

# Exact Augmented Lagrangian Functions for Nonlinear Semidefinite Programming\*

Ellen H. Fukuda<sup>†</sup>      Bruno F. Lourenço<sup>‡</sup>

June 20, 2018

## Abstract

In this paper, we study augmented Lagrangian functions for nonlinear semidefinite programming (NSDP) problems with exactness properties. The term *exact* is used in the sense that the penalty parameter can be taken appropriately, so a single minimization of the augmented Lagrangian recovers a solution of the original problem. This leads to reformulations of NSDP problems into unconstrained nonlinear programming ones. Here, we first establish a unified framework for constructing these exact functions, generalizing Di Pillo and Lucidi's work from 1996, that was aimed at solving nonlinear programming problems. Then, through our framework, we propose a practical augmented Lagrangian function for NSDP, proving that it is continuously differentiable and exact under the so-called nondegeneracy condition. We also present some preliminary numerical experiments.

**Keywords:** Differentiable exact merit functions, generalized augmented Lagrangian functions, nonlinear semidefinite programming.

## 1 Introduction

The following *nonlinear semidefinite programming* (NSDP) problem is considered:

$$\begin{aligned} & \underset{x}{\text{minimize}} && f(x) \\ & \text{subject to} && G(x) \in \mathbb{S}_+^m, \end{aligned} \tag{NSDP}$$

where  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  and  $G: \mathbb{R}^n \rightarrow \mathbb{S}^m$  are twice continuously differentiable functions,  $\mathbb{S}^m$  is the linear space of all real symmetric matrices of dimension  $m \times m$ , and  $\mathbb{S}_+^m$  is the cone

---

\*This is a pre-print of an article published in Computational Optimization and Applications. The final authenticated version is available online at: <https://doi.org/10.1007/s10589-018-0017-z>. This work was supported by the Grant-in-Aid for Young Scientists (B) (26730012) and for Scientific Research (B) (15H02968) from Japan Society for the Promotion of Science.

<sup>†</sup>Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan ([ellen@i.kyoto-u.ac.jp](mailto:ellen@i.kyoto-u.ac.jp)).

<sup>‡</sup>Department of Mathematical Informatics, Graduate School of Information Science & Technology, University of Tokyo, Tokyo 113-8656, Japan ([lourenco@mist.i.u-tokyo.ac.jp](mailto:lourenco@mist.i.u-tokyo.ac.jp)).

of all positive semidefinite matrices in  $\mathbb{S}^m$ . For simplicity, here we do not take equality constraints into consideration. The above problem extends the well-known nonlinear programming (NLP) and the linear semidefinite programming (linear SDP) problems. NLP and linear SDP models are certainly important, but they may be insufficient in applications where more general constraints are necessary. In particular, in the recent literature, some applications of NSDP are considered in different fields, such as control theory [3, 15], structural optimization [22, 24], truss design problems [4], and finance [23]. However, compared to NLP and linear SDP models, there are still few methods available to solve NSDP, and the theory behind them requires more investigation.

Some theoretical issues associated to NSDP, like optimality conditions, are discussed in [7, 17, 21, 26, 28]. There are, in fact, some methods for NSDPs proposed in the literature such as primal-dual interior-point, augmented Lagrangian, filter-based, sequential quadratic programming, and exact penalty methods. Nevertheless, there are few implementations and, as far as we know, only two general-purpose solvers are able to handle nonlinear semidefinite constraints: PENLAB/PENNON [16] and NuOpt [32]. For a complete survey, see Yamashita and Yabe [31], and references therein.

Here, our main object of interest is the so-called *augmented Lagrangian functions* and this work can be seen as a stepping stone towards new algorithms for NSDPs. An augmented Lagrangian function is basically the usual Lagrangian function with an additional term that depends on a positive coefficient, called the *penalty parameter*. When there exists an appropriate choice of the parameter, such that a single minimization of the augmented Lagrangian recovers a solution to the original problem, then we say that this function is *exact*. This is actually the same definition of the so-called *exact penalty* function. The difference is that an exact augmented Lagrangian function is defined on the product space of the problem's variables and the Lagrange multipliers, and an exact penalty function is defined on the same space of the original problem's variables. Both exact functions, which are also called *exact merit* functions, have been studied quite extensively when the original problem is an NLP.

The first proposed exact merit functions were nondifferentiable, and the basic idea was to incorporate terms into the objective function that penalize constraint violations. However, unconstrained minimization of nondifferentiable functions demands special methods, and so, continuously differentiable exact functions were considered subsequently. For NLPs, both exact penalty and exact augmented Lagrangian functions were studied. The first one has an advantage of having to deal with less variables, but it tends to have a more complicated formula, because the information of the Lagrange multipliers is, in some sense, hidden in the formula. In most exact penalty functions, this is done by using a function that estimates the value of the Lagrange multipliers associated to a point [11]. The evaluation of this estimate is, however, computationally expensive.

To overcome such a drawback, exact augmented Lagrangian functions can be considered, with a price of increasing the number of variables. The choice between these two types of exact functions depends, of course, on the optimization problem at hand. So, exact augmented Lagrangian functions were proposed in [9] and [10], by Di Pillo and Grippo for NLP

problems with equality and inequality constraints, respectively. They were further investigated in [6, 12, 14, 27], with additional theoretical issues and schemes for box-constrained NLP problems. However, as far as we know, there are no proposals for exact augmented Lagrangian functions for more general conic constrained problems, in particular, for NSDP. The augmented Lagrangian function considered by Correa and Ramírez [8], and Shapiro and Sun [29], for example, is not exact.

In this paper, we introduce a continuously differentiable exact augmented Lagrangian function for NSDP problems. We also give a unified framework for constructing such functions. More precisely, we propose a generalized augmented Lagrangian function for NSDP, and give conditions for it to be exact. The main difference between the classical (and not exact) augmented Lagrangian and this exact version is the addition of a term, that we define in Section 3 as  $\gamma$ . This is a continuously differentiable function defined in the product space of problem's variables and the Lagrange multipliers, with key properties that guarantee the exactness of the augmented Lagrangian function. A general framework with such  $\gamma$  term was also given by Di Pillo and Lucidi in [13] for the NLP case. Besides the optimization problem, a difference between [13] and our work is that, here, we propose the generalization first, and then construct one particular exact augmented Lagrangian function. We believe that the generalized function can be used in the future to easily build other exact merit functions, together with possibly useful methods. Meanwhile, we make some preliminary numerical experiments with the particular exact function, using a quasi-Newton method.

The paper is organized as follows. In Section 2, we start with basic definitions and necessary results associated to NSDP problems. In Section 3, a general framework for constructing augmented Lagrangian with exactness properties is given. A practical exact augmented Lagrangian as well as its exactness results are given in Section 4. This particular function is used in Section 5, where some numerical examples are presented. We conclude in Section 6, with some final remarks.

## 2 Preliminaries

Let us first present some basic notations that will be used throughout the paper. Let  $x \in \mathbb{R}^r$  be a  $r$ -dimensional column vector and  $Z \in \mathbb{S}^s$  a symmetric matrix with dimension  $s \times s$ . We use  $x_i$  and  $Z_{ij}$  to denote the  $i$ th element of  $x$  and  $(i, j)$  entry ( $i$ th row and  $j$ th column) of  $Z$ , respectively. We also use the notation  $[x_i]_{i=1}^r$  and  $[Z_{ij}]_{i,j=1}^s$  to denote  $x$  and  $Z$ , respectively. The trace of  $Z$  is denoted by  $\text{tr}(Z) := \sum_{i=1}^s Z_{ii}$ . Moreover, if  $Y \in \mathbb{S}^s$ , then the inner product of  $Y$  and  $Z$  is written as  $\langle Y, Z \rangle := \text{tr}(YZ)$ , and the Frobenius norm of  $Z$  is given by  $\|Z\|_F := \langle Z, Z \rangle^{1/2}$ . The identity matrix, with dimension defined in each context, is denoted by  $I$ , and  $P_{\mathbb{S}_+^m}$  denotes the projection onto the cone  $\mathbb{S}_+^m$ .

For a function  $p: \mathbb{R}^s \rightarrow \mathbb{R}$ , its gradient and Hessian at a point  $x \in \mathbb{R}^s$  are given by  $\nabla p(x) \in \mathbb{R}^s$  and  $\nabla^2 p(x) \in \mathbb{R}^{s \times s}$ , respectively. For  $q: \mathbb{S}^\ell \rightarrow \mathbb{R}$ ,  $\nabla q(Z)$  denotes the matrix with  $(i, j)$  term given by the partial derivatives  $\partial q(Z)/\partial Z_{ij}$ . If  $\psi: \mathbb{R}^s \times \mathbb{S}^\ell \rightarrow \mathbb{R}$ , then its gradient at  $(x, Z) \in \mathbb{R}^s \times \mathbb{S}^\ell$  with respect to  $x$  and  $Z$  are denoted by  $\nabla_x \psi(x, Z) \in \mathbb{R}^s$  and  $\nabla_Z \psi(x, Z) \in \mathbb{S}^\ell$ , respectively. Similarly, the Hessian of  $\psi$  at  $(x, Z)$  with respect to  $x$  is

written as  $\nabla_{xx}^2 \psi(x, Z)$ . For any linear operator  $\mathcal{G}: \mathbb{R}^s \rightarrow \mathbb{S}^\ell$  defined by  $\mathcal{G}v = \sum_{i=1}^s v_i \mathcal{G}_i$  with  $\mathcal{G}_i \in \mathbb{S}^\ell$ ,  $i = 1, \dots, s$ , and  $v \in \mathbb{R}^s$ , the adjoint operator  $\mathcal{G}^*$  is defined by

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

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

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

where  $\partial \mathcal{G}(x)/\partial x_i \in \mathbb{S}^\ell$  are the partial derivative matrices.

One important operator that is necessary when dealing with NSDP problems is the *Jordan product* associated to the space  $\mathbb{S}^m$ . For any  $Y, Z \in \mathbb{S}^m$ , it is defined by

$$Y \circ Z := \frac{YZ + ZY}{2}.$$

Taking  $Y \in \mathbb{S}^m$ , we also denote by  $\mathcal{L}_Y: \mathbb{S}^m \rightarrow \mathbb{S}^m$  the linear operator given by

$$\mathcal{L}_Y(Z) := Y \circ Z.$$

Since we are only considering the space  $\mathbb{S}^m$  of symmetric matrices, we have  $\mathcal{L}_Y(Z) = \mathcal{L}_Z(Y)$ . In the following lemmas, we present some useful results associated to this Jordan product and the projection operator  $P_{\mathbb{S}_+^m}$ .

**Lemma 2.1.** *For any matrix  $Z \in \mathbb{R}^{m \times m}$ , the following statements hold:*

- (a)  $P_{\mathbb{S}_+^m}(-Z) = P_{\mathbb{S}_+^m}(Z) - Z$ ;
- (b)  $P_{\mathbb{S}_+^m}(Z) \circ P_{\mathbb{S}_+^m}(-Z) = 0$ .

*Proof.* See [30, Section 1]. □

**Lemma 2.2.** *If  $Y, Z \in \mathbb{S}^m$ , then the following statements are equivalent:*

- (a)  $Y, Z \in \mathbb{S}_+^m$  and  $Y \circ Z = 0$ ;
- (b)  $Y, Z \in \mathbb{S}_+^m$  and  $\langle Y, Z \rangle = 0$ ;
- (c)  $Y - P_{\mathbb{S}_+^m}(Y - Z) = 0$ .

