On the use of Jordan Algebras for improving global convergence of an Augmented Lagrangian method in nonlinear semidefinite programming

R. Andreani* E.H. Fukuda† G. Haeser‡ D.O. Santos§ L.D. Secchin¶

May 11, 2020 (revised September 18, 2020)

Abstract

Jordan Algebras are an important tool for dealing with semidefinite programming and optimization over symmetric cones in general. In this paper, a judicious use of Jordan Algebras in the context of sequential optimality conditions is done in order to generalize the global convergence theory of an Augmented Lagrangian method for nonlinear semidefinite programming. An approximate complementarity measure in this context is typically defined in terms of the eigenvalues of the constraint matrix and the eigenvalues of an approximate Lagrange multiplier. By exploiting the Jordan Algebra structure of the problem, we show that a simpler complementarity measure, defined in terms of the Jordan product, is stronger than the one defined in terms of eigenvalues. Thus, besides avoiding a tricky analysis of eigenvalues, a stronger necessary optimality condition is presented. We then prove the global convergence of an Augmented Lagrangian algorithm to this improved necessary optimality condition. The optimality conditions we present are sequential ones, and no constraint qualification is employed; in particular, a global convergence result is available even when Lagrange multipliers are unbounded.

Keywords: nonlinear semidefinite programming; symmetric cones; optimality conditions; constraint qualifications; augmented Lagrangian method

1 Introduction

Optimization problems on a symmetric cone have attracted a lot of attention in recent years. The reason for this is the fact that the non-negative orthant of $\mathbb{R}^m$, the second-order cone (Lorentz cone), and the positive semidefinite cone of symmetric matrices are examples of symmetric cones. The approach via symmetric cones allows unifying many results for all these relevant problems (see, e.g., [25]), but when particularizing for each specific cone, one may exploit the specific structure of the cone to obtain new results. In this paper we are particularly interested in the nonlinear semidefinite programming (NSDP) problem due to its large number of applications such as material optimization [22, 28], control theory [15, 16] and others [17, 24, 29].

Our goal in this paper is to consider the algebraic structure of NSDPs, via the Jordan product, to improve the global convergence result of an Augmented Lagrangian method for NSDPs. Typically, the global convergence is done proving that a feasible limit point of a sequence generated by the Augmented Lagrangian satisfies the KKT conditions under Robinson's constraint qualification (RCQ) (or Mangasarian-Fromovitz constraint qualification, in the context of nonlinear programming). However, in [8], it was shown that such feasible limit points satisfy the so-called Approximate-KKT (AKKT) necessary optimality condition, which is a strictly stronger result. In particular, the dual sequence generated by the algorithm may be unbounded, which is ruled out when RCQ is assumed.

In the context of nonlinear programming, several constraint qualifications weaker than RCQ were defined such that a point satisfying AKKT is in addition a KKT point. These have been called strict constraint qualifications, which gives new global convergence results to KKT points under weaker constraint qualifications. This has been a fruitful area of

*Department of Applied Mathematics, University of Campinas, Campinas-SP, Brazil. Email: andreani@ime.unicamp.br
†Graduate School of Informatics, Kyoto University, Kyoto, Japan. Email: ellen@i.kyoto-u.ac.jp
‡Department of Applied Mathematics, University of São Paulo, São Paulo-SP, Brazil. Email: ghaeser@ime.usp.br
§Institute of Science and Technology, Federal University of São Paulo, São José dos Campos-SP, Brazil. Email: daiana@ime.usp.br
¶Department of Applied Mathematics, Federal University of Espírito Santo, São Mateus, ES, Brazil. E-mail: leonardo.secchin@ufes.br

research in the past 10 years, where several constraint qualifications have been defined for this purpose. See, for instance, [4, 9].

One may say that weaker constraint qualifications, such as Abadie's [1] or Guignard's [18] conditions, also imply the validity of the KKT conditions at solutions, even though they are not strict. This is true; however, the importance of sequential optimality conditions such as AKKT is not simply theoretical, as they are linked to the fact that the sequences in its definition can be typically generated by several primal-dual algorithms. In the context of nonlinear optimization, for instance, linear constraints satisfy a strict constraint qualification, hence there is no need for a separate analysis of degenerate linear constraints. Even when the KKT conditions do not hold, one may say that sequential optimality conditions give an important notion of stationarity; this is true in particular for a class of problems where derivatives are absent at the solution [20] and approximate KKT conditions are the only notion of stationarity available. Extensions of these ideas to several classes of problems have been conducted (such as Nash equilibria [12], variational inequalities [21], quasi-equilibrium problems [13], complementarity constraints [7], Banach spaces [23], among others), together with extensions to second-order KKT conditions [5, 19].

Even though we analyze only an Augmented Lagrangian method inspired by [2], several other algorithms generate sequences that satisfy sequential optimality conditions such as AKKT. For instance, for nonlinear optimization, Interior Point methods [6], Inexact Restoration methods [26], Sequential Quadratic Programming methods [27], and others. See [6]. For nonlinear semidefinite programming, besides the Augmented Lagrangian method from [8], the Stabilized Sequential Quadratic programming method from [30] is also based on the AKKT optimality condition.

In this paper, we present an improvement of the global convergence of the Augmented Lagrangian method for NSDP from [8]. We prove that by using the structure of the Jordan product, inherent to the semidefinite cone, one may measure complementarity in a simpler and stronger way. This follows a previous work done in [3] where, similarly, the structure of the Jordan product was exploited in an Augmented Lagrangian algorithm for nonlinear second-order cone programming.

