BURER-MONTEIRO GUARANTEES FOR GENERAL SEMIDEFINITE PROGRAMS

DIEGO CIFUENTES

ABSTRACT. Consider a semidefinite program (SDP) involving an $n \times n$ positive semidefinite matrix X. The Burer-Monteiro method consists in solving a nonconvex program in Y, where Y is an $n \times p$ matrix such that $X = YY^T$. Despite nonconvexity, Boumal et al. showed that the method provably solves generic equality-constrained SDP's when $p \gtrsim \sqrt{2m}$, where m is the number of constraints. We extend this result to arbitrary SDP's, possibly involving inequalities or multiple semidefinite constraints. We illustrate applications to sensor network localization and to matrix sensing.

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):

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

where $C\in\mathbb{S}^n$, $b\in\mathbb{R}^m$ and $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$, $X \mapsto (A_1\bullet X, \ldots, A_m\bullet X)$ is a linear map. We assume that $\mathfrak{C}$ is nonempty and that the minimum is achieved. Though interior point methods can solve (SDP) in polynomial time, they typically run into memory problems for large values of n. This has motivated the development of several new methods with lower memory requirements, see e.g., [7–9, 11, 20]. Among them is the low rank factorization method, pioneered by Burer and Monteiro [7, 8].

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:

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

Let $\tau(k) := \binom{k+1}{2}$ denote the k-th triangular number. Barvinok [1] and Pataki [15] independently showed that (SDP) has an optimal solution of rank r, with $\tau(r) \leq m$. Consequently, problems (SDP) and (BM) have the same optimal value for any p such that $\tau(p) \geq m$. But due to nonconvexity, local optimization methods may not always recover the global optimum of (BM). Nonetheless, the Burer-Monteiro method performs very well in several applications, see e.g., [7, 12, 18].

Key words and phrases. Semidefinite programming, Burer-Monteiro, Low rank factorization, Global convergence.

There has been much recent work in proving global guarantees for (BM). Most remarkably, Boumal et al. [5,6] showed that generic equality-constrained SDP's (m_2=0) have no spurious local minima when $\tau(p) > m$, under some regularity conditions. Their result relies on a previous characterization of the local minima of (BM) by Burer and Monteiro [7]. These guarantees have been recently extended to approximate local minima [2, 16], and the bound on p was shown to be optimal [19]. Though other global guarantees for (BM) exist, e.g. [10, 14], their setting is more restrictive.

In this note we generalize the result from Boumal et al. [5, 6] to arbitrary SDP's, possibly involving inequalities or multiple positive semidefinite (PSD) constraints. Though we focus on real valued SDP's, all the results extend to the complex case. The structure of this note is as follows.

Section 2 analyzes the Burer-Monteiro method for problem (SDP). We show in Theorem 1 that if $\tau(p) > m$ and the cost is generic, then any second-order critical point of (BM) is globally optimal. If we further assume regularity conditions (constraint qualifications), we may conclude that no spurious local minima exist. Similar results might be proved even when the cost matrix is not generic, see Theorem 5. We show applications to sensor network localization and PSD matrix sensing.

Section 3 considers an SDP involving multiple PSD constraints. We study the Burer-Monteiro method applied to a given subset of these constraints. We prove in Theorem 7 that, generically, any critical point is globally optimal when p satisfies a bound due to Pataki [15]. We illustrate an application to symmetric matrix sensing (the restricted isometry property is not needed).

2. Inequality constrained SDP's

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

Theorem 1. *Let p such that $\tau(p) > m'$. For a generic C, any critical point Y of problem (BM) is globally optimal, and YY^T is optimal for (SDP).*

By generic, we mean the following. For fixed $\mathcal{A}$, b, the set of all $C \in \mathbb{S}^n$ for which (BM) has a spurious critical point has measure zero. Moreover, the set of all such C is contained in a proper algebraic set of $\mathbb{S}^n$, see Corollary 3.