*Proof.* It follows from [5, Section 8.12] and [30, Lemma 2.1(b)]. □

**Lemma 2.3.** *The following statements hold.*

- (a) *Let  $Q: \mathbb{R}^n \rightarrow \mathbb{S}^m$  be a differentiable function, and define  $\psi: \mathbb{R}^n \rightarrow \mathbb{R}$  as  $\psi(x) := \|P_{\mathbb{S}_+^m}(Q(x))\|_F^2$ . Then, the gradient of  $\psi$  at  $x \in \mathbb{R}^n$  is given by*

$$\nabla \psi(x) = 2\nabla Q(x)^* P_{\mathbb{S}_+^m}(Q(x)).$$

*A similar result holds when the domain of the functions  $Q$  and  $\psi$  is changed to  $\mathbb{S}^m$ .*

(b) Let  $R_1, R_2: \mathbb{R}^n \rightarrow \mathbb{S}^m$  be differentiable functions, and define  $P: \mathbb{R}^n \rightarrow \mathbb{S}^m$  as  $P(x) := \mathcal{L}_{R_1(x)}(R_2(x)) = R_1(x) \circ R_2(x)$ . Then, we have

$$\nabla P(x)^* Z = \left[ \left\langle \frac{\partial R_1(x)}{\partial x_i} \circ R_2(x) + R_1(x) \circ \frac{\partial R_2(x)}{\partial x_i}, Z \right\rangle \right]_{i=1}^n \text{ for all } Z \in \mathbb{S}^m.$$

(c) Let  $\xi: \mathbb{R}^n \rightarrow \mathbb{R}$  be a differentiable function, and define  $S: \mathbb{R}^n \rightarrow \mathbb{S}^m$  as  $S(x) := \xi(x)W$ , with  $W \in \mathbb{S}^m$ . Then, we obtain

$$\nabla S(x)^* Z = \langle W, Z \rangle \nabla \xi(x) \text{ for all } Z \in \mathbb{S}^m.$$

(d) Let  $\eta: \mathbb{S}^m \rightarrow \mathbb{R}$  be a differentiable function, and define  $T: \mathbb{S}^m \rightarrow \mathbb{S}^m$  as  $T(Y) := \eta(Y)Y$ . Then, we have

$$\nabla T(Y)^* Z = \langle Y, Z \rangle \nabla \eta(Y) + \eta(Y)Z \text{ for all } Z \in \mathbb{S}^m.$$

*Proof.* Item (a) follows from [25, Corollary 3.2] and item (b) follows easily from the definitions of adjoint operator and Jordan product. Item (c) holds also from the definition of adjoint operator, and because  $\partial S(x)/\partial x_i = (\partial \xi(x)/\partial x_i)W$  for all  $i$ . For item (d), observe that for all  $W \in \mathbb{S}^m$ , we obtain

$$\begin{aligned} \nabla T(Y)(W) &= \lim_{t \downarrow 0} \frac{T(Y + tW) - T(Y)}{t} \\ &= \lim_{t \downarrow 0} \frac{\eta(Y + tW)(Y + tW) - \eta(Y)Y}{t} \\ &= \lim_{t \downarrow 0} \frac{(\eta(Y + tW) - \eta(Y))Y}{t} + \eta(Y + tW)W \\ &= \eta'(Y; W)Y + \eta(Y)W, \end{aligned}$$

where  $\eta'(Y; W)$  is the directional derivative of  $\eta$  at  $Y$  in the direction  $W$ . From the differentiability of  $\eta$ , we have  $\nabla T(Y)(W) = \langle \nabla \eta(Y), W \rangle Y + \eta(Y)W$ . Recalling that  $\nabla T(Y)^*$  denotes the adjoint of  $\nabla T(Y)$ , this equality yields

$$\langle W, \nabla T(Y)^* Z \rangle = \langle \nabla T(Y)(W), Z \rangle = \langle \nabla \eta(Y), W \rangle \langle Y, Z \rangle + \eta(Y) \langle W, Z \rangle$$

for all  $W, Z \in \mathbb{S}^m$ , which completes the proof.  $\square$

Let us return to problem (NSDP). Define  $L: \mathbb{R}^n \times \mathbb{S}^m \rightarrow \mathbb{R}$  as the Lagrangian function associated to problem (NSDP), that is,

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

The pair  $(x, \Lambda) \in \mathbb{R}^n \times \mathbb{S}^m$  satisfies the *KKT conditions* of problem (NSDP) (or, it is a KKT pair) if the following conditions hold:

$$\begin{aligned} \nabla_x L(x, \Lambda) &= 0, \\ \Lambda \circ G(x) &= 0, \\ G(x) &\in \mathbb{S}_+^m, \\ \Lambda &\in \mathbb{S}_+^m, \end{aligned} \tag{2.1}$$

where

$$\nabla_x L(x, \Lambda) = \nabla f(x) - \nabla G(x)^* \Lambda.$$

The above conditions are necessary for optimality under a constraint qualification. Moreover, Lemma 2.2 shows that the condition  $\Lambda \circ G(x) = 0$  can be replaced by  $\langle \Lambda, G(x) \rangle = 0$  because  $G(x) \in \mathbb{S}_+^m$  and  $\Lambda \in \mathbb{S}_+^m$  hold. Furthermore, it can be shown that this condition can also be replaced by  $\Lambda G(x) = 0$  [31, Section 2].

Now, consider the nonlinear programming below:

$$\begin{aligned} &\underset{x, \Lambda}{\text{minimize}} && \Psi_c(x, \Lambda) \\ &\text{subject to} && (x, \Lambda) \in \mathbb{R}^n \times \mathbb{S}^m, \end{aligned} \tag{2.2}$$

where  $\Psi_c: \mathbb{R}^n \times \mathbb{S}^m \rightarrow \mathbb{R}$ , and  $c > 0$  is a penalty parameter. Observe that the above problem is unconstrained, with both the original variable  $x$  and the Lagrange multiplier  $\Lambda$  as variables. As usual, we say that  $(x, \Lambda)$  is *stationary* of  $\Psi_c$  (or for problem (2.2)) when  $\nabla \Psi_c(x, \Lambda) = 0$ . We use  $G_{\text{NLP}}(c)$  and  $L_{\text{NLP}}(c)$  to denote the sets of global and local minimizers, respectively, of problem (2.2). We also define  $G_{\text{NSDP}}$  and  $L_{\text{NSDP}}$  as the set of global and local minimizers of problem (NSDP), respectively. Using such notations, we present the formal definition of exact augmented Lagrangian functions.

**Definition 2.4.** *A function  $\Psi_c: \mathbb{R}^n \times \mathbb{S}^m \rightarrow \mathbb{R}$  is called an exact augmented Lagrangian function associated to (NSDP) if, and only if, there exists  $\hat{c} > 0$  satisfying the following:*

- (a) *For all  $c \geq \hat{c}$ , if  $(x, \Lambda) \in G_{\text{NLP}}(c)$ , then  $x \in G_{\text{NSDP}}$  and  $\Lambda$  is a corresponding Lagrange multiplier. Conversely, if  $x \in G_{\text{NSDP}}$  with  $\Lambda$  as a corresponding Lagrange multiplier, then  $(x, \Lambda) \in G_{\text{NLP}}(c)$  for all  $c \geq \hat{c}$ .*
- (b) *For all  $c \geq \hat{c}$ , if  $(x, \Lambda) \in L_{\text{NLP}}(c)$ , then  $x \in L_{\text{NSDP}}$  and  $\Lambda$  is a corresponding Lagrange multiplier.*

Basically, the above definition shows that  $\Psi_c$  is an exact augmented Lagrangian function when, without considering Lagrange multipliers, there are equivalence between the global minimizers, and if all local solutions of (2.2) are local solutions of (NSDP), for penalty parameters greater than a threshold value. It means that the original constrained conic problem (NSDP) can be replaced with an unconstrained nonlinear programming problem (2.2) when the penalty parameter is chosen appropriately. Note that the definition of exact penalty functions is similar. The only difference is that in the exact penalty case, the objective function of problem (2.2) does not involve Lagrange multipliers explicitly.

### 3 A general framework

In this section, we propose a general formula for continuously differentiable augmented Lagrangian functions associated to NSDP problems, with exactness properties. It can be seen as a generalization of the one proposed by Di Pillo and Lucidi in [13] for NLP problems. With this purpose, let us first define the following function  $\varphi: \mathbb{S}^m \times \mathbb{S}^m \rightarrow \mathbb{R}$ :

$$\varphi(Y, Z) := \left\| P_{\mathbb{S}_+^m} \left( \frac{Z}{2} - Y \right) \right\|_F^2 - \frac{\|Z\|_F^2}{4}. \quad (3.1)$$

Observe that this function is continuously differentiable because  $\|\cdot\|_F^2$  and  $\|P_{\mathbb{S}_+^m}(\cdot)\|_F^2$  are both continuously differentiable. Moreover, it has the properties below.

**Lemma 3.1.** *Let  $\varphi: \mathbb{S}^m \times \mathbb{S}^m \rightarrow \mathbb{R}$  be defined by (3.1). Then, the following statements hold.*

- (a) *If  $Y, Z \in \mathbb{S}_+^m$  and  $\langle Y, Z \rangle = 0$ , then  $\varphi(Y, Z) = 0$ .*
- (b) *If  $Y \in \mathbb{S}_+^m$ , then  $\varphi(Y, Z) \leq 0$  for all  $Z \in \mathbb{S}^m$ .*

*Proof.* (a) Clearly,  $Z/2 \in \mathbb{S}_+^m$  because  $\mathbb{S}_+^m$  is a cone. From Lemma 2.2, we have  $Z/2 = P_{\mathbb{S}_+^m}(Z/2 - Y)$ . Thus, taking the square of the Frobenius norm in both sides of this expression gives the result.

(b) Since  $Y \in \mathbb{S}_+^m$ , we obtain  $P_{\mathbb{S}_+^m}(-Y) = 0$ . Using this fact and the nonexpansive property of the projection, we get

$$\left\| P_{\mathbb{S}_+^m} \left( \frac{Z}{2} - Y \right) \right\|_F = \left\| P_{\mathbb{S}_+^m} \left( \frac{Z}{2} - Y \right) - P_{\mathbb{S}_+^m}(-Y) \right\|_F \leq \left\| \frac{Z}{2} \right\|_F.$$

Thus, the result follows by squaring both sides of the above inequality.  $\square$

We propose a *generalized augmented Lagrangian function*  $\mathcal{A}_c: \mathbb{R}^n \times \mathbb{S}^m \rightarrow \mathbb{R}$  as follows:

$$\mathcal{A}_c(x, \Lambda) := f(x) + \alpha_c(x, \Lambda)\varphi(G(x), \beta_c(x, \Lambda)\Lambda) + \gamma(x, \Lambda), \quad (3.2)$$

where  $c > 0$  is a penalty parameter,  $\alpha_c, \beta_c, \gamma: \mathbb{R}^n \times \mathbb{S}^m \rightarrow \mathbb{R}$ , and  $\varphi$  is given in (3.1), namely

$$\varphi(G(x), \beta_c(x, \Lambda)\Lambda) = \left\| P_{\mathbb{S}_+^m} \left( \frac{\beta_c(x, \Lambda)}{2}\Lambda - G(x) \right) \right\|_F^2 - \frac{\beta_c(x, \Lambda)^2}{4}\|\Lambda\|_F^2.$$

We will show now that  $\mathcal{A}_c$  is an exact augmented Lagrangian function associated to (NSDP) in the sense of Definition 2.4, when certain assumptions for  $\alpha_c$ ,  $\beta_c$ , and  $\gamma$  are satisfied.

**Assumption 3.2.** *The functions  $\alpha_c, \beta_c, \gamma: \mathbb{R}^n \times \mathbb{S}^m \rightarrow \mathbb{R}$  satisfy the following conditions.*

- (a)  *$\alpha_c, \beta_c, \gamma$  are continuously differentiable for all  $c > 0$ .*
- (b)  *$\alpha_c(x, \Lambda) > 0$  for all  $x$  feasible for (NSDP),  $\Lambda \in \mathbb{S}^m$ , and all  $c > 0$ .*

Moreover, if  $(\bar{x}, \bar{\Lambda}) \in \mathbb{R}^n \times \mathbb{S}^m$  is a KKT pair of (NSDP), then the conditions below hold.

(c)  $\alpha_c(\bar{x}, \bar{\Lambda})\beta_c(\bar{x}, \bar{\Lambda}) = 1$  for all  $c > 0$ .

