

MIT Open Access Articles

On the Burer–Monteiro method for general semidefinite programs

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

As Published: <https://doi.org/10.1007/s11590-021-01705-4>

Publisher: Springer Berlin Heidelberg

Persistent URL: <https://hdl.handle.net/1721.1/136891>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.



On the Burer–Monteiro method for general semidefinite programs

Cite this article as: Diego Cifuentes, On the Burer–Monteiro method for general semidefinite programs, Optimization Letters <https://doi.org/10.1007/s11590-021-01705-4>

This Author Accepted Manuscript is a PDF file of an unedited peer-reviewed manuscript that has been accepted for publication but has not been copyedited or corrected. The official version of record that is published in the journal is kept up to date and so may therefore differ from this version.

Terms of use and reuse: academic research for non-commercial purposes, see here for full terms. <https://www.springer.com/aam-terms-v1>

Author accepted manuscript

| |
|--|
| Noname manuscript No. (will be inserted by the editor) |
|--|

On the Burer-Monteiro method for general semidefinite programs

Diego Cifuentes

the date of receipt and acceptance should be inserted later

Abstract Consider a semidefinite program involving an $n \times n$ positive semidefinite matrix X . The Burer-Monteiro method uses the substitution $X = YY^T$ to obtain a nonconvex optimization problem in terms of an $n \times p$ matrix Y . Boumal et al. showed that this nonconvex method provably solves equality-constrained semidefinite programs with a generic cost matrix when $p \gtrsim \sqrt{2m}$, where m is the number of constraints. In this note we extend their result to arbitrary semidefinite programs, possibly involving inequalities or multiple semidefinite constraints. We derive similar guarantees for a fixed cost matrix and generic constraints. We illustrate applications to matrix sensing and integer quadratic minimization.

Keywords Semidefinite programming, Burer-Monteiro method, Low rank factorization, Nonconvex optimization, Spurious local minima

1 Introduction

Consider a *semidefinite program* (SDP) in \mathbb{S}^n , the space of $n \times n$ symmetric matrices, with $m = m_1 + m_2$ constraints (m_1 equalities and m_2 inequalities):

$$\min_{X \in \mathcal{X}} C \bullet X, \quad \mathcal{X} := \{X \in \mathbb{S}_+^n : \mathcal{A}(X) - b \in \{0\}^{m_1} \times \mathbb{R}_+^{m_2}\} \quad (P)$$

where $C \in \mathbb{S}^n$, $b \in \mathbb{R}^m$, the symbol \bullet denotes the trace inner product, and $\mathcal{A} : \mathbb{S}^n \rightarrow \mathbb{R}^m$, $X \mapsto (A_1 \bullet X, \dots, A_m \bullet X)$ is a linear map. We assume that \mathcal{X} is nonempty. Though interior point methods can solve (P) up to an arbitrary precision in polynomial time, they often run into memory problems for large values of n . This has motivated a surge of newer, more scalable techniques;

D. Cifuentes
 Massachusetts Institute of Technology, Cambridge, MA, USA
 E-mail: diegcif@mit.edu

see the recent survey [19]. We study here the low rank factorization method, pioneered by Burer and Monteiro [11, 12].

The Burer-Monteiro method consists in writing $X = YY^T$ for some $Y \in \mathbb{R}^{n \times p}$, and solving the following nonconvex optimization problem:

$$\min_{Y \in \mathbb{R}^{n \times p}} C \bullet YY^T \quad \text{such that} \quad YY^T \in \mathcal{X}. \quad (BM)$$

Let $\tau(k) := \binom{k+1}{2}$ be the k -th triangular number. Barvinok [3] and Pataki [23] independently showed that (P) has an optimal solution of rank r , with $\tau(r) \leq m$. Consequently, problems (P) and (BM) are equivalent for any p with $\tau(p) \geq m$. But due to nonconvexity, local optimization methods may not always recover the global optimum of (BM) . Nonetheless, the Burer-Monteiro performs very well in several applications, see e.g., [10, 11, 17, 25].

There has been much recent work in understanding when local optimization leads to the global optimum of (BM) . Most remarkably, Boumal et al. [8, 9] showed that equality-constrained SDPs ($m_2=0$) have no spurious 2nd-order critical points when $\tau(p) > m$ under certain assumptions. Concretely, they require that the cost matrix C is generic and that the feasible set of (BM) is sufficiently regular (i.e., satisfies a constraint qualification). By *generic* we mean that the result holds outside a set of measure zero. Though other global guarantees for (BM) exist, e.g. [14, 21], their setting is more restrictive.

In this note we generalize the result from Boumal et al. [8, 9] to arbitrary SDPs, possibly involving inequalities or multiple positive semidefinite (PSD) constraints. For the inequality-constrained problem (P) , we show in Theorem 1 that if $\tau(p) > m$ and the cost is generic, then any 2nd-order critical point of (BM) is globally optimal. Similar guarantees might be derived even when the cost matrix is fixed, see Theorem 3. We show applications to integer quadratic minimization and PSD matrix sensing.