We now recall the notion of critical points. Consider the nonlinear program $\min_y \{f(y) : h(y) \in \{0\}^{m_1} \times \mathbb{R}_+^{m_2}\}$. We say that y is a critical point if there exist multipliers $\lambda \in \mathbb{R}^m$ satisfying the KKT conditions. Let $L(y, \lambda) = f(y) - \lambda \cdot h(y)$ be the Lagrangian function, and let $I(y) \subset [m]$ be the indices of the active constraints at y. The first-order and second-order KKT conditions are:

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

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

Remark (Local minima). The KKT conditions are necessary for local optimality under suitable *regularity* assumptions [17]. Consequently, if we further assume in Theorem 1 that the feasible set of (BM) is regular everywhere, then we may conclude that there are no spurious local minima. We explain this more carefully in Appendix A. As opposed to [6], we do not need regularity to argue the absence of spurious critical points.

The optimality conditions for (BM) are obtained by specializing (1). We have $h(Y) = \mathcal{A}(YY^T) - b$ and $L(Y, \lambda) = S(\lambda) \bullet YY^T$, where

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

and $\mathcal{A}^* : \mathbb{R}^m \to \mathbb{S}^n$, $\lambda \mapsto \sum_i \lambda_i A_i$ is the adjoint of $\mathcal{A}$. The first-order and second-order conditions for (BM) are:

$$YY^T \in \mathfrak{C}, \quad \lambda \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}_+, \quad S(\lambda)Y = 0, \quad \lambda_i = 0 \text{ for } i \notin I(Y), \tag{2a}$$

$$S(\lambda) \bullet UU^T \geq 0, \quad \forall U \in \mathbb{R}^{n \times p} \text{ such that } A_i \bullet UY^T = 0 \text{ for } i \in I(Y). \tag{2b}$$

We say that Y is a *critical point* of (BM) if there is some $\lambda \in \mathbb{R}^m$ as above. The critical point is *spurious* if YY^T is not optimal for (SDP).

We proceed to prove Theorem 1. The next proposition says that any spurious critical point must be full rank. This proposition was originally shown in [8] and later in [6, 12]. Our assumptions are slightly different, since we allow inequalities. We follow the proof strategy from [6, 12], that relies on constructing a dual certificate of optimality.

Proposition 2. *Let Y be a critical point of (BM). If Y is rank deficient then YY^T is optimal for (SDP).*

Proof. The conic dual of (SDP) is $\max_\lambda \{b^T\lambda : S(\lambda) \in \mathbb{S}^n_+,\ \lambda \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}_+\}$. Let $\bar{\lambda}$ be the multiplier of Y, and let $\bar{X} := YY^T$. We will show that the primal/dual pair $(\bar{X}, \bar{\lambda})$ is optimal for the SDP. It suffices to show that the following three conditions are met: $\bar{X}$ is primal feasible, $\bar{\lambda}$ is dual feasible, and complementary slackness holds (i.e., $S(\bar{\lambda})\bar{X} = 0$ and $\bar{\lambda}_i = 0$ for $i \notin I(\bar{X})$). Primal feasibility and complementary slackness follow from (2a). As for dual feasibility, we need to show that $S(\bar{\lambda}) \in \mathbb{S}^n_+$. Let $x \in \mathbb{R}^n$, and let us see that $x^T S(\bar{\lambda}) x \geq 0$. 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 $S(\bar{\lambda}) \bullet UU^T \geq 0$ by (2b). Since $S(\bar{\lambda}) \bullet UU^T = \|z\|^2 (x^T S(\bar{\lambda}) x)$, then $x^T S(\bar{\lambda}) x \geq 0$. □

A consequence of the above proposition is that spurious critical points may only exist when the cost matrix C lies in a certain algebraic set.

Corollary 3. *If (BM) has a spurious critical point then $C \in \mathbb{S}^n_{n-p} + \mathcal{L}$, with*

$$\mathbb{S}^n_{n-p} := \{X : \operatorname{rank} X \leq n-p\} \subset \mathbb{S}^n, \tag{3}$$

$$\mathcal{L} := \bigcup_I \{\mathcal{A}^*(\lambda) : \lambda \in \mathbb{R}^m,\ \lambda_i = 0 \textit{ for } i \notin I\} \subset \mathbb{S}^n, \tag{4}$$

where the union is over the possible subsets of constraints $I \subset [m]$ that can be simultaneously active.