This paper is organized as follows. In Section 2 we present some basic ideas about symmetric cones that can be seen in details in [14], and we prove that the new optimality condition is stronger than AKKT in the context of NSDP. A discussion for general symmetric cones is also presented. In Section 3 we show the improved global convergence result of the Augmented Lagrangian for NSDP. We conclude with our final remarks.

2 Complementarity measures on symmetric cones

Let us consider the nonlinear optimization problem over a symmetric cone below

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Minimize}} \quad & f(x), \\ \text{subject to} \quad & g(x) \in \mathscr{K}, \end{aligned} \tag{NSCP}$$

where $f:\mathbb{R}^n\to\mathbb{R}$ and $g:\mathbb{R}^n\to\mathscr{E}$ are continuously differentiable functions, $\mathscr{E}$ is a finite dimensional real inner product space and $\mathscr{K}\subseteq\mathscr{E}$ is a symmetric cone, that is, a self-dual, homogeneous cone with non-empty interior. It is well known that $\mathscr{K}$ induces an Euclidean Jordan Algebra $(\mathscr{E},\circ)$ such that $\mathscr{K}=\{u\circ u : u\in\mathscr{E}\}$, where $\circ:\mathscr{E}\times\mathscr{E}\to\mathscr{E}$ is a bilinear operator such that:

1. $u\circ v = v\circ u$,
2. $u\circ(u^2\circ v) = u^2\circ(u\circ v)$,
3. $\langle u\circ v, w\rangle = \langle u, v\circ w\rangle$ for all $u,v,w\in\mathscr{E}$, where $u^2=u\circ u$ and $\langle\cdot,\cdot\rangle$ is the inner product of $\mathscr{E}$.

It is well known that when the vector space $\mathscr{E}$ is real, all symmetric cones are the Cartesian product of semidefinite cones $\mathbb{S}^m_+\subset\mathbb{S}^m$, or second-order cones $K_m := \{z=(z_0,\bar{z})\in\mathbb{R}\times\mathbb{R}^{m-1} : z_0\geq\|\bar{z}\|\}\subset\mathbb{R}^m$. Here, $\mathbb{S}^m$ denotes the set of $m\times m$ symmetric matrices and $\mathbb{S}^m_+$ denotes the semidefinite ones, while $\|\cdot\|$ is the Euclidean norm. When $m=1$, both cones reduce to the set of non-negative real numbers $\mathbb{R}_+$. In the case of the semidefinite cone $\mathbb{S}^m_+$, the Jordan product is given by $X\circ Y=(XY+YX)/2, X,Y\in\mathbb{S}^m$ whereas in the case of the second-order cone K_m, the Jordan product is given by $z\circ w = (\langle z,w\rangle, z_0\overline{w}+w_0\overline{z}), z,w\in\mathbb{R}^m$. When $m=1$ both products reduce to the usual multiplication of real numbers. Note that the Jordan product over a Cartesian product of Jordan algebras can be defined componentwise. Similarly for the inner product. In particular, one Jordan product associated with the symmetric cone $\mathbb{R}^m_+$ is the Hadamard product. We refer the reader to [14] and [11] for more details on Euclidean Jordan Algebras and symmetric cones. In particular, an important tool is the spectral decomposition theorem below. To state it, let r be the rank of $(\mathscr{E},\circ)$ and e its unity. A Jordan frame is a set of idempotents $\{c_1,\ldots,c_r\}\subset\mathscr{E}$, that is, $c_i^2=c_i$ for all i, such that $c_i\circ c_j=0$ for all $i\neq j$ and $\sum_{i=1}^r c_i = e$.

Theorem 2.1 (Theorem III.1.2 in [14]). *For every $u\in\mathscr{E}$, there exists a Jordan frame $\{c_1(u),\ldots,c_r(u)\}$ and so-called eigenvalues $\lambda_1(u),\ldots,\lambda_r(u)\in\mathbb{R}$ such that*
$u=\sum_{i=1}^r \lambda_i(u)c_i(u)$. The decomposition is unique in the sense that if $u=\sum_{i=1}^r \eta_i c_i$ with a Jordan frame $\{c_1,\ldots,c_r\}$, and

$\{\eta_i\}$, $\{\lambda_i(u)\}$ are chosen in increasing order, then $\eta_i = \lambda_i(u)$ for all i and $\sum_{\{j:\eta_j=\xi\}} c_j = \sum_{\{j:\eta_j=\xi\}} c_j(u)$, for all $\xi \in \mathbb{R}$. Also, fixing the ordering, the eigenvalues are continuous functions of u.

In [3], the following necessary optimality conditions for (NSCP) was proved. Let us denote the Lagrangian function of (NSCP) by $\mathscr{L}(x,\mu)$, where $(x,\mu) \in \mathbb{R}^n \times \mathscr{K} \to \mathscr{L}(x,\mu) := f(x) - \langle g(x), \mu \rangle$.

Theorem 2.2. *([3]) Let $x^* \in \mathbb{R}^n$ be a local minimizer of (NSCP). Then, there exists a primal sequence $\{x^k\} \subset \mathbb{R}^n, x^k \to x^*$, and a dual sequence $\{\mu^k\} \subset \mathscr{K}$ such that*

$$\nabla_x \mathscr{L}(x^k, \mu^k) \to 0, \tag{1}$$

$$\lambda_i(g(x^*)) > 0 \Rightarrow \lambda_i(\mu_i^k) \to 0, \text{ for all } i = 1, \ldots, r, \tag{2}$$

$$c_i(\mu^k) \to c_i(g(x^*)), \text{ for all } i = 1, \ldots, r, \tag{3}$$

$$g(x^k) \circ \mu^k \to 0, \tag{4}$$

where $\mu_i^k = \sum_{i=1}^r \lambda_i(\mu_i^k) c_i(\mu^k)$ and $g(x^*) = \sum_{i=1}^r \lambda_i(g(x^*)) c_i(g(x^*))$ are spectral decompositions.

