The Use of Squared Slack Variables in Nonlinear Second-Order Cone Programming*

Ellen H. Fukuda† Masao Fukushima‡

December 15, 2015

Abstract

In traditional nonlinear programming, the technique of converting a problem with inequality constraints into a problem containing only equality constraints, by the addition of squared slack variables, is well-known. Unfortunately, it is considered to be an avoided technique in the optimization community, since the advantages usually do not compensate for the disadvantages, like the increase of the dimension of the problem, the numerical instabilities, and the singularities. However, in the context of nonlinear second-order cone programming, the situation changes, because the reformulated problem with squared slack variables has no longer conic constraints. This fact allows us to solve the problem by using a general-purpose nonlinear programming solver. The objective of this work is to establish the relation between Karush-Kuhn-Tucker points of the original and the reformulated problems by means of the second-order sufficient conditions and regularity conditions. We also present some preliminary numerical experiments.

Keywords: Karush-Kuhn-Tucker conditions, nonlinear second-order cone programming, second-order sufficient condition, slack variables.

1 Introduction

A well-known technique in constrained optimization is the introduction of slack variables, which converts inequality constraints into equality constraints. It is widely used in *linear programming*, in order to transform a polyhedron into an isomorphic one in standard form, i.e., defined by a system of equations in nonnegative variables [5, 10]. It is also used in many algorithms for *nonlinear programming* (NLP), such as the reduced-gradient-type methods like MINOS [18], the augmented Lagrangian-type methods like LANCELOT [9], and the

*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.

†Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan (`ellen@i.kyoto-u.ac.jp`).

‡Department of Systems and Mathematical Science, Faculty of Science and Engineering, Nanzan University, Nagoya 466-8673, Japan (`fuku@nanzan-u.ac.jp`).

Newton-type methods [21, 22, 23]. As a particular case, the squared slack variables may also be used in NLP, so that the nonnegativity constraints of the variables can be eliminated, as considered, for example, in [4, 17, 21, 22, 23]. However, this strategy is usually avoided in the optimization community, because it increases considerably the dimension of the problem and may lead to numerical instabilities or singularities [20].

The situation changes in *nonlinear second-order cone programming* (NSOCP). An NSOCP problem is an optimization problem with second-order (or Lorentz) cone constraints, where the involved functions are nonlinear. Various problems can be formulated as an NSOCP problem [1, 16]. Moreover, it can be viewed as an extension of the NLP problem and a particular case of the *nonlinear semidefinite programming* problem [1]. However, it is usually desirable to treat an NSOCP problem directly, instead of treating it as a semidefinite programming problem, in order to exploit its particular structure and, consequently, to avoid high computational costs. Although there exist many methods for the convex case where the involved functions are linear [1, 8, 16], the research for the general nonlinear, possibly nonconvex, case is relatively scarce and only a few methods have been developed [12, 13, 14, 15, 24].

Here we observe that the use of squared slack variables is more interesting and helpful for NSOCP problems than NLP problems. In fact, the reformulated problem with the additional squared slack variables is no longer an NSOCP problem, but only an NLP problem. This is important in practice, especially if one regards the second-order cone as an object that is not so easy to handle. Moreover, this fact indicates that a general-purpose NLP solver may be used to find a solution, or at least a stationary point, of the original NSOCP problem. One may still claim that the use of squared slack variables is not a good practice, but we believe that it is worth a try in many situations, because it is trivial to reformulate an NSOCP as an NLP using this technique, and since NLP solvers are widely available.

However, the aim of this work is neither to solve NSOCP problems using NLP solvers, nor to propose an efficient method for NSOCP problems. In fact, the contribution here is in the analysis between the original NSOCP problem and the reformulated NLP problem, in terms of Karush-Kuhn-Tucker (KKT) points. We will show that the second-order sufficient conditions are the key to establish the equivalence of the optimality conditions of both problems. Although the results are intuitive, especially when the original problem is an NLP, we will observe that the NSOCP case is difficult to prove. We will also solve numerically some instances of NSOCP problems, but this is merely for investigating the validity of the squared slack variables approach.

The paper is organized as follows. In Section 2, we introduce the definition of the problem, the KKT conditions, and other preliminary results. In Section 3, we prove that the original problem is equivalent to the reformulated one, in terms of KKT points, and by means of the second-order sufficient conditions. In Section 4, we show the equivalence between regularity conditions satisfied by KKT points of the original and the reformulated problems. In Section 5, we present some numerical experiments using the squared slack variables technique. We conclude in Section 6, with some remarks and future works.

2 Preliminaries

We start this section with the following notations, which will be used in the whole paper. The Euclidean inner product and norm are denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively. The identity matrix with dimension s is denoted by $I_s \in \mathbb{R}^{s \times s}$, and for any matrix $A \in \mathbb{R}^{s \times \ell}$, its transpose is denoted by $A^\top \in \mathbb{R}^{\ell \times s}$. A vector $x \in \mathbb{R}^s$ with entries $x_i \in \mathbb{R}$, $i = 1, \ldots, s$, is written as $(x_1, \ldots, x_s)^\top$, or simply $(x_1, \ldots, x_s)$. We also use the notation $x = (x_0, \bar{x})$, where x_0 is the first entry of x and $\bar{x}$ is the subvector that consists of the remaining entries. Also, for a vector $x = (x_1, \ldots, x_s) \in \mathbb{R}^s$ and an index set $\mathcal{I} \subseteq \{1, \ldots, s\}$, we denote by $\operatorname{diag}(x)$ and $\operatorname{diag}(x_i)_{i \in \mathcal{I}}$ the diagonal matrices with diagonal entries x_i, where $i = 1, \ldots, s$ and $i \in \mathcal{I}$, respectively. If $\{A_i\}_{i=1}^{\ell}$ and $\{B_i\}_{i \in \mathcal{I}}$ are two sequences of matrices, with $\mathcal{I}$ being an index set, then $\operatorname{diag}(A_i)_{i=1}^{\ell}$ and $\operatorname{diag}(B_i)_{i \in \mathcal{I}}$ denote the block diagonal matrices with block diagonal entries A_i, $i = 1, \ldots, \ell$, and B_i, $i \in \mathcal{I}$, respectively. Moreover, the gradient and the Hessian of a function $p\colon \mathbb{R}^s \to \mathbb{R}$ at $x \in \mathbb{R}^s$ are denoted by $\nabla p(x)$ and $\nabla^2 p(x)$, respectively. If $p\colon \mathbb{R}^{s+\ell} \to \mathbb{R}$, then the gradient and the Hessian at $(x, y) \in \mathbb{R}^{s+\ell}$ with respect to x are denoted by $\nabla_x p(x, y)$ and $\nabla_x^2 p(x, y)$, respectively. Also, the Jacobian of a mapping $q\colon \mathbb{R}^s \to \mathbb{R}^\ell$ at $x \in \mathbb{R}^s$ is denoted by $Jq(x) \in \mathbb{R}^{\ell \times s}$. Finally, for a closed convex and pointed cone $\mathcal{K} \subset \mathbb{R}^\ell$, its interior, boundary, and boundary excluding the origin are denoted by $\operatorname{int}(\mathcal{K})$, $\operatorname{bd}(\mathcal{K})$, and $\operatorname{bd}^+(\mathcal{K})$, respectively.

In this work, we consider the following NSOCP problem:

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & g(x) \in \mathcal{K}, \end{array} \tag{P1}$$

where $f\colon \mathbb{R}^n \to \mathbb{R}$ and $g\colon \mathbb{R}^n \to \mathbb{R}^m$ are twice continuously differentiable functions, and $\mathcal{K} := \mathcal{K}_1 \times \cdots \times \mathcal{K}_r$ is a Cartesian product of *second-order cones* (or *Lorentz cones*),

$$\mathcal{K}_i := \left\{ (z_0, \bar{z}) \in \mathbb{R} \times \mathbb{R}^{m_i - 1} \colon z_0 \geq \|\bar{z}\| \right\} \subset \mathbb{R}^{m_i}, \quad i = 1, \ldots, r,$$

with $m_1 + \cdots + m_r = m$. It is well-known that $\mathcal{K}_i$ is a *cone of squares* with respect to the *Jordan product* operator [1, Section 4]. For any vectors w, z with the same dimension, their Jordan product is given by

$$w \circ z := \big(\langle w, z \rangle, w_0 \bar{z} + z_0 \bar{w} \big).$$

So, the second-order cones can be written as

$$\mathcal{K}_i = \left\{ z \circ z \colon z \in \mathbb{R}^{m_i} \right\}, \quad i = 1, \ldots, r. \tag{2.1}$$

Now, let $g := (g_1, \ldots, g_r)$ with $g_i\colon \mathbb{R}^n \to \mathbb{R}^{m_i}$, $i = 1, \ldots, r$. Then, the constraints in (P1) can be rewritten as $g_i(x) \in \mathcal{K}_i$, $i = 1, \ldots, r$. Adding slack variables $y := (y_1, \ldots, y_r) \in \mathbb{R}^m$, with $y_i \in \mathbb{R}^{m_i}$ for all $i = 1, \ldots, r$ in problem (P1), we have the following NLP formulation:

$$\begin{array}{ll} \underset{x,y}{\text{minimize}} & f(x) \\ \text{subject to} & g_i(x) - y_i \circ y_i = 0, \quad i = 1, \ldots, r. \end{array} \tag{P2}$$

This problem is equivalent to (P1) in the sense that if (x^*, y^*) is a global (local) minimizer of (P2), then x^* is a global (local) minimizer of (P1). Conversely, if x^* is a global (local) minimizer of (P1), then there exists y^* such that (x^*, y^*) is a global (local) minimizer of (P2). From the practical viewpoint, it is more important to examine the relation between stationary points, or KKT points, of the two problems, because we can only expect to compute such points in practice. However, the relation between stationary points is less clear than that between optimal solutions. Once again, we emphasize that the use of slack variables is interesting and significant here, because the formulation (P2) is no longer an NSOCP problem, but only an NLP problem.

We now show the relation between KKT points of (P1) and (P2). Before presenting their KKT conditions, we first show some useful results concerning the Jordan algebra associated with the second-order cones. For further details, we refer to [1].

Lemma 2.1. *For any vectors $u, w, z \in \mathbb{R}^{m_i}$, the following properties hold:*

(a) (commutativity 1) $u \circ z = z \circ u$;

(b) (commutativity 2) $u \circ ((u \circ u) \circ z) = (u \circ u) \circ (u \circ z)$;

(c) (identity) $\mathbf{e} \circ u = u \circ \mathbf{e} = u$, *where* $\mathbf{e} := (1, 0, \dots, 0) \in \mathbb{R}^{m_i}$;

(d) (distributivity) $(w + u) \circ z = (w \circ z) + (u \circ z)$;

(e) (inner product) $\langle w \circ u, z\rangle = \langle u \circ z, w\rangle = \langle w \circ z, u\rangle$.

We also recall that the Jordan product is not associative. Besides, for any vector $z \in \mathbb{R}^{m_i}$, the *arrow matrix* of z is defined by

$$\mathrm{Arw}(z) := \begin{bmatrix} z_0 & \bar{z}^\top \\ \bar{z} & z_0 I_{m_i - 1} \end{bmatrix}.$$

Then, for any vector $w \in \mathbb{R}^{m_i}$, we have $z \circ w = \mathrm{Arw}(z)w$. It is well-known [1] that $\mathrm{Arw}(z)$ is positive definite if and only if $z \in \mathrm{int}(\mathcal{K}_i)$. Moreover, from [1, Theorem 3], $\mathrm{Arw}(z)$ is singular if and only if either $z \in \mathrm{bd}^+(\mathcal{K}_i)$ or $z \in \mathrm{bd}^+(-\mathcal{K}_i)$ or $z = 0$ or z satisfies $z_0 = 0$ with $m_i > 2$.

