SIAM J. OPTIM.
Vol. 30, No. 3, pp. 2577–2602

RANK OPTIMALITY FOR THE BURER–MONTEIRO FACTORIZATION*

IRÈNE WALDSPURGER† AND ALDEN WATERS‡

Abstract. When solving large-scale semidefinite programs that admit a low-rank solution, an efficient heuristic is the Burer–Monteiro factorization: instead of optimizing over the full matrix, one optimizes over its low-rank factors. This reduces the number of variables to optimize but destroys the convexity of the problem, thus possibly introducing spurious second-order critical points. The article [N. Boumal, V. Voroninski, and A. S. Bandeira, *Deterministic Guarantees for Burer-Monteiro Factorizations of Smooth Semidefinite Programs*, https://arxiv.org/abs/1804.02008, 2018] shows that when the size of the factors is of the order of the square root of the number of linear constraints, this does not happen: for almost any cost matrix, second-order critical points are global solutions. In this article, we show that this result is essentially tight: for smaller values of the size, second-order critical points are not generically optimal, even when the global solution is rank 1.

Key words. semidefinite programming, Burer–Monteiro factorization

AMS subject classifications. 90C25, 90C26, 65K10

DOI. 10.1137/19M1255318

1. Introduction.

We consider a semidefinite program

$$\text{(SDP)} \qquad \begin{aligned} &\text{minimize } \operatorname{Trace}(CX) \\ &\text{such that } \mathcal{A}(X) = b, \\ &X \succeq 0, \end{aligned}$$

where the variable X and the fixed matrix C are symmetric and of size $n \times n$, and $\mathcal{A}$ is a linear operator capturing m equality constraints.

Various iterative algorithms have been developed to solve such a problem at a given precision level but tend to be computationally demanding. For example, in full generality, each iteration may cost $O((m+n)mn^2)$ arithmetic operations with an interior-point solver [5, p. 357] (assuming $m \leq \dim(\mathrm{Sym}(n)) = \frac{n(n+1)}{2}$) and $O((m+n)n^2)$ with first-order techniques applied to a smoothed version of the problem [20, section 3].

Improvements are possible if $\mathcal{A}$ has some structure that can be exploited, but they often do not suffice to make large-scale semidefinite programs computationally easy. Another property can then be used: semidefinite programs tend to have a low-rank minimizer (in many applications, there is one with rank $O(1)$ and, in any case, always one with rank $\sim \sqrt{2m}$ [21, Theorem 2.1]). Low-rank matrices can be stored and manipulated in a much more efficient way than full-rank ones, which allows for less computationally demanding algorithms.

Frank–Wolfe methods, in particular, take advantage of this [14, 16, 29]. Here, we are interested in another approach, the Burer–Monteiro factorization [10]. Its principle is that a semidefinite matrix with rank $p \ll n$ can be factorized as $X = UU^T$, with

*Received by the editors April 9, 2019; accepted for publication (in revised form) July 21, 2020; published electronically September 23, 2020.
https://doi.org/10.1137/19M1255318

†CNRS, Université Paris Dauphine, Inria Mokaplan, Paris 75016, France (waldspurger@ceremade.dauphine.fr).
‡Bernoulli Institute, Rijksuniversiteit Groningen, Groningen, 9747 AG, Netherlands (a.m.s.waters@rug.nl).

$U \in \mathbb{R}^{n\times p}$. Assuming that a low-rank solution X_{opt} exists, if $p \geq \mathrm{rank}(X_{opt})$, problem (SDP) is then equivalent to

$$\text{(Factorized SDP)} \qquad \begin{aligned} &\text{minimize } \mathrm{Trace}(CUU^T) \\ &\text{such that } \mathcal{A}(UU^T) = b, \\ &\qquad U \in \mathbb{R}^{n\times p}. \end{aligned}$$

Now the unknown U has np coordinates, fewer than the n^2 coordinates of X. Consequently, we can run on problem (Factorized SDP) local optimization algorithms that would be too slow on problem (SDP). The caveat is that, since the factorized problem is not convex, they are not guaranteed to find a global minimizer, at best a second-order critical point. Nevertheless, they work extremely well in many applications. Typically, as soon as p is slightly larger than $\mathrm{rank}(X_{opt})$, local optimization algorithms seem to globally solve problem (Factorized SDP). Numerical examples where this phenomenon occurs can be found in [9], [15, section 5], [6, section 5], or [24, section 5].

The article [3] gives a rigorous explanation of this behavior for instances of (Factorized SDP) coming from $\mathbb{Z}_2$-synchronization and community detection. It notably establishes, in particular statistical regimes where problem (SDP) has a rank-1 solution, that all second-order critical points of problem (Factorized SDP) with $p = 2$ are global minimizers. Hence, suitable local optimization algorithms globally solve problem (Factorized SDP). Similarly, [12, 25, 19] show, in related settings, that all second-order critical points of the Burer–Monteiro factorization are the optimal solution as soon as $p \geq \mathrm{rank}(X_{opt})$.

But these works apply in very specific settings only. They provide no general theory on when local optimization algorithms solve problem (Factorized SDP). With no restrictive assumptions, essentially the only result is [8]. Building on [10] and [6], it shows that, under simple hypotheses, all second-order critical points of problem (Factorized SDP) are global minimizers, for almost any matrix C, as soon as

$$\frac{p(p+1)}{2} > m, \tag{1.1}$$

that is, $p > \lfloor\sqrt{2m + 1/4} - 1/2\rfloor$. Extensions can be found in [23, 4].

As a result there is a gap in the literature: in all the concrete settings that could be studied, all second-order critical points of problem (Factorized SDP) are global minimizers as soon as $p \gtrsim \mathrm{rank}(X_{opt})$, in line with numerical experiments, but in the general case, the only guarantees at our disposal state that we need p to be at least as large as $\sim \sqrt{2m}$. In many applications, $\mathrm{rank}(X_{opt}) = O(1)$ while $m = O(n)$, hence these two estimates are far apart, making a huge difference on the computational cost of certifiable algorithms.