Our proof of Theorem 1 is simpler than the one in [8, 9], as it relies on nonlinear programming instead of Riemannian optimization. This simplicity is reflected in the fact that Theorem 1 does not require any regularity assumptions on the domain (constraint qualifications). Nevertheless, regularity conditions might still be needed to prevent the existence of local minima that do not satisfy the 2nd-order criticality conditions.

We also consider SDPs involving multiple PSD variables, and study the Burer-Monteiro method applied to a subset of these variables. We prove in Theorem 4 that, for a generic cost, any 2nd-order critical point is globally optimal when p satisfies a bound due to Pataki [23]. We present an application to symmetric matrix sensing (the restricted isometry property is not needed).

The structure of this note is as follows. Section 2 reviews the notion of 2nd-order critical points in nonlinear programming. Section 3 analyzes the Burer-Monteiro method for the inequality-constrained problem (P) . Section 4 studies SDPs involving multiple PSD constraints.

Related work. The guarantees from Boumal et al. [8, 9] have been further studied in [5, 13, 24, 27], but all these papers focus on the equality-constrained case. The bound $\tau(p) > m$ was shown to be optimal up to lower order terms

in [27]. Guarantees for approximate 2-critical points were derived in [5, 13, 24]. The first polynomial time bounds for the Burer-Monteiro method were recently proved in [13]. We hope that the techniques developed in this paper may lead to polynomial time guarantees for arbitrary SDPs.

2 Criticality conditions

We review the notion of critical points. Consider the *nonlinear program*

$$\min_y \{f(y) : h(y) \in \{0\}^{m_1 \times \mathbb{R}_+^{m_2}}\}, \quad (NLP)$$

where $f : \mathbb{R}^k \rightarrow \mathbb{R}$, $h : \mathbb{R}^k \rightarrow \mathbb{R}^{m_1+m_2}$ are twice continuously differentiable. Let $L(y, \lambda) = f(y) - \lambda \cdot h(y)$ be the Lagrangian function, where \cdot is the dot product. We use the notation $[m] := \{1, 2, \dots, m\}$ throughout the paper. Let $I(y) \subset [m]$ be the indices of the *active* constraints at y , i.e., the indices for which $h_i(y) = 0$. The 1st-order and 2nd-order necessary conditions for optimality are:

$$y \text{ feasible, } \lambda \in \mathbb{R}^{m_1 \times \mathbb{R}_+^{m_2}}, \quad \lambda_i = 0 \text{ for } i \notin I(y), \quad \nabla_y L(y, \lambda) = 0, \quad (1a)$$

$$u^T \nabla_{yy}^2 L(y, \lambda) u \geq 0, \quad \forall u \text{ such that } \nabla_y h_i(y) u = 0 \text{ for } i \in I(y). \quad (1b)$$

A point y is *1st-order critical* for (NLP), abbreviated *1-critical*, if there exist multipliers λ satisfying (1a). The point is *2nd-order critical*, abbreviated *2-critical*, if (1b) also holds. A critical point is *spurious* if it is not the global minimum of (NLP).

Given a local minimum y of (NLP), it is known that y satisfies (1) under suitable *regularity* assumptions. Different regularity conditions, known as *constraint qualifications*, have been proposed [4]. One of the simplest is:

$$\{\nabla h_i(y) : i \in I(y)\} \text{ are linearly independent.} \quad (LICQ)$$

Various algorithms with provable convergence guarantees to 2-critical points are known, see e.g., [2, 6, 15] and the references therein. These results rely either on (LICQ) or a weaker constraint qualification.

We will show in Section 3 that any feasible point for (BM) that satisfies (1) is globally optimal when p is sufficiently large. We point out that our proof remains valid even if we replace (1b) by the weaker 2nd-order condition:

$$u^T \nabla_{yy}^2 L(y, \lambda) u \geq 0, \quad \forall u \text{ such that } \nabla_y h(y) u = 0.$$

For clarity, we focus our discussion in Section 3 on conditions (1a) and (1b).

More generally, consider the *nonlinear conic program*

$$\min_{x,y} \{f(x,y) : h(x,y) = 0, x \in \mathcal{K}\}, \quad (NLCP)$$

where $f : \mathbb{R}^{k_1+k_2} \rightarrow \mathbb{R}$, $h : \mathbb{R}^{k_1+k_2} \rightarrow \mathbb{R}^m$ are twice continuously differentiable and $\mathcal{K} \subset \mathbb{R}^{k_1}$ is a closed convex cone. The Lagrangian function is $L(x, y, \lambda, s) =$

$f(x, y) - \lambda \cdot h(x, y) - s \cdot x$. The following 1st-order conditions are necessary for optimality under suitable regularity conditions, see e.g., [7, §3.1]:

$$(x, y) \text{ feasible, } s \in \mathcal{K}^*, \quad s \cdot x = 0, \quad \nabla_{x,y} L(x, y, \lambda, s) = 0, \quad (2a)$$

where $\mathcal{K}^* \subset \mathbb{R}^{k_1}$ is the dual cone of \mathcal{K} . Though there exist 2nd-order conditions for conic programs (e.g., [7, §3.2]), it suffices for us to restrict the domain to pairs (x, y) with a fixed value of x . We get a nonlinear program in y , with 2nd-order condition:

$$u^T \nabla_{yy}^2 L(x, y, \lambda, s) u \geq 0, \quad \forall u \text{ such that } \nabla_y h(x, y) u = 0. \quad (2b)$$

A point (x, y) is *1-critical* for (NLCP) if it satisfies (2a) for some λ, s . The point is *2-critical* if (2b) also holds.