Lemma 2.2. *The following statements hold.*

(a) Let $p \colon \mathbb{R}^{m_i} \to \mathbb{R}^{m_i}$ be the function defined by $p(z) := z \circ z$. Then, $Jp(z) = 2\mathrm{Arw}(z)$.

(b) Let $q \colon \mathbb{R}^{m_i} \to \mathbb{R}$ be the function defined by $q(z) := \langle s, z \circ z\rangle$, where $s \in \mathbb{R}^{m_i}$ is a given vector. Then, $\nabla q(z) = 2z \circ s$ and $\nabla^2 q(z) = 2\mathrm{Arw}(s)$.

(c) Let $z := w \circ w$ with $w \in \mathbb{R}^{m_i}$ and $z \in \mathrm{bd}^+(\mathcal{K}_i)$. Then, we have $|w_0| = \|\bar{w}\| \neq 0$ and $z = 2w_0 w$. Moreover, $w \in \mathrm{bd}^+(\mathcal{K}_i)$ or $w \in \mathrm{bd}^+(-\mathcal{K}_i)$, and $\mathrm{Arw}(w)$ is singular.

(d) Let $z := w \circ w$ with $w \in \mathbb{R}^{m_i}$ and $z \in \mathrm{int}(\mathcal{K}_i)$. Then, we have $|w_0| \neq \|\bar{w}\|$. Moreover, $w \neq 0$, $w \notin \mathrm{bd}^+(\mathcal{K}_i)$, and $w \notin \mathrm{bd}^+(-\mathcal{K}_i)$. If in addition, we have $w_{i0} \neq 0$, then $\mathrm{Arw}(w)$ is nonsingular.

Proof. Items (a) and (b) follow directly from the definition of Jordan product. To prove item (c), observe that the definition of z and $z \in \mathrm{bd}^+(\mathcal{K}_i)$ imply $w_0^2 + \|\bar{w}\|^2 = 2|w_0|\|\bar{w}\| \neq 0$. But this is equivalent to $(|w_0| - \|\bar{w}\|)^2 = 0$ and $w_0 \neq 0$, which in turn is equivalent to $|w_0| = \|\bar{w}\| \neq 0$. Thus, we have $z = (w_0^2 + \|\bar{w}\|^2, 2w_0\bar{w}) = 2w_0 w$, that is, w is a nonzero multiple of z. Clearly, if $w_0 > 0$ then $w \in \mathrm{bd}^+(\mathcal{K}_i)$, and if $w_0 < 0$ then $w \in \mathrm{bd}^+(-\mathcal{K}_i)$. Item (d) can be shown in a similar manner. $\square$

Next, we introduce the *spectral decomposition* of vectors associated with a second-order cone. Let $z = (z_0, \bar{z}) \in \mathcal{K}_i$. Then, z can be decomposed as

$$z = \eta_1 c^{(1)} + \eta_2 c^{(2)}, \tag{2.2}$$

where $\eta_1 := z_0 - \|\bar{z}\|$ and $\eta_2 := z_0 + \|\bar{z}\|$ are the spectral values of z, and

$$c^{(1)} := \begin{cases} (1/2)(1, -\bar{z}/\|\bar{z}\|), & \text{if } \bar{z} \neq 0, \\ (1/2)(1, -\bar{w}), & \text{if } \bar{z} = 0, \end{cases} \qquad c^{(2)} := \begin{cases} (1/2)(1, \bar{z}/\|\bar{z}\|), & \text{if } \bar{z} \neq 0, \\ (1/2)(1, \bar{w}), & \text{if } \bar{z} = 0 \end{cases}$$

are the spectral vectors of z, with $\bar{w} \in \mathbb{R}^{m_i - 1}$ satisfying $\|\bar{w}\| = 1$. We briefly mention some properties of the spectral vectors: (i) $c^{(1)} \circ c^{(2)} = 0$, (ii) $c^{(1)} \circ c^{(1)} = c^{(1)}$ and $c^{(2)} \circ c^{(2)} = c^{(2)}$, (iii) $c^{(1)} + c^{(2)} = \mathbf{e}$, and (iv) $c^{(1)}, c^{(2)} \in \mathrm{bd}(\mathcal{K}_i)$. We also refer to $\{c^{(1)}, c^{(2)}\}$ as the *Jordan frame* associated to z. If two vectors z and w share a Jordan frame, i.e., $z = \eta_1 c^{(1)} + \eta_2 c^{(2)}$ and $w = \gamma_1 c^{(1)} + \gamma_2 c^{(2)}$ for a Jordan frame $\{c^{(1)}, c^{(2)}\}$, then z and w *operator commute*, or equivalently, $\mathrm{Arw}(z)\mathrm{Arw}(w) = \mathrm{Arw}(w)\mathrm{Arw}(z)$. In particular, from Lemma 2.1 and [1, Theorem 6], z and $z \circ z$ operator commute. The following lemma is also useful and straightforward to prove.

Lemma 2.3. *Let $z \in \mathbb{R}^{m_i}$ be a vector with spectral decomposition* (2.2)*. Then, the following statements hold.*

(a) $z = 0$ if and only if $\eta_1 = 0$ and $\eta_2 = 0$.

(b) $z \in \mathrm{bd}^+(\mathcal{K}_i)$ if and only if $\eta_1 = 0$ and $\eta_2 > 0$.

(c) $z \in \mathrm{int}(\mathcal{K}_i)$ if and only if $\eta_1 > 0$.

(d) $z \in \mathrm{bd}^+(-\mathcal{K}_i)$ if and only if $\eta_1 < 0$ and $\eta_2 = 0$.

(e) $z \in \mathrm{int}(-\mathcal{K}_i)$ if and only if $\eta_2 < 0$.

Let us return to problems (P1) and (P2). We say that $(x, \lambda) \in \mathbb{R}^{n+m}$ is a KKT pair of problem (P1), with $\lambda := (\lambda_1, \ldots, \lambda_r) \in \mathbb{R}^m$ and $\lambda_i \in \mathbb{R}^{m_i}$, $i = 1, \ldots, r$, if the following conditions are satisfied:

$$\nabla f(x) - \sum_{i=1}^{r} J g_i(x)^\top \lambda_i = 0, \tag{P1.1}$$

$$\lambda_i \in \mathcal{K}_i, \quad i = 1, \ldots, r, \tag{P1.2}$$

$$g_i(x) \in \mathcal{K}_i, \quad i = 1, \ldots, r, \tag{P1.3}$$

$$\lambda_i \circ g_i(x) = 0, \quad i = 1, \ldots, r. \tag{P1.4}$$

Notice that under a constraint qualification, the above conditions are necessary for optimality [7]. For the equality constrained problem (P2), a triple $(x, y, \lambda) \in \mathbb{R}^{n+2m}$ satisfies the KKT conditions if

$$\begin{aligned} \nabla_{(x,y)} \mathcal{L}(x, y, \lambda) = 0, \\ g_i(x) - y_i \circ y_i = 0, \quad i = 1, \ldots, r, \end{aligned}$$

where $\mathcal{L}$ is the Lagrangian function associated with (P2):

$$\mathcal{L}(x, y, \lambda) := f(x) - \sum_{i=1}^{r} \langle \lambda_i, g_i(x) - y_i \circ y_i \rangle. \tag{2.3}$$

Using Lemma 2.2(b), we observe that the above conditions can be rewritten as follows:

$$\nabla f(x) - \sum_{i=1}^{r} J g_i(x)^\top \lambda_i = 0, \tag{P2.1}$$

$$y_i \circ \lambda_i = 0, \quad i = 1, \ldots, r, \tag{P2.2}$$

$$g_i(x) - y_i \circ y_i = 0, \quad i = 1, \ldots, r. \tag{P2.3}$$

These conditions are necessary for optimality for (P2) under a constraint qualification.

Let (x, λ) be a KKT pair of (P1). We define the following sets of indices:

$$\begin{aligned} \mathcal{I}_0 &:= \{ i \in \{1, \ldots, r\} \colon g_i(x) = 0 \}, \\ \mathcal{I}_B &:= \{ i \in \{1, \ldots, r\} \colon g_i(x) \in \mathrm{bd}^+(\mathcal{K}_i) \}, \\ \mathcal{I}_I &:= \{ i \in \{1, \ldots, r\} \colon g_i(x) \in \mathrm{int}(\mathcal{K}_i) \}. \end{aligned} \tag{2.4}$$

Clearly, the sets $\mathcal{I}_0$, $\mathcal{I}_B$ and $\mathcal{I}_I$ constitute a partition of $\{1, \ldots, r\}$. We also define below some subsets of these sets:

$$\begin{aligned} \mathcal{I}_{00} &:= \{ i \in \{1, \ldots, r\} \colon g_i(x) = 0, \lambda_i = 0 \}, \\ \mathcal{I}_{0I} &:= \{ i \in \{1, \ldots, r\} \colon g_i(x) = 0, \lambda_i \in \mathrm{int}(\mathcal{K}_i) \}, \\ \mathcal{I}_{0B} &:= \{ i \in \{1, \ldots, r\} \colon g_i(x) = 0, \lambda_i \in \mathrm{bd}^+(\mathcal{K}_i) \}, \\ \mathcal{I}_{B0} &:= \{ i \in \{1, \ldots, r\} \colon g_i(x) \in \mathrm{bd}^+(\mathcal{K}_i), \lambda_i = 0 \}, \\ \mathcal{I}_{BB} &:= \{ i \in \{1, \ldots, r\} \colon g_i(x) \in \mathrm{bd}^+(\mathcal{K}_i), \lambda_i \in \mathrm{bd}^+(\mathcal{K}_i) \}, \\ \mathcal{I}_{I0} &:= \{ i \in \{1, \ldots, r\} \colon g_i(x) \in \mathrm{int}(\mathcal{K}_i), \lambda_i = 0 \}. \end{aligned} \tag{2.5}$$

Observe that all these sets are suitable for a KKT triple (x, y, λ) of (P2) as well. However, in the latter case, since λ_i can lie outside the cone $\mathcal{K}_i$, we also have to define the following index sets:

$$\begin{aligned} \mathcal{I}_{0N} &:= \{ i \in \{1, \ldots, r\} \colon g_i(x) = 0, \lambda_i \notin \mathcal{K}_i \}, \\ \mathcal{I}_{BN} &:= \{ i \in \{1, \ldots, r\} \colon g_i(x) \in \mathrm{bd}^+(\mathcal{K}_i), \lambda_i \notin \mathcal{K}_i \}, \\ \mathcal{I}_{IN} &:= \{ i \in \{1, \ldots, r\} \colon g_i(x) \in \mathrm{int}(\mathcal{K}_i), \lambda_i \neq 0 \}, \end{aligned} \tag{2.6}$$

Note that $\mathcal{I}_{00}, \mathcal{I}_{0I}, \mathcal{I}_{0B}$ and $\mathcal{I}_{0N}$ constitute a partition of $\mathcal{I}_0$, and that $\mathcal{I}_{I0}$ and $\mathcal{I}_{IN}$ constitute a partition of $\mathcal{I}_I$. Moreover, from (P2.3) and Lemma 2.2(c), if $g_i(x) \in \mathrm{bd}^+(\mathcal{K}_i)$, then $y_i \in \mathrm{bd}^+(\mathcal{K}_i) \cup \mathrm{bd}^+(-\mathcal{K}_i)$. From (P2.2), in particular it implies that $\lambda_i \notin \mathrm{int}(\mathcal{K}_i)$. So, $\mathcal{I}_{B0}$, $\mathcal{I}_{BB}$ and $\mathcal{I}_{BN}$ constitute a partition of $\mathcal{I}_B$.

3 Equivalence Between KKT Points

We now establish the equivalence between KKT points of the original NSOCP problem (P1) and the reformulated NLP problem (P2). One of the implications is simple, as shown in the next proposition.