(d)  $\gamma(\bar{x}, \bar{\Lambda}) = 0$ ,  $\nabla_x \gamma(\bar{x}, \bar{\Lambda}) = 0$ , and  $\nabla_{\Lambda} \gamma(\bar{x}, \bar{\Lambda}) = 0$ .

(e) There exist neighborhoods  $V_{\bar{x}}$  and  $V_{\bar{\Lambda}}$  of  $\bar{x}$  and  $\bar{\Lambda}$ , respectively, and a continuous function  $\Gamma: V_{\bar{x}} \rightarrow V_{\bar{\Lambda}}$  such that  $\Gamma(\bar{x}) = \bar{\Lambda}$  and  $\gamma(x, \Gamma(x)) = 0$  for all  $x \in V_{\bar{x}}$ .

**Proposition 3.3.** *Suppose that Assumption 3.2(a) holds. Then, the function  $\mathcal{A}_c$  defined in (3.2) is continuously differentiable. Moreover, its gradient with respect to  $x$  and  $\Lambda$ , respectively, can be written as follows:*

$$\begin{aligned} \nabla_x \mathcal{A}_c(x, \Lambda) &= \nabla f(x) + \varphi(G(x), \beta_c(x, \Lambda)\Lambda) \nabla_x \alpha_c(x, \Lambda) \\ &\quad - 2\alpha_c(x, \Lambda) \nabla G(x)^* P_{\mathbb{S}_+^m} \left( \frac{\beta_c(x, \Lambda)}{2} \Lambda - G(x) \right) \\ &\quad + \alpha_c(x, \Lambda) \left\langle \Lambda, P_{\mathbb{S}_+^m} \left( \frac{\beta_c(x, \Lambda)}{2} \Lambda - G(x) \right) \right\rangle \nabla_x \beta_c(x, \Lambda) \\ &\quad - \frac{1}{2} \alpha_c(x, \Lambda) \beta_c(x, \Lambda) \|\Lambda\|_F^2 \nabla_x \beta_c(x, \Lambda) + \nabla_x \gamma(x, \Lambda), \\ \nabla_{\Lambda} \mathcal{A}_c(x, \Lambda) &= \varphi(G(x), \beta_c(x, \Lambda)\Lambda) \nabla_{\Lambda} \alpha_c(x, \Lambda) \\ &\quad + \alpha_c(x, \Lambda) \left[ \left\langle \Lambda, P_{\mathbb{S}_+^m} \left( \frac{\beta_c(x, \Lambda)}{2} \Lambda - G(x) \right) \right\rangle \nabla_{\Lambda} \beta_c(x, \Lambda) \right. \\ &\quad \left. + \beta_c(x, \Lambda) P_{\mathbb{S}_+^m} \left( \frac{\beta_c(x, \Lambda)}{2} \Lambda - G(x) \right) \right. \\ &\quad \left. - \frac{1}{2} \beta_c(x, \Lambda) \|\Lambda\|_F^2 \nabla_{\Lambda} \beta_c(x, \Lambda) - \frac{1}{2} \beta_c(x, \Lambda)^2 \Lambda \right] + \nabla_{\Lambda} \gamma(x, \Lambda). \end{aligned}$$

*Proof.* The continuous differentiability of  $\mathcal{A}_c$  follows from Assumption 3.2(a) and the fact that  $f$ ,  $G$ , and  $\varphi$  are continuously differentiable. For the gradient's formula, we use Lemma 2.3(a),(c),(d) and some simple calculations.  $\square$

Before proving the exactness results, we will first show the relation between the function  $\mathcal{A}_c$  and the objective function  $f$  of (NSDP). As we can see in the next propositions, the values of  $\mathcal{A}_c$  and  $f$  at KKT points coincide, but if a point is only feasible, then a simple inequality holds.

**Proposition 3.4.** *Suppose that Assumption 3.2(b) holds. Let  $x \in \mathbb{R}^n$  be a feasible point of (NSDP). Then,  $\mathcal{A}_c(x, \Lambda) \leq f(x) + \gamma(x, \Lambda)$  for all  $\Lambda \in \mathbb{S}^m$  and all  $c > 0$ .*

*Proof.* Let  $\Lambda \in \mathbb{S}^m$  and  $c > 0$  be taken arbitrarily. Since  $x$  is feasible for (NSDP), we have  $G(x) \in \mathbb{S}_+^m$ . Thus, Lemma 3.1(b) shows that  $\varphi(G(x), \beta_c(x, \Lambda)\Lambda) \leq 0$  is satisfied. The proof is complete because  $\alpha_c(x, \Lambda) > 0$  also holds from Assumption 3.2(b).  $\square$

**Proposition 3.5.** *Suppose that Assumption 3.2 holds. Let  $(x, \Lambda) \in \mathbb{R}^n \times \mathbb{S}^m$  be a KKT pair of (NSDP). Then,  $(x, \Lambda)$  is also stationary of  $\mathcal{A}_c$ , and  $\mathcal{A}_c(x, \Lambda) = f(x)$  for all  $c > 0$ .*

*Proof.* Let  $c > 0$  be arbitrarily given and recall the formulas of  $\nabla_x \mathcal{A}_c(x, \Lambda)$  and  $\nabla_\Lambda \mathcal{A}_c(x, \Lambda)$  given in Proposition 3.3. From Assumption 3.2(b),(c), we have  $\beta_c(x, \Lambda) > 0$ . So, from the KKT conditions (2.1),  $G(x) \in \mathbb{S}_+^m$ ,  $\beta_c(x, \Lambda)\Lambda \in \mathbb{S}_+^m$ , and  $\langle G(x), \beta_c(x, \Lambda)\Lambda \rangle = 0$  also hold, which imply that

$$\varphi(G(x), \beta_c(x, \Lambda)\Lambda) = 0 \quad (3.3)$$

from Lemma 3.1(a). Moreover, Lemma 2.2 shows that

$$P_{\mathbb{S}_+^m} \left( \frac{\beta_c(x, \Lambda)}{2} \Lambda - G(x) \right) = \frac{\beta_c(x, \Lambda)}{2} \Lambda. \quad (3.4)$$

The equalities (3.3), (3.4) and Assumption 3.2(d) yield  $\nabla_\Lambda \mathcal{A}_c(x, \Lambda) = 0$ . Moreover, from (3.4) and Assumption 3.2(c), we have

$$\nabla f(x) - 2\alpha_c(x, \Lambda)\nabla G(x)^* P_{\mathbb{S}_+^m} \left( \frac{\beta_c(x, \Lambda)}{2} \Lambda - G(x) \right) = \nabla f(x) - \nabla G(x)^* \Lambda.$$

So, once again using Assumption 3.2(d), equalities (3.3), (3.4) and the KKT condition  $\nabla f(x) - \nabla G(x)^* \Lambda = 0$ , we can conclude that  $\nabla_x \mathcal{A}_c(x, \Lambda) = 0$  holds. Finally, (3.3) and Assumption 3.2(d) also yields  $\mathcal{A}_c(x, \Lambda) = f(x)$ , and the proof is complete.  $\square$

The above proposition shows that a KKT pair of (NSDP) is stationary of  $\mathcal{A}_c$ , and this assertion does not depend on the parameter  $c$ . The exactness properties of  $\mathcal{A}_c$  can be shown only if the other implication also holds, that is, a stationary point of  $\mathcal{A}_c$  should be a KKT pair of (NSDP), at least when  $c$  is greater than some threshold value. If such a statement holds, then the exactness of  $\mathcal{A}_c$  is guaranteed, as it can be seen below. Before that, we recall that the global (local) minimizers of problems (NSDP) and (2.2) with  $\Psi_c := \mathcal{A}_c$  are, respectively, denoted by  $G_{\text{NSDP}}$  ( $L_{\text{NSDP}}$ ) and  $G_{\text{NLP}}(c)$  ( $L_{\text{NLP}}(c)$ ). We also consider the following assumption:

**Assumption 3.6.** *The sets  $G_{\text{NSDP}}$  and  $G_{\text{NLP}}(c)$  are nonempty for all  $c > 0$ . Moreover, for every  $x \in G_{\text{NSDP}}$  there is at least one  $\Lambda \in \mathbb{S}^m$  such that  $(x, \Lambda)$  is a KKT pair of (NSDP).*

The existence of optimal solutions of the unconstrained problem is guaranteed if an extra-aneous compact set is considered [11], or by exploiting some properties of the problem, as coercivity and monotonicity [1]. Furthermore, we can ensure the existence of a Lagrange multiplier by imposing some constraint qualification. Now, if we define

$$\begin{aligned} \tilde{G}_{\text{NSDP}} := \{ & (x, \Lambda) : x \in G_{\text{NSDP}} \text{ and} \\ & \Lambda \text{ is a corresponding Lagrange multiplier} \}, \end{aligned} \quad (3.5)$$

then, using Assumption 3.6, we obtain  $\tilde{G}_{\text{NSDP}} \neq \emptyset$ .

**Lemma 3.7.** *Suppose that Assumptions 3.2 and 3.6 hold. Then for all  $c > 0$ ,  $G_{\text{NLP}}(c) \subseteq \tilde{G}_{\text{NSDP}}$  implies  $G_{\text{NLP}}(c) = \tilde{G}_{\text{NSDP}}$ .*

*Proof.* Let  $c > 0$  be arbitrarily given and  $(\bar{x}, \bar{\Lambda}) \in G_{\text{NLP}}(c)$ . By assumption,  $(\bar{x}, \bar{\Lambda}) \in \tilde{G}_{\text{NSDP}}$  also holds, and thus,  $(\bar{x}, \bar{\Lambda})$  is a KKT pair of (NSDP). From Proposition 3.5, we obtain

$$f(\bar{x}) = \mathcal{A}_c(\bar{x}, \bar{\Lambda}). \quad (3.6)$$

Recall that  $\tilde{G}_{\text{NSDP}} \neq \emptyset$  because of Assumption 3.6. So, take  $(\tilde{x}, \tilde{\Lambda}) \in \tilde{G}_{\text{NSDP}}$ , with  $(\tilde{x}, \tilde{\Lambda}) \neq (\bar{x}, \bar{\Lambda})$ . Since  $(\tilde{x}, \tilde{\Lambda})$  satisfies the KKT conditions of (NSDP), once again from Proposition 3.5, we have  $f(\tilde{x}) = \mathcal{A}_c(\tilde{x}, \tilde{\Lambda})$ . This fact, together with (3.6), and the definition of global solutions, gives

$$\mathcal{A}_c(\tilde{x}, \tilde{\Lambda}) = f(\tilde{x}) \leq f(\bar{x}) = \mathcal{A}_c(\bar{x}, \bar{\Lambda}) \leq \mathcal{A}_c(\tilde{x}, \tilde{\Lambda}),$$

which shows that the whole expression above holds with equalities. Therefore,  $(\tilde{x}, \tilde{\Lambda}) \in G_{\text{NLP}}(c)$ , which completes the proof.  $\square$

**Theorem 3.8.** *Suppose that Assumption 3.2 and 3.6 hold. Assume also that there exists  $\hat{c} > 0$  such that every stationary point of  $\mathcal{A}_c$  is also a KKT pair of (NSDP) for all  $c \geq \hat{c}$ . Then,  $\mathcal{A}_c$  is an exact augmented Lagrangian function associated to (NSDP), in other words:*

$$(a) \quad G_{\text{NLP}}(c) = \tilde{G}_{\text{NSDP}} \text{ for all } c \geq \hat{c}.$$

$$(b) \quad L_{\text{NLP}}(c) \subseteq \{(x, \Lambda) : x \in L_{\text{NSDP}} \text{ and } \Lambda \text{ is a corresponding multiplier}\} \text{ for all } c \geq \hat{c}.$$

*Proof.* (a) Let  $c \geq \hat{c}$  be arbitrarily given. From Lemma 3.7, we only need to prove that  $G_{\text{NLP}}(c) \subseteq \tilde{G}_{\text{NSDP}}$ . Let  $(\bar{x}, \bar{\Lambda}) \in G_{\text{NLP}}(c)$ . From (3.5), we need to show that  $\bar{x} \in G_{\text{NSDP}}$ , with  $\bar{\Lambda}$  as a corresponding Lagrange multiplier. Then,  $(\bar{x}, \bar{\Lambda})$  is a stationary point of  $\mathcal{A}_c$ . From this theorem's assumption, it is also a KKT pair of (NSDP), which implies  $f(\bar{x}) = \mathcal{A}_c(\bar{x}, \bar{\Lambda})$  from Proposition 3.5. Now, assume that there exists  $\tilde{x} \in G_{\text{NSDP}}$  such that  $\tilde{x} \neq \bar{x}$ . Since  $\tilde{x}$  satisfies a constraint qualification from Assumption 3.6, there exists  $\tilde{\Lambda}$  such that  $(\tilde{x}, \tilde{\Lambda})$  satisfies the KKT conditions of (NSDP). Once again by Proposition 3.5, we have  $f(\tilde{x}) = \mathcal{A}_c(\tilde{x}, \tilde{\Lambda})$ . So, the definition of global minimizers gives

$$\mathcal{A}_c(\tilde{x}, \tilde{\Lambda}) = f(\tilde{x}) \leq f(\bar{x}) = \mathcal{A}_c(\bar{x}, \bar{\Lambda}) \leq \mathcal{A}_c(\tilde{x}, \tilde{\Lambda}),$$

which shows that the whole expression above holds with equalities. Therefore,  $\bar{x} \in G_{\text{NSDP}}$ , with  $\bar{\Lambda}$  as a corresponding Lagrange multiplier.