There are several algorithms for (NLCP) for the case $\mathcal{K} = \mathbb{S}_+^n$, see the survey paper [28]. Symmetric cones (e.g., products of PSD cones) were studied in [18]. These methods are provably convergent to 1-critical points. In order to escape from points that do not satisfy (2b) we may rely on 2nd-order methods for the (NLP) given by fixing the x coordinate.

The conditions in (2) will be used in Section 4 to analyze the Burer-Monteiro method for SDPs with multiple PSD constraints.

3 Inequality constrained SDPs

Consider problems (P) and (BM). For $X \in \mathcal{X}$, recall that the i -th constraint is active at X if $A_i \bullet X = b_i$. Let $m' = m'(\mathcal{X})$ be the largest number of linearly independent constraints that can be simultaneously active. For instance, if $m_2 = 0$ then $m' = \text{rank } \mathcal{A}$. We will show the following theorem.

Theorem 1 *Let p such that $\tau(p) > m'$. For a generic C , problem (BM) has no spurious 2-critical points. This means that any 2-critical point Y for (BM) is also globally optimal, and hence YY^T is optimal for (P).*

Example 1 (Integer quadratic minimization) Consider the optimization problem $\min\{f(x) : x \in \mathbb{Z}^n\}$ where $f(x)$ is a convex quadratic function. Denoting $\tilde{x} := (x, 1) \in \mathbb{Z}^{n+1}$, we may write $f(x) = \tilde{x}^T C \tilde{x}$ for some $C \in \mathbb{S}^{n+1}$. The following SDP relaxation for this problem was proposed in [22]:

$$\min_X C \bullet X \quad \text{s.t.} \quad X_{i,i} \geq X_{i,n+1} \text{ for } i \in [n], \quad X_{n+1,n+1} = 1, \quad X \in \mathbb{S}_+^{n+1}.$$

By Theorem 1, for a generic cost function any 2-critical point of the Burer-Monteiro problem is globally optimal when $\tau(p) > n+1$.

By generic, we mean the following. For fixed \mathcal{A}, b , the set of all cost matrices $C \in \mathbb{S}^n$ for which (BM) has a spurious 2-critical point has measure zero. We can provide an explicit characterization of this measure-zero set in \mathbb{S}^n . This

set is contained in the Minkowski sum of two special algebraic sets. The first algebraic set is given by a rank constraint:

$$\mathbb{S}_{n-p}^n := \{X : \text{rank } X \leq n-p\} \subset \mathbb{S}^n. \quad (3)$$

It is known that $\dim \mathbb{S}_{n-p}^n = \tau(n) - \tau(p)$, see e.g., [16, Prop.2.1]. The second algebraic set is a union of linear subspaces:

$$\mathcal{L} := \bigcup_I \mathcal{L}_I \subset \mathbb{S}^n, \quad \text{with} \quad \mathcal{L}_I := \text{span}\{A_i : i \in I\}, \quad (4)$$

where the union is over the possible subsets of constraints $I \subset [m]$ that can be simultaneously active. Note that $\dim \mathcal{L} = m'$ by definition of m' .

Theorem 2 *If (BM) has a spurious 2-critical point then $C \in \mathbb{S}_{n-p}^n + \mathcal{L}$.*

Theorem 1 follows directly from Theorem 2. Indeed, if $\tau(p) > m'$ then

$$\dim(\mathbb{S}_{n-r}^n + \mathcal{L}) \leq (\tau(n) - \tau(p)) + m' < \tau(n) = \dim \mathbb{S}^n. \quad (5)$$

Therefore the closure of $\mathbb{S}_{n-r}^n + \mathcal{L}$ is a proper algebraic set in \mathbb{S}^n , and has measure zero.

We proceed to prove Theorem 2. We first derive the criticality conditions for (BM). This is a special instance of (NLP), so we need to specialize (1). We have $h(Y) = \mathcal{A}(YY^T) - b$ and $L(Y, \lambda) = S(\lambda) \bullet YY^T + b^T \lambda$, where

$$S(\lambda) := C - \mathcal{A}^*(\lambda) \in \mathbb{S}^n \text{ is the slack matrix,}$$

and $\mathcal{A}^* : \mathbb{R}^m \rightarrow \mathbb{S}^n$, $\lambda \mapsto \sum_i \lambda_i A_i$ is the adjoint of \mathcal{A} . The 1st-order and 2nd-order necessary conditions for optimality are:

$$YY^T \in \mathcal{X}, \quad \lambda \in \mathbb{R}^{m_1} \times \mathbb{R}_+^{m_2}, \quad \lambda_i = 0 \text{ for } i \notin I(Y), \quad S(\lambda)Y = 0, \quad (6a)$$

$$S(\lambda) \bullet UU^T \geq 0, \quad \forall U \in \mathbb{R}^{n \times p} \text{ such that } A_i \bullet UY^T = 0 \text{ for } i \in I(Y). \quad (6b)$$

The following lemma establishes sufficient conditions for a critical point to be global optimal. The lemma is known, see [9, 12, 17], but our assumptions are slightly different since we allow inequalities.