The natural question is, "can the gap be reduced?" In this article, we negatively answer this question and show that inequality (1.1) is essentially optimal. It can be slightly improved, to

$$\frac{p(p+1)}{2} + p > m. \tag{1.2}$$

This is Theorem 3.2. But Theorem 3.3 (our main result) states that, under reasonable assumptions on $\mathcal{A}, b$, if p is such that

$$\frac{p(p+1)}{2} + pr_* \leq m,$$

where $r_* = \min\{\mathrm{rank}(X), X \succeq 0, \mathcal{A}(X) = b\}$, there exists a set of cost matrices C with nonzero Lebesgue measure on which problem (SDP) admits a global minimizer with rank r_*, but problem (Factorized SDP) has second-order critical points which are not global minimizers. In particular, if $r_* = 1$ (as is the case in *MaxCut* relaxations, for instance), inequality (1.2) is exactly optimal. Therefore without specific assumptions on C, when running a local optimization algorithm on problem (Factorized SDP) with p smaller than $\sim \sqrt{2m}$, we cannot be sure not to run into a spurious second-order critical point, even if there exists a global minimizer with rank $O(1)$.

Regarding the organization of this article, section 2 contains basic definitions (subsection 2.1) and properties (subsection 2.2), and defines and discusses an important assumption for our main result (subsection 2.3). Section 3 presents the main results: Theorems 3.2 and 3.3 are respectively stated in subsections 3.1 and 3.2. Subsection 3.3 discusses their application to *MaxCut* relaxations. The other sections contain the proofs: Theorem 3.2 is proved in section 4, and Theorem 3.3 in section 5.

1.1. Notation. For any $p, q \in \mathbb{N}^*$, we denote by I_p the $p \times p$ identity matrix and by $0_{p,q}$ the zero $p \times q$ matrix. For any $p \in \mathbb{N}^*$, we denote by $\mathbb{S}^{p\times p}$ the set of real symmetric $p \times p$ matrices, by $\mathrm{Anti}(p)$ the set of antisymmetric $p \times p$ matrices, and by $O(p)$ the set of orthogonal $p \times p$ matrices. For any n_1, n_2, we equip $\mathbb{R}^{n_1\times n_2}$, the set of $n_1 \times n_2$ matrices, with the usual scalar product:

$$\forall M_1, M_2 \in \mathbb{R}^{n_1\times n_2}, \quad \langle M_1, M_2\rangle \overset{def}{=} \mathrm{Tr}\big(M_1^T M_2\big).$$

The same formula also defines a scalar product on $\mathbb{S}^{p\times p}$, for any $p \in \mathbb{N}^*$. In both cases, the associated norm is the Frobenius norm, which we denote by $||.||_F$. For any $p \in \mathbb{N}^*$, we define $\mathrm{diag} : \mathbb{R}^{p\times p} \to \mathbb{R}^p$ as the operator which associates to a matrix the vector of its diagonal elements.

For any element x of a metric space, and any positive ϵ, we denote $B(x, \epsilon)$ the open ball with radius ϵ, and $\overline{B}(x, \epsilon)$ the closed ball. When $\mathcal{M}$ is a manifold, and x an element of $\mathcal{M}$, we denote by $T_x\mathcal{M}$ the tangent space of $\mathcal{M}$ at x.

2. Preliminaries.

2.1. Definitions. We consider a problem of the following form:

$$\text{(SDP)} \qquad \begin{aligned} &\text{minimize } \langle C, X\rangle \\ &\text{such that } \mathcal{A}(X) = b, \\ &X \succeq 0. \end{aligned}$$

Here, $\mathcal{A} : \mathbb{S}^{n\times n} \to \mathbb{R}^m$ is a fixed linear map, b a fixed element of $\mathbb{R}^m$, and C an element of $\mathbb{S}^{n\times n}$, which is called the *cost matrix*.

We denote by $\mathcal{C}$ the set of feasible points for this problem:

$$\mathcal{C} = \{X \in \mathbb{S}^{n\times n}, \mathcal{A}(X) = b, X \succeq 0\}.$$

As explained in the introduction, if we assume that problem (SDP) has an optimal solution X_{opt} with rank r, and fix some $p \geq r$, it is equivalent to its *rank p Burer–Monteiro factorization*:

$$\text{(Factorized SDP)} \qquad \begin{aligned} &\text{minimize } \left\langle C, VV^T\right\rangle \\ &\text{such that } \mathcal{A}(VV^T) = b, \\ &V \in \mathbb{R}^{n\times p}. \end{aligned}$$

We denote by $\mathcal{M}_p$ the set of feasible points for the factorized problem:

$$\mathcal{M}_p = \left\{ V \in \mathbb{R}^{n\times p}, \mathcal{A}\left(VV^T\right) = b \right\}.$$

It is invariant under multiplication by elements of $O(p)$. We assume that it is sufficiently regular so that we can apply smooth optimization algorithms to problem (Factorized SDP). More precisely, all our results require that $(\mathcal{A}, b)$ is *p-regular.*

Definition 2.1. *For some $p \in \mathbb{N}^*$, $(\mathcal{A}, b)$ is said to be* p-regular *if, for all $V \in \mathcal{M}_p$, the following linear map is surjective:*

$$\dot{V} \in \mathbb{R}^{n\times p} \to \mathcal{A}\left(V\dot{V}^T + \dot{V}V^T\right) \in \mathbb{R}^m.$$

This assumption is of the same style as [8, Assumption 1.1]. It notably guarantees [2, Proposition 3.3.3] that $\mathcal{M}_p$ is a submanifold of $\mathbb{R}^{n\times p}$, with dimension $\dim(\mathcal{M}_p) = np - m$, whose tangent space at any point V is

$$T_V\mathcal{M}_p = \left\{ \dot{V} \in \mathbb{R}^{n\times p}, \mathcal{A}\left(V\dot{V}^T + \dot{V}V^T\right) = 0 \right\}.$$

The scalar product of $\mathbb{R}^{n\times p}$ defines a metric on the manifold $\mathcal{M}_p$, which we then view as a Riemannian manifold. Many algorithms exist for attempting to minimize a smooth function on a Riemannian manifold; a classical reference on this topic is [2].