(b) Let  $c \geq \hat{c}$  and  $(\bar{x}, \bar{\Lambda}) \in L_{\text{NLP}}(c)$ . Since  $(\bar{x}, \bar{\Lambda})$  is a stationary point of  $\mathcal{A}_c$ , it is also a KKT pair of (NSDP) by theorem's assumption. So, from Proposition 3.5, we have

$$\mathcal{A}_c(\bar{x}, \bar{\Lambda}) = f(\bar{x}). \quad (3.7)$$

Moreover, from the definition of local minimizer, there exist neighborhoods  $V_{\bar{x}}$  and  $V_{\bar{\Lambda}}$  of  $\bar{x}$  and  $\bar{\Lambda}$ , respectively, such that

$$\mathcal{A}_c(\bar{x}, \bar{\Lambda}) \leq \mathcal{A}_c(x, \Lambda) \quad \text{for all } (x, \Lambda) \in V_{\bar{x}} \times V_{\bar{\Lambda}}.$$

Here, we suppose that  $V_{\bar{x}}$  and  $V_{\bar{\Lambda}}$  are sufficiently small, which guarantees the existence of a function  $\Gamma$  as in Assumption 3.2(e). In particular, we obtain  $\mathcal{A}_c(\bar{x}, \bar{\Lambda}) \leq \mathcal{A}_c(x, \Gamma(x))$  for all  $x \in V_{\bar{x}}$ . This inequality, together with (3.7), shows that

$$f(\bar{x}) \leq \mathcal{A}_c(x, \Gamma(x)) \quad \text{for all } x \in V_{\bar{x}}. \quad (3.8)$$

Thus, from Proposition 3.4 and Assumption 3.2(e), we get

$$f(\bar{x}) \leq \mathcal{A}_c(x, \Gamma(x)) \leq f(x) + \gamma(x, \Gamma(x)) = f(x)$$

for all  $x \in V_{\bar{x}}$  that is feasible for (NSDP). So, we conclude that  $\bar{x} \in L_{\text{NSDP}}$ .  $\square$

The above result shows that the generalized function  $\mathcal{A}_c$  is an exact augmented Lagrangian function if a finite penalty parameter  $\hat{c} > 0$  satisfying

$$\text{“stationary of } \mathcal{A}_c \text{ with parameter } c \implies \text{KKT of (NSDP)”} \quad \text{for all } c \geq \bar{c}$$

is guaranteed to exist. However, even if Assumption 3.2 holds, we usually cannot expect that such  $\hat{c}$  exists. In the next section, we will observe that the functions  $\alpha_c$ ,  $\beta_c$ , and  $\gamma$ , used in the formula of  $\mathcal{A}_c$ , should be taken carefully for such a purpose.

## 4 The proposed exact augmented Lagrangian function

Here, we construct a particular exact augmented Lagrangian function by choosing the functions  $\alpha_c$ ,  $\beta_c$ , and  $\gamma$ , used in  $\mathcal{A}_c$  (formula (3.2)) appropriately. Before that, let us note that by defining

$$\alpha_c(x, \Lambda) = \frac{c}{2}, \quad \beta_c(x, \Lambda) = \frac{2}{c}, \quad \gamma(x, \Lambda) = 0,$$

where  $c > 0$  is the penalty parameter, we obtain the augmented Lagrangian function for NSDP given in [8, 29]. This function is actually an extension of the classical augmented Lagrangian function for NLP (see [6] for instance), and it is equal to the Lagrangian function with some additional terms. However, it is not exact in the sense of Definition 2.4.

In order to construct an augmented Lagrangian function with exactness property, we choose a more complex  $\gamma$ , that satisfies Assumption 3.2(d),(e). As in [13], the function  $\Gamma$  of item (e) can be taken as a function that estimates the value of the Lagrange multipliers associated to a point. One possibility for such an estimate for NSDP problems is given in [20], which in turn extends the ones proposed in [18, 19]. Basically, given  $x \in \mathbb{R}^n$ , we consider the following unconstrained problem:

$$\begin{aligned} & \underset{\Lambda}{\text{minimize}} && \|\nabla_x L(x, \Lambda)\|^2 + \zeta_1^2 \|\mathcal{L}_{G(x)}(\Lambda)\|_F^2 + \zeta_2^2 r(x) \|\Lambda\|_F^2 \\ & \text{subject to} && \Lambda \in \mathbb{S}^m, \end{aligned} \quad (4.1)$$

where  $\zeta_1, \zeta_2 \in \mathbb{R}$  are positive scalars, and  $r: \mathbb{R}^n \rightarrow \mathbb{R}$  denotes the residual function associated to the feasible set, that is,

$$r(x) := \frac{1}{2} \|P_{\mathbb{S}_+^m}(-G(x))\|_F^2 = \frac{1}{2} \|P_{\mathbb{S}_+^m}(G(x)) - G(x)\|_F^2.$$

Observe that  $r(x) = 0$  if, and only if,  $x$  is feasible for (NSDP). The idea underlying problem (4.1) is to force KKT conditions (2.1) to hold, except for the feasibility of the Lagrange multiplier. Actually, this problem can be seen as a linear least squares problem, and so its solution can be written explicitly, when the so-called *nondegeneracy* assumption holds. It is well-known that the nondegeneracy condition, defined below, extends the classical linear independence constraint qualification for nonlinear programming [7, 28], see also Section 4 and Corollary 2 in [26]. In particular, under nondegeneracy, Lagrange multipliers are ensured to exist at optimal points.

**Assumption 4.1.** *Every  $x \in \mathbb{R}^n$  feasible for (NSDP) is nondegenerate, that is,*

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

where  $\mathcal{T}_{\mathbb{S}_+^m}(G(x))$  denotes the tangent cone of  $\mathbb{S}_+^m$  at  $G(x)$ ,  $\text{Im } \nabla G(x)$  is the image of the linear map  $\nabla G(x)$ , and  $\text{lin}$  means lineality space.

**Lemma 4.2.** *Suppose that Assumption 4.1 holds. For a given  $x \in \mathbb{R}^n$ , define  $N: \mathbb{R}^n \rightarrow \mathbb{S}^m$  as*

$$N(x) := \nabla G(x) \nabla G(x)^* + \zeta_1^2 \mathcal{L}_{G(x)}^2 + \zeta_2^2 r(x) I. \quad (4.2)$$

Then, the following statements are true.

- (a)  $N(\cdot)$  is continuously differentiable and for all  $x \in \mathbb{R}^n$ , the matrix  $N(x)$  is positive definite.
- (b) The solution of problem (4.1) is unique and it is given by

$$\Lambda(x) = N(x)^{-1} \nabla G(x) \nabla f(x). \quad (4.3)$$

- (c) If  $(x, \Lambda) \in \mathbb{R}^n \times \mathbb{S}^m$  is a KKT pair of (NSDP), then  $\Lambda(x) = \Lambda$ .
- (d) The operator  $\Lambda(\cdot)$  is continuously differentiable, and  $\nabla \Lambda(x) = N(x)^{-1} Q(x)$ , where

$$\begin{aligned} Q(x) &:= \nabla^2 G(x) \nabla_x L(x, \Lambda(x)) + \nabla G(x) \nabla_{xx}^2 L(x, \Lambda(x)) \\ &\quad - \zeta_1^2 \nabla_x [\mathcal{L}_{G(x)}^2(\Lambda(x))] - \zeta_2^2 \nabla r(x) \Lambda(x). \end{aligned}$$

*Proof.* See [20, Lemma 2.2 and Proposition 2.3]. □

The augmented Lagrangian function  $L_c: \mathbb{R}^n \times \mathbb{S}^m \rightarrow \mathbb{R}$  that we propose is given by

$$L_c(x, \Lambda) := f(x) + \frac{1}{2c} \left( \|P_{\mathbb{S}_+^m}(\Lambda - cG(x))\|_F^2 - \|\Lambda\|_F^2 \right) + \|N(x)(\Lambda(x) - \Lambda)\|_F^2, \quad (4.4)$$

where  $\Lambda(\cdot)$  and  $N(\cdot)$  are given in Lemma 4.2. It is equivalent to the usual augmented Lagrangian function for NSDP, except for the last term. So, comparing to the generalized one (3.2), we have

$$\alpha_c(x, \Lambda) = \frac{c}{2}, \quad \beta_c(x, \Lambda) = \frac{2}{c}, \quad \gamma(x, \Lambda) = \|N(x)(\Lambda(x) - \Lambda)\|_F^2. \quad (4.5)$$

Observe that the functions  $\alpha_c, \beta_c, \gamma$  defined in such a way satisfy Assumption 3.2. In fact, items (a), (b), and (c) of this assumption hold trivially, and item (d) is satisfied because of Lemma 4.2(c). The function  $\Gamma$  of item (e) corresponds to  $\Lambda(\cdot)$ , with the necessary properties described in Lemma 4.2(c),(d). Note that  $\gamma(x, \Lambda(x)) = 0$  for all  $x$  in this case.