Lemma 1 *Either of the following conditions imply global optimality:*

- (i) Y is 1-critical and the multiplier λ satisfies $S(\lambda) \in \mathbb{S}_+^n$,
- (ii) or Y is 2-critical and Y is column rank deficient.

Proof (i) The conic dual of (P) is $\max_{\lambda} \{b^T \lambda : S(\lambda) \in \mathbb{S}_+^n, \lambda \in \mathbb{R}^{m_1} \times \mathbb{R}_+^{m_2}\}$. Let (Y, λ) satisfy (6a), and let $X := YY^T$. We will show that the primal/dual pair (X, λ) is optimal for the SDP. It suffices to verify three conditions: X is primal feasible, λ is dual feasible, and complementary slackness holds (i.e., $\lambda_i = 0$ for $i \notin I(X)$ and $S(\lambda)X = 0$). Primal feasibility and complementary slackness follow from (6a), while dual feasibility corresponds to $S(\lambda) \in \mathbb{S}_+^n$.

(ii) Let (Y, λ) satisfy (6). By the above item, it suffices to show that $S(\lambda) \in \mathbb{S}_+^n$. Let $x \in \mathbb{R}^n$, and let us see that $x^T S(\lambda)x \geq 0$. Since Y is rank deficient, there is a nonzero vector $z \in \mathbb{R}^p$ such that $Yz = 0$. The matrix $U := xz^T$ satisfies $UY^T = 0$, so $S(\lambda) \bullet UU^T \geq 0$ by (6b). Since $S(\lambda) \bullet UU^T = \|z\|^2 (x^T S(\lambda)x)$, then $x^T S(\lambda)x \geq 0$. □

We are ready to prove Theorem 2 (which implies Theorem 1).

Proof (of Theorem 2) Let (Y, λ) be a spurious point satisfying (6). Observe that $\text{rank } Y = p$, as otherwise Lemma 1(ii) would imply that Y is not spurious. By (6a) we have $S(\lambda)Y = 0$, which implies $S(\lambda) \in \mathbb{S}_{n-p}^n$, and also $\lambda_i = 0$ for $i \notin I(Y)$. Thus $C = S(\lambda) + \mathcal{A}^*(\lambda) \in \mathbb{S}_{n-p}^n + \mathcal{L}$. \square

To finish this section, we observe that Theorem 2 can be used even if the cost matrix C is not generic. For instance, the next theorem assumes that both b, C are fixed and \mathcal{A} is generic (i.e., A_1, \dots, A_m are generic).

Theorem 3 *Let p such that $\tau(p) > m$ and $\text{rank } C > n - p$. For a generic \mathcal{A} , problem (BM) has no spurious 2-critical points.*

Proof By Theorem 2, it suffices to see that $C \notin \mathbb{S}_{n-p}^n + \mathcal{L}$. Fix $I \subset [m]$, and let $\mathcal{L}_I \subset \mathbb{S}^n$ as in (4). Note that \mathcal{L}_I is generic among the subspaces of dimension $|I|$, as it depends on the generic matrices A_i . Recall that $\dim(\mathbb{S}_{n-p}^n + \mathcal{L}_I) < \dim \mathbb{S}^n$ by (5). Since $C \notin \mathbb{S}_{n-p}^n$, then $C \notin \mathbb{S}_{n-p}^n + \mathcal{L}_I$ for a generic \mathcal{L}_I . The result follows from $\mathcal{L} = \bigcup_I \mathcal{L}_I$. \square

An additional advantage of having generic constraints is that regularity is always satisfied. Therefore any local minimum of (BM) is also 2-critical, and hence is subject to Theorem 3. The next proposition is shown in Appendix A.

Proposition 1 *Assume that the entries of b are nonzero. For a generic \mathcal{A} , any feasible point of (BM) satisfies (LICQ).*

Example 2 (Matrix sensing) Given a linear map $\mathcal{A} : \mathbb{S}^n \rightarrow \mathbb{R}^m$ and a vector $b \in \mathbb{R}^m$, consider finding a low rank matrix $X \in \mathbb{S}^n$ such that $\mathcal{A}(X) = b$. A standard technique to promote low rank is to minimize the nuclear norm:

$$\min_{X \in \mathbb{S}^n} \|X\|_* \quad \text{such that} \quad \mathcal{A}(X) = b. \quad (7)$$

If we further assume that X is PSD, the cost function is $I_n \bullet X$. By Theorem 3, if \mathcal{A} is generic and $\tau(p) > m$, then any local minimum of (BM) is globally optimal. The PSD assumption will be relaxed in the next section.

Remark 1 Different guarantees about the Burer-Monteiro method for matrix sensing were obtained in [21], relying on the restricted isometry property.