However, they are a priori not guaranteed to find a global minimizer of problem (Factorized SDP) but only an (approximate) first- or second-order critical point of the cost function $V \in \mathcal{M}_p \to \left\langle C, VV^T \right\rangle$ [7]. These points are defined as follows.

Definition 2.2. *Let $\mathcal{N}$ be a Riemannian manifold and $f : \mathcal{N} \to \mathbb{R}$ a smooth function. We denote ∇ and* Hess *its gradient and Hessian with respect to the manifold.*

For any $x_0 \in \mathcal{N}$, we say that x_0 is a first-order critical point *of f if $\nabla f(x_0) = 0$ and a* second-order critical point *of f if $\nabla f(x_0) = 0$ and $\mathrm{Hess} f(x_0) \succeq 0$.*

The goal of this article is to study for which values of p the set of second-order critical points coincides with the set of global minimizers of problem (Factorized SDP).

We note that there are pairs $(\mathcal{A}, b)$ which are not p-regular, regardless of the value of p (an example is when $0_{n,n} \in \mathcal{C}$). This setting is significantly different from the one that we consider in this article: $\mathcal{M}_p$ may then have singularities, and classical Riemannian tools are a priori not applicable to problem (Factorized SDP).

2.2. Basic properties. It is convenient to be able to describe the solutions of problem (SDP) in terms of Karush–Kuhn–Tucker conditions. This is a priori possible only if strong duality holds but, fortunately for us, strong duality always holds when $(\mathcal{A}, b)$ is p-regular for some p, yielding the following proposition (whose proof is in Appendix A.1).

Proposition 2.3. *We assume that there exists $p \in \mathbb{N}^*$ such that $(\mathcal{A}, b)$ is p-regular and $\mathcal{M}_p \neq \emptyset$. Then a matrix $X_0 \in \mathcal{C}$ is a solution of problem* (SDP) *if and only if there exist $g_1 \in \mathbb{R}^m, C_1 \in \mathbb{S}^{n\times n}$ such that*

- $C = \mathcal{A}^*(g_1) + C_1$*;*
- $C_1 \succeq 0$*;*
- $C_1 X_0 = 0$.

When g_1, C_1 satisfy $\mathrm{rank}(C_1) = n - \mathrm{rank}(X_0)$ in addition to the above three conditions, we say that *strict complementary slackness* holds. The following proposition (whose proof is in Appendix A.2) states that, under an additional condition on X_0, it implies that the solution of problem (SDP) is unique.

PROPOSITION 2.4. *If strict complementary slackness holds and* X_0 *is an extremal point of* $\mathcal{C}$*, then* X_0 *is the unique solution of problem* (SDP).

The next proposition characterizes, in a similar way as Proposition 2.3, the first-order critical points of problem (Factorized SDP). Its proof is in Appendix A.3.

PROPOSITION 2.5. *We assume that* $(\mathcal{A}, b)$ *is* p*-regular for some* $p \in \mathbb{N}^*$*. A matrix* $V \in \mathcal{M}_p$ *is a first-order critical point of problem* (Factorized SDP) *if and only if there exist* $g_2 \in \mathbb{R}^m, C_2 \in \mathbb{S}^{n\times n}$ *such that*

- $C = \mathcal{A}^*(g_2) + C_2$;
- $C_2 V = 0$.

When it exists, the pair (g_2, C_2) *is unique.*

Finally, we also provide a reformulation of second-order criticality; the proof is in Appendix A.4.

PROPOSITION 2.6. *We assume that* $(\mathcal{A}, b)$ *is* p*-regular for some* $p \in \mathbb{N}^*$*. Let* $V \in \mathcal{M}_p$ *be a first-order critical point of problem* (Factorized SDP)*, whose cost function we denote* f_C*. For any* $\dot{V} \in T_V\mathcal{M}_p$*,*

$$\mathrm{Hess} f_C(V) \cdot (\dot{V}, \dot{V}) = 2 \left\langle C_2, \dot{V}\dot{V}^T \right\rangle,$$

with C_2 *defined as in Proposition* 2.5*. Thus,* V *is second-order critical if and only if*

$$\forall \dot{V} \in T_V\mathcal{M}_p, \quad \left\langle C_2, \dot{V}\dot{V}^T \right\rangle \geq 0. \tag{2.1}$$

Using the notation of the previous proposition, we observe that, since f_C is invariant under right multiplication by elements of $O(p)$, $\mathrm{Hess} f_C(V) \cdot (\dot{V}, \dot{V}) = 0$ for any $\dot{V}$ tangent to the orbit of V under the action of $O(p)$, that is, $\dot{V} = VA$ for some $A \in \mathrm{Anti}(p)$ (recall that $\mathrm{Anti}(p)$ is the set of $p \times p$ antisymmetric matrices). This motivates the following definition.

DEFINITION 2.7. *A second-order critical point* V *of problem* (Factorized SDP) *is* nondegenerate *if, in* (2.1)*, the equality is attained exactly for matrices* $\dot{V}$ *of the form* $\dot{V} = VA, A \in \mathrm{Anti}(p)$.

Remark 2.8. Equivalently, a second-order critical point is nondegenerate if

$$\begin{aligned}\mathrm{rank}(\mathrm{Hess} f_C(V)) &= \dim(\mathcal{M}_p) - \dim\{VA, A \in \mathrm{Anti}(p)\} \\ &= \dim(\mathcal{M}_p) - \frac{p(p-1)}{2}.\end{aligned}$$

2.3. Definition of "minimally secant." The following technical property is needed for our main theorem.

DEFINITION 2.9. *Let* $p \in \mathbb{N}^*$ *be such that* $(\mathcal{A}, b)$ *is* p*-regular. Let* r *be in* $\mathbb{N}^*$*.*

Let X_0 *be a rank* r *element of* $\mathcal{C}$ *and* V *be in* $\mathcal{M}_p$*. We say that* $\mathcal{M}_p$ *is* X_0-minimally secant at V *if the following three conditions hold:*

1. $\mathrm{rank}(V) = p$;
2. $\mathrm{Range}(X_0) \cap \mathrm{Range}(V) = \{0\}$;
3. *for any* $\dot{V} \in T_V\mathcal{M}_p$*, if* $\mathrm{Range}(\dot{V}) \subset \mathrm{Range}(X_0) + \mathrm{Range}(V)$*, then* $\dot{V} = VA$ *for some* $A \in \mathrm{Anti}(p)$.