When the cone $\mathscr{K}$ in (NSCP) is the product of second-order cones (what we refer as *nonlinear second-order cone programming* (NSOCP)), an Augmented Lagrangian method was proposed in [3] such that its feasible limit points satisfy the optimality condition given in Theorem 2.2. In particular, when a feasible point x^* admits the existence of a sequence $\{(x^k, \mu^k)\} \subset \mathbb{R}^n \times \mathscr{K}$ with $x^k \to x^*$ such that (1), (2), and (3) hold, we say that x^* satisfies the Approximate-KKT (AKKT) necessary optimality condition, whereas when the sequence is such that only (1) and (4) hold, x^* is said to satisfy the Complementarity-AKKT (CAKKT) necessary optimality condition.

These are extensions of necessary optimality conditions well known in nonlinear programming. Condition AKKT was introduced in [4] while CAKKT was introduced in [10]. We note that these are genuine necessary optimality conditions without the need of assuming a constraint qualification.

In the context of nonlinear programming, the fact that CAKKT implies AKKT follows trivially from the spectral decomposition of $x \in \mathbb{R}^m$ as $x = \sum_{i=1}^m x_i e_i$, where $x_i \in \mathbb{R}$ is the i-th component of x and e_i is the i-th vector of the canonical basis, and from the fact that the Jordan product resumes to the Hadamard product.

In the context of NSOCPs it was proved in [3] that CAKKT also implies AKKT. In this case, a spectral decomposition of $z \in \mathbb{R}^m$ is given by $z = \lambda_-(z)c_-(z) + \lambda_+(z)c_+(z)$, where $\lambda_\pm(z) := z_0 \pm \|\bar{z}\|$ and $c_\pm(z) := 1/2(1, \pm \frac{\bar{z}}{\|\bar{z}\|})$, when $\bar{z} \neq 0$, and when $\bar{z} = 0$, the term $\frac{\bar{z}}{\|\bar{z}\|}$ can be replaced by any unit norm vector. In particular, it was shown in [3] that the convergence of the Jordan product (4) implies the convergence of the Jordan frames (3) in reverse order, namely, $c_\pm(\mu^k) \to c_\mp(g(x^*))$.

Let us now show that CAKKT also implies AKKT in the context of *nonlinear semidefinite programming* (NSDP). A discussion of CAKKT in this context was considered in [8] but no adequate definition was available. Here, the spectral decomposition coincides with the usual spectral decomposition of symmetric matrices, where the Jordan frame $\{c_i(X)\}$ of $X \in \mathbb{S}^m$ is given by $c_i(X) := q_i(X)q_i(X)^T$, where $\{q_i(X)\}$ forms a basis of $\mathbb{R}^m$ of orthonormal eigenvectors of X. We consider a single semidefinite cone for simplicity of notation.

Theorem 2.3. *If $\mathscr{K} = \mathbb{S}^m_+$ then CAKKT implies AKKT.*

Proof. Let x^* be a CAKKT point, that is, there exist $\{x^k\} \subset \mathbb{R}^n$ and $\{\mu^k\} \subset \mathbb{S}^m_+$ such that $x^k \to x^*$, $\nabla_x \mathscr{L}(x^k, \mu^k) \to 0$, and

$$2g(x^k) \circ \mu^k = g(x^k)\mu^k + \mu^k g(x^k) \to 0.$$

Consider the following decomposition

$$\mu^k = \sum_{i=1}^m \lambda_i(\mu^k) c_i(\mu^k), \tag{5}$$

where $\lambda_i^k := \lambda_i(\mu^k)$ and v_i^k such that $c_i(\mu^k) = v_i^k (v_i^k)^T$ denote the eigenvalues and the unitary eigenvectors of μ^k, respectively, for all $i = 1, \ldots, m$. Given $K' \subset \mathbb{N}$ an infinite set, let us define

$$\Lambda_0^{K'} = \{i \mid \lim_{k \in K'} \lambda_i^k = 0\}.$$

Let us fix K maximal in the sense that $|\Lambda_0^K|$ is maximum. Thus,

$$j \notin \Lambda_0^K \Rightarrow \liminf_{k \in K} \lambda_j^k > 0, \tag{6}$$

otherwise we could take a subsequence with indexes in K in order to increase the cardinality of Λ_0^K. In this sense, Λ_0^K "captures" all eigenvalues of μ^k that converge to zero. This allows us to define a new sequence of multipliers $\{\tilde{\mu}^k\}_{k \in K}$ as follows: for each $k \in K$, we take the decomposition (5) of μ^k given by

$$\mu^k = S_k D_k S_k^T.$$

Defining $\tilde{\mu}^k = S_k \tilde{D}_k S_k^T$, $k \in K$, where

$$(\tilde{D}_k)_{ij} = \begin{cases} 0, & i = j \in \Lambda_0^K \\ (D_k)_{ij}, & \text{otherwise} \end{cases}$$

is the eigenvalue matrix obtained from D_k making equal to zero the diagonal elements that converge to zero. Note that, $\tilde{\mu}^k \in \mathbb{S}_+^m$, $\lim_{k\in K}(\tilde{D}_k - D_k) = 0$ and hence,

$$g(x^k) \circ \tilde{\mu}^k = g(x^k) \circ \mu^k + g(x^k) S_k (\tilde{D}_k - D_k) S_k^T + S_k (\tilde{D}_k - D_k) S_k^T g(x^k) \to_{k\in K} 0. \tag{7}$$

Also note that the same $v_1^k, \ldots, v_m^k$ are the eigenvectors of $\tilde{\mu}^k$ ($k \in K$), associated with eigenvalues λ_j^k, $j \notin \Lambda_j^K$, and zero for $j \in \Lambda_0^K$.