Proof. Let (Y, λ) satisfy (2a) and $\operatorname{rank} Y = p$. Then $S(\lambda)Y = 0$, which implies $S(\lambda) \in \mathbb{S}^n_{n-p}$, and also $\lambda_i = 0$ for $i \notin I(Y)$. Thus $C = S(\lambda) + \mathcal{A}^*(\lambda) \in \mathbb{S}^n_{n-p} + \mathcal{L}$. □

The above corollary characterizes the matrices C that may admit a spurious critical point. It remains to show that such a set has measure zero.

Proof of Theorem 1. Note that $\dim \mathbb{S}^n_{n-p} = \tau(n) - \tau(p)$, and that $\dim \mathcal{L} = m'$ by definition of m'. Since $\tau(p) > m'$, then $\dim \mathbb{S}^n_{n-r} + \dim \mathcal{L} < \tau(n) = \dim \mathbb{S}^n$. Therefore $\mathbb{S}^n_{n-r} + \mathcal{L}$ is a proper algebraic set in $\mathbb{S}^n$, and has measure zero. □

We now illustrate an application of Theorem 1.

Example 4 (Noisy sensor network localization)**.** Let $\{x_i\}_{i=1}^n \subset \mathbb{R}^\ell$ be positions of some unknown sensors, and let $\{a_k\}_{k=1}^m \subset \mathbb{R}^\ell$ be some known anchors. Consider finding $\{x_i\}_i$ satisfying the interval constraints:

$$\underline{d}_{ij} \leq \|x_i - x_j\| \leq \overline{d}_{ij} \text{ for } (i,j) \in N_x, \qquad \underline{d}_{kj} \leq \|a_k - x_j\| \leq \overline{d}_{kj} \text{ for } (k,j) \in N_a,$$

for given $\underline{d}_{ij}, \overline{d}_{ij}, \underline{d}_{kj}, \overline{d}_{kj}$. The following SDP relaxation was proposed in [3]:

$$\text{find } Z \in \mathbb{S}^{n+\ell}_+ \text{ s.t. } \underline{d}_{ij}^2 \leq \begin{pmatrix} 0 \\ e_i - e_j \end{pmatrix} \begin{pmatrix} 0 & e_i^T - e_j^T \end{pmatrix} \bullet Z \leq \overline{d}_{ij}^2, \ \ \underline{d}_{kj}^2 \leq \begin{pmatrix} a_k \\ -e_j \end{pmatrix} \begin{pmatrix} a_k^T & -e_j^T \end{pmatrix} \bullet Z \leq \overline{d}_{kj}^2.$$

Since there is no cost function, we may choose a generic matrix C. By Theorem 1, any critical point of the Burer-Monteiro problem is globally optimal when $\tau(p) > m'$. In particular, $p \geq \sqrt{2|N_x \cup N_a|}$ suffices.

To finish this section, we observe that Corollary 3 can be used even if the cost matrix C is not generic. For instance, the next theorem assumes that C is fixed and $\mathcal{A}$ is generic. Moreover, regularity is guaranteed when $\mathcal{A}$ is generic, as shown in Proposition 13 of Appendix A, so we may conclude that (BM) has no spurious local minima.

Theorem 5. *Let p such that $\tau(p) > m$ and* $\operatorname{rank} C > n - p$*. For a generic $\mathcal{A}$, any local minimum Y of problem (BM) is such that YY^T is optimal for (SDP).*

Proof. By Corollary 3 and Proposition 13, it suffices to see that $C \notin \mathbb{S}^n_{n-p} + \mathcal{L}$. Fix $I \subset [m]$, and let $\mathcal{L}_I \subset \mathbb{S}^n$ be the I-th subspace in (4). Observe that $\mathcal{L}_I$ is generic among the subspaces of dimension $|I|$. Also note that $\dim \mathbb{S}^n_{n-p} + \dim \mathcal{L}_I < \dim \mathbb{S}^n$ by the calculation in the proof of Theorem 1. As $C \notin \mathbb{S}^n_{n-p}$, then $C \notin \mathbb{S}^n_{n-p} + \mathcal{L}_I$ for a generic $\mathcal{L}_I$. The result follows from $\mathcal{L} = \bigcup_I \mathcal{L}_I$. □

Example 6 (Matrix sensing)**.** Given a linear map $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$ 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:

$$\min_{X \in \mathbb{S}^n} \|X\|_* \quad \text{such that} \quad \mathcal{A}(X) = b. \tag{5}$$

If we further assume that X that is PSD, the cost function is $I_n \bullet X$. By Theorem 5, 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. Different guarantees about the Burer-Monteiro method for matrix sensing were obtained in [14], relying on the restricted isometry property.

