数理解析研究所講究録
第 1981 巻 2016 年 143-148

A note on nonlinear semidefinite programming and the squared slack variables technique*

Ellen Hidemi Fukuda
Graduate School of Informatics, Kyoto University

Bruno Figueira Lourenço
Department of Mathematical and Computing Sciences, Tokyo Institute of Technology

Masao Fukushima
Faculty of Science and Engineering, Nanzan University

Abstract

In this work, we consider the squared slack variables approach for nonlinear semidefinite programming (NSDP) problems. We establish the relation between the Karush-Kuhn-Tucker points and the regularity conditions of the general NSDP problem and its reformulation via slack variables. We will observe that the key to establish the equivalence of the problems is the second-order sufficient condition.

Keywords: Nonlinear semidefinite programming, optimality conditions, squared slack variables, second-order sufficient conditions.

1 Introduction

We consider the following *nonlinear semidefinite programming* (NSDP) problem:

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & G(x) \in \mathcal{S}^m_+, \end{array} \tag{P1}$$

where $f\colon \mathbb{R}^n \to \mathbb{R}$ and $G\colon \mathbb{R}^n \to \mathcal{S}^m$ are twice continuously differentiable functions, $\mathcal{S}^m$ is the linear space of all real symmetric matrices of dimension $m \times m$, and $\mathcal{S}^m_+$ is the cone of all positive semidefinite matrices in $\mathcal{S}^m$. For simplicity, we restrict our discussion to problems without equality constraints. Moreover, it is important to observe that this paper is just a simple note concerning the relation between (P1) with its reformulation using squared slack variables. For more details, including proofs of all the results, we refer to [4].

*This work was supported by Grant-in-Aid for Young Scientists (B) (26730012) and for Scientific Research (C) (26330029) from Japan Society for the Promotion of Science.

In the *nonlinear programming* (NLP) context, it is well-known that the squared slack variables can easily convert inequality constraints into equality constraints. However, this strategy is hardly considered nowadays because it increases the dimension of the problem and may lead to numerical instabilities [5]. Recently, Fukuda and Fukushima [2] showed that the same approach can be done for *nonlinear second-order cone programming* (NSOCP) problems. In this case, the reformulated problem does not contain conic constraints, which is an advantage, if we consider second-order cones as objects that are difficult to deal with. In fact, although some methods for NSOCP exist, it is nothing compared to the NLP solvers, which tend to be more efficient and widely available. Similarly, comparing to the NLP case, there are few efficient methods for the general nonlinear (and nonconvex) semidefinite programming problems [7].

The squared slack variables approach can be used in the NSDP case because, like the nonnegative orthant and the second-order cone, the cone of positive semidefinite matrices is also a *cone of squares*. More precisely, $\mathcal{S}_+^m$ can be written as

$$\mathcal{S}_+^m = \{Z \circ Z \colon Z \in \mathcal{S}^m\},$$

where $\circ$ is the *Jordan product* associated with the space $\mathcal{S}^m$, which is defined as

$$W \circ Z := \frac{WZ + ZW}{2}$$

for any $W, Z \in \mathcal{S}^m$. Note that actually $Z \circ Z = Z^2$ for any $Z \in \mathcal{S}^m$.

Using the above fact, we can introduce a slack variable $Y \in \mathcal{S}^m$ in (P1), so we obtain the following reformulation:

$$\begin{array}{ll} \underset{x,Y}{\text{minimize}} & f(x) \\ \text{subject to} & G(x) - Y \circ Y = 0. \end{array} \tag{P2}$$

Note that it is an NLP problem with only equality constraints. Moreover, it is equivalent to (P1) in terms of optimal points. In fact, if $(x, Y) \in \mathbb{R}^n \times \mathcal{S}^m$ is a global (local) minimizer of (P2), then x is a global (local) minimizer of (P1). On the other hand, if $x \in \mathbb{R}^n$ is a global (local) minimizer of (P1), then there exists $Y \in \mathcal{S}^m$ such that (x, Y) is a global (local) minimizer of (P2). Here, we observe that in practice, we can only expect to compute stationary points, or *Karush-Kuhn-Tucker* (KKT) points of the problems. Because of this, it is more important to establish the relation between the KKT points, instead of optimal points, of (P1) and (P2), although in this case the relation is not so trivial.