Proposition 3.1. *Let* $(x, \lambda) \in \mathbb{R}^{n+m}$ *be a KKT pair of problem* (P1)*. Then, there exists* $y \in \mathbb{R}^m$ *such that* (x, y, λ) *is a KKT triple of* (P2)*.*

Proof. It is clear that condition (P2.1) holds. Also, from (2.1), there exists $y_i \in \mathbb{R}^{m_i}$ such that $g_i(x) = y_i \circ y_i$ for all $i = 1, \ldots, r$. Thus, the condition (P2.3) is satisfied. To prove (P2.2), we consider three cases:

Case 1. If $i \in \mathcal{I}_0$, (P2.3) shows that $y_i = 0$ and thus, $y_i \circ \lambda_i = 0$.

Case 2. If $i \in \mathcal{I}_I$, (P1.2) and (P1.4) imply $\lambda_i = 0$. Therefore, $y_i \circ \lambda_i = 0$.

Case 3. Let $i \in \mathcal{I}_B$. Once again from (P1.2) and (P1.4), we conclude that $\lambda_i \in \mathrm{bd}^+(\mathcal{K}_i)$ or $\lambda_i = 0$. In the latter case, $y_i \circ \lambda_i = 0$ holds trivially. Thus, suppose that $\lambda_i \in \mathrm{bd}^+(\mathcal{K}_i)$. From (P1.2)–(P1.4) and Lemma 2.3(c), we can write

$$g_i(x) = \eta_1 c^{(1)} \quad \text{and} \quad \lambda_i = \eta_2 c^{(2)},$$

where $\{c^{(1)}, c^{(2)}\} \subset \mathbb{R}^{m_i}$ is a Jordan frame and $\eta_1, \eta_2 > 0$. Moreover, from (P2.3), it can be seen that $g_i(x)$ and y_i operator commute, which implies that they share the same Jordan frame. Then, we can deduce that $y_i = \eta_3 c^{(1)}$ with $\eta_3^2 = \eta_1$. This shows that

$$y_i \circ \lambda_i = \eta_2 \eta_3 \big(c^{(1)} \circ c^{(2)}\big) = 0,$$

and the result follows.

We then conclude that (x, y, λ) is a KKT triple of (P2). □

The converse implication is not always true. Even if (x, y, λ) is a KKT triple of (P2), it does not necessarily imply that (x, λ) is a KKT pair of (P1), because λ_i may lie outside $\mathcal{K}_i$ for some i. We will show now that the converse is true when the second-order sufficient condition is considered. First, we recall the definition of such a condition for the NSOCP problem [6].

Definition 3.2. *Let* $(x, \lambda) \in \mathbb{R}^{n+m}$ *be a KKT pair of problem* (P1)*. The* second-order sufficient condition *(SOSC-NSOCP)*[1] *holds if*

$$\left\langle \left(\nabla_x^2 L(x, \lambda) + \sum_{i \in \mathcal{I}_{BB}} H_i(x, \lambda) \right) d, d \right\rangle > 0$$

for all nonzero $d \in \mathcal{C}(x)$*, where*[2] $\nabla_x^2 L(x, \lambda) = \nabla^2 f(x) - \sum_{i=1}^r \sum_{j=1}^{m_i} \lambda_{i,j} \nabla^2 g_{i,j}(x)$ *is the Hessian of the Lagrangian associated with* (P1)*,*

$$\mathcal{C}(x) := \left\{ d \in \mathbb{R}^n : \begin{array}{l} \langle \nabla f(x), d \rangle = 0; \\ Jg(x) d \in \mathcal{T}_{\mathcal{K}}(g(x)) \end{array} \right\}$$

[1] We refer to this condition as SOSC-NSOCP in order to distinguish it from SOSC for NLP.

[2] Notice that $g_i(x) = \big(g_{i0}(x), \overline{g_i(x)}\big) = \big(g_{i,1}(x), \ldots, g_{i,m_i}(x)\big)$, i.e., $g_{i0}(x) = g_{i,1}(x)$. Similarly, we have $\lambda_i = (\lambda_{i0}, \bar{\lambda}_i) = (\lambda_{i,1}, \ldots, \lambda_{i,m_i})$, i.e., $\lambda_{i0} = \lambda_{i,1}$.

is the critical cone at x, $\mathcal{T}_{\mathcal{K}}(g(x))$ denotes the tangent cone of $\mathcal{K}$ at $g(x)$, and

$$H_i(x,\lambda) := -\kappa_i Jg_i(x)^\top \Gamma_i Jg_i(x), \tag{3.1}$$

with

$$\kappa_i := \frac{\lambda_{i0}}{g_{i0}(x)} \quad \text{and} \quad \Gamma_i := \begin{bmatrix} 1 & 0^\top \\ 0 & -I_{m_i-1} \end{bmatrix}, \quad \text{for all } i \in \mathcal{I}_{BB}. \tag{3.2}$$

Definition 3.3. *Let $(x,\lambda) \in \mathbb{R}^{n+m}$ be a KKT pair of problem* (P1)*. The strict complementarity condition holds if*

$$g_i(x) + \lambda_i \in \operatorname{int}(\mathcal{K}_i), \quad i = 1, \dots, r.$$

In terms of the index sets defined in (2.5)*, this condition means that*

$$\mathcal{I}_{B0} = \mathcal{I}_{0B} = \mathcal{I}_{00} = \emptyset.$$

Lemma 3.4. *Let $(x,\lambda) \in \mathbb{R}^{n+m}$ be a KKT pair of problem* (P1)*. The critical cone can be written as follows:*

$$\mathcal{C}(x) = \left\{ d \in \mathbb{R}^n : \begin{array}{ll} Jg_i(x)d \in \mathcal{K}_i, & i \in \mathcal{I}_{00}; \\ \langle \overline{Jg_i(x)d}, \overline{g_i(x)} \rangle - g_{i0}(x)\langle \nabla g_{i0}(x), d\rangle \le 0, & i \in \mathcal{I}_{B0}; \\ Jg_i(x)d = 0, & i \in \mathcal{I}_{0I}; \\ Jg_i(x)d \in \mathbb{R}_+(\lambda_{i0}, -\bar{\lambda}_i), & i \in \mathcal{I}_{0B}; \\ \langle Jg_i(x)d, \lambda_i \rangle = 0, & i \in \mathcal{I}_{BB} \end{array} \right\}. \tag{3.3}$$

Proof. It follows from [6, Lemma 25] and [6, Corollary 26]. □

Recalling that (P2) is just an NLP problem, the following lemma shows the second-order sufficient condition (SOSC-NLP) associated with it.

Lemma 3.5. *Let $(x,y,\lambda) \in \mathbb{R}^{n+2m}$ be a KKT triple of problem* (P2)*. The SOSC-NLP can be written as*

$$\left\langle \nabla_x^2 L(x,\lambda)v, v \right\rangle + 2\sum_{i=1}^{r} \langle w_i \circ w_i, \lambda_i \rangle > 0 \tag{3.4}$$

for all nonzero $(v,w) \in \mathbb{R}^{n+m}$ such that

$$\begin{array}{ll} Jg_i(x)v - 2y_i \circ w_i = 0, & i \in \mathcal{I}_I \cup \mathcal{I}_B, \\ Jg_i(x)v = 0, & i \in \mathcal{I}_0. \end{array}$$

Proof. Recall the definition of $\mathcal{L}$ given in (2.3). Then, from [4, Section 3.3] or [19, Section 12.4], SOSC-NLP for (P2) holds if

$$\left\langle \nabla^2_{(x,y)} \mathcal{L}(x,y,\lambda)d, d \right\rangle > 0$$

for all nonzero $d \in \mathbb{R}^{n+m}$ such that

$$\left[Jg_i(x),\ 0^{m_i \times \ell_i}, -2\operatorname{Arw}(y_i),\ 0^{m_i \times (m - m_i - \ell_i)} \right] d = 0, \quad i = 1, \dots, r,$$

where $\ell_i := \sum_{1 \le j < i} m_j$, and $0^{m_i \times s}$ denotes the zero matrix with dimension $m_i \times s$, $i = 1, \ldots, r$. Notice that the Hessian of the Lagrangian can be written as

$$\nabla^2_{(x,y)} \mathcal{L}(x, y, \lambda) = \begin{bmatrix} \nabla^2_x L(x, \lambda) & 0 \\ 0 & 2 \operatorname{diag}(\operatorname{Arw}(\lambda_i))_{i=1}^r \end{bmatrix}.$$

Also, for any $w_i \in \mathbb{R}^{m_i}$, we have

$$\langle w_i, \operatorname{Arw}(\lambda_i) w_i \rangle = \langle w_i, \lambda_i \circ w_i \rangle = \langle w_i \circ w_i, \lambda_i \rangle,$$

where the first equality holds since $\operatorname{Arw}(\lambda_i) w_i = \lambda_i \circ w_i$ and the second one follows by Lemma 2.1(e). Then, by letting $d = (v, w)$ with $v \in \mathbb{R}^n$ and $w \in \mathbb{R}^m$ and observing that $y_i = 0$ for $i \in \mathcal{I}_0$, we have the desired result. $\square$

Proposition 3.6. *Let $(x, y, \lambda) \in \mathbb{R}^{n+2m}$ be a KKT triple of* (P2) *and assume that it satisfies SOSC-NLP. Then, (x, λ) is a KKT pair of* (P1).

Proof. Observe that the condition (P1.1) holds trivially. The condition (P1.3) is also satisfied from (P2.3) and (2.1). For each $i = 1, \ldots, r$, let us prove (P1.2) and (P1.4). We consider the following three cases:

Case 1. Let $i \in \mathcal{I}_0$. Then, (P1.4) holds trivially. To prove (P1.2), assume, to the contrary, that $\lambda_i \notin \mathcal{K}_i$. Then, there exists $u_i \circ u_i \in \mathcal{K}_i$ such that $\langle u_i \circ u_i, \lambda_i \rangle < 0$. Since there is no restriction for w_i, $i \in \mathcal{I}_0$ in the SOSC-NLP, the inequality (3.4) should be satisfied in particular for w_i with $w_i \circ w_i = \alpha u_i \circ u_i$ with any $\alpha > 0$ arbitrarily large. However, this is impossible since $\langle u_i \circ u_i, \lambda_i \rangle < 0$. Thus, (P1.2) holds.

Case 2. Let $i \in \mathcal{I}_B$. From (P2.3), $g_i(x)$ and y_i operator commute. Then, they share a Jordan frame $\{c^{(1)}, c^{(2)}\} \subset \mathbb{R}^{m_i}$. From Lemmas 2.2(c) and 2.3(c), and equality (P2.3), we can write $y_i = \eta_1 c^{(1)}$ and $g_i(x) = \eta_1^2 c^{(1)}$, where $\eta_1 \neq 0$. Thus, we have

$$y_i \circ \lambda_i = 0 \quad \Leftrightarrow \quad \eta_1 c^{(1)} \circ \lambda_i = 0 \quad \Leftrightarrow \quad \eta_1^2 c^{(1)} \circ \lambda_i = 0 \quad \Leftrightarrow \quad g_i(x) \circ \lambda_i = 0,$$

which shows that (P1.4) holds. Since $g_i(x) \in \operatorname{bd}^+(\mathcal{K}_i)$ and $g_i(x) \circ \lambda_i = 0$, we have $\lambda_i = 0$ or $\lambda_i \in \operatorname{bd}^+(\mathcal{K}_i)$ or $\lambda_i \in \operatorname{bd}^+(-\mathcal{K}_i)$. Let us assume, to the contrary, that $\lambda_i \in \operatorname{bd}^+(-\mathcal{K}_i)$. From (P2.3) and (P1.4), we can deduce that y_i and λ_i share the same Jordan frame $\{c^{(1)}, c^{(2)}\}$. Moreover, from Lemma 2.3, $y_i = \eta_1 c^{(1)}$ and $\lambda_i = \eta_2 c^{(2)}$ with $\eta_1 \neq 0$ and $\eta_2 < 0$. Observe that the inequality (3.4) should hold in particular when $v = 0$, $w_j = 0$ for all $j \neq i$, and w_i is a nonzero vector that satisfies $y_i \circ w_i = 0$. Then, we can choose $w_i = \eta_3 c^{(2)}$ for some $\eta_3 \neq 0$ and obtain

$$\langle w_i \circ w_i, \lambda_i \rangle = \langle \eta_3^2 c^{(2)}, \eta_2 c^{(2)} \rangle = \eta_2 \eta_3^2 \|c^{(2)}\|^2 < 0.$$

The inequality (3.4) should hold for any $w_i = \eta_3 c^{(2)}$ with arbitrarily large η_3. Since this is impossible, we conclude that $\lambda_i \notin \operatorname{bd}^+(-\mathcal{K}_i)$. Therefore, $\lambda_i \in \mathcal{K}_i$.