If $\Lambda_0^K \neq \emptyset$, take $j \notin \Lambda_0^K$. Let us show that the accumulation points of eigenvector sequences v_j^k of $\tilde{\mu}^k$ associated with λ_j^k are eigenvectors of $g(x^*)$ associated with zero. In this sense, it is possible to decompose $g(x^*)$ so that $\lambda_j(g(x^*)) > 0 \Rightarrow \lambda_j(\tilde{\mu}^k) \to_{k\in K} 0$ worth for these indexes j, and also for the pairing of these eigenvectors.

For each $k \in K$, consider (λ_j^k, v_j^k) of $\tilde{\mu}^k$. Equation (7) gives us

$$[g(x^k)\tilde{\mu}^k + \tilde{\mu}^k g(x^k)] v_j^k = (\lambda_j^k I + \tilde{\mu}^k)(g(x^k) v_j^k) \to_{k\in K} 0.$$

Since $\lambda_j^k I + \tilde{\mu}^k - 1/2(\liminf_{l\in K} \lambda_j^l) I \in \mathbb{S}_+^m$ and $1/2(\liminf_{l\in K} \lambda_j^l) I$ is positive definite for all $k \in K$ large enough, we have $\lim_{k\in K} g(x^k) v_j^k = 0$. Thus, $(0, v_j^*)$ is a pair of eigenvalue and eigenvector of $g(x^*)$ where v_j^* is *any* point of accumulation of the unit sequence $\{v_j^k\}_{k\in K}$.

The above argument holds true for all $j \notin \Lambda_0^K$. Let us consider for simplicity that $\Lambda_0^K = \{d+1, \ldots, m\}$. Let us take $K_1 \subset K$ so that $\lim_{k\in K_1} v_1^k = v_1^*$; $K_2 \subset K_1$ so that $\lim_{k\in K_2} v_2^k = v_2^*$; and so on until K_d. Note that $v_i^k v_j^k = 0$ for all $k \in K_d$, $i \neq j$, and then $v_i^* v_j^* = 0$. Thus, we build an orthonormal set of eigenvectors

$$V_+ = \{v_1^*, \ldots, v_d^*\}$$

obtained as limits of eigenvectors of $\tilde{\mu}^k$associated with eigenvalues with indexes out of Λ_0^K *which is also an orthonormal set of eigenvectors of* $g(x^*)$ *associated to zero*. This provides the pairing of the eigenvectors of $\tilde{\mu}^k$ e $g(x^*)$ required in AKKT for indices $j \notin \Lambda_0^K$.

We will now build a complete and paired basis of eigenvectors for $g(x^*)$. The following argument serves the case $\Lambda_0^K = \emptyset$.

V_+ can be completed to an orthonormal basis of $\mathbb{R}^m$ by taking eigenvector limits not only on their first d (those with indexes outside Λ_0^K), but on

$$\{v_1^k, \ldots, v_d^k, v_{d+1}^k, \ldots, v_m^k\}, \tag{8}$$

in a construction similar to the previous one. This does not affect the previous discussion as it does not depend on the accumulation points we take. Let us say that a subsequence $K_m \subset K$ is obtained in this way. Each set $\{v_1^*, \ldots, v_d^*, v_{d+1}^*, \ldots, v_m^*\}$ obtained in this way will be orthonormal. Note that it is trivial that

$$\lambda_j(g(x^*)) > 0 \Rightarrow \lambda_j(\tilde{\mu}^k) \to_{k\in K_n} 0$$

for $j \in \Lambda_0^K$, since $\lambda_j^k = 0$ for all $k \in K_m$ and all $j \in \Lambda_0^K$

It remains to be shown that the pairing of eigenvectors is possible. We will show that, completing a basis of eigenvectors of $g(x^*)$ from eigenvectors in V_+, associated with null eigenvalues, we managed to change the eigenvector basis (8) of the $\tilde{\mu}^k$'s correspondingly.

If for some $\Lambda_0^K \ni j \geq d+1$ we have $\liminf_{k\in K_m} g(x^k) v_j^k = 0$, then we extract a subsequence if necessary to conclude that $(0, v_j^*)$ is an eigenvalue and eigenvector pair of $g(x^*)$. The remaining case is when $\|g(x^k) v_j^k\| \geq c > 0$, $\forall k \gg 1$, $k \in K_m$. Suppose without loss of generality that this occurs with eigenvectors $v_r^*, \ldots, v_m^*$, $r \geq d+1$. Let $\tilde{v}_r^*, \ldots, \tilde{v}_m^*$ be unitary eigenvectors of $g(x^*)$ associated with *positive* eigenvalues, taken in a way that

$$\{v_1^*, \ldots, v_d^*, \ldots, v_{r-1}^*, \tilde{v}_r^*, \ldots, \tilde{v}_m^*\}$$

is an orthonormal basis of $\mathbb{R}^m$ (which is possible since we can take vectors successively in each orthogonal autospace, and orthonormalize them). In particular, each $\tilde{v}_j^*$, $j \geq r$, is combination of $v_r^*, \ldots, v_m^*$, that is,

$$\tilde{v}_j^* = \sum_{i=r}^{m} \alpha_i^j v_i^*.$$

For each $k \in K_m$ we define

$$\tilde{v}_j^k := \sum_{i=r}^{m} \alpha_i^j v_i^k$$

(α_i^j's are constants here). Note that, for $j \in \Lambda_0^K$, in particular $j \geq r$, $(0, v_j^k)$ is a pair of eigenvalue and eigenvector $\tilde{\mu}^k$ and then

$$\tilde{\mu}^k \tilde{v}_j^k = \sum_{i=r}^{m} \alpha_i^j (\tilde{\mu}^k v_i^k) = 0.$$