2 Notations

Throughout the paper, the following notations will be used. For $x \in \mathbb{R}^s$ and $Y \in \mathbb{R}^{\ell \times s}$, $x_i \in \mathbb{R}$ and $Y_{ij} \in \mathbb{R}$ denote the ith entry of x and the (i, j) entry (ith row and jth column) of Y, respectively. The transpose, the Moore-Penrose pseudo-inverse, and the rank of $Y \in \mathbb{R}^{\ell \times s}$ are denoted by $Y^\top \in \mathbb{R}^{s \times \ell}$, $Y^\dagger \in \mathbb{R}^{s \times \ell}$, and $\operatorname{rank} Y$, respectively. If Y is a square matrix, its trace is denoted by $\mathrm{tr}(Y) := \sum_i Y_{ii}$. For square matrices W and Y of the same dimension, their inner product is denoted by $\langle W, Y\rangle := \mathrm{tr}(W^\top Y)$. We will also use the same notation $\langle\cdot,\cdot\rangle$ for the Euclidean inner product of two vectors. Moreover, given a matrix $Z \in \mathcal{S}^m_+$, we will denote by $\sqrt{Z}$ the positive semidefinite square root of Z, that is, $\sqrt{Z}$ satisfies $\sqrt{Z} \in \mathcal{S}^m_+$ and $\sqrt{Z}^2 = Z$.

For any linear operator $\mathcal{G}\colon \mathbb{R}^s \to \mathcal{S}^\ell$ defined by $\mathcal{G}v = \sum_{i=1}^s v_i \mathcal{G}_i$ with $\mathcal{G}_i \in \mathcal{S}^\ell$, $i = 1, \ldots, s$, and $v \in \mathbb{R}^s$, the adjoint operator $\mathcal{G}^*\colon \mathcal{S}^\ell \to \mathbb{R}^s$ is defined by

$$\mathcal{G}^* Z = (\langle \mathcal{G}_1, Z\rangle, \ldots, \langle \mathcal{G}_s, Z\rangle)^\top, \quad Z \in \mathcal{S}^\ell.$$

Given a mapping $\mathcal{H}\colon \mathbb{R}^s \to \mathcal{S}^\ell$, its derivative at a point $x \in \mathbb{R}^s$ is denoted by $\nabla\mathcal{H}(x)\colon \mathbb{R}^s \to \mathcal{S}^\ell$ and defined by

$$\nabla\mathcal{H}(x)v = \sum_{i=1}^s v_i \frac{\partial \mathcal{H}(x)}{\partial x_i}, \quad v \in \mathbb{R}^s,$$

where $\partial\mathcal{H}(x)/\partial x_i \in \mathcal{S}^\ell$ are the partial derivative matrices. Finally, for a closed convex cone $\mathcal{K}$, we will denote by $\mathrm{lin}\,\mathcal{K}$ the largest subspace contained in $\mathcal{K}$. Note that in this case $\mathrm{lin}\,\mathcal{K} = \mathcal{K} \cap -\mathcal{K}$.

3 KKT conditions

Let us first recall the KKT conditions of the NSDP problem (P1). Define the Lagrangian function $L\colon \mathbb{R}^n \times \mathcal{S}^m \to \mathbb{R}$ associated with problem (P1) as

$$L(x, \Lambda) := f(x) - \langle G(x), \Lambda\rangle,$$

where $\Lambda \in \mathcal{S}^m$ corresponds to the Lagrange multiplier. We say that $(x, \Lambda) \in \mathbb{R}^n \times \mathcal{S}^m$ is a KKT pair of problem (P1) if the following conditions are satisfied:

$$\begin{aligned} \nabla f(x) - \nabla G(x)^*\Lambda &= 0, \\ \Lambda &\in \mathcal{S}^m_+, \\ G(x) &\in \mathcal{S}^m_+, \\ \Lambda \circ G(x) &= 0, \end{aligned} \tag{1}$$

where we have $\nabla f(x) - \nabla G(x)^*\Lambda = \nabla_x L(x, \Lambda)$. It can be shown that the last equality can be replaced by $\langle \Lambda, G(x)\rangle = 0$ or $\Lambda G(x) = 0$.

Similarly, for the equality constrained NLP problem (P2), we define the Lagrangian function $\mathcal{L}\colon \mathbb{R}^n \times \mathcal{S}^m \times \mathcal{S}^m \to \mathbb{R}$ associated with it as

$$\mathcal{L}(x, Y, \Lambda) := f(x) - \langle G(x) - Y \circ Y, \Lambda\rangle.$$

So, $(x, Y, \Lambda) \in \mathbb{R}^n \times \mathcal{S}^m \times \mathcal{S}^m$ is a KKT triple of problem (P2) if the conditions below are satisfied:

$$\begin{aligned} \nabla_{(x,Y)}\mathcal{L}(x, Y, \Lambda) &= 0, \\ G(x) - Y \circ Y &= 0. \end{aligned}$$

Some calculations show that these conditions can be written as follows:

$$\begin{aligned} \nabla f(x) - \nabla G(x)^* \Lambda &= 0, \\ \Lambda \circ Y &= 0, \\ G(x) - Y \circ Y &= 0. \end{aligned}$$

Using the above definitions, we can prove the following result. Again, we refer to [4] for the details of the proofs.