Now, from (4.2) and (4.3), observe that

$$\begin{aligned} N(x)(\Lambda(x) - \Lambda) &= \nabla G(x) \nabla f(x) - \nabla G(x) \nabla G(x)^* \Lambda - \zeta_1^2 \mathcal{L}_{G(x)}^2(\Lambda) - \zeta_2^2 r(x) \Lambda \\ &= \nabla G(x) \nabla_x L(x, \Lambda) - \zeta_1^2 \mathcal{L}_{G(x)}^2(\Lambda) - \zeta_2^2 r(x) \Lambda. \end{aligned} \quad (4.6)$$

Also, consider the following auxiliary function  $Y_c: \mathbb{R}^n \times \mathbb{S}^m \rightarrow \mathbb{S}^m$  defined by

$$Y_c(x, \Lambda) := P_{\mathbb{S}_+^m} \left( \frac{\Lambda}{c} - G(x) \right) - \frac{\Lambda}{c}. \quad (4.7)$$

The gradient of  $L_c(x, \Lambda)$  with respect to  $x$  is given by

$$\begin{aligned} \nabla_x L_c(x, \Lambda) &= \nabla f(x) - \nabla G(x)^* P_{\mathbb{S}_+^m}(\Lambda - cG(x)) \\ &\quad + 2K(x, \Lambda)^* N(x)(\Lambda(x) - \Lambda) \\ &= \nabla_x L(x, \Lambda) - c \nabla G(x)^* Y_c(x, \Lambda) + 2K(x, \Lambda)^* N(x)(\Lambda(x) - \Lambda), \end{aligned} \quad (4.8)$$

with

$$\begin{aligned} K(x, \Lambda) &:= \nabla_x [N(x)(\Lambda(x) - \Lambda)] \\ &= \nabla_x [\nabla G(x) \nabla_x L(x, \Lambda) - \zeta_1^2 \mathcal{L}_{G(x)}^2(\Lambda) - \zeta_2^2 r(x) \Lambda], \end{aligned}$$

where the second equality follows from (4.6). Using Lemma 2.3(a), as well as some additional calculations, we obtain

$$\begin{aligned} \nabla_x L_c(x, \Lambda) &= \\ &\nabla_x L(x, \Lambda) - c \nabla G(x)^* Y_c(x, \Lambda) + 2 \nabla_{xx}^2 L(x, \Lambda) \nabla G(x)^* N(x)(\Lambda(x) - \Lambda) \\ &\quad + 2 \left[ \left\langle \frac{\partial^2 G(x)}{\partial x_i \partial x_j}, N(x)(\Lambda(x) - \Lambda) \right\rangle \right]_{i,j=1}^n \nabla_x L(x, \Lambda) \\ &\quad - 2 \zeta_1^2 \left[ \left\langle \frac{\partial G(x)}{\partial x_i} \circ (G(x) \circ \Lambda) + G(x) \circ \left( \frac{\partial G(x)}{\partial x_i} \circ \Lambda \right), N(x)(\Lambda(x) - \Lambda) \right\rangle \right]_{i=1}^n \\ &\quad - 2 \zeta_2^2 \langle \Lambda, N(x)(\Lambda(x) - \Lambda) \rangle \nabla r(x). \end{aligned}$$

Moreover, the gradient of  $L_c(x, \Lambda)$  with respect to  $\Lambda$  can be written as follows:

$$\nabla_\Lambda L_c(x, \Lambda) = Y_c(x, \Lambda) - 2N(x)^2(\Lambda(x) - \Lambda). \quad (4.9)$$

Here, we point out that the formulas of  $L_c$ ,  $\nabla_x L_c$  and  $\nabla_\Lambda L_c$ , presented respectively in (4.4), (4.8) and (4.9), do not require explicit computation of the multiplier estimate  $\Lambda(x)$ . In fact, the estimate only appears in the expression  $N(x)(\Lambda(x) - \Lambda)$ , that can be written as (4.6). It means that both  $L_c$  and their gradients do not require solving the linear least squares problem (4.1), which is computational expensive.

## 4.1 Exactness results

In the whole section, we suppose that Assumptions 3.6 and 4.1 hold. Indeed, it can be noted that the assertion about constraint qualifications in Assumption 3.6 holds automatically from Assumption 4.1. Here, we will show that the particular augmented Lagrangian  $L_c$ , defined in (4.4), is in fact exact. With this purpose, we will first establish the relation between the KKT points of the original (NSDP) problem and the stationary points of the unconstrained problem:

$$\begin{aligned} & \underset{x, \Lambda}{\text{minimize}} && L_c(x, \Lambda) \\ & \text{subject to} && (x, \Lambda) \in \mathbb{R}^n \times \mathbb{S}^m. \end{aligned} \quad (4.10)$$

**Proposition 4.3.** *Let  $(x, \Lambda) \in \mathbb{R}^n \times \mathbb{S}^m$  be a KKT pair of (NSDP). Then, for all  $c > 0$ ,  $L_c(x, \Lambda) = f(x)$  and  $(x, \Lambda)$  is stationary of  $L_c$ , that is,  $\nabla_x L_c(x, \Lambda) = 0$  and  $\nabla_\Lambda L_c(x, \Lambda) = 0$ .*

*Proof.* Recalling that the functions defined in (4.5) satisfy Assumption 3.2, the result follows from Proposition 3.5.  $\square$

**Proposition 4.4.** *Let  $\hat{x} \in \mathbb{R}^n$  be feasible for (NSDP) and  $\hat{\Lambda} \in \mathbb{S}^m$ . So, there exist  $\hat{c}, \hat{\delta}_1, \hat{\delta}_2 > 0$  such that if  $(x, \Lambda) \in \mathbb{R}^n \times \mathbb{S}^m$  is stationary of  $L_c$  with  $\|x - \hat{x}\| \leq \hat{\delta}_1$ ,  $\|\Lambda - \hat{\Lambda}\|_F \leq \hat{\delta}_2$  and  $c \geq \hat{c}$ , then  $(x, \Lambda)$  is a KKT pair of (NSDP).*

*Proof.* Let us first consider an arbitrary pair  $(x, \Lambda) \in \mathbb{R}^n \times \mathbb{S}^m$  and  $c > 0$ . For convenience, we also define the following function:

$$\hat{Y}_c(x, \Lambda) := Y_c(x, \Lambda) + G(x) = P_{\mathbb{S}_+^m} \left( -\frac{\Lambda}{c} + G(x) \right), \quad (4.11)$$

with the last equality following from Lemma 2.1(a). Lemma 2.1(b) and the definition of  $Y_c(x, \Lambda)$  in (4.7) show that

$$0 = \hat{Y}_c(x, \Lambda) \circ P_{\mathbb{S}_+^m} \left( \frac{\Lambda}{c} - G(x) \right) = \hat{Y}_c(x, \Lambda) \circ \left( Y_c(x, \Lambda) + \frac{\Lambda}{c} \right).$$

The above expression can be rewritten using the distributivity of the Jordan product:

$$\hat{Y}_c(x, \Lambda) \circ Y_c(x, \Lambda) = -\hat{Y}_c(x, \Lambda) \circ \frac{\Lambda}{c}.$$

Moreover, from (4.11), the above equality, the distributivity and the commutativity of the Jordan product, we obtain

$$\begin{aligned} G(x) \circ \Lambda &= \hat{Y}_c(x, \Lambda) \circ \Lambda - Y_c(x, \Lambda) \circ \Lambda \\ &= -c\hat{Y}_c(x, \Lambda) \circ Y_c(x, \Lambda) - Y_c(x, \Lambda) \circ \Lambda \\ &= -(c\hat{Y}_c(x, \Lambda) + \Lambda) \circ Y_c(x, \Lambda). \end{aligned}$$

Using this expression, we have

$$\begin{aligned}\frac{1}{c}\mathcal{L}_{G(x)}^2(\Lambda) &= -\frac{1}{c}G(x) \circ \left[ (c\hat{Y}_c(x, \Lambda) + \Lambda) \circ Y_c(x, \Lambda) \right] \\ &= -\mathcal{L}_{G(x)}\mathcal{L}_{\hat{Y}_c(x, \Lambda) + \Lambda/c}(Y_c(x, \Lambda)).\end{aligned}\tag{4.12}$$

Now, the formula of  $\nabla_x L_c(x, \Lambda)$  in (4.8) and the equality (4.6) show that

$$\begin{aligned}\frac{1}{c}\nabla G(x)\nabla_x L_c(x, \Lambda) &= \frac{1}{c}\nabla G(x)\nabla_x L(x, \Lambda) - \nabla G(x)\nabla G(x)^*Y_c(x, \Lambda) \\ &\quad + \frac{2}{c}\nabla G(x)K(x, \Lambda)^*N(x)(\Lambda(x) - \Lambda) \\ &= \frac{1}{c}(I + 2\nabla G(x)K(x, \Lambda)^*)N(x)(\Lambda(x) - \Lambda) \\ &\quad + \frac{1}{c}\zeta_1^2\mathcal{L}_{G(x)}^2(\Lambda) + \frac{1}{c}\zeta_2^2r(x)\Lambda - \nabla G(x)\nabla G(x)^*Y_c(x, \Lambda).\end{aligned}$$

Thus, from (4.12), we have

$$\begin{aligned}\frac{1}{c}\nabla G(x)\nabla_x L_c(x, \Lambda) &= \frac{1}{c}(I + 2\nabla G(x)K(x, \Lambda)^*)N(x)(\Lambda(x) - \Lambda) \\ &\quad - N_c(x, \Lambda)Y_c(x, \Lambda) + \frac{1}{c}\zeta_2^2r(x)\Lambda,\end{aligned}\tag{4.13}$$

where

$$N_c(x, \Lambda) := \nabla G(x)\nabla G(x)^* + \zeta_1^2\mathcal{L}_{G(x)}\mathcal{L}_{\hat{Y}_c(x, \Lambda) + \Lambda/c}.$$

Let us now consider  $(x, \Lambda) \in \mathbb{R}^n \times \mathbb{S}^m$  that is stationary of  $L_c$ . Since  $\nabla_\Lambda L_c(x, \Lambda) = 0$ , we obtain, from (4.9),

$$\frac{1}{2}N(x)^{-1}Y_c(x, \Lambda) = (N(x)(\Lambda(x) - \Lambda)),\tag{4.14}$$

because  $N(x)$  is nonsingular from Lemma 4.2(a). Recalling that  $\nabla_x L_c(x, \Lambda) = 0$  also holds, then, from (4.13) we obtain

$$\tilde{N}_c(x, \Lambda)Y_c(x, \Lambda) - \frac{1}{c}\zeta_2^2r(x)\Lambda = 0,\tag{4.15}$$

where

$$\tilde{N}_c(x, \Lambda) := N_c(x, \Lambda) - \frac{1}{2c}(I + 2\nabla G(x)K(x, \Lambda)^*)N(x)^{-1}.$$

Using the fact that  $\|W\|_F^2/2 - \|Z\|_F^2 \leq \|W - Z\|_F^2$  for any matrices  $W, Z$ , from (4.15), we can write

$$\frac{1}{2}\|\tilde{N}_c(x, \Lambda)Y_c(x, \Lambda)\|_F^2 - \frac{1}{c^2}\zeta_2^4r(x)^2\|\Lambda\|_F^2 \leq 0.\tag{4.16}$$

Moreover, the definition of projection and Lemma 2.1(a) yield

$$\begin{aligned}
r(x) &\leq \frac{1}{2} \left\| P_{\mathbb{S}_+^m} \left( -\frac{\Lambda}{c} + G(x) \right) - G(x) \right\|_F^2 \\
&= \frac{1}{2} \left\| P_{\mathbb{S}_+^m} \left( -\frac{\Lambda}{c} + G(x) \right) - \left( -\frac{\Lambda}{c} + G(x) \right) - \frac{\Lambda}{c} \right\|_F^2 \\
&= \frac{1}{2} \|Y_c(x, \Lambda)\|_F^2.
\end{aligned}$$

The above inequality, together with (4.16) implies

$$\left[ \frac{1}{2} \sigma_{\min}^2(\tilde{N}_c(x, \Lambda)) - \frac{1}{2c^2} \zeta_2^4 r(x) \|\Lambda\|_F^2 \right] \|Y_c(x, \Lambda)\|_F^2 \leq 0, \quad (4.17)$$

where  $\sigma_{\min}(\cdot)$  denotes the smallest singular value function.

Recalling that  $\hat{x}$  is feasible for (NSDP), we observe that if  $c \rightarrow \infty$ , then  $\hat{Y}_c(\hat{x}, \Lambda) \rightarrow P_{\mathbb{S}_+^m}(G(\hat{x})) = G(\hat{x})$  for all  $\Lambda$ . Since  $r(\hat{x}) = 0$ , this also shows that

$$c \rightarrow \infty \quad \Rightarrow \quad N_c(\hat{x}, \Lambda) \rightarrow N(\hat{x}) \quad \text{for all } \Lambda \in \mathbb{S}^m,$$

with  $N(\hat{x})$  defined in (4.2). Now, note that from Lemma 4.2(a),  $N(\hat{x})$  is positive definite. Also, define

$$M(x) := \nabla G(x) \nabla G(x)^* + \zeta_1^2 \mathcal{L}_{G(x)} \mathcal{L}_{P_{\mathbb{S}_+^m}(G(x))}.$$

Observe that  $M$  is continuous because all functions involved in its formula are continuous, and that  $M(\hat{x}) = N(\hat{x})$ , which is positive definite. Therefore, there is  $\delta_1 > 0$  such that  $\|x - \hat{x}\| \leq \delta_1$  implies that  $M(x)$  is also positive definite.