That is, for $j \geq r$, we have $(0, \tilde{v}_j^k)$ a pair of eigenvalue and eigenvector pf $\tilde{\mu}^k$. Moreover,

$$(v_l^k)^T \tilde{v}_j^k = \sum_{i=r}^{m} \alpha_i^j [(v_l^k)^T v_i^k] = 0$$

for all $l < r$ and $k \in K_m$, and

$$(\tilde{v}_l^k)^T \tilde{v}_j^k = \left(\sum_{i=r}^{m} \alpha_i^l (v_i^k)^T \right) \left(\sum_{p=r}^{m} \alpha_p^j v_p^k \right) = \sum_{i=r}^{m} \alpha_i^l \alpha_i^j = (\tilde{v}_l^*)^T \tilde{v}_j^* = \begin{cases} 1, & l = j \\ 0, & l \neq j \end{cases}$$

for all $l \geq r$ and $k \in K_m$. We can then replace the basis (8) of eigenvectors of $\tilde{\mu}^k$ by its other orthonormal basis of eigenvectors

$$\{v_1^k, \ldots, v_d^k, \ldots, v_{r-1}^k, \tilde{v}_r^k, \ldots, \tilde{v}_m^k\}.$$

Pairing eigenvectors with indexes $j = r, \ldots, m$ follows from the convergence

$$\lim_{k \in K_m} \tilde{v}_j^k = \tilde{v}_j^*.$$

□

It is somewhat surprising that (4) is enough to ensure (2) and (3) in the context of semidefinite programming. Note that (3) is needed in order for (2) to make sense, since (3) provides a correspondence of the eigenvalues of the Lagrange multipliers with the eigenvalues of the constraint function.

Let us show that when (3) holds with an additional continuity property of the Jordan frame, we may provide a simple proof of this implication for general symmetric cones.

Theorem 2.4. *Let $\mathscr{K} \subseteq \mathscr{E}$ be a symmetric cone and $x^* \in \mathscr{K}$ satisfying CAKKT with a primal-dual sequence $\{(x^k, \mu^k)\}$. Let us assume that there is a way of ordering the idempotents of $\{g(x^k)\}$ and $\{\mu^k\}$ such that*

$$c_i(\mu^k) \to c_i(g(x^*)), \text{ for all } i = 1, \ldots, r, \tag{9}$$

$$c_i(g(x^k)) \to c_i(g(x^*)), \text{ for all } i = 1, \ldots, r. \tag{10}$$

Then AKKT also holds.

Proof. We have

$$g(x^k) \circ \mu^k = \sum_{i=1}^{r} \sum_{j=1}^{r} \lambda_i(g(x^k)) \lambda_j(\mu^k) c_i(g(x^k)) \circ c_j(\mu^k) \to 0.$$

Given $p = 1, \ldots, r$, we take the inner product with $c_p(\mu^k)$. Since

$$\langle c_i(g(x^k)) \circ c_j(\mu^k), c_p(\mu^k) \rangle = \langle c_i(g(x^k)), c_j(\mu^k) \circ c_p(\mu^k) \rangle$$

and using the properties of the Jordan frame, we arrive at

$$\sum_{i=1}^{r} \lambda_p(\mu^k) \lambda_i(g(x^k)) \langle c_i(g(x^k)), c_p(\mu^k) \rangle \to 0.$$

That is, $\lambda_p(\mu^k) \langle g(x^k), c_p(\mu^k) \rangle \to 0$. Let us take p such that $\lambda_p(g(x^*)) > 0$. Since

$$\langle g(x^k), c_p(\mu^k) \rangle = \sum_{i=1}^{r} \lambda_i(g(x^k)) \langle c_i(x^k), c_p(\mu^k) \rangle,$$

$\lambda_i(g(x^k)) \to \lambda_i(g(x^*))$, and $\langle c_i(x^k), c_p(\mu^k) \rangle \to 0$ if $i \neq p$ and converges to

$$\|c_p(g(x^*))\|^2 > 0$$

otherwise, we arrive at $\langle g(x^k), c_p(\mu^k) \rangle \to \lambda_p(g(x^*)) \|c_p(g(x^*))\|^2 > 0$. This implies that $\lambda_p(\mu^k) \to 0$ and AKKT follows.
□

The following example shows that this extra assumption may not hold in general.

Example 2.1. *Let us consider the sequence of matrices* $g(x^k) := \begin{pmatrix} 1/k & 1/k \\ 1/k & 1/k \end{pmatrix}$ *and* $\mu^k := \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$. *Then,*

$$2g(x^k) \circ \mu^k = \begin{pmatrix} 2/k & 1/k \\ 1/k & 0 \end{pmatrix} \to 0.$$

Computing the spectral decompositions we have:

$$\begin{aligned} g(x^k) &= (2/k)\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}(1/\sqrt{2}, 1/\sqrt{2}) + 0\begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}(1/\sqrt{2}, -1/\sqrt{2}) \\ \mu^k &= 1\begin{pmatrix} 1 \\ 0 \end{pmatrix}(1,0) + 0\begin{pmatrix} 0 \\ 1 \end{pmatrix}(0,1). \end{aligned}$$

Thus, since $c_1(g(x^k)), c_2(g(x^k)), c_1(\mu^k), c_2(\mu^k))$ *are different constant vectors, it is not the case that* (9) *and* (10) *hold together, although the decomposition of* $g(x^*)$ *may be chosen such that one of these limits holds.*