Case 3. Let $i \in \mathcal{I}_I$. From (P2.2) and (P2.3), we obtain

$$\langle g_i(x), \lambda_i \rangle = \langle y_i \circ y_i, \lambda_i \rangle = \langle y_i, y_i \circ \lambda_i \rangle = 0. \tag{3.5}$$

Also, since $g_i(x)$ and y_i operator commute, by Lemma 2.1(b), we have

$$0 = g_i(x) \circ (y_i \circ \lambda_i) = y_i \circ (g_i(x) \circ \lambda_i).$$

This equality, together with (3.5), shows that

$$\overline{y_i \circ (g_i(x) \circ \lambda_i)} = y_{i0}(\overline{g_i(x) \circ \lambda_i}) + \langle g_i(x), \lambda_i \rangle \bar{y}_i = 0 \quad \Rightarrow \quad y_{i0}(\overline{g_i(x) \circ \lambda_i}) = 0,$$

that is, $y_{i0} = 0$ or $\overline{g_i(x) \circ \lambda_i} = 0$.

Let us suppose that $y_{i0} = 0$. Then, from (P2.3), we have $g_{i0}(x) = \|\bar{y}_i\|^2$ and $\overline{g_i(x)} = 0$, which, together with (3.5), shows that $\|\bar{y}_i\|^2 \lambda_{i0} = 0$. Since $i \in \mathcal{I}_I$, $\bar{y}_i \neq 0$ and thus $\lambda_{i0} = 0$. Now, observe that the inequality (3.4) should hold for any (v, w) such that $v = 0$, $w_j = 0$, $j \in (\mathcal{I}_I \cup \mathcal{I}_B) \setminus \{i\}$ and $w_i \neq 0$ satisfying $y_i \circ w_i = 0$. But from this equality, we have

$$y_{i0}\bar{w}_i + w_{i0}\bar{y}_i = 0 \quad \Rightarrow \quad w_{i0}\bar{y}_i = 0 \quad \Rightarrow \quad w_{i0} = 0.$$

Thus, $\langle w_i \circ w_i, \lambda_i \rangle = \|w_i\|^2 \lambda_{i0} + \langle 2w_{i0}\bar{w}_i, \bar{\lambda}_i \rangle = 0$. It means that with such choice of (v, w), the inequality (3.4) is not satisfied. Therefore, $y_{i0} \neq 0$ and we conclude that $\overline{g_i(x) \circ \lambda_i} = 0$ holds. Recalling (3.5), we obtain $g_i(x) \circ \lambda_i = 0$, which means that (P1.4) holds. Moreover, since $g_i(x) \in \operatorname{int}(\mathcal{K}_i)$, $\operatorname{Arw}(g_i(x))$ is nonsingular. Thus, $g_i(x) \circ \lambda_i = \operatorname{Arw}(g_i(x))\lambda_i = 0$ implies that $\lambda_i = 0$ and so (P1.2) is also satisfied.

We then conclude that (x, λ) is a KKT pair of (P1). □

Corollary 3.7. *Let $(x, y, \lambda) \in \mathbb{R}^{n+2m}$ be a KKT triple of* (P2) *and assume that it satisfies SOSC-NLP. If there exists $i \in \mathcal{I}_I$, then we have $y_{i0} \neq 0$.*

Proof. It can be seen easily in the proof of Proposition 3.6, case 3. □

Next lemma shows that the condition SOSC-NLP for a KKT triple of (P2) makes the index sets $\mathcal{I}_{B0}$, $\mathcal{I}_{0B}$ and $\mathcal{I}_{00}$ empty. From Proposition 3.6, this means that the corresponding KKT pair (P1) satisfies the strict complementarity condition.

Lemma 3.8. *Let $(x, y, \lambda) \in \mathbb{R}^{n+2m}$ be a KKT triple of* (P2) *and assume that it satisfies SOSC-NLP. Then, we have $\mathcal{I}_{B0} = \mathcal{I}_{0B} = \mathcal{I}_{00} = \emptyset$.*

Proof. Let us first prove that $\mathcal{I}_{0B} = \mathcal{I}_{00} = \emptyset$. Assume that there exists an index j such that $g_j(x) = 0$, which also implies $y_j = 0$. Since (x, λ) is a KKT pair of (P1) from Proposition 3.6, we know that $\lambda_j \in \mathcal{K}_j$. So, we need to show that $\lambda_j \in \operatorname{int}(\mathcal{K}_j)$. Recalling SOSC-NLP of (P2) and taking $v = 0$ in (3.4), we obtain

$$\langle w_j \circ w_j, \lambda_j \rangle + \sum_{i \neq j} \langle w_i \circ w_i, \lambda_i \rangle > 0 \tag{3.6}$$

for all nonzero $w \in \mathbb{R}^m$ satisfying

$$y_i \circ w_i = 0, \quad i \in \mathcal{I}_I \cup \mathcal{I}_B. \tag{3.7}$$

In particular, the condition (3.6) holds when $w_j \neq 0$ and $w_i = 0$ for all $i \neq j$. Such a choice of w shows that $\langle w_j \circ w_j, \lambda_j \rangle > 0$. Since $w_j \circ w_j \in \mathcal{K}_j$ from (2.1), it turns out that $\lambda_j \in \mathrm{int}(\mathcal{K}_j)$.

Now, let us show that $\mathcal{I}_{B0} = \emptyset$. Assume that there exists an index j such that $g_j(x) \in \mathrm{bd}^+(\mathcal{K}_j)$. We need to prove that $\lambda_j \neq 0$. In the same way as before, taking $v = 0$ in (3.4), we have (3.6) for all nonzero $w \in \mathbb{R}^m$ satisfying (3.7). From Lemma 2.2(c), it follows that either $y_j \in \mathrm{bd}^+(\mathcal{K}_j)$ or $y_j \in \mathrm{bd}^+(-\mathcal{K}_j)$. This means that $\mathrm{Arw}(y_j)$ is singular, and so there exists a nonzero w_j such that $\mathrm{Arw}(y_j)w_j = y_j \circ w_j = 0$. Thus, the condition (3.6) holds particularly for such $w_j \neq 0$ and $w_i = 0$ for all $i \neq j$. This choice of w shows that $\langle w_j \circ w_j, \lambda_j \rangle > 0$, which clearly implies $\lambda_j \neq 0$. □

We can actually show that the KKT pair (x, λ) of (P1), corresponding to a KKT triple of (P2), also satisfies the SOSC-NSOCP.

Proposition 3.9. *Let $(x, y, \lambda) \in \mathbb{R}^{n+2m}$ be a KKT triple of* (P2) *and assume that it satisfies SOSC-NLP. Then, (x, λ) is a KKT pair of* (P1) *satisfying SOSC-NSOCP and the strict complementarity.*

Proof. From Proposition 3.6, (x, λ) is a KKT pair of (P1) and the strict complementarity follows from Lemma 3.8. So, we only have to prove that SOSC-NSOCP holds. From (3.3) and the strict complementarity condition, we have to prove that

$$\langle \nabla_x^2 L(x, \lambda)v, v \rangle + \sum_{i \in \mathcal{I}_{BB}} \langle H_i(x, \lambda)v, v \rangle > 0 \tag{3.8}$$

for all nonzero $v \in \mathbb{R}^n$ such that

$$\begin{aligned} &\langle Jg_i(x)v, \lambda_i \rangle = 0, && i \in \mathcal{I}_{BB}, \\ &Jg_i(x)v = 0, && i \in \mathcal{I}_{0I}. \end{aligned} \tag{3.9}$$

From the SOSC-NLP of (P2), the fact that $\lambda_i = 0$ for $i \in \mathcal{I}_{I0}$, and the assumption $\mathcal{I}_{00} = \mathcal{I}_{B0} = \mathcal{I}_{0B} = \emptyset$, we have, by Lemma 3.5,

$$\langle \nabla_x^2 L(x, \lambda)v, v \rangle + 2 \sum_{i \in \mathcal{I}_{BB} \cup \mathcal{I}_{0I}} \langle w_i \circ w_i, \lambda_i \rangle > 0$$

for all nonzero $(v, w) \in \mathbb{R}^{n+m}$ such that

$$\begin{aligned} &Jg_i(x)v - 2y_i \circ w_i = 0, && i \in \mathcal{I}_{BB} \cup \mathcal{I}_{I0}, \\ &Jg_i(x)v = 0, && i \in \mathcal{I}_{0I}. \end{aligned} \tag{3.10}$$

Note that there is no restriction for w_i when $i \in \mathcal{I}_{0I}$. Thus, in particular, we have

$$\langle \nabla_x^2 L(x, \lambda)v, v \rangle + 2 \sum_{i \in \mathcal{I}_{BB}} \langle w_i \circ w_i, \lambda_i \rangle > 0 \tag{3.11}$$

for all nonzero $(v, w) \in \mathbb{R}^{n+m}$ such that (3.10) holds and $w_i = 0$ for all $i \in \mathcal{I}_{0I}$.

Now, let $v \in \mathbb{R}^n$ be an arbitrary nonzero vector satisfying (3.9). We must prove that (3.8) holds. To this end, we proceed with the following three steps:

- Step 1: First, we choose $w \in \mathbb{R}^m$ such that

$$\begin{array}{ll} w_i = 0 & \text{for all } i \in \mathcal{I}_{0I}, \\ Jg_i(x)v - 2y_i \circ w_i = 0 & \text{for all } i \in \mathcal{I}_{I0}, \\ Jg_i(x)v - 2y_i \circ w_i = 0 \text{ and } \langle \Gamma_i y_i, w_i \rangle = 0 & \text{for all } i \in \mathcal{I}_{BB}. \end{array} \tag{3.12}$$

  We will actually show that such w_i exist for all i.

- Step 2: Next, we will prove that, with w_i satisfying (3.12), the condition $\langle Jg_i(x)v, \lambda_i \rangle = 0$, $i \in \mathcal{I}_{BB}$ remains true. This means that the inequality (3.11) holds for an arbitrary v satisfying (3.9), along with w as chosen above.

- Step 3: Finally, we will prove that

$$\sum_{i \in \mathcal{I}_{BB}} 2\langle w_i \circ w_i, \lambda_i \rangle - \sum_{i \in \mathcal{I}_{BB}} \langle H_i(x, \lambda)v, v \rangle = 0 \tag{3.13}$$

  for any (v, w) satisfying (3.12). Then, we can conclude that (3.8) holds for any v satisfying (3.9).