We observe that, for any $V \in \mathcal{M}_p$, the intersection

$$T_V\mathcal{M}_p \cap \{\dot{V} \in \mathbb{R}^{n\times p}, \mathrm{Range}(\dot{V}) \subset \mathrm{Range}(X_0) + \mathrm{Range}(V)\}$$

necessarily contains $\{VA, A \in \text{Anti}(p)\}$. Therefore, the third property in the above definition amounts to requiring that the intersection is "as small as possible" (hence the name "minimally secant").

Remark 2.10. Condition 3 in Definition 2.9 notably implies that if $\text{Range}(\dot{V}) \subset \text{Range}(X_0) + \text{Range}(V)$, then, actually, $\text{Range}(\dot{V}) \subset \text{Range}(V)$.

Our main result, Theorem 3.3, contains the assumption that there exists X_0, V such that $\mathcal{M}_p$ is X_0-minimally secant at V. This assumption is crucial for the proof: to show the existence of at least one cost matrix C for which a spurious second-order critical point exists, our strategy is to fix X_0 and V, and construct C for which a second-order critical point exists and is precisely V, while the global minimizer corresponds to X_0. To ensure first- and second-order criticality for V, and global optimality for X_0, we essentially need that the quadratic form defined by C satisfies some properties when restricted to $\text{Range}(X_0)$, some other properties on $\text{Range}(V)$, and still some other ones on $\text{Range}(\dot{V})$ for $\dot{V} \in T_V\mathcal{M}_p$. In order for these properties to be compatible with each other, the spaces $\text{Range}(X_0), \text{Range}(V), \text{Range}(\dot{V})$ must "not intersect too much." The formal content behind "not intersecting too much" is precisely Definition 2.9.

However, as explained in Appendix B, when $\frac{p(p+1)}{2} + pr \leq m$, we expect such X_0, V to almost always exist. This is notably the case for *MaxCut* problems (see subsection 3.3 and Appendix E) and also for *Orthogonal-Cut* and optimization over a product of spheres [27].

3. Main results.

3.1. Regime where critical points are global minimizers. As previously stated, most smooth optimization algorithms, applied to problem (Factorized SDP), are only guaranteed to find a critical point of this problem, not a global minimizer. Fortunately, [8] shows that, when p is large enough, second-order critical points are always global minimizers, for almost all cost matrices C. Therefore, algorithms able to find second-order critical points (for instance, the trust-region method) actually solve problem (Factorized SDP) to optimality, provided that C is "generic." A restated version of the theorem in [8], under minor modifications, is the following.

THEOREM 3.1 (see [8, Theorem 1.4]). *Let $p \in \mathbb{N}^*$ be fixed. We assume that*
1. *the set $\mathcal{C}$ of feasible points for problem* (SDP) *is compact;*
2. *$(\mathcal{A}, b)$ is p-regular.*

If

$$\frac{p(p+1)}{2} > m, \qquad \left(\iff p > \left\lfloor \sqrt{2m + \frac{1}{4}} - \frac{1}{2} \right\rfloor\right), \tag{3.1}$$

then, for almost all cost matrices $C \in \mathbb{S}^{n\times n}$, if $V \in \mathcal{M}_p$ is a second-order critical point of problem (Factorized SDP)*, then*
- *V is a global minimizer of problem* (Factorized SDP)*;*
- *$X = VV^T$ is a global minimizer of problem* (SDP).

It is natural to ask whether condition (3.1) is optimal or whether the same guarantees hold for smaller ranks p, allowing further reductions in the computational complexity of solving problem (Factorized SDP). Our first result is that condition (3.1) can be slightly relaxed.

THEOREM 3.2. *Let $p \in \mathbb{N}^*$ be fixed. We assume that*
1. *the set $\mathcal{C}$ of feasible points for problem* (SDP) *is compact;*
2. *$(\mathcal{A}, b)$ is p-regular.*

If $\frac{p(p+1)}{2} + p > m$, then the same conclusion holds as in the previous theorem.

The proof of Theorem 3.2 is in section 4.

3.2. Regime where there may be bad critical points. We can now address our main question: How optimal is the result of the previous section? When problem (SDP) has a unique global minimizer, with rank r of the order of $\sqrt{2m}$, the result cannot be significantly improved: $p \geq r$ is a necessary condition for problems (SDP) and (Factorized SDP) to have the same minimum. However, as said in the introduction, problem (SDP) often admits a solution with rank $r \ll \sqrt{2m}$, and the Burer–Monteiro factorization is numerically observed to work when $p = O(r)$.

Our main theorem, however, states that even if we assume $r \ll \sqrt{2m}$, our previous result is essentially not improvable without additional hypotheses on C: under reasonable assumptions on $(\mathcal{A}, b)$, if $\frac{p(p+1)}{2} + pr \leq m$, there is a set of cost matrices with nonzero Lebesgue measure for which problem (SDP) has a rank r optimal solution, but problem (Factorized SDP) has a nonoptimal second-order critical point. In particular, for $r = 1$, the inequality $\frac{p(p+1)}{2} + p > m$ in Theorem 3.2 is exactly optimal.