Proposition 1. *Let $(x, \Lambda) \in \mathbb{R}^n \times \mathcal{S}^m$ be a KKT pair of problem* (P1). *Then, there exists $Y \in \mathcal{S}^m$ such that (x, Y, Λ) is a KKT triple of* (P2).

However, the converse implication is not always true. That is, even if (x, Y, Λ) is a KKT triple of (P2), (x, Λ) may fail to be a KKT pair of (P1), because the condition (1) may not hold, i.e., Λ is not necessarily positive semidefinite. This is actually the only obstacle for establishing the equivalence.

Proposition 2. *If $(x, Y, \Lambda) \in \mathbb{R}^n \times \mathcal{S}^m \times \mathcal{S}^m$ is a KKT triple of* (P2) *such that Λ is positive semidefinite, then (x, Λ) is a KKT pair of* (P1).

4 Equivalence between KKT conditions

Proposition 2 leads us to consider conditions which ensure that the multiplier Λ is positive semidefinite. One of these conditions follows from the lemma below, which gives a new characterization of positive semidefinite matrices, taking into account the rank information.

Lemma 3. *Let $\Lambda \in \mathcal{S}^m$. The following statements are equivalent:*

(a) $\Lambda \in \mathcal{S}^m_+$;

(b) There exists $Y \in \mathcal{S}^m$ such that $Y \circ \Lambda = 0$ and $Y \in \Phi(\Lambda)$, where

$$\Phi(\Lambda) := \big\{ Y \in \mathcal{S}^m \colon \langle W \circ W, \Lambda \rangle > 0 \text{ for all } 0 \neq W \in \mathcal{S}^m \text{ with } Y \circ W = 0 \big\}. \tag{2}$$

For any Y satisfying the conditions in (b), we have $\operatorname{rank} \Lambda = m - \operatorname{rank} Y$. *Moreover, if σ and $\hat{\sigma}$ are nonzero eigenvalues of Y, then $\sigma + \hat{\sigma} \neq 0$.*

Proposition 4. *Assume that $(x, Y, \Lambda) \in \mathbb{R}^n \times \mathcal{S}^m \times \mathcal{S}^m$ is a KKT triple of* (P2) *such that $Y \in \Phi(\Lambda)$, where $\Phi(\Lambda)$ is the set defined in* (2). *Then, (x, Λ) is a KKT pair of* (P1) *satisfying the* strict complementarity *condition, i.e.,*

$$\operatorname{rank} G(x) + \operatorname{rank} \Lambda = m.$$

The above result shows that the KKT pair of (P1) also satisfies the strict complementarity condition. Moreover, it can be shown that the second-order sufficient condition of (P2) can also guarantee the same result. In fact, the idea for the characterization of positive definiteness given in Lemma 3 comes from the second-order conditions of (P2).

Lemma 5. *Let $(x, Y, \Lambda) \in \mathbb{R}^n \times \mathcal{S}^m \times \mathcal{S}^m$ be a KKT triple of problem* (P2). *The* second-order sufficient condition *(SOSC-NLP) holds if*

$$\langle \nabla_x^2 L(x, \Lambda)v, v\rangle + 2\langle W \circ W, \Lambda\rangle > 0$$

for every nonzero $(v, W) \in \mathbb{R}^n \times \mathcal{S}^m$ such that $\nabla G(x)v - 2Y \circ W = 0$. Recall that $\nabla_x^2 L(x, \Lambda)$ is the Hessian of the Lagrangian function associated to problem (P1) *with respect to x.*

Corollary 6. *Assume that SOSC-NLP is satisfied at a KKT triple $(x, Y, \Lambda) \in \mathbb{R}^n \times \mathcal{S}^m \times \mathcal{S}^m$ of problem* (P2). *Then (x, Λ) is a KKT pair for* (P1) *which satisfies the strict complementarity condition.*

The converse implication is also true from Proposition 1, and under the strict complementarity condition.

Proposition 7. *Suppose that $(x, \Lambda) \in \mathbb{R}^n \times \mathcal{S}^m$ is a KKT pair of* (P1) *which satisfies the strict complementarity condition. Then there exists some $Y \in \Phi(\Lambda)$ such that (x, Y, Λ) is a KKT triple of* (P2).