Finally, we end this section with a discussion of the relation of CAKKT with the optimality condition Trace-AKKT (TAKKT) introduced in [8] as a tentative to avoid the eigenvalue computation in AKKT. A feasible point $x^* \in \mathbb{R}^n$ of (NSCP) with $\mathscr{K} = \mathbb{S}^m_+$ satisfies TAKKT when there are sequences $\{(x^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{S}^m_+, x^k \to x^*$, such that (1) holds and the complementarity condition (4) of CAKKT is replaced by $\langle g(x^k), \mu^k \rangle \to 0$. In [3], the relation of AKKT and TAKKT was clarified as being independent conditions. However, CAKKT is strictly stronger than both conditions in this context. To see this, it is sufficient to see that $\langle g(x^k), \mu^k \rangle = \mathrm{Tr}(g(x^k) \circ \mu^k)$, where $\mathrm{Tr}(\cdot)$ denotes the trace operator. Since CAKKT implies both TAKKT and AKKT, being the latter conditions independent [3, Example 3.1], it must be the case that CAKKT is strictly stronger than both conditions.

3 Extended global convergence of an Augmented Lagrangian algorithm

In this section we will show that the Augmented Lagrangian method proposed in [8] for NSDPs generates CAKKT sequences. This extends this known result in nonlinear programming [10] and nonlinear second-order cone programming [3] to NSDPs. Let us recall the definition of the algorithm. We use $[A]_+$ to denote the projection of $A \in \mathbb{S}^m$ onto $\mathbb{S}^m_+$.

Given a penalty parameter $\rho > 0$, the Powell-Hestenes-Rockafellar Augmented Lagrangian function $L_\rho : \mathbb{R}^n \times \mathbb{S}^m_+ \to \mathbb{R}^n$ for problem (NSCP) when $\mathscr{K} = \mathbb{S}^m_+$ is given by

$$L_\rho(x, \mu) = f(x) + \frac{1}{2\rho}\{\|[\mu - \rho g(x)]_+\|^2 - \|\mu\|^2\}, \tag{11}$$

where $x \in \mathbb{R}^n$, $\mu \in \mathbb{S}^m_+$, and $[\cdot]_+$ is the orthogonal projection onto $\mathbb{S}^m_+$. The partial derivative with respect to x is given by

$$\nabla L_\rho(x, \mu) = \nabla f(x) - Dg(x)^*[\mu - \rho g(x)]_+, \tag{12}$$

where $Dg(x)^*$ is the adjoint of the derivative operator $Dg(x) : \mathbb{R}^n \to \mathbb{S}^m$. The formal definition of the algorithm is given in Algorithm 3.1.

Similarly to the previously known cases [3, 8, 10], the proof is based on the assumption below. This is a weak assumption on the smoothness of the function g. See [10].

Assumption 1. *All feasible points* $x^* \in \mathbb{R}^n$ *that are limit points of* $\{x^k\}$ *generated by Algorithm 3.1 satisfy the generalized Lojasiewicz inequality below, that is, there exist* $\delta > 0$ *and a continuous function* $\phi : B(x^*; \delta) \to \mathbb{R}$, *with* $\phi(x) \to 0$ *when* $x \to x^*$ *and*

$$|P(x) - P(x^*)| \le \phi(x)\|\nabla P(x)\|, \quad \forall x \in B(x^*; \delta)$$

where $P(x)$ *is the square of the Frobenius norm of* $[-g(x)]_+$ *and* $B(x^*; \delta)$ *is the Euclidean ball of radius* δ *around* x^*.

In [8], the authors proved that the Augmented Lagrangian algorithm tends to find feasible points in the sense that all limit points are stationary points of the problem of minimizing $P(x)$. Now let us to show that Algorithm 3.1 generates CAKKT sequences.

Theorem 3.1. *Let Assumption 1 hold. If* $x^* \in \mathbb{R}^n$ *is a feasible limit point of a sequence* $\{x^k\}$ *generated by Algorithm 3.1, then* x^* *satisfies CAKKT.*

Algorithm 3.1 Augmented Lagrangian Algorithm

Let $\rho_1 > 0$, $\tau \in (0,1)$, $\gamma > 1$ and $M > 0$ be given. Define $\bar{\mu}^1 \in \mathbb{S}^m_+$. Choose a positive sequence $\{\varepsilon_k\}$ such that $\varepsilon_k \to 0$. Initialize $k := 1$.

(i) Determine x^k by the approximate minimization of $L_{\rho_k}(x, \bar{\mu}^k)$, that is, a point x^k such that $\|\nabla L_{\rho_k}(x^k, \mu^k)\| \le \varepsilon_k$.

(ii) Define $V^k := \left[\frac{\bar{\mu}^k}{\rho_k} - g(x^k)\right]_+ - \frac{\bar{\mu}^k}{\rho_k}$, which is used for updating the penalty parameter as follows: if $k > 1$ and $\|V^k\| \le \tau \|V^{k-1}\|$, define $\rho_{k+1} := \rho_k$, otherwise, define $\rho_{k+1} := \gamma \rho_k$.

(iii) Update Lagrange multipliers by computing $\mu^k := \left[\bar{\mu}^k - \rho_k g(x^k)\right]_+$, and defining $\bar{\mu}^{k+1} := proj_S(\mu^k)$, the orthogonal projection of μ^k onto S, where $S \subset \mathbb{S}^m_+$ is the set of matrices with spectral radius bounded by M. Set $k := k+1$, and go to (i).