3. General SDP's

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

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

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

We apply the Burer-Monteiro method to the first k matrices. Let $Y := (Y_1, \ldots, Y_k)$, with $Y_j \in \mathbb{R}^{n_j \times p}$, and let $q(Y) := (Y_1 Y_1^T, \ldots, Y_k Y_k^T)$. Notice that we use the same rank p for all matrices. We denote

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

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

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

Pataki [15] showed that $(SDP_{\mathbf{n}})$ always has an optimal solution such that $\sum_j \tau(r_j) \le m - d$, where $r_j := \operatorname{rank} X_j$. It follows that $(BM_{\mathbf{n}})$ and $(SDP_{\mathbf{n}})$ have the same optimal value when $\tau(p) \ge m'_k$, with

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

where the maximum is over the possible ranks $r_{k+1}, \ldots, r_\ell$. We will show that if $\tau(p) > m'_k$ then, generically, there are no spurious critical points.

Theorem 7. *Assume that $\tau(p) > m'_k$. For a generic C, any critical point $(Y, \overline{X}, x)$ of problem $(BM_{\mathbf{n}})$ is such that $(q(Y), \overline{X}, x)$ is optimal for $(SDP_{\mathbf{n}})$.*

Example 8. Consider the inequality constrained problem (SDP). We may view each of the m_2 inequalities as a PSD constraint on a 1×1 matrix. Therefore, this is a special instance of $(SDP_{\mathbf{n}})$ with $k=1$, $\ell = m_2 + 1$, $d = 0$, and $n_2 = \ldots = n_\ell = 1$. Note that $r_{i+1} = 1$ when the i-th inequality is inactive, and is zero otherwise. Hence $m'_k = m - \#(\text{inactive constrs}) = \#(\text{active constrs})$. This is consistent with the results from Section 2.

Let us derive the optimality conditions for $(BM_{\mathbf{n}})$. More generally, consider the conic nonlinear program $\min_{x,y}\{f(x,y) : h(x,y) = 0,\ x \in \mathcal{K}\}$, where $\mathcal{K}$ 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 first-order conditions are necessary for optimality under suitable regularity conditions, see e.g., [4, §3.1]:

$$(6) \qquad (x, y) \text{ feasible}, \quad s \in \mathcal{K}^*, \quad \langle s, x \rangle = 0, \quad \nabla_{x,y} L(x, y, \lambda, s) = 0.$$

Though it is possible to obtain second-order conditions for conic programs, 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 second-order condition:

$$u^T \nabla^2_{yy} L(x,y,\lambda,s)\, u \geq 0, \quad \forall\, u \;\text{ such that } \nabla_y h(x,y)\, u = 0. \tag{7}$$

Optimality conditions for $(BM_{\mathbf{n}})$ are derived by specializing (6) and (7). 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 first-order and second-order conditions for $(BM_{\mathbf{n}})$ are:

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

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

We say that $(Y, \overline{X}, x)$ is a *critical point* if it there is some $\lambda \in \mathbb{R}^m$ as above. The critical point is *spurious* if $(q(Y), \overline{X}, x)$ is not optimal for $(SDP_{\mathbf{n}})$.

The strategy to prove Theorem 7 is identical to Theorem 1. We first generalize Proposition 2, and show that a spurious critical point $(Y, \overline{X}, x)$ must be have at least one matrix Y_j of full rank. The proof is analogous.

Proposition 9. *Let $(Y, \overline{X}, x)$ be a critical point of (BM). If Y_j is rank deficient for all $1 \leq j \leq k$, then $(q(Y), \overline{X}, x)$ is optimal for (SDP).*

Recall the set $\mathbb{S}^n_r := \{X \in \mathbb{S}^n : \operatorname{rank} X \leq r\}$ of dimension $\tau(n) - \tau(n-r)$.