Letting  $\hat{\Lambda} \in \mathbb{S}^m$ , there exist  $c_0, \hat{\delta}_1, \hat{\delta}_2 > 0$  with  $\hat{\delta}_1 < \delta_1$  such that both  $M(x)$  and  $N_{c_0}(x, \Lambda)$  are positive definite for all  $(x, \Lambda)$  in the set

$$\mathcal{V} := \{(x, \Lambda) \in \mathbb{R}^n \times \mathbb{S}^m : \|x - \hat{x}\| \leq \hat{\delta}_1, \|\Lambda - \hat{\Lambda}\|_F \leq \hat{\delta}_2\}.$$

Now, we would like to prove that there is  $\hat{c}_0 > 0$  such that  $N_c(x, \Lambda)$  is positive definite for all  $c \geq \hat{c}_0$  and all  $(x, \Lambda)$  in  $\mathcal{V}$ . To do so, suppose that this statement is false. Then, there are sequences  $\{c_k\} \subset \mathbb{R}_{++}$  and  $\{(x^k, \Lambda_k)\} \in \mathcal{V}$  such that  $c_k \rightarrow \infty$  and  $N_{c_k}(x^k, \Lambda_k)$  is not positive definite for all  $k$ . Since  $\mathcal{V}$  is compact, we may assume that  $\{(x^k, \Lambda_k)\}$  converges to some  $(\tilde{x}, \tilde{\Lambda}) \in \mathcal{V}$ . However, we have

$$M(\tilde{x}) = \lim_{k \rightarrow \infty} N_{c_k}(x^k, \Lambda_k).$$

Since  $M(\tilde{x})$  is positive definite,  $N_{c_k}(x^k, \Lambda_k)$  should also be positive definite for  $k$  sufficiently large. This contradicts the fact that no  $N_{c_k}(x^k, \Lambda_k)$  is positive definite, by construction. We conclude that there is  $\hat{c}_0$  such that  $N_c(x, \Lambda)$  is positive definite for all  $c \geq \hat{c}_0$  and all  $(x, \Lambda)$  in  $\mathcal{V}$ .

Considering one such  $(x, \Lambda) \in \mathcal{V}$ , we now seek some  $c(x, \Lambda) \geq \hat{c}_0$  such that  $\tilde{N}_c(x, \Lambda)$  is nonsingular for all  $c \geq c(x, \Lambda)$ . We remark that we already know that  $\sigma_{\min}(N_c(x, \Lambda))$  is positive over  $\mathcal{V}$ . Denote by  $\sigma_{\max}(\cdot)$  the maximum singular value function. Then, recalling the formula for  $\tilde{N}_c(x, \Lambda)$  and elementary properties of singular values<sup>1</sup>, we have that if

$$c(x, \Lambda) = \max \left( 1 + \frac{\sigma_{\max}((I + 2\nabla G(x)K(x, \Lambda)^*)N(x)^{-1})}{2\sigma_{\min}(N_c(x, \Lambda))}, \hat{c}_0 \right)$$

then  $\tilde{N}_c(x, \Lambda)$  will be nonsingular for all  $c \geq c(x, \Lambda)$ . As  $c(x, \Lambda)$  is a continuous function of  $(x, \Lambda)$ , we have that  $c_1 := \sup_{\mathcal{V}} c(x, \Lambda)$  is finite. Thus,  $\tilde{N}_c(x, \Lambda)$  is positive definite, and hence  $\sigma_{\min}(\tilde{N}_c(x, \Lambda)) > 0$  for all  $c \geq c_1$  and  $(x, \Lambda) \in \mathcal{V}$ . Similarly, there exists  $\hat{c} \geq c_1$  such that

$$\frac{1}{2}\sigma_{\min}(\tilde{N}_c(x, \Lambda)) - \frac{1}{2c^2}\zeta_2^4 r(x)\|\Lambda\|_F^2 > 0 \quad \text{for all } c \geq \hat{c}, \quad (x, \Lambda) \in \mathcal{V}.$$

Finally, consider  $(x, \Lambda) \in \mathbb{R}^n \times \mathbb{S}^m$  and  $c \in \mathbb{R}_{++}$  such that  $(x, \Lambda)$  is stationary of  $L_c$ ,  $\|x - \hat{x}\| \leq \hat{\delta}_1$ ,  $\|\Lambda - \hat{\Lambda}\|_F \leq \hat{\delta}_2$  and  $c \geq \hat{c}$ . From (4.17) and the above inequality, it means that  $Y_c(x, \Lambda) = 0$ . Also, from Lemma 2.2, and the fact that  $c > 0$ , it yields

$$G(x) \in \mathbb{S}_+^m, \quad \Lambda \in \mathbb{S}_+^m \quad \text{and} \quad \Lambda \circ G(x) = 0.$$

Moreover, from (4.14) and the fact that  $N(x)$  is nonsingular by Lemma 4.2(a), we have  $\Lambda(x) = \Lambda$ . Since  $\nabla_x L_c(x, \Lambda) = 0$  also holds, from (4.8), we obtain  $\nabla_x L(x, \Lambda) = 0$ . Therefore,  $(x, \Lambda)$  is a KKT pair of (NSDP).  $\square$

**Proposition 4.5.** *Let  $\{x^k\} \subset \mathbb{R}^n$ ,  $\{\Lambda_k\} \subset \mathbb{S}^m$ , and  $\{c_k\} \subset \mathbb{R}_{++}$  be sequences such that  $c_k \rightarrow \infty$  and  $(x^k, \Lambda_k)$  is stationary of  $L_{c_k}$  for all  $k$ . Assume that there are subsequences  $\{x^{k_j}\}$  and  $\{\Lambda_{k_j}\}$  of  $\{x^k\}$  and  $\{\Lambda_k\}$ , respectively, such that  $x^{k_j} \rightarrow \hat{x}$  and  $\Lambda_{k_j} \rightarrow \hat{\Lambda}$  for some  $(\hat{x}, \hat{\Lambda}) \in \mathbb{R}^n \times \mathbb{S}^m$ . Then, either there exists  $\hat{k} > 0$  such that  $(x^{k_j}, \Lambda_{k_j})$  is a KKT pair of (NSDP) for all  $k_j \geq \hat{k}$ , or  $\hat{x}$  is a stationary point of the residual function  $r$  that is infeasible for (NSDP).*

*Proof.* We first show that  $\hat{x}$  is a stationary point of  $r$ , in other words,

$$\nabla r(\hat{x}) = -\nabla G(\hat{x})^* P_{\mathbb{S}_+^m}(-G(\hat{x})) = 0.$$

In fact, using (4.8) and dividing the equation  $\nabla_x L_{c_{k_j}}(x^{k_j}, \Lambda_{k_j}) = 0$  by  $c_{k_j}$ , we have

$$\begin{aligned} & \frac{1}{c_{k_j}} \nabla_x L(x^{k_j}, \Lambda_{k_j}) - \nabla G(x^{k_j})^* Y_{c_{k_j}}(x^{k_j}, \Lambda_{k_j}) + \\ & \frac{2}{c_{k_j}} K(x^{k_j}, \Lambda_{k_j})^* N(x^{k_j})(\Lambda(x^{k_j}) - \Lambda_{k_j}) = 0. \end{aligned}$$

Recalling Lemma 4.2, we observe that all the functions involved in the above equation are continuous. Thus, taking the limit  $k_j \rightarrow \infty$ , from the definition of  $Y_c$  in (4.7), we obtain

<sup>1</sup>Namely, that  $\inf_{\|x\|=1} \|Ax - Bx\| = \sigma_{\min}(A - B) \geq \sigma_{\min}(A) - \sigma_{\max}(B)$ .

$-\nabla G(\hat{x})^* P_{\mathbb{S}_+^m}(-G(\hat{x})) = 0$ , as we claimed. Now, assume that  $\hat{x}$  is feasible. Then, from Proposition 4.4, there exists  $\hat{k} > 0$  such that  $(x^{k_j}, \Lambda_{k_j})$  is a KKT pair of (NSDP) for all  $k_j \geq \hat{k}$ , which completes the proof.  $\square$

Now, recalling Definition 2.4, once again, we use the notations  $G_{\text{NSDP}}$  ( $L_{\text{NSDP}}$ ) and  $G_{\text{NLP}}(c)$  ( $L_{\text{NLP}}(c)$ ) to denote the sets of global (local) minimizers of problems (NSDP) and (4.10), respectively. The following theorems show that the proposed function  $L_c$  given in (4.4) is in fact an exact augmented Lagrangian function. However, the results are established as in [2], where it is admitted that we can end up with a stationary point of the residual function  $r$  that is infeasible for (NSDP).

**Theorem 4.6.** *Let  $\{x^k\} \subset \mathbb{R}^n$ ,  $\{\Lambda_k\} \subset \mathbb{S}^m$ , and  $\{c_k\} \subset \mathbb{R}_{++}$  be sequences such that  $c_k \rightarrow \infty$  and  $(x^k, \Lambda_k) \in L_{\text{NLP}}(c_k)$  for all  $k$ . Assume that there are subsequences  $\{x^{k_j}\}$  and  $\{\Lambda_{k_j}\}$  of  $\{x^k\}$  and  $\{\Lambda_k\}$ , respectively, such that  $x^{k_j} \rightarrow \hat{x}$  and  $\Lambda_{k_j} \rightarrow \hat{\Lambda}$  for some  $(\hat{x}, \hat{\Lambda}) \in \mathbb{R}^n \times \mathbb{S}^m$ . Then, either there exists  $\hat{k} > 0$  such that  $x^{k_j} \in L_{\text{NSDP}}$ , with an associated Lagrange multiplier  $\Lambda_{k_j}$  for all  $k_j \geq \hat{k}$ , or  $\hat{x}$  is a stationary point of the residual function  $r$  that is infeasible for (NSDP).*

*Proof.* From Proposition 4.5, either there exists  $\hat{k} > 0$  such that  $(x^{k_j}, \Lambda_{k_j})$  is a KKT pair of (NSDP) for all  $k_j \geq \hat{k}$ , or  $\hat{x}$  is a stationary point of  $r$  that is infeasible for (NSDP). So, the result follows in the same way as the proof of Theorem 3.8(b).  $\square$

For the above result, that concerns local minimizers, we note that the existence of the subsequence  $\{x^{k_j}\}$  is guaranteed, for example, when the whole sequence  $\{x^k\}$  is bounded. Moreover, if the constraint function  $G$  is convex with respect to the cone  $\mathbb{S}_+^m$ , then the residual function  $r$  is also convex, which means that all stationary points of  $r$  are feasible for (NSDP). In the case of global minimizers, it is possible to prove full equivalence, and we do not have to concern about stationary points of  $r$  that are infeasible for (NSDP).