Proof. By item (i) of Algorithm 3.1 and $\varepsilon_k \to 0$ we have that

$$\nabla f(x^k) - Dg(x^k)^*[\bar{\mu}^k - \rho_k g(x^k)]_+ \to 0, \tag{13}$$

where $\mu^k = [\bar{\mu}^k - \rho_k g(x^k)]_+$. It remains to prove that $2g(x^k) \circ \mu^k = g(x^k)\mu^k + \mu^k g(x^k) \to 0$. We will consider two cases: when $\rho_k \to +\infty$ and when the sequence $\{\rho_k\}$ is bounded.

(i) Assuming that $\rho_k \to +\infty$, let us first prove that $g(x^k)\mu^k \to 0$. Consider the following spectral decomposition

$$\frac{\bar{\mu}^k}{\rho_k} - g(x^k) = S_k D_k S_k^T,$$

where S_k is an orthogonal matrix and D_k is a diagonal matrix with all eigenvalues of $\left(\frac{\bar{\mu}^k}{\rho_k} - g(x^k)\right)$. Thus, we have that

$$\mu^k = [\bar{\mu}^k - \rho_k g(x^k)]_+ = \rho_k S_k [D_k]_+ S_k^T.$$

Since $g(x^k) = \frac{\bar{\mu}^k}{\rho_k} + S_k D_k S_k^T$ we have

$$g(x^k)\mu^k = \left(\frac{\bar{\mu}^k}{\rho_k} + S_k D_k S_k^T\right) \rho_k S_k [D_k]_+ S_k^T \tag{14}$$

$$= \bar{\mu}^k S_k [D_k]_+ S_k^T + \rho_k S_k D_k [D_k]_+ S_k^T. \tag{15}$$

Note that $\bar{\mu}^k S_k [D_k]_+ S_k^T = \bar{\mu}^k \left[\frac{\bar{\mu}^k}{\rho_k} - g(x^k)\right]_+ \to 0$. Then, it is necessary to show only that

$$\rho_k S_k D_k [D_k]_+ S_k^T = \rho_k \sum_{i=1}^m \lambda_i^k [\lambda_i^k]_+ s_i^k [s_i^k]^T \to 0,$$

where $\lambda_i^k, i = 1, \dots, m$ are the diagonal elements of D_k with correspondent column s_i^k of S_k. Since S_k are orthogonal matrices for all k, there exists a subsequence of $\{S_k\}$ that converges to some orthogonal matrix S. Hence, it is enough to show that $\rho_k \lambda_i^k [\lambda_i^k]_+ \to 0$. This follows from the proof of [8, Theorem 4.2], where Assumption 1 is used to ensure that $\rho_k [\lambda_i^k]_+^2 \to 0$, which is essential to complete the proof.

(ii) Supposing now that $\{\rho_k\}$ is bounded, that is, there exists k_0 such that for $k \ge k_0$ the penalty parameter ρ_k remains unchanged. Since $V^k = \left[\frac{\bar{\mu}^k}{\rho_{k_0}} - g(x^k)\right]_+ - \frac{\bar{\mu}^k}{\rho_{k_0}} \to 0$ and $\{\bar{\mu}^k\}$ is bounded, one can take a subsequence such that $\bar{\mu}^k$ converges to some $\mu \in \mathbb{S}^m_+$ with

$$[\mu - \rho_{k_0} g(x^*)]_+ = \mu.$$

Considering the spectral decomposition $\bar{\mu}^k - \rho_{k_0} g(x^k) = S_k D_k S_k^T$ with $S_k \to S$ and $D_k \to D$, we have that

$$\mu = S[D]_+ S^T \text{ and } g(x^*) = \frac{1}{\rho_{k_0}} S(D - [D]_+) S^T.$$

Hence,

$$g(x^k)\mu^k \to g(x^*)\mu = \frac{1}{\rho_{k_0}} S(D-[D]_+)[D]_+ S^T,$$

with $(D-[D]_+)[D]_+ = 0$. With a similar argument used to prove that $g(x^k)\mu^k \to 0$, one can show that $\mu^k g(x^k) \to 0$. Thus CAKKT follows.

□

4 Final Remarks

In [8], an Augmented Lagrangian method for NSDPs was introduced with global convergence theory based on a constraint qualification strictly weaker than Robinson's constraint qualification. Thus, as far as we know, the case of an unbounded Lagrange multiplier could be treated for the first time. There, two necessary optimality conditions were introduced which are satisfied by feasible limit points of the algorithm. In one of them, complementarity is measured in terms of the eigenvalues of the constraints and an approximate Lagrange multiplier matrix; and in the other, one relies on the inner product structure of $\mathbb{S}^m$. It was shown in [3] that these are independent global convergence results, in the sense that no optimality condition is implied by the other. In this paper we show that the optimality condition CAKKT presented in [3] is strictly stronger than both optimality conditions previously defined in [8], and we show that the Augmented Lagrangian method still enjoys global convergence to points satisfying this renewed condition. The result is obtained by exploiting the Jordan algebraic structure of NSDPs, and points to a more general global convergence result in the context of optimization over a general symmetric cone, which will be the subject of further studies.

Acknowledgement

This work was supported by the Grant-in-Aid for Scientific Research (C) (19K11840) from Japan Society for the Promotion of Science, FAPESP grants 2013/07375-0, 2017/18308-2 and 2018/24293-0, FAPES grant 116/2019, CNPq.

References

[1] J. Abadie, *On the Kuhn-Tucker Theorem*, in Nonlinear Programming, J. Abadie, ed., John Wiley, New York, 1967, pp. 21–36.

[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 (2007), pp. 1286–1309.

