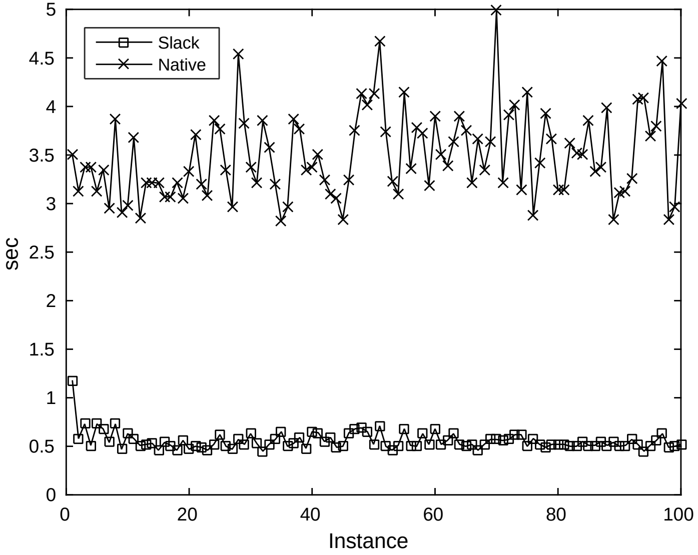
$$\begin{array}{ll} \underset{x}{\text{minimize}} & \langle X - H, X - H \rangle \\ \text{subject to} & X_{ii} = 1 \qquad \forall i, \\ & X \in \mathcal{K}. \end{array} \tag{Cor}$$

Generated 100 symmetric matrices H such that the diagonal elements are all 1 and other elements are uniform random numbers between -1 and 1.

Table: Comparison between native and slack

	Slack			Native		
m	mean (s)	min (s)	max (s)	mean (s)	min (s)	max (s)
5	0.090	0.060	0.140	0.201	0.130	0.250
10	0.153	0.120	0.230	0.423	0.330	0.630
15	0.287	0.210	0.430	1.306	1.020	1.950
20	0.556	0.450	1.180	3.491	2.820	4.990

Case-by-case comparison (m = 20)



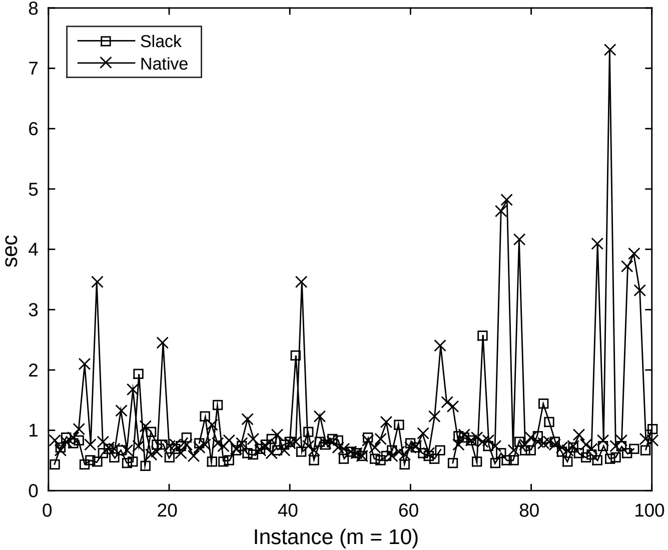
The closest correlation matrix problem - extended version

$$\begin{aligned} \underset{X,z}{\text{minimize}} \quad & \langle zX - H, zX - H \rangle \\ \text{subject to} \quad & zX_{ii} = 1 \qquad \forall i, \\ & I_m \preceq X \preceq \kappa I_m, \end{aligned} \tag{Cor-Ext}$$

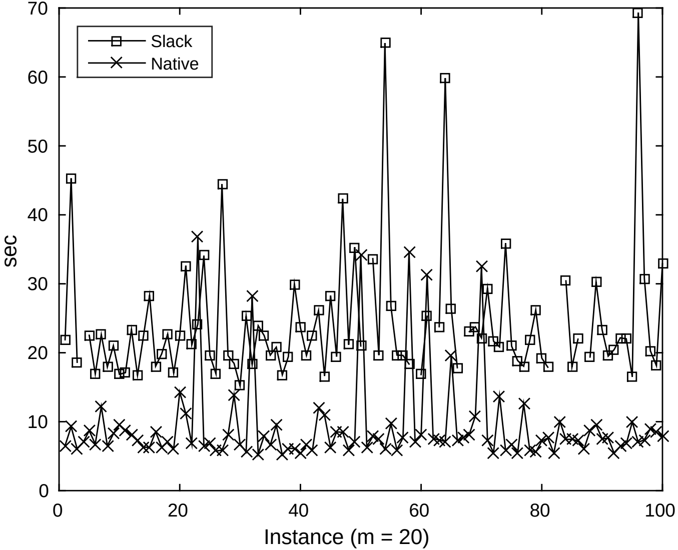
Table: Comparison between Native and Slack

	Slack				Native			
m	mean (s)	min (s)	max (s)	fail	mean (s)	min (s)	max (s)	fail
5	0.236	0.130	0.830	15	0.445	0.250	2.130	1
10	0.741	0.420	2.580	3	1.206	0.580	7.300	0
15	4.651	2.090	26.96	15	3.809	1.960	14.12	0
20	24.32	15.20	69.34	8	9.288	5.150	36.81	0

Case-by-case comparison (m = 10)



Case-by-case comparison (m = 20)