4 General SDPs

Let $\mathbf{n} := (n_1, \dots, n_\ell) \in \mathbb{N}^\ell$ and $d \in \mathbb{N}$. We consider an SDP involving PSD matrices of sizes n_1, \dots, n_ℓ and a free variable of dimension d . Let the Euclidean space $\mathbb{S}^{\mathbf{n}} := \mathbb{S}^{n_1} \times \dots \times \mathbb{S}^{n_\ell}$ and the convex cone $\mathbb{S}_+^{\mathbf{n}} := \mathbb{S}_+^{n_1} \times \dots \times \mathbb{S}_+^{n_\ell}$. Given $C \in \mathbb{S}^{\mathbf{n}} \times \mathbb{R}^d$, $b \in \mathbb{R}^m$, and a linear map $\mathcal{A} : \mathbb{S}^{\mathbf{n}} \times \mathbb{R}^d \rightarrow \mathbb{R}^m$, consider:

$$\min_{X \in \mathcal{X}} \langle C, X \rangle, \quad \mathcal{X} := \{X \in \mathbb{S}_+^{\mathbf{n}} \times \mathbb{R}^d : \mathcal{A}(X) = b\}, \quad (P_{\mathbf{n}})$$

where $X := (X_1, \dots, X_\ell, x)$ with $X_j \in \mathbb{S}^{n_j}$, $x \in \mathbb{R}^d$. As before, we assume that \mathcal{X} is nonempty and that the minimum is achieved.

We apply the Burer-Monteiro method to the first k matrices. Let $Y := (Y_1, \dots, Y_k)$, with $Y_j \in \mathbb{R}^{n_j \times p_j}$, and let $q(Y) := (Y_1 Y_1^T, \dots, Y_k Y_k^T)$. We denote

$$\underline{\mathbf{n}} := (n_1, \dots, n_k), \quad \overline{\mathbf{n}} := (n_{k+1}, \dots, n_\ell), \quad \overline{X} := (X_{k+1}, \dots, X_\ell).$$

In particular, $\mathbb{S}^{\mathbf{n}} = \mathbb{S}^{\underline{\mathbf{n}}} \times \mathbb{S}^{\overline{\mathbf{n}}}$. The Burer-Monteiro problem is:

$$\min_{Y, \overline{X}, x} \langle C, (q(Y), \overline{X}, x) \rangle \quad \text{such that} \quad (q(Y), \overline{X}, x) \in \mathcal{X}. \quad (BM_{\mathbf{n}})$$

Pataki [23] showed that $(P_{\mathbf{n}})$ always has an optimal solution such that $\sum_{j=1}^{\ell} \tau(r_j) \leq m - d$, where $r_j := \text{rank } X_j$. We can ensure that there is a solution with $r_j \leq p_j$ for all $j \in [k]$ if either $p_j \geq n_j$ or $\tau(p_j) \geq m'$, with

$$m' := \max_{r_{k+1}, \dots, r_\ell} m - d - \tau(r_{k+1}) - \tau(r_{k+2}) - \dots - \tau(r_\ell),$$

where the maximum is over the possible ranks r_{k+1}, \dots, r_ℓ . Hence, problems $(P_{\mathbf{n}})$ and $(BM_{\mathbf{n}})$ agree when $\tau(p_j) \geq \min\{m', \tau(n_j)\}$ for $j \in [k]$.

Theorem 4 *Assume that $\tau(p_j) > \min\{m', \tau(n_j)\}$ for $j \in [k]$. For a generic C , problem $(BM_{\mathbf{n}})$ has no spurious 2-critical points.*

Example 3 (Inequalities) Consider the inequality constrained problem (P) . We may view each of the m_2 inequalities as a PSD constraint on a 1×1 matrix. So this is a special instance of $(P_{\mathbf{n}})$ with $k=1$, $\ell=m_2+1$, $d=0$, and $n_2 = \dots = n_\ell = 1$. Note that $r_{i+1}=1$ when the i -th inequality constraint is inactive, and is zero otherwise. Hence $m' = m - \#(\text{inactive constra}) = \#(\text{active constra})$. This is consistent with the results from Section 3.

Example 4 (Second-order cone) Let $\mathcal{Q}^n := \{x \in \mathbb{R}^n : \|(x_2, \dots, x_n)\| \leq x_1\}$ be the second-order cone. Consider minimizing a linear cost on $\mathbb{S}_+^{n_1} \times \mathcal{Q}^{n_2}$ subject to m_1 linear equalities. Apply the Burer-Monteiro factorization to the matrix in $\mathbb{S}_+^{n_1}$. We can embed \mathcal{Q}^{n_2} inside $\mathbb{S}_+^{n_2}$ by adding $\tau(n_2-1)$ new linear equalities, see [1, pg.7]. So this is a special case of $(BM_{\mathbf{n}})$ with $\ell=2$, $k=1$, $d=0$, $m=m_1+\tau(n_2-1)$. Given $x \in \mathcal{Q}^{n_2}$, the rank of the corresponding PSD matrix is $r_2=0$ if $x=0$, $r_2=n_2-1$ if x lies in the boundary, and $r_2=n_2$ if x lies in the interior. So Theorem 4 applies when $\tau(p_1) > m_1+\tau(n_2-1)-\tau(r_2)$, where r_2 is the smallest feasible rank. We point out that embedding \mathcal{Q}^{n_2} inside $\mathbb{S}_+^{n_2}$ is used for the analysis, but we do not need to do this in practice. The reason is that the embedding preserves critical points.