Step 1: Let us show that w_i satisfying (3.12) exist for all i. When $i \in \mathcal{I}_{0I}$, the existence of w_i is trivial. Next, notice that $Jg_i(x)v - 2y_i \circ w_i = 0$ is equivalent to $\mathrm{Arw}(y_i)w_i = (Jg_i(x)v)/2$. Then, for each $i \in \mathcal{I}_{I0}$, we have $w_i = (\mathrm{Arw}(y_i)^{-1} Jg_i(x)v)/2$, since $\mathrm{Arw}(y_i)$ is nonsingular by Lemma 2.2(d) and Corollary 3.7. Now, let $i \in \mathcal{I}_{BB}$. Recall that $\mathrm{Arw}(y_i)$ is singular because $y_i \in \mathrm{bd}^+(\mathcal{K}_i)$ or $y_i \in \mathrm{bd}^+(-\mathcal{K}_i)$ by Lemma 2.2(c). Then, we have to show that there exists w_i satisfying the overdetermined system of equations

$$\begin{array}{rcl} \mathrm{Arw}(y_i)w_i & = & (Jg_i(x)v)/2, \\ (\Gamma_i y_i)^\top w_i & = & 0. \end{array} \tag{3.14}$$

To this end, define $z_i := (Jg_i(x)v)/2$ and let us show first that

$$\langle z_i, \Gamma_i y_i \rangle = 0. \tag{3.15}$$

Let $\{c^{(1)}, c^{(2)}\} \subset \mathbb{R}^{m_i}$ be a Jordan frame associated to λ_i. Note that $c^{(1)} = \Gamma_i c^{(2)}$. Observing that $\lambda_i \in \mathrm{bd}^+(\mathcal{K}_i)$, $y_i \in \mathrm{bd}^+(\mathcal{K}_i)$ or $y_i \in \mathrm{bd}^+(-\mathcal{K}_i)$, and $y_i \circ \lambda_i = 0$ by (P2.2), from Lemma 2.3, we can write

$$\lambda_i = \eta_1 c^{(1)} \text{ with } \eta_1 > 0 \quad \text{and} \quad y_i = \eta_2 c^{(2)} \text{ with } \eta_2 \neq 0.$$

Since v was chosen to satisfy (3.9), we also have $\langle z_i, \lambda_i \rangle = 0$, which implies $\langle z_i, c^{(1)} \rangle = 0$ because $\eta_1 > 0$. Thus, we obtain

$$\langle z_i, \Gamma_i y_i \rangle = \eta_2 \langle z_i, \Gamma_i c^{(2)} \rangle = \eta_2 \langle z_i, c^{(1)} \rangle = 0,$$

that is, (3.15) holds. Recall now that $z_i = (z_{i0}, \bar{z}_i) = (z_{i,1}, \dots, z_{i,m_i})$ and so $z_{i0} = z_{i,1}$. Similarly, we have $y_{i0} = y_{i,1}$ and $w_{i0} = w_{i,1}$. Rewriting (3.14), we have to show that there exists w_i such that

$$\begin{aligned} y_{i0}w_{i0} + \sum_{j=2}^{m_i} y_{i,j}w_{i,j} &= z_{i0}, \\ y_{i,j}w_{i0} + y_{i0}w_{i,j} &= z_{i,j}, \quad j = 2, \dots, m_i, \\ y_{i0}w_{i0} - \sum_{j=2}^{m_i} y_{i,j}w_{i,j} &= 0. \end{aligned} \tag{3.16}$$

Since $y_{i0} \neq 0$ by Lemma 2.2(c), the second line shows that

$$w_{i,j} = \frac{z_{i,j} - y_{i,j}w_{i0}}{y_{i0}}, \quad j = 2, \dots, m_i, \tag{3.17}$$

that is, $w_{i,j}$, $j = 2, \dots, m_i$, are determined by the value of w_{i0}. Once again from Lemma 2.2(c), we have $y_{i0}^2 = \|\bar{y}_i\|^2 = \sum_{j=2}^{m_i} y_{i,j}^2 \neq 0$. This fact, together with (3.15), yields

$$y_{i0}z_{i0} - \sum_{j=2}^{m_i} y_{i,j}z_{i,j} + \left(\sum_{j=2}^{m_i} y_{i,j}^2 - y_{i0}^2 \right) w_{i0} = 0 \quad \text{for any } w_{i0} \in \mathbb{R}.$$

Dividing the above equality by $y_{i0} \neq 0$, we have

$$y_{i0}w_{i0} + \sum_{j=2}^{m_i} \frac{y_{i,j}(z_{i,j} - y_{i,j}w_{i0})}{y_{i0}} = z_{i0} \quad \text{for any } w_{i0} \in \mathbb{R},$$

which, together with (3.17), shows that the first equality of (3.16) holds for any $w_{i0} \in \mathbb{R}$. In particular, if w_{i0} is given by

$$w_{i0} = \frac{1}{y_{i0}} \sum_{j=2}^{m_i} y_{i,j}w_{i,j},$$

then, the last equality in (3.16) holds. Hence, we conclude that there exists w_i satisfying all equations of (3.16), or equivalently (3.14). Finally, since $i \in \mathcal{I}_{BB}$ was taken arbitrarily, we can say that w_i satisfying (3.12) exists for all i.

Step 2: Let $i \in \mathcal{I}_{BB}$ be arbitrary. Then,

$$\langle Jg_i(x)v, \lambda_i \rangle = 2\langle y_i \circ w_i, \lambda_i \rangle = 2\langle w_i, y_i \circ \lambda_i \rangle = 0,$$

where the second equality holds from Lemma 2.1(e), and the third one follows from (P2.2). Thus, the condition in (3.9) remains true.

Step 3: Let $i \in \mathcal{I}_{BB}$ be arbitrary. Recalling the formula of κ_i and Γ_i in (3.2), we have

$$\lambda_i = \kappa_i \left(g_{i0}(x), \frac{g_{i0}(x)}{\lambda_{i0}} \bar{\lambda}_i \right) = \kappa_i \big(g_{i0}(x), -\overline{g_i(x)} \big) = \kappa_i \Gamma_i \, g_i(x) = \kappa_i \Gamma_i (y_i \circ y_i),$$

where the second equality holds because (P1.4) implies $\lambda_{i0}\overline{g_i(x)} + g_{i0}(x)\bar{\lambda}_i = 0$ and $\lambda_i \in \mathrm{bd}^+(\mathcal{K}_i)$ implies $\lambda_{i0} \neq 0$, and the last equality holds from (P2.3). Then, we have

$$\begin{aligned}
& 2\langle w_i \circ w_i, \lambda_i\rangle - \langle H_i(x,\lambda)v, v\rangle \\
= \;& 2\kappa_i\big(\langle w_i \circ w_i, \Gamma_i(y_i \circ y_i)\rangle + 2\langle y_i \circ w_i, \Gamma_i(y_i \circ w_i)\rangle\big) \\
= \;& 2\kappa_i\big(\|w_i\|^2\|y_i\|^2 - 8w_{i0}y_{i0}\langle \bar{w}_i, \bar{y}_i\rangle + 2\langle w_i, y_i\rangle^2 - 2y_{i0}^2\|\bar{w}_i\|^2 - 2w_{i0}^2\|\bar{y}_i\|^2\big) \\
= \;& 4\kappa_i\big(\langle w_i, y_i\rangle^2 - 4w_{i0}y_{i0}\langle \bar{w}_i, \bar{y}_i\rangle\big) \\
= \;& 4\kappa_i\big(\langle \bar{w}_i, \bar{y}_i\rangle - w_{i0}y_{i0}\big)^2 \\
= \;& 4\kappa_i\langle \Gamma_i y_i, w_i\rangle^2 \qquad (3.18) \\
= \;& 0,
\end{aligned}$$

where the first equality follows from the definition (3.1) of $H_i(x,\lambda)$ with $Jg_i(x)v = 2y_i \circ w_i$, the third equality follows from the fact that $y_{i0}^2 = \|\bar{y}_i\|^2$ by Lemma 2.2(c), and the last equality follows from the last condition in (3.12). Hence, we conclude that (3.13) holds.

The steps 1, 2 and 3 show that (3.8) holds for all nonzero v satisfying (3.9). Thus, the proof is complete. □

The above result shows that if we find a KKT point of the reformulated problem (P2) that satisfies SOSC-NLP, then it must be a KKT point of the original problem (P1). Moreover, such a KKT point also satisfies the SOSC-NSOCP of (P1) and the strict complementarity condition. We now show that the converse implication also holds under the assumption below.

Assumption 3.10. *Let $(x, y, \lambda) \in \mathbb{R}^{n+2m}$ be a KKT triple of* (P2)*. For any $i \in \mathcal{I}_I = \mathcal{I}_{I0}$, we have $y_{i0} \neq 0$.*

We observe that from Corollary 3.7, the above assumption is necessary to ensure SOSC-NLP of (P2).

Proposition 3.11. *Let $(x, \lambda) \in \mathbb{R}^{n+m}$ be a KKT pair of* (P1)*, and suppose that SOSC-NSOCP and the strict complementarity hold. Then, there exists $y \in \mathbb{R}^m$ such that (x, y, λ) is a KKT triple of* (P2) *satisfying SOSC-NLP whenever Assumption 3.10 holds.*

Proof. From Proposition 3.1, there exists $y \in \mathbb{R}^m$ such that (x, y, λ) is a KKT triple of (P2). Suppose y satisfies the condition in Assumption 3.10. Let us prove that SOSC-NLP of (P2) also holds. Choose an arbitrary nonzero vector $(v, w) \in \mathbb{R}^{n+m}$ such that

$$\begin{aligned}
& Jg_i(x)v - 2y_i \circ w_i = 0, && i \in \mathcal{I}_I \cup \mathcal{I}_B, \qquad (3.19)\\
& Jg_i(x)v = 0, && i \in \mathcal{I}_0.
\end{aligned}$$

In view of Lemma 3.5, we have to show that (3.4) holds.

Let us show first that $v \in \mathcal{C}(x)$, that is, v satisfies the conditions in (3.3). Since $Jg_i(x)v = 0$ for all $i \in \mathcal{I}_0 = \mathcal{I}_{00} \cup \mathcal{I}_{0I} \cup \mathcal{I}_{0B}$, the first, third and fourth conditions in (3.3) hold. For the case $i \in \mathcal{I}_{BB}$, we have

$$\langle Jg_i(x)v, \lambda_i\rangle = 2\langle y_i \circ w_i, \lambda_i\rangle = 2\langle w_i, y_i \circ \lambda_i\rangle = 0, \qquad (3.20)$$

where the second equality follows from Lemma 2.1(e), and the last equality follows from (P2.2). It means that the last condition of (3.3) holds. Now, let $i \in \mathcal{I}_{B0}$. Recall from (P2.3), that $g_{i0}(x) = \|y_i\|^2$ and $\overline{g_i(x)} = 2y_{i0}\bar{y}_i$. Moreover, from the choice of w_i, $\overline{Jg_i(x)v} = 2(y_{i0}\bar{w}_i + w_{i0}\bar{y}_i)$ and $\langle \nabla g_{i0}(x), v\rangle = (Jg_i(x)v)_0 = 2\langle y_i, w_i\rangle$. Thus, we have

$$\begin{aligned}
& \langle \overline{Jg_i(x)v}, \overline{g_i(x)}\rangle - g_{i0}(x)\langle \nabla g_{i0}(x), v\rangle \\
= \;& 4\langle y_{i0}\bar{w}_i + w_{i0}\bar{y}_i, y_{i0}\bar{y}_i\rangle - 2\|y_i\|^2\langle y_i, w_i\rangle \\
= \;& 4y_{i0}^2\langle \bar{y}_i, \bar{w}_i\rangle + 4y_{i0}w_{i0}\|\bar{y}_i\|^2 - 2(y_{i0}^2 + \|\bar{y}_i\|^2)(y_{i0}w_{i0} + \langle \bar{y}_i, \bar{w}_i\rangle) \\
= \;& 2y_{i0}^2\langle \bar{y}_i, \bar{w}_i\rangle + 2y_{i0}w_{i0}\|\bar{y}_i\|^2 - 2\|\bar{y}_i\|^2\langle \bar{y}_i, \bar{w}_i\rangle - 2y_{i0}^2(y_{i0}w_{i0}) \\
= \;& 0,
\end{aligned}$$

where the last equality holds because $y_{i0}^2 = \|\bar{y}_i\|^2$ by Lemma 2.2(c). So, the second condition of (3.3) also holds, and we conclude that $v \in \mathcal{C}(x)$.