[3] R. Andreani, E. H. Fukuda, G. Haeser, D. O. Santos, and L. D. Secchin, *Optimality conditions for nonlinear second-order cone programming and symmetric cone programming*, Optimization Online, (2019).

[4] R. Andreani, G. Haeser, and J. M. Martínez, *On sequential optimality conditions for smooth constrained optimization*, Optimization, 60 (2011), pp. 627–641.

[5] R. Andreani, G. Haeser, A. Ramos, and P. J. S. Silva, *A second-order sequential optimality condition associated to the convergence of algorithms*, IMA Journal of Numerical Analysis, 37 (2017), pp. 1902–1929.

[6] R. Andreani, G. Haeser, M. L. Schuverdt, and P. J. S. Silva, *Two new weak constraint qualifications and applications*, SIAM Journal on Optimization, 22 (2012), pp. 1109–1135.

[7] R. Andreani, G. Haeser, L. Secchin, and P. Silva, *New sequential optimality conditions for mathematical programs with complementarity constraints and algorithmic consequences*, SIAM Journal on Optimization, 29 (2019), pp. 3201–3230, https://doi.org/10.1137/18M121040X.

[8] R. Andreani, G. Haeser, and D. Viana, *Optimality conditions and global convergence for nonlinear semidefinite programming*, Mathematical Programming, 180 (2020), pp. 203–235.

[9] R. Andreani, J. M. Martínez, A. Ramos, and P. J. S. Silva, *Strict constraint qualifications and sequential optimality conditions for constrained optimization*, Mathematics of Operations Research, 43 (2018), pp. 693–717.

[10] R. Andreani, J. M. Martínez, and B. F. Svaiter, *A new sequencial optimality condition for constrained optimization and algorithmic consequences*, SIAM Journal on Optimization, 20 (2010), pp. 3533–3554.

[11] M. Baes, *Convexity and differentiability properties of spectral functions and spectral mappings on Euclidean Jordan algebras*, Linear Algebra and its Applications, 422 (2007), pp. 664–700.

[12] L. Bueno, G. Haeser, and F. Rojas, *Optimality conditions and constraint qualifications for generalized Nash equilibrium problems and their practical implications*, SIAM Journal on Optimization, 29 (2019), pp. 31–54, https://doi.org/10.1137/17M1162524.

[13] L. F. Bueno, G. Haeser, F. Lara, and F. Rojas, *An augmented Lagrangian method for quasi-equilibrium problems*, Computational Optimization and Applications, (2020), https://doi.org/10.1007/s10589-020-00180-4.

[14] J. Faraut and A. Korànyi, *Analysis on symmetric Cones*, Oxford mathematical monographs, Clarendon Press, Oxford, 1994.

[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 (2001), pp. 348–360.

[16] B. Fares, D. Noll, and P. Apkarian, *Robust control via sequential semidefinite programming*, SIAM Journal on Control and Optimization, 40 (2002), pp. 1791–1820.

[17] R. W. Freund, F. Jarre, and C. H. Vogelbusch, *Nonlinear semidefinite programming: sensitivity, convergence, and an application in passive reduced-order modeling*, Mathematical Programming, 109 (2007), pp. 581–611.

[18] M. Guignard, *Generalized Kunh-Tucker conditions for mathematical programming in a banach space*, SIAM Journal of Control, 7 (1969), pp. 232–241.

[19] G. Haeser, *A second-order optimality condition with first- and second-order complementarity associated with global convergence of algorithms*, Computational Optimization and Applications, 70 (2018), pp. 615–639.

[20] G. Haeser, H. Liu, and Y. Ye, *Optimality condition and complexity analysis for linearly-constrained optimization without differentiability on the boundary*, Mathematical Programming, 178 (2019), pp. 263–299, https://doi.org/10.1007/s10107-018-1290-4.

[21] G. Haeser and M. L. Schuverdt, *On approximate KKT condition and its extension to continuous variational inequalities*, Journal of Optimization Theory and Applications, 149 (2011), pp. 528–539.

[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 (2006), pp. 265–287.

[23] C. Kanzow, D. Steck, and D. Wachsmuth, *An augmented Lagrangian methods for optimization problems in Banach spaces*, SIAM Journal Control Optimization, 56 (2018), pp. 272–291, https://doi.org/10.1137/16M1107103.

[24] H. Konno, N. Kawadai, and D. Wu, *Estimation of failure probability using semi-definite Logit model*, Computational Management Science, 1 (2003), pp. 59–73.

[25] B. F. Lourenço, E. H. Fukuda, and M. Fukushima, *Optimality conditions for problems over symmetric cones and a simple augmented Lagrangian method*, Mathematics of Operations Research, 43 (2018), pp. 1233–1251.

[26] J. M. Martínez and E. Pilotta, *Inexact restoration algorithm for constrained optimization*, Journal of Optimization Theory and Applications, 104 (2000), pp. 135–163.

[27] L. Qi and Z. Wei, *On the constant positive linear dependence conditions and its application to SQP methods*, SIAM Journal on Optimization, 10 (2000), pp. 963–981.

[28] M. Stingl, M. Kočvara, and G. Leugering, *A sequential convex semidefinite programming algorithm with an application to multiple-load free material optimization*, SIAM Journal on Optimization, 20 (2009), pp. 130–155.

[29] L. Vandenberghe, S. Boyd, and S. P. Wu, *Determinant maximization with linear matrix inequality constraints*, SIAM Journal on Matrix Analysis and Applications, 19 (1998), pp. 499–533.

[30] Y. Yamakawa and T. Okuno, *Global convergence of a stabilized sequential quadratic semidefinite programming method for nonlinear semidefinite programs without constraint qualifications*, ArXiv:1909.13544, (2019).