We also provide an explicit characterization of the costs C for which spurious 2-critical points may exist. These costs lie in the Minkowski sum of two algebraic sets, which are closely related to the ones in (3) and (4).

Theorem 5 *If $(BM_{\mathbf{n}})$ has a spurious 2-critical point, then C lies in the algebraic set $\underline{\mathcal{V}} \times \overline{\mathcal{V}} \times \{0^d\} + \text{Im } \mathcal{A}^* \subset \mathbb{S}^{\mathbf{n}} \times \mathbb{R}^d$, with*

$$\begin{aligned} \underline{\mathcal{V}} &:= \bigcup_{j \in [k]: p_j \leq n_j} (\mathbb{S}^{n_1} \times \cdots \times \mathbb{S}^{n_{j-1}} \times \mathbb{S}_{n_j - p_j}^{n_j} \times \mathbb{S}^{n_{j+1}} \times \cdots \times \mathbb{S}^{n_\ell}) \subset \mathbb{S}^{\mathbf{n}}, \\ \overline{\mathcal{V}} &:= \bigcup_{r_{k+1}, \dots, r_\ell} (\mathbb{S}_{n_{k+1} - r_{k+1}}^{n_{k+1}} \times \cdots \times \mathbb{S}_{n_\ell - r_\ell}^{n_\ell}) \subset \mathbb{S}^{\overline{\mathbf{n}}}, \end{aligned}$$

where the last union is over the possible ranks r_{k+1}, \dots, r_ℓ in $(P_{\mathbf{n}})$.

Theorem 4 follows from Theorem 5 by counting dimensions. Let p_{\min} be the minimum of p_1, \dots, p_k , ignoring the values with $p_j > n_j$. Note that

$$\dim \underline{\mathcal{V}} = \sum_{j \leq k} \tau(n_j) - \tau(p_{\min}), \quad \dim \overline{\mathcal{V}} = \max_{r_{k+1}, \dots, r_\ell} \sum_{j > k} \tau(n_j) - \tau(r_j).$$

Let $D := \dim(\mathbb{S}^{\mathbf{n}} \times \mathbb{R}^d) = \tau(n_1) + \cdots + \tau(n_\ell) + d$. If $\tau(p_{\min}) > m'$, then

$$\begin{aligned} \dim(\underline{\mathcal{V}} \times \overline{\mathcal{V}} \times \{0\} + \text{Im } \mathcal{A}^*) &= m + \sum_{j \leq k} \tau(n_j) - \tau(p_{\min}) + \max_{r_{k+1}, \dots, r_\ell} \sum_{j > k} \tau(n_j) - \tau(r_j) \\ &= D - \tau(p_{\min}) + \max_{r_{k+1}, \dots, r_\ell} \left\{ m - d - \sum_{j > k} \tau(r_j) \right\} = D - \tau(p_{\min}) + m' < D, \end{aligned}$$

Hence $\underline{\mathcal{V}} \times \overline{\mathcal{V}} \times \{0\} + \text{Im } \mathcal{A}^*$ has measure zero.

We proceed to prove Theorem 5. We first derive the optimality conditions for $(BM_{\mathbf{n}})$. This is a special instance of $(NLCP)$, so we need to specialize (2). For $\lambda \in \mathbb{R}^m$, consider the slack variable $S(\lambda) := C - \mathcal{A}^*(\lambda) \in \mathbb{S}^{\mathbf{n}} \times \mathbb{R}^n$. Let $S_j(\lambda) \in \mathbb{S}^{n_j}$ be the j -th component of $S(\lambda)$. Similarly define $\overline{S}(\lambda) \in \mathbb{S}^{\overline{\mathbf{n}}}$ and $s(\lambda) \in \mathbb{R}^d$. The necessary conditions for optimality are:

$$(q(Y), \overline{X}, x) \in \mathcal{X}, \quad \overline{S}(\lambda) \in \mathbb{S}_+^{\overline{\mathbf{n}}}, \quad \langle \overline{S}(\lambda), \overline{X} \rangle = 0, \quad s(\lambda) = 0, \quad S_j(\lambda) Y_j = 0, \quad (8a)$$

$$S_j(\lambda) \bullet U_j U_j^T \geq 0, \quad \forall U_j \in \mathbb{R}^{n_j \times p_j} \text{ s.t. } \mathcal{A}_j(U_j Y_j^T) = 0 \quad (\text{for } j \in [k]). \quad (8b)$$

We now provide sufficient conditions for global optimality of critical points.

Lemma 2 *Either of the following conditions imply global optimality:*

- (i) (Y, \overline{X}, x) is 1-critical and $S_j(\lambda) \in \mathbb{S}_+^{n_j}$ for $j \in [k]$.
- (ii) or (Y, \overline{X}, x) is 2-critical and Y_j is column rank deficient for $j \in [k]$.