THEOREM 3.3. *Let $r \in \mathbb{N}^*$ be fixed. Let $p \geq r$ be such that*

$$\frac{p(p+1)}{2} + pr \leq m.$$

We make the following hypotheses:
1. *$\mathcal{C}$ has at least one extreme point with rank r, denoted by X_0;*
2. *$(\mathcal{A}, b)$ is p-regular;*
3. *there exists $V \in \mathcal{M}_p$ such that $\mathcal{M}_p$ is X_0-minimally secant at V.*

Then there exists a subset $\mathcal{E}_{bad}$ of $\mathbb{S}^{n\times n}$ with nonzero Lebesgue measure such that, for any cost matrix $C \in \mathcal{E}_{bad}$,
- *problem* (SDP) *has a unique global minimizer, which has rank r;*
- *problem* (Factorized SDP) *has at least one second-order critical point that is not a global minimizer.*

An example of a problem where the three hypotheses hold true is provided in the next subsection and two other ones are in the long version of this article [27]. The proof of Theorem 3.3 is in section 5.

Remark 3.4. Theorem 3.3 stays valid if one replaces "second-order critical point" with "local minimizer." Indeed, it turns out that the second-order critical points constructed in our proof are nondegenerate and, therefore, local minimizers.

3.3. MaxCut. In this subsection, we apply our results to the most famous instance of a problem with the form (SDP), the *MaxCut* relaxation:

$$\text{(SDP-Maxcut)} \qquad \begin{aligned} &\text{minimize } \langle C, X\rangle \\ &\text{such that } \operatorname{diag}(X) = 1, \\ &\qquad X \succeq 0. \end{aligned}$$

This problem is a relaxation of the "maximum cut" problem from graph theory [11, 22], made famous by the work [13]. It also appears in phase retrieval [26] and $\mathbb{Z}/2\mathbb{Z}$

synchronization [1, 3] (in which cases its global optimizer is known, both theoretically and numerically, to often have very low rank, typically 1).

Theorems 3.2 and 3.3 exactly describe when its Burer–Monteiro factorization has no nonoptimal second-order critical point for almost any cost matrix, even if we assume that the global minimizer has rank 1.

COROLLARY 3.5. *If $p \in \mathbb{N}$ is such that $\frac{p(p+1)}{2} + p > n$, then, for almost any cost matrix C, all second-order critical points of the Burer–Monteiro factorization of problem* (SDP-Maxcut) *are globally optimal.*

On the other hand, for any p such that $\frac{p(p+1)}{2} + p \leq n$, the set of cost matrices admits a subset with nonzero Lebesgue measure on which

- *problem* (SDP-Maxcut) *has a unique global minimizer, which has rank* 1*;*
- *its Burer–Monteiro factorization with rank p has at least one nonoptimal second-order critical point.*

This result is proved in Appendix E.

4. Proof of Theorem 3.2.

Let m be such that

$$m < \frac{p(p+1)}{2} + p. \tag{4.1}$$

From [21, Theorem 2.1], problem (SDP) has a minimizer with rank at most p. Consequently, problems (SDP) and (Factorized SDP) have the same minimum, and if V is a global minimizer of problem (Factorized SDP), $X = VV^T$ is a minimizer of problem (SDP). It therefore suffices to show that, for almost all cost matrices, problem (Factorized SDP) has no second-order critical point which is not a global minimizer.

4.1. Overview of the proof.

Our proof starts in a similar way as the one in [8]. Namely, we use the first- and second-order properties of critical points to parametrize the set of "bad" cost matrices: we define an appropriate manifold $\mathcal{M}_{param}$ and a smooth map $\phi : \mathcal{M}_{param} \to \mathbb{S}^{n\times n}$ such that the set of bad cost matrices is included in $\phi(\mathcal{M}_{param})$.

Then the proofs differ. The authors of [8] show that when $\frac{p(p+1)}{2} > m$, the dimension of their manifold $\mathcal{M}_{param}$ is strictly smaller than $\dim(\mathbb{S}^{n\times n})$, hence $\phi(\mathcal{M}_{param})$ has zero Lebesgue measure in $\mathbb{S}^{n\times n}$. On our side, we use additional properties of critical points to show that the set of bad cost matrices is actually included in the critical values of ϕ, and not only in the range of ϕ. This set has zero Lebesgue measure in $\mathbb{S}^{n\times n}$, from Sard's theorem.

4.2. Details.

We define $\mathcal{M}_p^{full} = \{V \in \mathcal{M}_p, \mathrm{rank}(V) = p\}$. It is an open subset of $\mathcal{M}_p$ and therefore also an $(np - m)$-dimensional Riemannian manifold. We also define

$$\mathcal{E} = \left\{(V, C_2) \in \mathcal{M}_p^{full} \times \mathbb{S}^{n\times n} \text{ such that } C_2 V = 0_{n,p}\right\}.$$

This set is a manifold, as stated in the following proposition, whose proof is in Appendix C.1.

PROPOSITION 4.1. *The set $\mathcal{E}$ is a manifold, with dimension $np-m+\frac{(n-p)(n-p+1)}{2}$. Additionally, for any $(V, C_2) \in \mathcal{E}$,*

$$T_{(V,C_2)}\mathcal{E} = \left\{(\dot{V}, \dot{C}_2) \in T_V\mathcal{M}_p \times \mathbb{S}^{n\times n} \text{ such that } \dot{C}_2 V + C_2 \dot{V} = 0_{n,p}\right\}. \tag{4.2}$$

We define

$$\begin{array}{rccc} \phi : & \mathcal{E} \times \mathbb{R}^m & \to & \mathbb{S}^{n\times n}, \\ & ((V, C_2), \mu) & \to & C_2 + \mathcal{A}^*(\mu). \end{array}$$

We observe that ϕ does not explicitly depend on V, only through C_2. Intuitively, V represents the possible bad critical point, C_2 is a kind of corresponding dual variable, and the "problematic" cost matrix depends on the dual variable only, not directly on the critical point.

The following lemma, whose proof is in subsection 4.3, says that any cost matrix for which a nonoptimal second-order critical point exists is a critical value of ϕ.

LEMMA 4.2. *For any cost matrix $C \in \mathbb{S}^{n\times n}$, if problem* (Factorized SDP) *has a nonoptimal second-order critical point, then there exist $(V, C_2) \in \mathcal{E}, \mu \in \mathbb{R}^m$ such that*

$$C = \phi((V, C_2), \mu),$$

and the mapping $d\phi((V, C_2), \mu) : T_{(V,C_2)}\mathcal{E} \times \mathbb{R}^m \to \mathbb{S}^{n\times n}$ is not surjective.