5 Analysis of second-order sufficient conditions

In this section, we examine the relations between KKT points of problems (P1) and (P2) that satisfy second-order sufficient conditions. Before that, we also show that the so-called nondegeneracy condition of (P1) is equivalent to the *linear independence constraint qualification* (LICQ) of (P2).

Proposition 8. *If $(x, Y) \in \mathbb{R}^n \times \mathcal{S}^m$ satisfies LICQ for* (P2), *then x is* nondegenerate *for* (P1), *i.e., the following equality holds:*

$$\mathcal{S}^m = \operatorname{lin} \mathcal{T}_{\mathcal{S}_+^m}(G(x)) + \operatorname{Im} \nabla G(x),$$

where $\operatorname{Im} \nabla G(x)$ *denotes the image of the linear map $\nabla G(x)$, and $\mathcal{T}_{\mathcal{S}_+^m}(G(x))$ denotes the tangent cone of $\mathcal{S}_+^m$ at $G(x)$. On the other hand, if x satisfies nondegeneracy and if $Y = \sqrt{G(x)}$, then (x, Y) satisfies LICQ for* (P2).

Next, we will refine the result of Corollary 6, taking into account the second-order sufficient condition of (P1), which is defined below.

Definition 9. *Let $(x, \Lambda) \in \mathbb{R}^n \times \mathcal{S}^m$ be a KKT pair of problem* (P1). *Define $H(x, \Lambda) \in \mathcal{S}^m$ as a matrix with elements*

$$H(x, \Lambda)_{ij} := 2 \operatorname{tr}\left(\frac{\partial G(x)}{\partial x_i} G(x)^\dagger \frac{\partial G(x)}{\partial x_j} \Lambda \right)$$

for $i, j = 1, \ldots, n$. We say that the second-order sufficient condition *(SOSC-SDP) holds if*

$$\big\langle (\nabla_x^2 L(x, \Lambda) + H(x, \Lambda))d, d \big\rangle > 0$$

for all nonzero $d \in \mathbb{R}^n$ such that $\nabla G(x)d \in \mathcal{T}_{\mathcal{S}_+^m}(G(x))$ and $\langle \nabla f(x), d\rangle = 0$.

Theorem 10. *Assume that* $(x, Y, \Lambda) \in \mathbb{R}^n \times \mathcal{S}^m \times \mathcal{S}^m$ *is a KKT triple of* (P2) *satisfying SOSC-NLP. Then* (x, Λ) *is a KKT pair of* (P1) *that satisfies strict complementarity and SOSC-SDP.*

Returning to the converse implication, we have the following result from Proposition 7 and by using the second-order sufficient condition.

Theorem 11. *Assume that* $(x, \Lambda) \in \mathbb{R}^n \times \mathcal{S}^m$ *is a KKT pair for* (P1) *satisfying SOSC-SDP and the strict complementarity condition. Then, there exists* $Y \in \mathcal{S}^m$ *such that* (x, Y, Λ) *is a KKT triple for* (P2) *satisfying SOSC-NLP.*

6 Final remarks

We have analyzed the use of squared slack variables in the context of NSDP. We have shown that, under the second-order sufficient conditions and the regularity conditions, KKT points of the original and the reformulated problems are essentially equivalent. It is worth mentioning that the point of view considered in [4] is different from the one given here. In fact, in such a paper, the squared slack variables reformulation was used as a tool to derive second-order optimality conditions of NSDP problems. This point of view also makes sense if we consider important to find alternative ways for obtaining these not trivial conditions. See the starting work from Shapiro [6], and the subsequent ones from Forsgren [1] and Jarre [3].

References

[1] A. Forsgren. Optimality conditions for nonconvex semidefinite programming. *Mathematical Programming*, 88(1):105–128, 2000.

[2] E. H. Fukuda and M. Fukushima. The use of squared slack variables in nonlinear second-order cone programming. *Submitted*, 2014.

[3] F. Jarre. An interior method for nonconvex semidefinite programs. *Optimization and Engineering*, 1(4):347–372, 2000.

[4] B. F. Lourenço, E. H. Fukuda, and M. Fukushima. Optimality conditions for nonlinear semidefinite programming via squared slack variables. *Submitted*, 2015.

[5] S. M. Robinson. Stability theory for systems of inequalities, part II: differentiable nonlinear systems. *SIAM Journal on Numerical Analysis*, 13(4):497–513, 1976.

[6] A. Shapiro. First and second order analysis of nonlinear semidefinite programs. *Mathematical Programming*, 77(1):301–320, 1997.

[7] H. Yamashita and H. Yabe. A survey of numerical methods for nonlinear semidefinite programming. *Journal of the Operations Research Society of Japan*, 58(1):24–60, 2015.