Proof The proof is analogous to Lemma 1. For (i) we compare (8a) with the primal/dual optimality conditions for $(P_{\mathbf{n}})$. For (ii) we use a vector $z_j \in \mathbb{R}^{p_j}$ in the right kernel of Y_j in order to show that $S_j(\lambda) \in \mathbb{S}_+^{n_j}$.

Proof (of Theorem 5) Let $(Y, \overline{X}, x, \lambda)$ a spurious point satisfying (8). By Lemma 2(ii) we have $\text{rank } Y_j = p_j$ for some $j \in [k]$. As $S_j(\lambda) Y_j = 0$ then $S_j(\lambda) \in \mathbb{S}_{n_j - p_j}^{n_j}$. Let (r_{k+1}, \dots, r_ℓ) be the ranks of \overline{X} . Since $\langle \overline{S}(\lambda), \overline{X} \rangle = 0$ and both lie in $\mathbb{S}_+^{\overline{\mathbf{n}}}$, then $\overline{S}(\lambda) \in \mathbb{S}_{n_{k+1} - r_{k+1}}^{n_{k+1}} \times \cdots \times \mathbb{S}_{n_\ell - r_\ell}^{n_\ell}$. Hence $S(\lambda) \in \underline{\mathcal{V}} \times \overline{\mathcal{V}} \times \{0\}$, as $s(\lambda) = 0$. The result follows from $C = S(\lambda) + \mathcal{A}^*(\lambda)$. \square

As illustrated next, Theorem 5 can be used even when C is not generic.

Example 5 (Matrix sensing) We revisit the problem of sensing symmetric matrices from Example 2. For $X \in \mathbb{S}^n$, its nuclear norm satisfies:

$$\|X\|_* = \min_Z I_n \bullet Z \quad \text{such that} \quad Z+X \in \mathbb{S}_+^n, \quad Z-X \in \mathbb{S}_+^n.$$

Let $X_1 := \frac{1}{2}(Z+X)$, $X_2 := \frac{1}{2}(Z-X)$. We can rewrite problem (7) as follows:

$$\min_{X_1, X_2} I_n \bullet X_1 + I_n \bullet X_2 \quad \text{such that} \quad \mathcal{A}(X_1) - \mathcal{A}(X_2) = b, \quad X_1 \in \mathbb{S}_+^n, \quad X_2 \in \mathbb{S}_+^n.$$

Consider the Burer-Monteiro method applied to both matrices X_1, X_2 , so that $k = \ell = 2$, using the same rank p for both matrices. We will prove that there are no spurious 2-critical points when $\mathcal{A} : \mathbb{S}^n \rightarrow \mathbb{R}^m$ is generic and $\tau(p) > m$. By Theorem 5, we need to show that

$$(I_n, I_n) \notin \mathcal{V} + (1, -1) \otimes \text{Im } \mathcal{A}^*, \quad \text{where} \quad \mathcal{V} := \mathbb{S}_{n-p}^n \times \mathbb{S}^n \cup \mathbb{S}^n \times \mathbb{S}_{n-p}^n.$$

It suffices to see that $I_n \notin \mathbb{S}_{n-p}^n + \text{Im } \mathcal{A}^*$. But this was shown in Theorem 3.

Acknowledgements The author thanks Nicolas Boumal, Ankur Moitra, Pablo Parrilo, and David Rosen for helpful discussions and comments.

A Regularity with generic constraints

Proposition 1 follows from classical results about the genericity of optimality conditions in nonlinear programming [26]. Nonetheless, we present here the proof for completeness. We rely on Sard's theorem from differential geometry, see e.g., [20, §2].

Theorem 6 (Sard) *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a smooth map, with $n \geq m$. Let $v \in \mathbb{R}^m$ be a generic point. Then $\text{rank}(\nabla f(y)) = m$ for any $y \in f^{-1}(v)$.*

Proof (of Proposition 1) Fix a set of indices $I \subset [m]$, and let

$$\mathcal{M}_I := \{Y \in \mathbb{R}^{n \times p} : A_i \bullet YY^T = b_i \text{ for } i \in I\}.$$

We claim that $(LICQ)$ holds at all points on \mathcal{M}_I (i.e., the gradients are linearly independent). If this happens for each $I \subset [m]$, then $(LICQ)$ also holds for the feasible set of (BM) . So it suffices to show the claim.

We prove the claim under a more restrictive genericity setting. We assume that each $b_i \neq 0$ and that $A_i = \alpha_i \bar{A}_i$, where $\{\bar{A}_i\}$ are fixed matrices and $\{\alpha_i\}$ are generic scalars. Let

$$f_I : \mathbb{R}^{n \times p} \rightarrow \mathbb{R}^I, \quad Y \mapsto (\bar{A}_i \bullet YY^T : i \in I).$$

The vector $v := (b_i/\alpha_i : i \in I)$ is generic since $\{\alpha_i\}$ are generic. By Theorem 6, $\nabla f_I(Y)$ is full rank for any $Y \in f_I^{-1}(v) = \mathcal{M}_I$. So $(LICQ)$ holds on \mathcal{M}_I . \square

Remark 2 Proposition 1 remains valid if $m_1 \geq 1$ and only the first entry of b is nonzero. Indeed, by adding a generic scalar multiple of the first equation to the other ones we get back to the case where all the entries of b are nonzero.