Corollary 10. *If $(BM_{\mathbf{n}})$ has a spurious critical point, then C lies in the algebraic set $\underline{V} \times \overline{V} \times \{0^d\} + \operatorname{Im} \mathcal{A}^* \subset \mathbb{S}^{\mathbf{n}} \times \mathbb{R}^d$, with*

$$\underline{V} \;:=\; \bigcup_{j \in [k]} (\mathbb{S}^{n_1} \times \cdots \times \mathbb{S}^{n_{j-1}} \times \mathbb{S}^{n_j}_{n_j - p} \times \mathbb{S}^{n_{j+1}} \times \cdots \times \mathbb{S}^{n_\ell}) \;\subset\; \mathbb{S}^{\underline{\mathbf{n}}},$$

$$\overline{V} \;:=\; \bigcup_{r_{k+1}, \ldots, r_\ell} (\mathbb{S}^{n_{k+1}}_{n_{k+1} - r_{k+1}} \times \cdots \times \mathbb{S}^{n_\ell}_{n_\ell - r_\ell}) \;\subset\; \mathbb{S}^{\overline{\mathbf{n}}},$$

where the last union is over the possible ranks $r_{k+1}, \ldots, r_\ell$ in $(SDP_{\mathbf{n}})$.

Proof. Let $(Y, \overline{X}, x, \lambda)$ satisfy (8a), and let $\operatorname{rank} Y_j = p$ for some $j \in [k]$. Since $S_j(\lambda) Y_j = 0$ then $S_j(\lambda) \in \mathbb{S}^{n_j}_{n_j - p}$. Let $(r_{k+1}, \ldots, r_\ell)$ 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 $\overline{S}(\lambda) \subset \mathbb{S}^{n_{k+1}}_{n_{k+1} - r_{k+1}} \times \cdots \times \mathbb{S}^{n_\ell}_{n_\ell - r_\ell}$. Then $S(\lambda) \in \underline{V} \times \overline{V} \times \{0\}$, as $s(\lambda) = 0$. The result follows from $C = S(\lambda) + \mathcal{A}^*(\lambda)$. ☐

The remaining part of Theorem 7 is a dimension counting.

Proof of Theorem 7. Let $D := \dim(\mathbb{S}^{\mathbf{n}} \times \mathbb{R}^d) = \tau(n_1) + \cdots + \tau(n_\ell) + d$. We will show that $\underline{V} \times \overline{V} \times \{0\} + \operatorname{Im} \mathcal{A}^*$ has strictly smaller dimension. Note that

$$\dim \operatorname{Im} \mathcal{A}^* + \dim \underline{V} + \dim \overline{V} = m + \sum_{j \leq k} \tau(n_j) - \tau(p) + \max_{r_{k+1} \cdots r_\ell} \sum_{j > k} \tau(n_j) - \tau(r_j)$$

$$= \; D - \tau(p) + \max_{r_{k+1}, \ldots, r_\ell} \Big\{ m - d - \sum_{j > k} \tau(r_j) \Big\} \;=\; D - \tau(p) + m'_k \;<\; D.$$

It follows that $\dim(\underline{V} \times \overline{V} \times \{0\} + \operatorname{Im} \mathcal{A}^*) < D$. ☐

As illustrated next, Corollary 10 can be used even when C is not generic.

Example 11 (Matrix sensing)**.** We revisit the problem of sensing symmetric matrices from Example 6. For $X \in \mathbb{S}^n$, its nuclear norm satisfies:

$$\|X\|_* \quad = \quad \min_Y \; I_n \bullet Y \quad \text{such that} \quad Y + X \in \mathbb{S}^n_+, \;\; Y - X \in \mathbb{S}^n_+.$$

Let $X_1 := \frac{1}{2}(Y+X)$, $X_2 := \frac{1}{2}(Y-X)$. We can rewrite problem (5) as follows:

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

Consider the Burer-Monteiro method applied to both matrices X_1, X_2, so that $k = \ell = 2$. We will prove that there are no spurious critical points when $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$ is generic and $\tau(p) > m$. By Corollary 10, we need to show that

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

It suffices to see that $I_n \notin \mathbb{S}^n_{n-p} + \operatorname{Im} \mathcal{A}^*$. But this was shown in Theorem 5.