We have the augmented Lagrangian

$$\mathfrak{L}_\rho^{\mathsf{Slack}}(x, y, \lambda) = f(x) - \langle g(x) - y \circ y, \lambda \rangle + \frac{\rho}{2}\|g(x) - y \circ y\|^2. \tag{AL-SLACK}$$

$$\min_y \mathfrak{L}_\rho^{\mathsf{Slack}}(x, y, \lambda) = f(x) + \frac{1}{2\rho}\left(-\|\lambda\|^2 + \||\lambda - \rho g(x)|_+\|^2\right),$$

where $|x|_+$ is the orthogonal projection of x on $\mathcal{K}$. This suggests this:

$$\mathfrak{L}_\rho^{\mathsf{Sym}}(x, \lambda) = f(x) + \frac{1}{2\rho}\left(-\|\lambda\|^2 + \||\lambda - \rho g(x)|_+\|^2\right) \tag{AL-CONE}$$

We have the augmented Lagrangian

$$\mathfrak{L}_\rho^{\text{Slack}}(x, y, \lambda) = f(x) - \langle g(x) - y \circ y, \lambda \rangle + \frac{\rho}{2}\|g(x) - y \circ y\|^2. \tag{AL-SLACK}$$

$$\min_y \mathfrak{L}_\rho^{\text{Slack}}(x, y, \lambda) = f(x) + \frac{1}{2\rho}\left(-\|\lambda\|^2 + \||\lambda - \rho g(x)|_+\|^2\right),$$

We have the augmented Lagrangian

$$\mathfrak{L}_{\rho}^{\mathrm{Slack}}(x, y, \lambda) = f(x) - \langle g(x) - y \circ y, \lambda \rangle + \frac{\rho}{2} \| g(x) - y \circ y \|^2. \tag{AL-SLACK}$$

$$\min_{y} \mathfrak{L}_{\rho}^{\mathrm{Slack}}(x, y, \lambda) = f(x) + \frac{1}{2\rho} \left(-\|\lambda\|^2 + \| |\lambda - \rho g(x)|_+ \|^2 \right),$$

where $|x|_+$ is the orthogonal projection of x on $\mathcal{K}$. This suggests this:

$$\mathfrak{L}_{\rho}^{\mathrm{Sym}}(x, \lambda) = f(x) + \frac{1}{2\rho} \left(-\|\lambda\|^2 + \| |\lambda - \rho g(x)|_+ \|^2 \right) \tag{AL-CONE}$$

Augmented Lagrangian Method for (Slack)

1. Choose initial points x_1, y_1, initial multipliers (μ_1, λ_1) and an initial penalty ρ_1.
2. $k \leftarrow 1$.
3. $(x_{k+1}, y_{k+1}) \leftarrow \mathrm{argmin}_{x,y}\, \mathfrak{L}^{\mathsf{Slack}}_{\rho_k}(\cdot, \cdot, \lambda_k)$.
4. $\lambda_{k+1} \leftarrow \lambda_k - \rho_k(g(x_{k+1}) - y_{k+1} \circ y_{k+1})$.
5. Choose a new penalty parameter ρ_{k+1} with $\rho_{k+1} \geq \rho_{k+1}$.
6. Let $k \leftarrow k + 1$ and return to Step 3.

Augmented Lagrangian Method for (NSCP)

1 Choose an initial point x_1, initial multipliers (μ_1, λ_1) and an initial penalty ρ_1.
2 $k \leftarrow 1$.
3 $x_{k+1} \leftarrow \operatorname{argmin}_x \mathfrak{L}^{\mathsf{Sym}}_{\rho_k}(\cdot, \lambda_k)$.
4 $\mu_{k+1} \leftarrow \mu_k - \rho_k h(x_{k+1})$.
5 $\lambda_{k+1} \leftarrow |\lambda_k - \rho_k g(x_{k+1})|_+$.
6 Choose a new penalty parameter ρ_{k+1} with $\rho_{k+1} \geq \rho_{k+1}$.
7 Let $k \leftarrow k + 1$ and return to Step 3.

- See (Bertsekas, 1982) for the case $\mathcal{K} = \mathbb{R}^m_+$
- For the case $\mathcal{K} = \mathcal{S}^m_+$, see *"The rate of convergence of the augmented Lagrangian method for nonlinear semidefinite programming"* by Defeng Sun, Jie Sun and Liwei Zhang.

Conclusion

- Using slack we can import results from NLP theory,

Conclusion

- Using slack we can import results from NLP theory, if you are not unhappy about strict complementarity.
- May prove insight into the structure of $\mathcal{K}$ and into the design of algorithms.
- It might not be that bad for problems where the conic constraints are small.

See

- B. F. L., Ellen H. Fukuda and Masao Fukushima. *Optimality conditions for nonlinear semidefinite programming via squared slack variables*. `arxiv:1512.05507` To appear in Math Prog, 2016
- Ellen H. Fukuda and Masao Fukushima. *The Use of Squared Slack Variables in Nonlinear Second-Order Cone Programming*. *Optimization Online December 2015* and JOTA, August 2016.

Conclusion

- Using slack we can import results from NLP theory, if you are not unhappy about strict complementarity.
- May prove insight into the structure of $\mathcal{K}$ and into the design of algorithms.
- It might not be that bad for problems where the conic constraints are small.

See

- B. F. L., Ellen H. Fukuda and Masao Fukushima. *Optimality conditions for nonlinear semidefinite programming via squared slack variables*. `arxiv:1512.05507` To appear in Math Prog, 2016
- Ellen H. Fukuda and Masao Fukushima. *The Use of Squared Slack Variables in Nonlinear Second-Order Cone Programming*. *Optimization Online December 2015* and JOTA, August 2016.

Thank you!