First, we consider the case that $v \neq 0$ and $w \in \mathbb{R}^m$ is an arbitrary vector. Since $v \in \mathcal{C}(x)$, by the SOSC-NSOCP of (P1), the inequality

$$\langle \nabla_x^2 L(x, \lambda)v, v\rangle + \sum_{i \in \mathcal{I}_{BB}} \langle H_i(x, \lambda)v, v\rangle > 0 \tag{3.21}$$

holds. Now, observe that adding the term

$$2\sum_{i=1}^{r} \langle w_i \circ w_i, \lambda_i\rangle - \sum_{i \in \mathcal{I}_{BB}} \langle H_i(x, \lambda)v, v\rangle \tag{3.22}$$

to the left-hand side of (3.21) yields that of the inequality (3.4). Recalling that (v, w) was taken arbitrarily so as to satisfy (3.19) and $v \neq 0$, if we prove that the above term is nonnegative, then, by Lemma 3.5, we obtain the SOSC-NLP of (P2). From (3.18) in the proof of Proposition 3.9, we obtain

$$2\langle w_i \circ w_i, \lambda_i\rangle - \langle H_i(x, \lambda)v, v\rangle = 4\kappa_i\langle \Gamma_i y_i, w_i\rangle^2 \geq 0$$

for all $i \in \mathcal{I}_{BB}$, since $\kappa_i > 0$ in this case. So, it remains to show that

$$2\langle w_i \circ w_i, \lambda_i\rangle \geq 0, \quad i = \{1, \dots, r\} \setminus \mathcal{I}_{BB}.$$

But this is true, because from (2.1) and (P1.2), $w_i \circ w_i \in \mathcal{K}_i$ and $\lambda_i \in \mathcal{K}_i$ for all $i = 1, \dots, r$, which implies that $\langle w_i \circ w_i, \lambda_i\rangle \geq 0$ for all $i = 1, \dots, r$. Therefore, the term (3.22) is nonnegative, and SOCP-NLP of (P2) is satisfied.

Now, consider the case that $v = 0$ and $w \in \mathbb{R}^m$ is an arbitrary nonzero vector. From Lemma 3.5, we have to show that

$$\sum_{i=1}^{r} \langle w_i \circ w_i, \lambda_i\rangle > 0$$

for all nonzero $w \in \mathbb{R}^m$ such that $y_i \circ w_i = 0$ with $i \in \mathcal{I}_I \cup \mathcal{I}_B$. Considering also the strict complementarity condition and $\lambda_i = 0$ for any $i \in \mathcal{I}_{I0}$, it means that we have to prove that

$$\sum_{i \in \mathcal{I}_{0I} \cup \mathcal{I}_{BB}} \langle w_i \circ w_i, \lambda_i \rangle > 0 \tag{3.23}$$

for all nonzero $w \in \mathbb{R}^m$ such that $y_i \circ w_i = 0$ with $i \in \mathcal{I}_{I0} \cup \mathcal{I}_{BB}$. In order to prove it, let us analyze the following three cases.

Case 1. Let $i \in \mathcal{I}_{0I}$. In this case, $\lambda_i \in \mathrm{int}(\mathcal{K}_i)$ and $w_i \circ w_i \in \mathcal{K}_i$ show that $\langle w_i \circ w_i, \lambda_i \rangle > 0$ unless $w_i = 0$.

Case 2. Let $i \in \mathcal{I}_{BB}$ and consider $\{c^{(1)}, c^{(2)}\} \subset \mathbb{R}^{m_i}$ a Jordan frame associated to y_i. Since $g_i(x) \in \mathrm{bd}^+(\mathcal{K}_i)$, we can write $y_i = \eta_1 c^{(2)}$ and $g(x) = \eta_1^2 c^{(2)}$ with $\eta_1 \neq 0$. Moreover, $\lambda_i \in \mathrm{bd}^+(\mathcal{K}_i)$ and (P1.4) gives $\lambda_i = \eta_2 c^{(1)}$ with $\eta_2 > 0$, and $y_i \circ w_i = 0$ implies that $w_i = \eta_3 c^{(1)}$ with $\eta_3 \in \mathbb{R}$. Thus, $\langle w_i \circ w_i, \lambda_i \rangle = \eta_2 \eta_3^2 > 0$ unless $w_i = 0$ (i.e., $\eta_3 = 0$).

Case 3. Let $i \in \mathcal{I}_{I0}$. In this case, $y_{i0} \neq 0$ by Assumption 3.10, and from Lemma 2.2(d), $\mathrm{Arw}(y_i)$ is nonsingular. Hence, $0 = y_i \circ w_i = \mathrm{Arw}(y_i) w_i$ implies that $w_i = 0$.

Recalling that the vector w is taken such that $w_i \neq 0$ for some i, we obtain the inequality (3.23), which concludes the proof. □

4 Equivalence Between the Regularity Conditions

We now proceed with results concerning the regularity conditions. For an NLP problem, we say that the *linear independence constraint qualification* (LICQ) holds at a point if the gradients of the equality constraints and the gradients of active inequality constraints are linearly independent (see, for example, [4, Section 3.3] or [19, Section 12.1]). Under the LICQ, the KKT conditions are necessary for optimality. For the NSOCP problem (P1), we recall the definition of nondegeneracy [6], which is shown to be a generalization of LICQ.

Definition 4.1. *Let $x \in \mathbb{R}^n$ be a feasible point of* (P1). *Then, it is called* nondegenerate *if and only if*

$$Jg_i(x)^\top \Gamma_i g_i(x),\ i \in \mathcal{I}_B \quad \text{and} \quad \nabla g_{i,j}(x),\ j = 1, \ldots, m_i,\ i \in \mathcal{I}_0$$

are linearly independent, where $\nabla g_{i,j}(x)$ is the gradient of the j-th entry of g_i at x, and Γ_i is defined as in (3.2).

For problem (P2), we observe that a feasible point $(x, y) \in \mathbb{R}^{n+m}$ satisfies LICQ if and only if the matrix

$$\begin{bmatrix} Jg_1(x) & -2\mathrm{Arw}(y_1) & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ Jg_r(x) & 0 & 0 & -2\mathrm{Arw}(y_r) \end{bmatrix}$$

has full row rank. Without loss of generality, we can write

$$Jg(x)^\top = [Jg_{\mathcal{I}_I}(x)^\top, Jg_{\mathcal{I}_B}(x)^\top, Jg_{\mathcal{I}_0}(x)^\top],$$

where $Jg_{\mathcal{I}_I}(x)$, $Jg_{\mathcal{I}_B}(x)$ and $Jg_{\mathcal{I}_0}(x)$ correspond to the submatrices of $Jg(x)$ consisting of $Jg_i(x)$ with $i \in \mathcal{I}_I$, $i \in \mathcal{I}_B$ and $i \in \mathcal{I}_0$, respectively. Recall that $\mathrm{Arw}(y_i) = 0$ for all $i \in \mathcal{I}_0$. Then, the LICQ condition of (P2) means that the matrix

$$\begin{bmatrix} Jg_{\mathcal{I}_I}(x) & -2\mathrm{diag}(\mathrm{Arw}(y_i))_{i\in\mathcal{I}_I} & 0 \\ Jg_{\mathcal{I}_B}(x) & 0 & -2\mathrm{diag}(\mathrm{Arw}(y_i))_{i\in\mathcal{I}_B} \\ Jg_{\mathcal{I}_0}(x) & 0 & 0 \end{bmatrix} \tag{4.1}$$

has full row rank.

Proposition 4.2. *Let* $(x, y, \lambda) \in \mathbb{R}^{n+2m}$ *be a KKT triple of* (P2) *and assume that it satisfies LICQ and SOSC-NLP. Then,* (x, λ) *is a KKT pair of* (P1) *that satisfies the nondegeneracy condition.*

Proof. From Proposition 3.6, it is sufficient to prove that (x, λ) is nondegenerate. First, recall that the LICQ condition of (P2), satisfied by (x, y, λ), means that the matrix (4.1) has full row rank. In particular, we can say that its submatrix

$$\begin{bmatrix} Jg_{\mathcal{I}_B}(x) & -2\mathrm{diag}(\mathrm{Arw}(y_i))_{i\in\mathcal{I}_B} \\ Jg_{\mathcal{I}_0}(x) & 0 \end{bmatrix}$$

has linearly independent rows. More precisely, denoting the j-th column of $\mathrm{Arw}(y_i)$ by $[\mathrm{Arw}(y_i)]_j$, if the equalities

$$\begin{aligned} &\sum_{i\in\mathcal{I}_B}\sum_{j=1}^{m_i} \tilde{\psi}_{i,j}\nabla g_{i,j}(x) + \sum_{i\in\mathcal{I}_0}\sum_{j=1}^{m_i} \tilde{\nu}_{i,j}\nabla g_{i,j}(x) = 0, \\ &\sum_{j=1}^{m_i} \tilde{\psi}_{i,j}[\mathrm{Arw}(y_i)]_j = 0, \quad \text{for all } i \in \mathcal{I}_B \end{aligned} \tag{4.2}$$

hold, with $\tilde{\psi}_{i,j} \in \mathbb{R}$, $j = 1, \ldots, m_i$, $i \in \mathcal{I}_B$, and $\tilde{\nu}_{i,j} \in \mathbb{R}$, $j = 1, \ldots, m_i$, $i \in \mathcal{I}_0$, then these scalars are all zero.

Now assume, for the purpose of contradiction, that the nondegeneracy condition of (P1) does not hold, which means that there exist scalars θ_i, $i \in \mathcal{I}_B$ and $\nu_{i,j}$, $j = 1, \ldots, m_i$, $i \in \mathcal{I}_0$, not all zero such that

$$\sum_{i\in\mathcal{I}_B} \theta_i Jg_i(x)^\top \Gamma_i g_i(x) + \sum_{i\in\mathcal{I}_0}\sum_{j=1}^{m_i} \nu_{i,j}\nabla g_{i,j}(x) = 0. \tag{4.3}$$

For all $i \in \mathcal{I}_B$, define

$$\psi_{i,1} := \theta_i g_{i0}(x) \quad \text{and} \quad \psi_{i,j} := -\theta_i g_{i,j}(x) = -\frac{\psi_{i,1} g_{i,j}(x)}{g_{i0}(x)}, \quad j = 2, \ldots, m_i. \tag{4.4}$$

Then, in view of the definition (3.2) of Γ_i, we can restate that the nondegeneracy condition does not hold if and only if there exist scalars $\psi_{i,j}$, $j = 1, \ldots, m_i$, $i \in \mathcal{I}_B$, and $\nu_{i,j}$, $j =$

$1, \ldots, m_i$, $i \in \mathcal{I}_0$, not all zero such that

$$\sum_{i \in \mathcal{I}_B} \sum_{j=1}^{m_i} \psi_{i,j} \nabla g_{i,j}(x) + \sum_{i \in \mathcal{I}_0} \sum_{j=1}^{m_i} \nu_{i,j} \nabla g_{i,j}(x) = 0. \tag{4.5}$$

Now, observe that for all $i \in \mathcal{I}_B$,

$$g_i(x) = [\mathrm{Arw}(g_i(x))]_1 = \sum_{j=2}^{m_i} \frac{g_{i,j}(x)}{g_{i0}(x)} [\mathrm{Arw}(g_i(x))]_j.$$

From Lemma 2.2(c), we have $\mathrm{Arw}(g_i(x)) = 2y_{i0}\mathrm{Arw}(y_i)$ and thus,

$$[\mathrm{Arw}(y_i)]_1 = \sum_{j=2}^{m_i} \frac{g_{i,j}(x)}{g_{i0}(x)} [\mathrm{Arw}(y_i)]_j,$$

which along with (4.4) yields

$$\sum_{j=1}^{m_i} \psi_{i,j} [\mathrm{Arw}(y_i)]_j = 0 \quad \text{for all } i \in \mathcal{I}_B.$$

This equality, together with (4.5), contradicts the fact that (4.2) holds only if $\tilde{\psi}_{i,j}$, $j = 1, \ldots, m_i$, $i \in \mathcal{I}_B$, and $\tilde{\nu}_{i,j}$, $j = 1, \ldots, m_i$, $i \in \mathcal{I}_0$, are all zero. Hence, the KKT triple (x, y, λ) is nondegenerate. $\square$

Unfortunately, the converse implication is not true. We give a counterexample to illustrate this fact.