From Sard's theorem, we can therefore conclude that the set of such cost matrices has zero measure in $\mathbb{S}^{n\times n}$.

4.3. Proof of Lemma 4.2. Let $C \in \mathbb{S}^{n\times n}$ be a cost matrix for which a nonoptimal second-order critical point exists. Let $V \in \mathcal{M}_p$ be such a critical point. From [8, Theorem 1.6], as V is nonoptimal, $\text{rank}(V) = p$, so V is in $\mathcal{M}_p^{full}$.

As V is first-order critical, there exist, from Proposition 2.5, $\mu \in \mathbb{R}^m, C_2 \in \mathbb{S}^{n\times n}$ such that $C_2 V = 0$ and

$$C = C_2 + \mathcal{A}^*(\mu) = \phi((V, C_2), \mu).$$

Let us now show that $d\phi((V, C_2), \mu)$ is not surjective. Again, from [8, Theorem 1.6], as V is nonoptimal, the dimension of the face of $\mathcal{C}$ containing VV^T is at least

$$\frac{p(p+1)}{2} - m + p \overset{\text{Eq. (4.1)}}{>} 0.$$

Let $X_{Face} \neq VV^T$ be an element of this face. Using the geometrical properties of $\{X \in \mathbb{S}^{n\times n}, X \succeq 0\}$, one can establish the following proposition, whose proof is in Appendix C.2.

PROPOSITION 4.3. *There exists $T \in \mathbb{S}^{p\times p}$ such $X_{Face} = VTV^T$.*

Let T be as in the proposition. For any $(\dot{V}, \dot{C}_2) \in T_{(V,C_2)}\mathcal{E}, \dot{\mu} \in \mathbb{R}^m$,

$$\begin{aligned}
&\left\langle d\phi((V, C_2), \mu) \cdot ((\dot{V}, \dot{C}_2), \dot{\mu}), X_{Face} - VV^T \right\rangle \\
&\qquad = \left\langle \dot{C}_2 + \mathcal{A}^*(\dot{\mu}), X_{Face} - VV^T \right\rangle \\
&\qquad = \left\langle \dot{C}_2, V(T - I_p)V^T \right\rangle + \left\langle \dot{\mu}, \mathcal{A}(X_{Face} - VV^T) \right\rangle \\
&\qquad \overset{(a)}{=} \left\langle \dot{C}_2, V(T - I_p)V^T \right\rangle \\
&\qquad = \left\langle \dot{C}_2 V, V(T - I_p) \right\rangle \\
&\qquad \overset{(b)}{=} -\left\langle C_2 \dot{V}, V(T - I_p) \right\rangle \\
&\qquad = -\left\langle \dot{V}, C_2 V(T - I_p) \right\rangle \\
&\qquad \overset{(c)}{=} 0.
\end{aligned}$$

Equality (a) is true because X_{face} belongs to $\mathcal{C}$, so $\mathcal{A}(X_{Face}) = b = \mathcal{A}(VV^T)$. Equality (b) is true because of (4.2), and equality (c) because $C_2V = 0_{n,p}$ from the definition of $\mathcal{E}$.

This shows that the range of $d\phi((V, C_2), \mu)$ in $\mathbb{S}^{n\times n}$ is included in $(X_{Face} - VV^T)^\perp$, so that $d\phi((V, C_2), \mu)$ cannot be surjective. ∎

5. Proof of Theorem 3.3. This section is devoted to the proof of the main theorem. The first two subsections, 5.1 and 5.2, each present the outline of one half of the proof, with the technical details hidden into lemmas. The remaining subsections contain the proofs of these lemmas.

5.1. First part. In the first part, we assume (proving this assumption is done in the second part) that there exists one cost matrix, C, for which problem (SDP) has a unique global minimizer, with rank r, but problem (Factorized SDP) has a spurious second-order critical point. We show that, for all matrices close enough to C, these properties still hold, hence they hold on a whole set, with nonzero Lebesgue measure.

As stated, this assertion may not be quite true (C might be an isolated "bad" cost matrix). However, it becomes true if we assume C to satisfy some additional nondegeneracy properties. Consequently, in this part of the proof, we use the following lemma, which will be proven later.

LEMMA 5.1. *There exists a cost matrix $C \in \mathbb{S}^{n\times n}$ such that*
- *problem* (SDP) *has a unique global minimizer, whose rank is r;*
- *strict complementary slackness holds;*
- *problem* (Factorized SDP) *has a second-order critical point, which is not a global minimizer;*
- *this second-order critical point is nondegenerate.*

We respectively denote X_0 and V the global minimizer and second-order critical point of Lemma 5.1. The two properties we must show are stated in the following lemmas.

LEMMA 5.2. *For any matrix C' close enough to C, problem* (SDP) *has a unique global minimizer, and this minimizer has rank r.*

LEMMA 5.3. *For any matrix C' close enough to C, problem* (Factorized SDP) *has a second-order critical point which is not a global minimizer.*

To prove Lemma 5.2, we use general convexity and continuity arguments to show that at least one minimizer exists and that it goes to X_0 when C' goes to C. In particular, it has rank at least r when C' is close enough to C. With another continuity argument, we show that, because strict complementary slackness holds for C, it also holds for any C' close enough to C. Therefore, the minimizer is unique (from Proposition 2.4), and strict complementary slackness also allows us to prove that it has rank exactly r. A detailed proof is in subsection 5.3.

For Lemma 5.3, it actually suffices to show that, for C' close to C, problem (Factorized SDP) has a second-order critical point close to V. Indeed, no matrix close enough to V can be a global minimizer. If the Hessian at V was positive definite, this would follow from general geometric arguments. But because of the invariance of the problem to multiplication by elements of $O(p)$, the Hessian is not positive definite. We must therefore consider the quotient manifold $\mathcal{M}_p/O(p)$ and a quotiented version of problem (Factorized SDP). For this version, as V is nondegenerate, the Hessian is positive definite, so the general arguments apply. The details are in subsection 5.4.