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

Appendix A. Local minima and regularity

Let $\mathfrak{M} := \{Y \in \mathbb{R}^{n \times p} : h(Y) \in \{0\}^{m_1} \times \mathbb{R}^{m_2}_+\}$ be the feasible set of (BM), where $h(Y) := \mathcal{A}(YY^T) - b$. Let $\bar{Y} \in \mathfrak{M}$ be a local minimum of (BM). To ensure that $\bar{Y}$ is a critical point (i.e., satisfies (2)), we need that $\bar{Y}$ is sufficiently regular. Different regularity conditions, known as *constraint qualifications*, have been proposed [17]. One of the simplest is:

(LICQ) $\qquad \{\nabla h_i(\bar{Y}) : i \in I(\bar{Y})\}$ are linearly independent,

where $I(\bar{Y}) \subset [m]$ are the active constraints. We say that $\mathfrak{M}$ is *regular* if all its points satisfy (LICQ), or some other constraint qualification. Regularity allows us to restate Theorem 1 in terms of local minima, as follows:

Corollary 12. *Assume that $\tau(p) > m'$ and that $\mathfrak{M}$ is regular. For generic C, any local minimum Y of (BM) is such that YY^T is optimal for (SDP).*

As shown next, regularity is guaranteed when $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$ is generic.

Proposition 13. *For fixed b and generic $\mathcal{A}$, then $\mathfrak{M}$ is regular.*

Proof. Fix $I \subset [m]$, and let $\mathfrak{M}_I := \{Y : h_i(Y) = 0 \text{ for } i \in I\}$. We view $\mathfrak{M}_I$ as an algebraic variety in $\mathbb{C}^{n \times p}$, parametrized by $\mathcal{A} \in \mathbb{C}^{m \times \tau(n)}$. Note that $\mathcal{A} \mapsto \mathfrak{M}_I$ is a linear system with no base locus. By Bertini's Theorem [13], the variety $\mathfrak{M}_I$ is regular for generic $\mathcal{A}$. So $\{\nabla h_i(Y)\}_{i \in I}$ are linearly independent for all $Y \in \mathfrak{M}_I$. Since this holds for any $I \subset [m]$, then (LICQ) also holds. $\square$

References

[1] A. I. Barvinok. Problems of distance geometry and convex properties of quadratic maps. *Discrete Comput. Geom.*, 13(2):189–202, 1995.

[2] 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.

[3] P. Biswas, T. Lian, T. Wang, and Y. Ye. Semidefinite programming based algorithms for sensor network localization. *ACM Trans. Sens. Netw. (TOSN)*, 2(2):188–220, 2006.

[4] J. F. Bonnans and A. Shapiro. *Perturbation analysis of optimization problems*. Springer Science & Business Media, 2013.

[5] 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.

[6] N. Boumal, V. Voroninski, and A. Bandeira. Deterministic guarantees for Burer-Monteiro factorizations of smooth semidefinite programs. *arXiv:1804.02008*, 2018.

[7] 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.

[8] S. Burer and R. D. Monteiro. Local minima and convergence in low-rank semidefinite programming. *Math. Program.*, 103(3):427–444, 2005.

[9] L. Ding, A. Yurtsever, V. Cevher, J. A. Tropp, and M. Udell. An optimal-storage approach to semidefinite programming using approximate complementarity. *arXiv:1902.03373*, 2019.

[10] 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.

[11] C. Helmberg and F. Rendl. A spectral bundle method for semidefinite programming. *SIAM J. Optim.*, 10(3):673–696, 2000.

[12] 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.

[13] S. L. Kleiman. Bertini and his two fundamental theorems. In *Studies in the History of Modern Mathematics III*, volume 55 of *Rend. Circ. Mat. Palermo (2) Suppl.*, pages 9–37, 1998.

[14] 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.

[15] 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.

[16] 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.

[17] T. Rockafellar. Lagrange multipliers and optimality. *SIAM Rev.*, 35(2):183–238, 1993.

[18] 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.

[19] I. Waldspurger and A. Waters. Rank optimality for the Burer-Monteiro factorization. *arXiv:1812.03046*, 2018.

[20] A. Yurtsever, M. Udell, J. Tropp, and V. Cevher. Sketchy decisions: Convex low-rank matrix optimization with optimal storage. In *Artif. Intell. Stat.*, pages 1188–1196, 2017.

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