References

1. F. Alizadeh and D. Goldfarb. Second-order cone programming. *Math. Program.*, 95:3–51, 2003.
2. R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. Second-order negative-curvature methods for box-constrained and general constrained optimization. *Comput. Optim. Appl.*, 45(2):209–236, 2010.
3. A. I. Barvinok. Problems of distance geometry and convex properties of quadratic maps. *Discrete Comput. Geom.*, 13(2):189–202, 1995.
4. M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear programming: theory and algorithms*. John Wiley & Sons, 2013.
5. S. Bhojanapalli, N. Boumal, P. Jain, and P. Netrapalli. Smoothed analysis for low-rank solutions to semidefinite programs in quadratic penalty form. In *Conf. Learn. Theory*, pages 3243–3270, 2018.
6. E. G. Birgin, G. Haeser, and A. Ramos. Augmented Lagrangians with constrained subproblems and convergence to second-order stationary points. *Comput. Optim. Appl.*, 69(1):51–75, 2018.
7. J. F. Bonnans and A. Shapiro. *Perturbation analysis of optimization problems*. Springer Science & Business Media, 2013.
8. N. Boumal, V. Voroninski, and A. Bandeira. The non-convex Burer-Monteiro approach works on smooth semidefinite programs. In *Adv. Neural Inf. Process. Syst.*, pages 2757–2765, 2016.
9. N. Boumal, V. Voroninski, and A. Bandeira. Deterministic guarantees for Burer-Monteiro factorizations of smooth semidefinite programs. *Commun. Pure Appl. Math.*, 2019.
10. S. Burer, R. Monteiro, and C. Choi. SDPLR 1.03. Available at <http://sburer.github.io/projects>.
11. S. Burer and R. D. Monteiro. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Math. Program.*, 95(2):329–357, 2003.
12. S. Burer and R. D. Monteiro. Local minima and convergence in low-rank semidefinite programming. *Math. Program.*, 103(3):427–444, 2005.
13. D. Cifuentes and A. Moitra. Polynomial time guarantees for the Burer-Monteiro method. *arXiv:1912.01745*, 2019.
14. R. Ge, J. D. Lee, and T. Ma. Matrix completion has no spurious local minimum. In *Adv. Neural Inf. Process. Syst.*, pages 2973–2981, 2016.
15. P. E. Gill, V. Kungurtsev, and D. P. Robinson. A stabilized SQP method: global convergence. *IMA J. Numer. Anal.*, 37(1):407–443, 2017.
16. U. Helmke and M. A. Shayman. Critical points of matrix least squares distance functions. *Linear Algebra Appl.*, 215:1–19, 1995.
17. M. Journée, F. Bach, P.-A. Absil, and R. Sepulchre. Low-rank optimization on the cone of positive semidefinite matrices. *SIAM J. Optim.*, 20(5):2327–2351, 2010.
18. Y.-J. Liu and L.-W. Zhang. On the approximate augmented Lagrangian for nonlinear symmetric cone programming. *Nonlinear Anal. Theory Methods Appl.*, 68(5):1210–1225, 2008.
19. A. Majumdar, G. Hall, and A. Ahmadi. Recent scalability improvements for semidefinite programming with applications in machine learning, control, and robotics. *Annu. Rev. Control Rob. Auton. Syst.*, 3, 2019.
20. J. Milnor and D. W. Weaver. *Topology from the differentiable viewpoint*. Princeton University Press, 1997.
21. D. Park, A. Kyrillidis, C. Carmanis, and S. Sanghavi. Non-square matrix sensing without spurious local minima via the Burer-Monteiro approach. In *Artif. Intell. Stat.*, pages 65–74, 2017.
22. J. Park and S. Boyd. A semidefinite programming method for integer convex quadratic minimization. *Optim. Lett.*, 12(3):499–518, 2018.
23. G. Pataki. On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues. *Math. Oper. Res.*, 23(2):339–358, 1998.
24. T. Pumir, S. Jelassi, and N. Boumal. Smoothed analysis of the low-rank approach for smooth semidefinite programs. In *Adv. Neural Inf. Process. Syst.*, pages 2287–2296, 2018.

-
25. D. Rosen, L. Carlone, A. Bandeira, and J. Leonard. A certifiably correct algorithm for synchronization over the special Euclidean group. In *Intl. Workshop Algorithmic Found. Rob. (WAFR)*, 2016.
 26. J. E. Spingarn and R. T. Rockafellar. The generic nature of optimality conditions in nonlinear programming. *Math. Oper. Res.*, 4(4):425–430, 1979.
 27. I. Waldspurger and A. Waters. Rank optimality for the Burer-Monteiro factorization. *arXiv:1812.03046*, 2018.
 28. H. Yamashita and H. Yabe. A survey of numerical methods for nonlinear semidefinite programming. *J. Oper. Res. Soc. Jpn.*, 58(1):24–60, 2015.

Author accepted manuscript