5.2. Second part: Proof of Lemma 5.1. We recall that we want to construct a cost matrix C such that

1. problem (SDP) has a unique global minimizer, with rank r;
2. strict complementary slackness holds;
3. problem (Factorized SDP) has a second-order critical point, which is not a global minimizer;
4. this second-order critical point is nondegenerate.

It turns out that, for any rank r matrix $X_0 \in \mathcal{C}$ and any $V \in \mathcal{M}_p$, provided that X_0 is extremal in $\mathcal{C}$ and $\mathcal{M}_p$ is X_0-minimally secant at V, it is possible to construct a matrix C as desired, and such that in addition, the unique global minimizer is precisely X_0, and the spurious critical point is precisely V.

Let us fix $X_0 \in \mathcal{C}, V \in \mathcal{M}_p$ with X_0 extremal and $\mathcal{M}_p$ X_0-minimally secant at V (we have made the hypothesis they existed) and explain how to construct C. First, the results in subsection 2.2 allow us to rephrase the desired conditions in more analytical terms: they are equivalent to the existence of $g_1, g_2 \in \mathbb{R}^m, C_1, C_2 \in \mathbb{S}^{n\times n}$ such that

$$\begin{aligned}
&\text{(Optim. SDP 1)} && C = \mathcal{A}^*(g_1) + C_1;\\
&\text{(Optim. SDP 2)} && C_1 \succeq 0;\\
&\text{(Optim. SDP 3)} && C_1 X_0 = 0;\\
&\text{(Strict complementarity)} && \operatorname{rank}(C_1) = n - \operatorname{rank}(X_0) = n - r;\\
&\text{(First-order optim. 1)} && \mathcal{A}^*(g_1) + C_1 = \mathcal{A}^*(g_2) + C_2;\\
&\text{(First-order optim. 2)} && C_2 V = 0;\\
& && \forall \dot{V} \in T_V\mathcal{M}_p, \quad \left\langle C_2, \dot{V}\dot{V}^T \right\rangle \geq 0, \text{ with equality}\\
&\text{(Second-order optim.)} && \text{iff } \dot{V} = VA, A \in \operatorname{Anti}(p).
\end{aligned}$$

The construction now proceeds as follows:

1. We set $g_2 = 0$.
2. We construct g_1, C_1, C_2 which ensure that (Optim. SDP 2), (Optim. SDP 3), (Strict complementarity), (First-order optim. 1), and (First-order optim. 2) hold.
3. From g_1, C_1, C_2, we construct $g_1^{(mod)}, C_1^{(mod)}, C_2^{(mod)}$ which satisfy property (Second-order optim.) in addition to the previous ones.
4. We set $C = \mathcal{A}^*(g_1^{(mod)}) + C_1^{(mod)}$; it satisfies all the required properties.

Points 1 and 4 are straightforward. For Points 2 and 3, see subsections 5.5 and 5.6. ∎

5.3. Proof of Lemma 5.2. To establish the lemma, it suffices to show that, for any sequence $(C'_k)_{k\in\mathbb{N}}$ of cost matrices converging to C, problem (SDP) with cost matrix C'_k has a unique minimizer, and this minimizer has rank r, as soon as k is large enough. Let $(C'_k)_{k\in\mathbb{N}}$ be such a sequence.

The following proposition shows that, for k large enough, at least one minimizer exists, and it is arbitrarily close to X_0. Its proof is in Appendix D.1.

PROPOSITION 5.4. *Let $\epsilon > 0$ be fixed. For k large enough,*

- *problem* (SDP) *(with cost matrix C'_k) admits at least one minimizer;*
- *all minimizers of problem* (SDP) *belong to the ball $B(X_0, \epsilon)$.*

For any k large enough, let X'_k be a minimizer corresponding to the cost matrix C'_k. If there are several of them, we choose X'_k as an extremal point of the set of minimizers (such a point exists because the set is bounded, from Proposition 5.4, convex, and closed); it is then also an extremal point of the feasible set $\mathcal{C}$. Let us show that, for k large enough,

$$\operatorname{rank}(X'_k) = r \text{ and } X'_k \text{ is the unique minimizer of problem (SDP).} \tag{5.1}$$

Let g_1, C_1 be defined as in proposition 2.3: $C = \mathcal{A}^*(g_1) + C_1, C_1 \succeq 0$, and $C_1 X_0 = 0$. Similarly, let, for any k, $h_k \in \mathbb{R}^m, D_k \in \mathbb{S}^{n\times n}$ be such that $C'_k = \mathcal{A}^*(h_k) + D_k, D_k \succeq 0$, and $D_k X'_k = 0$.

The following lemma states that $D_k \overset{k\to+\infty}{\to} C_1$ and $h_k \overset{k\to+\infty}{\to} g_1$. Its proof is in subsection D.2 and relies on the p-regularity of $(\mathcal{A}, b)$.

Lemma 5.5. *When k goes to infinity, $D_k \to C_1$ and $h_k \to g_1$.*

For any k, because $D_k X'_k = 0$,

$$\operatorname{rank}(D_k) + \operatorname{rank}(X'_k) \leq n. \tag{5.2}$$

From Proposition 5.4, $(X'_k)_{k\in\mathbb{N}}$ converges to X_0, and from Lemma 5.5, $(D_k)_{k\in\mathbb{N}}$ converges to C_1. In particular, for k large enough,

$$\begin{aligned} &\operatorname{rank}(X'_k) \geq \operatorname{rank}(X_0) = r \\ \text{and}\quad &\operatorname{rank}(D_k) \geq \operatorname{rank}(C_1) = n - r. \end{aligned}$$

Combined with (5.2), this proves that, for k large enough, $\operatorname{rank}(X'_k) = r$ and $\operatorname{rank}(D_k) = n - r$. This establishes the first part of property (5.1). The second part is a direct consequence of Proposition 2.4. ∎