**Proposition 4.7.** *Let  $\{x^k\} \subset \mathbb{R}^n$ ,  $\{\Lambda_k\} \subset \mathbb{S}^m$ , and  $\{c_k\} \subset \mathbb{R}_{++}$  be sequences such that  $x^k \rightarrow \hat{x}$  and  $\Lambda_k \rightarrow \hat{\Lambda}$  for some  $(\hat{x}, \hat{\Lambda}) \in \mathbb{R}^n \times \mathbb{S}^m$ ,  $c_k \rightarrow \infty$ , and  $(x^k, \Lambda_k) \in G_{\text{NLP}}(c_k)$  for all  $k$ . Then, there exists  $\hat{k} > 0$  such that  $x^k \in G_{\text{NSDP}}$  with an associated Lagrange multiplier  $\Lambda_k$  for all  $k \geq \hat{k}$ .*

*Proof.* Let  $\bar{x} \in G_{\text{NSDP}}$ , which exists by Assumption 3.6. Because of the nondegeneracy constraint qualification, there exists  $\bar{\Lambda} \in \mathbb{S}^m$  such that  $(\bar{x}, \bar{\Lambda})$  is a KKT pair of (NSDP). So, from Proposition 4.3,  $f(\bar{x}) = L_c(\bar{x}, \bar{\Lambda})$  holds for all  $c > 0$ . Moreover, since  $(x^k, \Lambda_k) \in G_{\text{NLP}}(c_k)$ , we have

$$L_{c_k}(x^k, \Lambda_k) \leq L_{c_k}(\bar{x}, \bar{\Lambda}) = f(\bar{x}) \quad (4.18)$$

for all  $k$ . Taking the supremum limit in this inequality, we obtain

$$\limsup_{k \rightarrow \infty} L_{c_k}(x^k, \Lambda_k) \leq f(\bar{x}). \quad (4.19)$$

Observe now that the formula of  $L_{c_k}(x^k, \Lambda_k)$  in (4.4) can be written equivalently as

$$\begin{aligned} L_{c_k}(x^k, \Lambda_k) &= f(x^k) + \frac{c_k}{2} \left( \left\| P_{\mathbb{S}_+^m} \left( \frac{\Lambda_k}{c_k} - G(x^k) \right) \right\|_F^2 - \left\| \frac{\Lambda_k}{c_k} \right\|_F^2 \right) \\ &\quad + \|N(x^k)(\Lambda(x^k) - \Lambda_k)\|_F^2. \end{aligned} \quad (4.20)$$

Recall from Lemma 4.2 that the functions involved in the above equality are all continuous. This fact, together with inequality (4.19), shows that  $P_{\mathbb{S}_+^m}(-G(\hat{x})) = 0$ , that is,  $\hat{x}$  is feasible. So, from Proposition 4.5, we conclude that there exists  $\hat{k} > 0$  such that  $(x^k, \Lambda_k)$  is a KKT pair of (NSDP) for all  $k \geq \hat{k}$ . Since  $c_k > 0$  and the norm is always nonnegative, (4.20) implies  $L_{c_k}(x^k, \Lambda_k) \geq f(x^k) - \|\Lambda_k\|_F^2 / (2c_k)$ . Again, taking the supremum limit in such an inequality, we have

$$\limsup_{k \rightarrow \infty} L_{c_k}(x^k, \Lambda_k) \geq f(\hat{x}),$$

which, together with (4.19) shows that  $f(\hat{x}) \leq f(\bar{x})$ . Thus,  $\hat{x} \in G_{\text{NSDP}}$  holds.

Now, since  $\hat{x}$  is feasible for (NSDP), there exist  $\hat{c}, \hat{\delta}_1, \hat{\delta}_2$  as in Proposition 4.4. Consider  $\hat{k}$  large enough so that  $\|x^k - \hat{x}\| \leq \hat{\delta}_1$ ,  $\|\Lambda_k - \hat{\Lambda}\|_F \leq \hat{\delta}_2$ ,  $c_k \geq \hat{c}$ , and  $(x^k, \Lambda_k) \in G_{\text{NLP}}(c_k)$  for all  $k \geq \hat{k}$ . Since  $(x^k, \Lambda_k)$  is stationary of  $L_{c_k}$ , from Proposition 4.4, we obtain that  $(x^k, \Lambda_k)$  is also a KKT pair of (NSDP) for all  $k \geq \hat{k}$ . Once again from Proposition 4.3 and (4.18), we have  $f(x^k) = L_{c_k}(x^k, \Lambda_k) \leq f(\bar{x})$ . Therefore,  $x^k \in G_{\text{NSDP}}$  for all  $k \geq \hat{k}$ .  $\square$

**Theorem 4.8.** *Assume that there exists  $\bar{c} > 0$  such that  $\bigcup_{c \geq \bar{c}} G_{\text{NLP}}(c)$  is bounded. Then, there exists  $\hat{c} > 0$  such that  $G_{\text{NLP}}(c) = \tilde{G}_{\text{NSDP}}$  for all  $c \geq \hat{c}$ , where  $\tilde{G}_{\text{NSDP}}$  is defined in (3.5).*

*Proof.* From Lemma 3.7, we only need to show the existence of  $\hat{c} > 0$  such that  $G_{\text{NLP}}(c) \subseteq \tilde{G}_{\text{NSDP}}$  for all  $c \geq \hat{c}$ . Assume that this statement is false. Then, there exist sequences  $\{(x^k, \Lambda_k)\} \subset \mathbb{R}^n \times \mathbb{S}^m$  and  $\{c_k\} \subset \mathbb{R}_{++}$  with  $c_k \rightarrow \infty$ ,  $c_k \geq \bar{c}$ , and  $(x^k, \Lambda_k) \in G_{\text{NLP}}(c_k)$ , but such that  $(x^k, \Lambda_k) \notin \tilde{G}_{\text{NSDP}}$ . Since  $\bigcup_{c \geq \bar{c}} G_{\text{NLP}}(c)$  is bounded, we can assume, without loss of generality, that  $x^k \rightarrow \hat{x}$  and  $\Lambda_k \rightarrow \hat{\Lambda}$  for some  $(\hat{x}, \hat{\Lambda}) \in \mathbb{R}^n \times \mathbb{S}^m$ . Thus, Proposition 4.7 shows that there exists  $\hat{k} > 0$  such that  $(x^k, \Lambda_k) \in \tilde{G}_{\text{NSDP}}$  for all  $k \geq \hat{k}$ , which is a contradiction.  $\square$

## 5 Preliminary numerical experiments

This work is focused on the theoretical aspects of exact augmented Lagrangian functions, however, we take a look at the numerical prospects of our approach by examining two simple problems in the next two subsections. We will now explain briefly our proposal. Given a problem (NSDP), the idea is to use some unconstrained optimization method to solve (4.10), i.e., minimization of  $L_c$  given in (4.4). First, an initial point  $x^0 \in \mathbb{R}^n$ , together with the initial penalty parameter  $c_0 > 0$  are selected. Then, we choose the initial Lagrange multiplier  $\Lambda_0 \in \mathbb{S}^m$ . One possibility is to use the multiplier estimate, i.e., to set  $\Lambda_0$  as  $\Lambda(x^0)$ , where  $\Lambda(\cdot)$  is defined in (4.3).

Then, we run the unconstrained optimization method of our choice. However, since the penalty values for which  $L_c$  becomes exact is not known beforehand, we attempt to adjust

the penalty parameter between iterations as follows. Let  $\tau \in (0, 1)$  and  $\rho > 1$ . Denote by  $x^k$  and  $\Lambda_k, c_k$  the values of  $x, \Lambda$  and  $c$  at the  $k$ th iteration, respectively. Recalling the function  $Y_c(\cdot, \cdot)$  defined in (4.7), if at the  $k$ th iteration we have

$$\|Y_{c_k}(x^k, \Lambda_k)\|_F = \left\| P_{\mathbb{S}_+^m} \left( \frac{\Lambda_k}{c_k} - G(x^k) \right) - \frac{\Lambda_k}{c_k} \right\|_F > \tau \|Y_{c_{k-1}}(x^{k-1}, \Lambda_{k-1})\|_F,$$

then we let  $c_k$  be  $\rho c_{k-1}$ . Otherwise, we let  $c_k$  be  $c_{k-1}$ . This is an idea that appears in many augmented Lagrangian methods, for example in [?]. The motivation is that  $\|Y_{c_k}(x^k, \Lambda_k)\|_F$  is zero if and only if  $G(x^k) \in \mathbb{S}_+^m, \Lambda_k \in \mathbb{S}_+^m$  and  $G(x^k) \circ \Lambda_k = 0$ . Therefore,  $\|Y_{c_k}(x^k, \Lambda_k)\|_F$  is a measure of the degree to which complementarity and feasibility are satisfied, taking into account the current penalty parameter. In summary, whenever there is not enough progress, we increase the penalty. In order to avoid the problem becoming too ill-conditioned, we never increase the penalty past some fixed value  $c_{\max}$ . We also point out that the update of the penalty parameter can be done by using the so-called test function, which is originally defined in [19]. However, as it can be seen in the paper about exact penalty functions [2], the above approach using  $Y_c(\cdot, \cdot)$  is more efficient, which justifies its use here.

In our implementation,  $\rho, \tau$  and  $c_{\max}$  is set to 1.1, 0.9 and 1000, respectively. The maximum number of iterations is 5000. The values of  $\zeta_1$  and  $\zeta_2$ , which control the behavior of  $N(x)$  in (4.2), are set to 1 and  $10^{-4}$ , respectively. The initial penalty parameter  $c_0$  is computed using the formula:

$$c_0 := \max \left\{ c_{\min}, \min \left\{ c_{\max}, \frac{10 \max(1, |f(x^0)|)}{\max(1, 0.5 \|G(x^0)\|_F^2)} \right\} \right\},$$

with  $c_{\min} = 0.1$ , which is similar to the one used in [?]. The unconstrained method of our choice is the BFGS method using the Armijo's condition for the line search. We stop the algorithm when the KKT conditions are satisfied within  $10^{-5}$  or when the norm of the gradient of  $L_c$  is less than  $10^{-5}$ .

We implemented the algorithm in Python and ran all the experiments on a Intel Core i7-6700 machine with 8 cores and 16GB of memory. As we already pointed out after (4.9), an important implementational aspect is that we never need to explicitly evaluate the function  $\Lambda(\cdot)$ , except in the optional way of computing the initial Lagrange multiplier.

## 5.1 Noll's example

As an initial example, we took a look at this simple instance by Noll [?]:

$$\begin{aligned} & \underset{x \in \mathbb{R}^2}{\text{minimize}} && 0.5(-x_1^2 - x_2^2) \\ & \text{subject to} && \begin{pmatrix} 1 & x_1 - 1 & 0 \\ x_1 - 1 & 1 & x_2 \\ 0 & x_2 & 1 \end{pmatrix} \in \mathbb{S}_+^3. \end{aligned} \tag{Noll}$$

The problem (Noll) is already in the format (NSDP), which can be seen by letting  $G$  be the function defined by

$$G(x_1, x_2) := \begin{pmatrix} 1 & x_1 - 1 & 0 \\ x_1 - 1 & 1 & x_2 \\ 0 & x_2 & 1 \end{pmatrix}.$$

In order to compute  $L_c$  and its gradient (see (4.4), (4.6), (4.8) and (4.9)), we need the partial derivative matrices of  $G$  and the adjoint of the gradient of  $G$ , which are given below:

$$\frac{\partial G(x)}{\partial x_1} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \frac{\partial G(x)}{\partial x_2} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \nabla G(x)^* V = 2(V_{12}, V_{23})^\top,$$

where  $V \in \mathbb{S}^3$  is an arbitrary matrix with  $(i, j)$  entry denoted by  $V_{ij}$ .

The optimal value of (Noll) is  $-2$  and it is achieved at  $(2, 0)$ . Starting at  $(1, 0)$ , our method found a solution satisfying the optimality criteria in 14 iterations and 0.01 seconds. The initial and final penalty parameters were 6.66 and 10.74, respectively. The objective function, the constraint function and their gradients were evaluated 41 times each.

## 5.2 The closest correlation matrix problem

Let  $H$  be a  $m \times m$  symmetric matrix. The goal is to find a correlation matrix  $X$  that is as close as possible to  $H$ . In other words, we seek a solution to the following problem:

$$\begin{aligned} & \underset{X}{\text{minimize}} && \langle X - H, X - H \rangle \\ & \text{subject to} && X_{ii} = 1 \quad \text{for all } i, \\ & && X \in \mathbb{S}_+^m. \end{aligned} \tag{Cor}$$

There are many variants of (Cor) where weighting factors are added, constraints on the eigenvalues are considered, and so on. With that, this family of problems has found of wealth of applications in statistics and finance [?].

In this example, it is possible to show that the nondegeneracy condition is satisfied at every feasible point (Assumption 4.1), which guarantee the theoretical properties of the exact augmented Lagrangian function  $L_c$ . This was proved by Qi and Sun in [?], but since there are some differences in notation, we will first take a look at this issue. Qi and Sun proved the following result.

**Proposition 5.1.** *Let  $Y \in \mathbb{S}_+^m$  be such that  $Y_{ii} = 1$ , for all  $i$ . Then,*

$$\text{diag}(\text{lin}\mathcal{T}_{\mathbb{S}_+^m}(Y)) = \mathbb{R}^m,$$

where  $\text{diag}: \mathbb{S}^m \rightarrow \mathbb{R}^m$  is the linear map that maps a symmetric matrix  $Z$  to its diagonal  $(Z_{11}, Z_{22}, \dots, Z_{mm})$ .