Example 4.3. *Let $r = 1$, $n = 3$, and $m = m_1 = 3$, with*

$$f(x) := x_1^2 + x_2^2 + x_3^2, \quad g(x) = g_1(x) := \begin{pmatrix} 2 + x_1 \\ x_1 - x_2^2 \\ -x_1 + x_3^3 \end{pmatrix}.$$

If $x^* = (0,0,0)$, $\lambda^* = (0,0,0)$, and $y^* = (0,1,-1)$, then it is easy to see that (x^*, λ^*) and (x^*, y^*, λ^*) satisfy the KKT conditions of problems (P1) *and* (P2)*, respectively. Moreover, we have $g(x^*) = (2,0,0)$, $\mathcal{I}_I = \{1\}$ and $\mathcal{I}_B = \mathcal{I}_0 = \emptyset$. Clearly, it satisfies the nondegeneracy condition of* (P1)*. On the other hand, the LICQ condition of* (P2) *means that the matrix*

$$\left[\, Jg(x^*), -2\mathrm{Arw}(y^*) \,\right] = \begin{bmatrix} 1 & 0 & 0 & 0 & -2 & 2 \\ 1 & 0 & 0 & -2 & 0 & 0 \\ -1 & 0 & 0 & 2 & 0 & 0 \end{bmatrix}$$

has linearly independent rows, which is not true. Observe that if we have $y^* = (0,1,1)$ instead of $y^* = (0,1,-1)$, then the KKT conditions and the nondegeneracy condition of (P1) *still hold; moreover, the LICQ condition of* (P2) *also holds in this case.*

If the original problem is an NLP problem instead of an NSOCP problem, then the situation is different. In such a case, the LICQ conditions of the original and reformulated problems at KKT points are equivalent [4, Section 3.3]. The equivalence in the NSOCP case is not guaranteed when the index set $\mathcal{I}_I$, which is equivalent to $\mathcal{I}_{I0}$ at a KKT point, is nonempty, and a slack variable y_i, $i \in \mathcal{I}_I$ satisfies $y_{i0} = 0$. However, the equivalence can be established if the second-order sufficient condition is assumed.

Proposition 4.4. *Let* $(x, \lambda) \in \mathbb{R}^{n+m}$ *be a KKT pair of* (P1) *that satisfies the nondegeneracy condition and SOSC-NSOCP. Then, there exists* $y \in \mathbb{R}^m$ *such that* (x, y, λ) *is a KKT triple of* (P2) *that satisfies LICQ.*

Proof. The existence of y such that (x, y, λ) is a KKT triple of (P2) follows from Proposition 3.1. From the nondegeneracy assumption of (P1), if the equality

$$\sum_{i\in\mathcal{I}_B} \tilde{\theta}_i J g_i(x)^\top \Gamma_i g_i(x) + \sum_{i\in\mathcal{I}_0}\sum_{j=1}^{m_i} \tilde{\nu}_{i,j} \nabla g_{i,j}(x) = 0$$

holds with $\tilde{\theta}_i \in \mathbb{R}$, $i \in \mathcal{I}_B$ and $\tilde{\nu}_{i,j} \in \mathbb{R}$, $j = 1, \dots, m_i$, $i \in \mathcal{I}_0$, then all these scalars are zero. For all $i \in \mathcal{I}_B$, define $\tilde{\psi}_{i,1} := \tilde{\theta}_i g_{i0}(x)$ and $\tilde{\psi}_{i,j} := -\tilde{\theta}_i g_{i,j}(x)$, $j = 2, \dots, m_i$. With these scalars, the nondegeneracy assumption of (P1) means that if

$$\sum_{i\in\mathcal{I}_B}\sum_{j=1}^{m_i} \tilde{\psi}_{i,j} \nabla g_{i,j}(x) + \sum_{i\in\mathcal{I}_0}\sum_{j=1}^{m_i} \tilde{\nu}_{i,j} \nabla g_{i,j}(x) = 0,$$

then the scalars $\tilde{\psi}_{i,j}, \tilde{\nu}_{i,j}$ are all zero.

Assume, for the purpose of contradiction, that (x, y, λ) does not satisfy LICQ of (P2), which means that the matrix (4.1) has linearly dependent rows. Then, there exist $\psi_{i,j} \in \mathbb{R}$, $j = 1, \dots, m_i$, $i \in \mathcal{I}_B$, $\nu_{i,j} \in \mathbb{R}$, $j = 1, \dots, m_i$, $i \in \mathcal{I}_0$, and $\mu_{i,j} \in \mathbb{R}$, $j = 1, \dots, m_i$, $i \in \mathcal{I}_I$, not all zero, such that

$$\sum_{i\in\mathcal{I}_B}\sum_{j=1}^{m_i} \psi_{i,j} \nabla g_{i,j}(x) + \sum_{i\in\mathcal{I}_0}\sum_{j=1}^{m_i} \nu_{i,j} \nabla g_{i,j}(x) + \sum_{i\in\mathcal{I}_I}\sum_{j=1}^{m_i} \mu_{i,j} \nabla g_{i,j}(x) = 0, \tag{4.6}$$

$$\sum_{j=1}^{m_i} \psi_{i,j} [\mathrm{Arw}(y_i)]_j = 0, \quad \text{for all } i \in \mathcal{I}_B,$$

$$\sum_{j=1}^{m_i} \mu_{i,j} [\mathrm{Arw}(y_i)]_j = 0, \quad \text{for all } i \in \mathcal{I}_I. \tag{4.7}$$

Let us consider two cases.

Case 1. Assume that $\mu_{i,j} = 0$ for all $j = 1, \dots, m_i$, $i \in \mathcal{I}_I$. Then, from (4.6) and the nondegeneracy assumption, $\psi_{i,j}$, $\nu_{i,j}$ are all zero, and we obtain a contradiction.

Case 2. Assume that there exists $i \in \mathcal{I}_I$ such that $\mu_{i,j} \neq 0$ for some j. For such i, it follows from (4.7) that $\mathrm{Arw}(y_i)$ is singular. This means that either $y_i \in \mathrm{bd}^+(\mathcal{K}_i)$ or $y_i \in \mathrm{bd}^+(-\mathcal{K}_i)$ or $y_i = 0$ or y_i satisfies $y_{i0} = 0$ with $m_i > 2$. Since $g_i(x) = y_i \circ y_i$, if one of the first three conditions holds, then $g_i(x) \in \mathrm{bd}^+(\mathcal{K}_i)$ or $g_i(x) = 0$, and we obtain a contradiction. So, we must have $m_i > 2$ and $y_{i0} = 0$. However, since SOSC-NSOCP of (P1) holds, by Proposition 3.11, SOSC-NLP of (P2) also holds. We then obtain a contradiction from Corollary 3.7.

Therefore, we can conclude that (x, y, λ) satisfies LICQ of (P2). □

Finally, summarizing the discussions of this section and the results of Section 3, we state the main result of this work.

Theorem 4.5. *The following statements hold.*

(a) Let $(x, \lambda) \in \mathbb{R}^{n+m}$ be a KKT pair of (P1)*. Assume that it satisfies the nondegeneracy condition, the strict complementarity and SOSC-NSOCP. Then, there exists $y \in \mathbb{R}^m$ such that (x, y, λ) is a KKT triple of* (P2) *satisfying LICQ and SOSC-NLP whenever Assumption 3.10 holds.*

(b) Let $(x, y, \lambda) \in \mathbb{R}^{n+2m}$ be a KKT triple of (P2)*. Assume that it satisfies LICQ and SOSC-NLP. Then, (x, λ) is a KKT pair of* (P1) *satisfying the nondegeneracy condition, SOSC-NSOCP and the strict complementarity.*

Proof. The item (a) follows from Propositions 3.11 and 4.4, and the item (b) follows from Propositions 3.9 and 4.2. □

It is clear that the above theorem also holds when the original problem is an NLP problem. In fact, in such a case, the nondegeneracy condition means LICQ, while SOSC-NSOCP reduces to SOSC-NLP. To the authors' knowledge, the same analysis for NLP problems has apparently not been published in the literature, except for some particular results shown, for example, in [4, Section 3.3]. A comprehensive treatment of the NLP case is also under investigation.

5 Numerical Experiments

In this section, we present some preliminary numerical results to examine the validity of squared slack variables technique. The problems data were all modeled in AMPL [11], and the NLP solver that we used to solve the reformulated problem is the ALGENCAN [2, 3], which is an augmented Lagrangian method implemented in Fortran. The experiments were carried out on a computer with core i7 3.4GHz, 16GB of RAM, and with Linux (Ubuntu 14.04) as the operating system.

Example 5.1. *Let us consider Example 4.1 from [13]. The problem has a nonconvex quadratic objective function and the single second-order cone constraint $x \in \mathcal{K}^3$, where $\mathcal{K}^3$ denotes*[3] *the second-order cone in $\mathbb{R}^3$.*

The solution of the above example is shown to be unique, satisfying the strict complementarity and the SOSC-NSOCP. Using squared slack variables, we obtain a reformulated NLP problem with six variables, i.e., $(x, y) \in \mathbb{R}^3 \times \mathbb{R}^3$. Using ALGENCAN as the NLP solver, we have the results shown in Table 1, where k denotes the outer iteration count. The corresponding objective function value, the infeasibility measure, and computed point $x^k = (x_1^k, x_2^k, x_3^k)$ at iteration k are presented, but the slack variables $y^k = (y_1^k, y_2^k, y_3^k)$ are omitted to save space. The starting point, shown as x^0 in Table 1, is feasible and chosen randomly from the box $[0, 2]^6 \subset \mathbb{R}^6$. We observe that the obtained solution is exactly the one shown in [13].

k	obj. function	infeasibility	x_1^k	x_2^k	x_3^k
0	1.716761e+00	4.440892e–16	1.782184e+00	8.129517e–01	1.517723e+00
1	6.844777e–01	2.217315e–01	8.068516e–01	1.150896e+00	7.786648e–02
2	1.002432e+00	2.595151e–02	1.001192e+00	9.987463e–01	-7.996669e–03
3	1.000359e+00	9.468084e–04	1.000180e+00	9.998198e–01	1.935266e–03
4	1.000025e+00	2.019148e–04	1.000017e+00	9.999921e–01	-3.236138e–04
5	1.000028e+00	1.706357e–05	1.000016e+00	9.999877e–01	-2.864587e–04
6	1.000005e+00	4.974331e–05	1.000005e+00	1.000000e+00	1.001607e–04
7	1.000000e+00	9.124102e–07	1.000000e+00	1.000000e+00	-2.181132e–06
8	1.000000e+00	2.458561e–08	1.000000e+00	1.000000e+00	4.929157e–08
9	1.000000e+00	2.515123e–10	1.000000e+00	1.000000e+00	3.920419e–08

Table 1: Numerical results for Example 5.1, a nonconvex NSOCP problem.

Example 5.2. *We now consider Example 4.2 from [13]. The problem has a nonlinear convex objective function, containing exponentials and polynomials of order up to 4, and the second-order cone constraint $x \in \mathcal{K}^3 \times \mathcal{K}^2$.*

Applying the squared slack variables technique, we obtain a reformulated NLP problem with ten variables, i.e., $(x, y) \in \mathbb{R}^5 \times \mathbb{R}^5$. Table 2 shows the results with ALGENCAN, where we omit x_4^k, x_5^k, as well as the slack variable y^k. The starting point is feasible and generated randomly from the box $[0, 2]^{10} \subset \mathbb{R}^{10}$. Here, we also could find the solution shown in [13].

Another experiment was done with the Example 4.3 from [13], which has a linear objective function, the second-order cone constraint $x \in \mathcal{K}^3 \times \mathcal{K}^3 \times \mathcal{K}^3$, and additional linear equality constraints. The strict complementarity condition does not hold for this problem. Also in this case, the squared slack variable technique using ALGENCAN could find successfully the unique solution in 11 iterations. We omit the details of the experimental results.

[3] Note the difference between $\mathcal{K}^\ell$ and $\mathcal{K}_\ell$. The former denotes the second-order cone in $\mathbb{R}^\ell$, and the latter means the ℓ-th second-order cone in the Cartesian product $\mathcal{K} = \mathcal{K}_1 \times \cdots \times \mathcal{K}_r$.

k	obj. function	infeasibility	x_1^k	x_2^k	x_3^k
0	4.142282e+00	7.993606e–15	5.114961e–01	4.075672e–01	1.711482e–01
1	2.363085e+00	2.251076e–01	1.882522e–01	-7.577021e–02	1.928192e–01
2	2.478229e+00	8.897084e–02	2.054357e–01	-8.573544e–02	2.210883e–01
3	2.532257e+00	5.650648e–02	2.194107e–01	-8.961111e–02	2.193602e–01
4	2.597236e+00	7.737408e–03	2.340394e–01	-9.711915e–02	2.156772e–01
5	2.607188e+00	1.178006e–03	2.360426e–01	-9.820783e–02	2.150642e–01
6	2.605668e+00	7.385796e–03	2.341680e–01	-7.099665e–02	2.203726e–01
7	2.597607e+00	2.284736e–05	2.324113e–01	-7.310262e–02	2.206101e–01
8	2.597576e+00	7.211111e–07	2.324027e–01	-7.307876e–02	2.206136e–01
9	2.597575e+00	6.413911e–09	2.324025e–01	-7.307919e–02	2.206136e–01
10	2.597575e+00	2.919317e–09	2.324025e–01	-7.307929e–02	2.206135e–01
11	2.597575e+00	1.181701e–11	2.324025e–01	-7.307929e–02	2.206135e–01

Table 2: Numerical results for Example 5.2, a convex NSOCP problem.

The above mentioned problems were also solved by replacing ALGENCAN with MINOS, a reduced-gradient-type method [18], implemented in Fortran. For all the problems, MINOS could also obtain the same solutions.

Next, we consider examples with larger dimensions that can be generated randomly. The following problem is presented in [14], and also used in [12].

Example 5.3. *Let us consider the following convex NSOCP:*

$$\begin{array}{ll} \underset{x}{\textit{minimize}} & \langle Cx, x\rangle + \sum_{i=1}^{n} (p_i x_i^4 + q_i x_i) \\ \textit{subject to} & A_i x + b_i \in \mathcal{K}_i, \quad i = 1, \ldots, r, \end{array} \tag{5.1}$$

where $C \in \mathbb{R}^{n\times n}$ is symmetric positive semidefinite, $p_i, q_i \in \mathbb{R}$ for all $i = 1, \ldots, n$, and $A_i \in \mathbb{R}^{m_i\times n}$, $b_i \in \mathbb{R}^{m_i}$ for all $i = 1, \ldots, r$. Also, we have $\mathcal{K} = \mathcal{K}_1 \times \cdots \times \mathcal{K}_r$ and $m_1 + \cdots + m_r = n$. Here, the scalars p_i, q_i, and the entries of A_i are generated randomly from the intervals $[0, 1]$, $[-1, 1]$ and $[0, 2]$, respectively. The matrix C is defined by $C := Z^\top Z$, where the entries of $Z \in \mathbb{R}^{n\times n}$ are chosen randomly from the interval $[0, 1]$. Moreover, the vectors b_i are defined as $b_{i0} = 1$ and $\bar{b}_i = 0$. Then, $x = 0$ is a feasible point of the problem.

Observe that this problem allows for a variable number of cone constraints. We consider problems with $\mathcal{K} = \mathcal{K}^5 \times \mathcal{K}^5$, $\mathcal{K} = \mathcal{K}^5 \times \mathcal{K}^5 \times \mathcal{K}^{20}$, and $\mathcal{K} = \mathcal{K}^5 \times \mathcal{K}^5 \times \mathcal{K}^{20} \times \mathcal{K}^{20}$. Thus, the squared slack variable technique gives NLP reformulations with 20, 60 and 100 variables, respectively. For each $\mathcal{K}$, we generated 10 problems randomly, with random feasible starting points. Once again, we used solver ALGENCAN, and the results are given in Table 3. It shows the median, minimum and maximum numbers of outer and inner iterations needed to solve the problems. We recall that an outer iteration of ALGENCAN consists of updating Lagrange multipliers and penalty parameters, while an inner iteration consists of solving an unconstrained (or box-constrained) minimization problem.

$\mathcal{K}$	outer iterations			inner iterations		
	median	min	max	median	min	max
$\mathcal{K}^5 \times \mathcal{K}^5$	7	6	8	112	63	3535
$\mathcal{K}^5 \times \mathcal{K}^5 \times \mathcal{K}^{20}$	7	6	8	246	113	2482
$\mathcal{K}^5 \times \mathcal{K}^5 \times \mathcal{K}^{20} \times \mathcal{K}^{20}$	7	5	13	308	173	9168

Table 3: Numerical results for Example 5.3, a convex NSOCP problem.

In this case, we were able to find solutions for all the generated problems. In order to check the validity of these solutions, we solve the same problems with the exact penalty method [12], developed for NSOCP problems, and implemented in Python. We should note here that the method deals with the original second-order cone problems directly, without reformulating them as NLP problems. So, the number of variables is equal to n, instead of $2n$. Since the objective functions of the problems are all convex, we can expect to obtain the same optimal values as those obtained by solving the reformulated problems using ALGENCAN. The result matches the expectation, showing the validity of the squared slack variables strategy. However, concerning the computational time, the exact penalty method performed better than ALGENCAN in 21 out of 30 problems.

We now consider nonconvex NSOCP problems generated randomly.

Example 5.4. *Let us consider problem* (5.1) *defined as in Example 5.3, except that* $C \in \mathbb{R}^{n \times n}$ *is a symmetric indefinite matrix, with entries chosen randomly from the interval* $[-1, 1]$. *The problem turns out to be a nonconvex NSOCP problem.*

In the same way as before, we use ALGENCAN as the NLP solver, and the results are shown in Table 4. We could also find solutions for all the generated problems, and most of them are exactly the same as the solutions found by the exact penalty method. Different solutions were found for some problems, but all of them have been confirmed to be, at least, KKT points of the original NSOCP problems. An interesting result here is that, in contrast with Example 5.3, ALGENCAN performed better than the exact penalty method in 23 out of 30 problems, in terms of computational time. This shows that, although the number of variables increases with the addition of squared slack variables, efficient NLP solvers, like ALGENCAN, could still outperform a recent method developed for NSOCP, like the exact penalty method.

$\mathcal{K}$	outer iterations			inner iterations		
	median	min	max	median	min	max
$\mathcal{K}^5 \times \mathcal{K}^5$	7	6	9	84.5	53	581
$\mathcal{K}^5 \times \mathcal{K}^5 \times \mathcal{K}^{20}$	7	6	8	162.5	91	1291
$\mathcal{K}^5 \times \mathcal{K}^5 \times \mathcal{K}^{20} \times \mathcal{K}^{20}$	7	7	7	231.5	175	2316

Table 4: Numerical results for Example 5.4, a nonconvex NSOCP problem.

In the above experiments, we observe that the squared slack variables approach can really solve NSOCP problems successfully. The comparison with the exact penalty method was made in order to see that efficient NLP solvers can cover the disadvantage of increasing the number of variables. Depending on the problem, as well as the solver and its programming language, one may prefer to adopt the squared slack variables approach. As usual, the squared variables technique may not be suitable for some problems, because of possible instabilities and singularities. Nevertheless, we believe that it should be a try for many of the users, because at least in the present moment, NSOCP algorithms and solvers are much less developed than the NLP ones.

6 Conclusions

We have analyzed the use of squared slack variables in the context of NSOCP. We have proved that, under the second-order sufficient conditions and the regularity conditions, KKT points of the original and the reformulated problems are essentially equivalent. Moreover, we have presented preliminary numerical experiments, showing that the approach can be competitive in practice, depending on the choice of the NLP solver. Concerning the theoretical analysis, we expect that the tools developed here is possibly valuable in other analysis involving NSOCP problems, or even in problems with other symmetric cones.

References

[1] F. Alizadeh and D. Goldfarb. Second-order cone programming. *Mathematical Programming*, 95(1):3–51, 2003.

[2] R. Andreani, E. G. Birgin, J. M. Martínez, and M.L. Schuverdt. On augmented Lagrangian methods with general lower-level constraints. *SIAM Journal on Optimization*, 18(4):1286–1309, 2007.

[3] R. Andreani, E. G. Birgin, J. M. Martínez, and M.L. Schuverdt. Augmented Lagrangian methods under the constant positive linear dependence constraint qualification. *Mathematical Programming*, 111(1-2):5–32, 2008.

[4] D. P. Bertsekas. *Nonlinear Programming.* Athena Scientific, 2nd edition, 1999.

[5] D. Bertsimas and J. N. Tsitsiklis. *Introduction to Linear Optimization.* Athena Scientific, 1st edition, 1997.

[6] J. F. Bonnans and H. Ramírez C. Perturbation analysis of second-order cone programming problems. *Mathematical Programming*, 104:205–227, 2005.

[7] J. F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems.* Springer-Verlag, New York, 2000.

[8] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, 2004.

[9] A. R. Conn, N. Gould, and Ph. L. Toint. A note on exploiting structure when using slack variables. *Mathematical Programming*, 67:89–97, 1994.

[10] G. B. Dantzig. *Linear Programming and Extensions*. Princeton University Press, Princeton, NJ, 1963.

[11] R. Fourer, D. M. Gay, and B. W. Kernighan. A modeling language for mathematical programming. *Management Science*, 36:519–554, 1990.

[12] E. H. Fukuda, P. J. S. Silva, and M. Fukushima. Differentiable exact penalty functions for nonlinear second-order cone programs. *SIAM Journal on Optimization*, 22(4):1607–1633, 2012.

[13] C. Kanzow, I. Ferenczi, and M. Fukushima. On the local convergence of semismooth Newton methods for linear and nonlinear second-order cone programs without strict complementarity. *SIAM Journal on Optimization*, 20(1):297–320, 2009.

[14] H. Kato and M. Fukushima. An SQP-type algorithm for nonlinear second-order cone programs. *Optimization Letters*, 1(2):129–144, 2007.

[15] Y. Z. Liu and L. W. Zhang. Convergence of the augmented Lagrangian method for nonlinear optimization problems over second-order cones. *Journal of Optimization Theory and Applications*, 139(3):557–575, 2008.

[16] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret. Applications of second-order cone programming. *Linear Algebra and Its Applications*, 284(1-3):193–228, 1998.

[17] O. L. Mangasarian. Unconstrained Lagrangians in nonlinear programming. *SIAM Journal on Control*, 13(4):772–791, 1975.

[18] B. A. Murtagh and M. A. Saunders. Large-scale linearly constrained optimization. *Mathematical Programming*, 14:41–72, 1978.

[19] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer Verlag, New York, 1st edition, 1999.

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

[21] E. Spedicato. On a Newton-like method for constrained nonlinear minimization via slack variables. *Journal of Optimization Theory and Applications*, 36(2):175–190, 1982.

[22] R. A. Tapia. A stable approach to Newton's method for general mathematical programming problems in $\mathbb{R}^n$. *Journal of Optimization Theory and Applications*, 14:453–476, 1974.

[23] R. A. Tapia. On the role of slack variables in quasi-Newton methods for constrained optimization. In L. C. W. Dixon and G. P. Szegö, editors, *Numerical Optimisation of Dynamic Systems*, pages 235–246. North-Holland Publishing Company, 1980.

[24] H. Yamashita and H. Yabe. A primal-dual interior point method for nonlinear optimization over second-order cones. *Optimization Methods & Software*, 24(3):407–426, 2009.