*Proof.* See Proposition 2.1 and Equation (2.2) in [?]. □ □

We now write (Cor) in a format similar to (NSDP). For that, denote by  $A^{ij}$  the  $m \times m$  the matrix that has 1 in the  $(i, j)$  and  $(j, i)$  entries and 0 elsewhere. Then, by discarding constant terms in the objective function, (Cor) can be reformulated equivalently as follows.

$$\begin{aligned} & \underset{x \in \mathbb{R}^{m(m-1)/2}}{\text{minimize}} && \sum_{1 \leq i < j \leq m} (H_{ij} - x_{ij})^2 && \text{(Cor2)} \\ & \text{subject to} && I + \sum_{1 \leq i < j \leq m} A^{ij} x_{ij} \in \mathbb{S}_+^m. \end{aligned}$$

Here,  $x$  can be thought as an upper triangular matrix without the diagonal. That is why the dimension of  $x$  is  $m(m-1)/2$  and we index  $x$  by using  $x_{ij}$  for  $1 \leq i < j \leq m$ .

**Proposition 5.2.** *Problem (Cor2) satisfies Assumption 4.1.*

*Proof.* We must show that

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

where  $G : \mathbb{R}^{m(m-1)/2} \rightarrow \mathbb{S}^m$  is the function such that

$$G(x) = I + \sum_{1 \leq i < j \leq m} A^{ij} x_{ij}.$$

Now, let  $Y \in \mathbb{S}^m$  be arbitrary. We can write  $Y$  as

$$Y = \sum_{i=1}^m A^{ii} Y_{ii} + \sum_{1 \leq i < j \leq m} A^{ij} Y_{ij}.$$

From Proposition 5.1, the first summation belongs to  $\text{lin} \mathcal{T}_{\mathbb{S}_+^m}(Y)$ . Then, noting that  $\text{Im} \nabla G(x)$  is the space spanned by  $\{A^{ij} : 1 \leq i < j \leq m\}$ , we conclude that the second summation belongs to  $\text{Im} \nabla G(x)$ . This shows that Assumption 4.1 is satisfied.  $\square$   $\square$

We now write some useful formulae which can be used in conjunction with (4.6), (4.8) and (4.9) to compute  $L_c$  and its gradient. Let  $V$  be an arbitrary  $m \times m$  symmetric matrix. Then,

$$\frac{\partial G(x)}{\partial x_{ij}} = A^{ij}, \quad \nabla G(x)^* V = 2v,$$

for  $1 \leq i < j \leq m$ , where  $v$  corresponds to the upper triangular part of  $V$  without the diagonal.

We now move on to the experiments. We generated 50 symmetric matrices  $H$  such that the diagonal entries are all 1 and non-diagonal elements are uniform random numbers between  $-1$  and  $1$ . This was repeated for  $m = 5, 10, 15, 20$ . We then ran our algorithm using as initial point the matrix having 1 in all its entries. The results can be seen in Table 1. All the values depicted in Table 1 are averages among 50 runs. The column ‘‘Iterations’’ correspond to average number of BFGS iterations. At each run, we recorded the number

of function evaluations for  $f$ , which is the same for  $G, \nabla f$  and  $\nabla G$ . Then, the column “Evaluations” in Table 1 is the average number of function evaluations. Columns “Initial  $c$ ” and “Final  $c$ ” correspond to the average of the initial and final penalty parameters, respectively. Finally, column “Time (s)” is the average running time, in seconds.

Table 1: Results for (Cor).

$m$	Iterations	Evaluations	Initial $c$	Final $c$	Time (s)
5	114.62	371.22	21.98	805.2	0.208
10	520.96	1844.62	23.31	1000.0	1.923
15	1191.62	4297.74	24.77	1000.0	10.170
20	2101.02	7801.00	25.39	1000.0	42.490

No failures were detected, that is, we obtained approximate KKT points within  $10^{-5}$  for all the instances. We also observed that, except for  $m = 5$ , the final penalty parameter climbed up to the maximum value. At first glance, this suggests that the penalty was not large enough. However, we were still able to solve the problems without increasing the maximum value. In fact, we noted that, in some cases, the performance degraded when the maximum penalty value was increased. At this moment, the method presented here is not competitive against the approach in [?], where we observed that an  $20 \times 20$  instance is typically solved in less than a second in our hardware. However, it should be emphasized that a second-order method is used in [?], where here we used BFGS. It would be interesting to apply and analyze a second-order method in combination with the exact augmented Lagrangian function  $L_c$ , but this investigation is beyond the scope of this paper.

## 6 Final remarks

We proposed a generalized augmented Lagrangian function  $\mathcal{A}_c$  for NSDP problems, giving conditions for it to be exact. After that, we considered a particular function  $L_c$ , and we proved that it is exact under the nondegeneracy condition and some reasonable assumptions as in Theorem 4.8. We also presented some preliminary numerical experiments using a quasi-Newton method with BFGS formula, showing the validity of the approach. One future work is to analyze more efficient methods that can solve the unconstrained minimization of  $L_c$ . From Lemma 4.2 and the formula of  $L_c$ , given in (4.4), we observe that  $L_c$  is an  $SC^1$  function, i.e., it is continuously differentiable and its gradient is semismooth. It means that the unconstrained problem can be solved with second-order methods, as the semismooth Newton. However, the gradient of  $L_c$ , given in (4.8), contains second-order terms of problem functions  $f$  and  $G$ , and thus, a second-order method would have to deal with third-order derivatives. We believe that we can use the idea proposed in [18], that avoids these third-order terms, but still guaranteeing the global superlinear convergence.

By using the generalized function  $\mathcal{A}_c$  as a tool, other practical exact augmented Lagrangian functions can be studied for NSDP, or other important conic optimization problems. In fact, as it can be seen in [13], many other exact augmented Lagrangian functions

exist, but only for the classical nonlinear programming. For example, recalling (4.5), we note that  $L_c$  is defined by choosing functions  $\alpha_c$  and  $\beta_c$  as constants. By considering more sophisticated formulas, it is possible to weaken the assumptions used here. This should be another matter of investigation.

## Acknowledgements

We would like to thank the anonymous referees for their suggestions which improved the original version of the paper. We are also thankful to Akiko Kobayashi for valuable discussions about exact augmented Lagrangian functions.

## References

- [1] T. A. André and P. J. S. Silva. Exact penalties for variational inequalities with applications to nonlinear complementarity problems. *Computational Optimization and Applications*, 47(3):401–429, 2010.
- [2] R. Andreani, E. H. Fukuda, and P. J. S. Silva. A Gauss-Newton approach for solving constrained optimization problems using differentiable exact penalties. *Journal of Optimization Theory and Applications*, 156(2):417–449, 2013.
- [3] P. Apkarian, D. Noll, and H. D. Tuan. Fixed-order  $H_\infty$  control design via a partially augmented Lagrangian method. *International Journal of Robust and Nonlinear Control*, 13(12):1137–1148, 2003.
- [4] A. Ben-Tal, F. Jarre, M. Kočvara, A. Nemirovski, and J. Zowe. Optimal design of trusses under a nonconvex global buckling constraint. *Optimization and Engineering*, 1(2):189–213, 2000.
- [5] D. S. Bernstein. *Matrix Mathematics: Theory, Facts, and Formulas*. Princeton University Press, 2nd edition, 2009.
- [6] D. P. Bertsekas. *Constrained Optimization and Lagrange Multipliers Methods*. Academic Press, New York, 1982.
- [7] J. F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer-Verlag, New York, 2000.
- [8] R. Correa and H. Ramírez C. A global algorithm for nonlinear semidefinite programming. *SIAM Journal on Optimization*, 15(1):303–318, 2004.
- [9] G. Di Pillo and L. Grippo. A new class of augmented Lagrangians in nonlinear programming. *SIAM Journal on Control and Optimization*, 17(5):618–628, 1979.

- [10] G. Di Pillo and L. Grippo. A new augmented Lagrangian function for inequality constraints in nonlinear programming. *Journal of Optimization Theory and Applications*, 36(4):495–519, 1982.
- [11] G. Di Pillo and L. Grippo. Exact penalty functions in constrained optimization. *SIAM Journal on Control and Optimization*, 27(6):1333–1360, 1989.
- [12] G. Di Pillo, G. Liuzzi, S. Lucidi, and L. Palagi. An exact augmented Lagrangian function for nonlinear programming with two-sided constraints. *Computational Optimization and Applications*, 25:57–83, 2003.
- [13] G. Di Pillo and S. Lucidi. On exact augmented Lagrangian functions in nonlinear programming. In G. Di Pillo and F. Giannessi, editors, *Nonlinear Optimization and Applications*, pages 85–100. Springer, Boston, MA, 1996.
- [14] G. Di Pillo and S. Lucidi. An augmented Lagrangian function with improved exactness properties. *SIAM Journal on Optimization*, 12(2):376–406, 2001.
- [15] B. Fares, P. Apkarian, and D. Noll. An augmented Lagrangian method for a class of LMI-constrained problems in robust control theory. *International Journal of Control*, 74(4):348–360, 2001.
- [16] J. Fiala, M. Kočvara, and M. Stingl. PENLAB: A MATLAB solver for nonlinear semidefinite optimization. *ArXiv e-prints*, November 2013.
- [17] A. Forsgren. Optimality conditions for nonconvex semidefinite programming. *Mathematical Programming*, 88(1):105–128, 2000.
- [18] E. H. Fukuda, Silva P. J. S., and M. Fukushima. Differentiable exact penalty functions for nonlinear second-order cone programs. *SIAM Journal on Optimization*, 22(4):1607–1633, 2012.
- [19] T. Glad and E. Polak. A multiplier method with automatic limitation of penalty growth. *Mathematical Programming*, 17(2):140–155, 1979.
- [20] L. Han. The differentiable exact penalty function for nonlinear semidefinite programming. *Pacific Journal of Optimization*, 10(2):285–303, 2014.
- [21] F. Jarre. Elementary optimality conditions for nonlinear SDPs. In *Handbook on Semidefinite, Conic and Polynomial Optimization*, volume 166 of *International Series in Operations Research & Management Science*, pages 455–470. Springer, 2012.
- [22] Y. Kanno and I. Takewaki. Sequential semidefinite program for maximum robustness design of structures under load uncertainty. *Journal of Optimization Theory and Applications*, 130(2):265–287, 2006.
- [23] H. Konno, N. Kawadai, and D. Wu. Estimation of failure probability using semi-definite logit model. *Computational Management Science*, 1(1):59–73, 2003.

- [24] M. Kočvara and M. Stingl. Solving nonconvex SDP problems of structural optimization with stability control. *Optimization Methods & Software*, 19(5):595–609, 2004.
- [25] A. S. Lewis. Convex analysis on the Hermitian matrices. *SIAM Journal on Optimization*, 6(1):164–177, 1996.
- [26] B. F. Lourenço, E. H. Fukuda, and M. Fukushima. Optimality conditions for nonlinear semidefinite programming via squared slack variables. *Mathematical Programming*, 2016. To appear.
- [27] S. Lucidi. New results on a class of exact augmented Lagrangians. *Journal of Optimization Theory and Applications*, 58(2), 1988.
- [28] A. Shapiro. First and second order analysis of nonlinear semidefinite programs. *Mathematical Programming*, 77(1):301–320, 1997.
- [29] A. Shapiro and J. Sun. Some properties of the augmented Lagrangian in cone constrained optimization. *Mathematics of Operations Research*, 29(3):479–491, 2004.
- [30] P. Tseng. Merit functions for semi-definite complementarity problems. *Mathematical Programming*, 83:159–185, 1998.
- [31] 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.
- [32] H. Yamashita, H. Yabe, and K. Harada. A primal-dual interior point method for nonlinear semidefinite programming. *Mathematical Programming*, 135(1-2):89–121, 2012.