5.4. Proof of Lemma 5.3. We recall that X_0 is a global minimizer of problem (SDP), but VV^T is not: $\langle C, X_0\rangle < \langle C, VV^T\rangle$. By continuity, there is actually a neighborhood $\mathcal{V}$ of V in $\mathcal{M}_p$ such that

$$\forall V' \in \mathcal{V}, \quad \langle C, X_0\rangle < \langle C, V'V'^T\rangle. \tag{5.3}$$

By continuity again, (5.3) stays true if we replace C by any close enough matrix C'. Therefore, for C' close enough to C, no matrix of the form $V'V'^T$ with $V' \in \mathcal{V}$ can be a global minimizer of problem (SDP), hence no matrix of $\mathcal{V}$ can be a global minimizer of problem (Factorized SDP). From this remark, if we show that, for any C' close enough to C, problem (Factorized SDP) has a second-order critical point in $\mathcal{V}$, we have proved the lemma. Let us do that.

For any cost matrix $C' \in \mathbb{S}^{n\times n}$, we denote $f_{C'} : W \in \mathcal{M}_p \to \langle C', WW^T\rangle \in \mathbb{R}$ the cost function of problem (Factorized SDP). If $\operatorname{Hess} f_C(V)$ was positive definite, we could apply the following general proposition (proved in Appendix D.3).

Proposition 5.6. *Let $\mathcal{M}$ be a Riemannian manifold, E a finite-dimensional vector space, and $f : E \times \mathcal{M} \to \mathbb{R}$ a smooth map. Let $c \in E, v \in \mathcal{M}$ be fixed. We assume that $f(c, .)$ has a second-order critical point at v and that* $\operatorname{Hess}(f(c, .))(v) \succ 0$.

Then, for any neighborhood $\mathcal{V}$ of v in $\mathcal{M}$, the map $f(c', .)$ has a second-order critical point in $\mathcal{V}$ for any $c' \in E$ close enough to c.

However, because f_C is invariant to right multiplication by elements of $O(p)$, the Hessian is degenerate. Therefore, before applying the proposition, we must explicitly factorize this invariance by introducing the corresponding quotient manifold. We refer to [2, section 3.4] for basic results on quotient manifolds. Specifically, let $\mathcal{M}_p^{full}$ be the open subset of $\mathcal{M}_p$ that contains its rank p elements.[1] The quotient $\mathcal{M}_p^{full}/O(p)$ is a manifold with dimension

$$\dim(\mathcal{M}_p) - \dim(O(p)) = \dim(\mathcal{M}_p) - \frac{p(p-1)}{2}.$$

Since $O(p)$ acts by isometries on $\mathcal{M}_p^{full}$, $\mathcal{M}_p^{full}/O(p)$ inherits from the Riemannian structure of $\mathcal{M}_p^{full}$. We denote $Q : \mathcal{M}_p^{full} \to \mathcal{M}_p^{full}/O(p)$ the canonical projection. It is a smooth map, with surjective differential everywhere.

For any C', since $f_{C'}$ is invariant to the action of $O(p)$, we can define its quotient, that is, the (also smooth) map $f_{C',O(p)} : \mathcal{M}_p^{full}/O(p) \to \mathbb{R}$ such that

$$f_{C',O(p)} \circ Q = f_{C'}.$$

The following proposition (proved in Appendix D.4) shows that there is a correspondance between the critical points of $f_{C',O(p)}$ and $f_{C'}$.

Proposition 5.7. *Let $\mathcal{N}_1, \mathcal{N}_2$ be two Riemannian manifolds and $f : \mathcal{N}_2 \to \mathbb{R}$ be a smooth function. Let $\phi : \mathcal{N}_1 \to \mathcal{N}_2$ be a smooth map with surjective differential at any point of $\mathcal{N}_1$.*

Then, for any $v \in \mathcal{N}_1$, v is a second-order critical point of $f \circ \phi$ if and only if $\phi(v)$ is a second-order critical point of f. Additionally,

$$\operatorname{rank}(\operatorname{Hess}(f \circ \phi)(v)) = \operatorname{rank}(\operatorname{Hess} f(\phi(v))).$$

This proposition, applied to $\mathcal{N}_1 = \mathcal{M}_p^{full}, \mathcal{N}_2 = \mathcal{M}_p^{full}/O(p)$, $f = f_{C',O(p)}$, and $\phi = Q$, shows that, because V is a second-order critical point of $f_C = f_{C,O(p)} \circ Q$, $Q(V)$ is a second-order critical point of $f_{C,O(p)}$ and

$$\begin{aligned}\operatorname{rank}(\operatorname{Hess} f_{C,O(p)}(Q(V))) &= \operatorname{rank}(\operatorname{Hess} f_C(V)) \\ &\overset{(Rem. 2.8)}{=} \dim(\mathcal{M}_p) - \frac{p(p-1)}{2} \\ &= \dim(\mathcal{M}_p^{full}/O(p)).\end{aligned}$$

In other words, $\operatorname{Hess} f_{C,O(p)}(Q(V))$ is positive definite.

We apply Proposition 5.6 to $E = \mathbb{S}^{n\times n}, \mathcal{M} = \mathcal{M}_p^{full}/O(p)$, and $f : (C', W) \in \mathbb{S}^{n\times n} \times \mathcal{M}_p^{full}/O(p) \to f_{C',O(p)}(W)$: for any neighborhood $\mathcal{V}_{O(p)}$ of $Q(V)$, $f_{C',O(p)}$ has a second-order critical point in $\mathcal{V}_{O(p)}$ if C' is close enough to C.

We use this property with $\mathcal{V}_{O(p)} = Q(\mathcal{V})$. For any C' close enough to C, $f_{C',O(p)}$ has a second-order critical point of the form $Q(W)$, with $W \in \mathcal{V}$. Then, from Proposition 5.7, $f_{C'}$ has a second-order critical point in $\mathcal{V}$. ∎

5.5. Construction of C: Point 2. We must show the existence of g_1, C_1, C_2 such that

$$C_1 \succeq 0, \quad C_1 X_0 = 0, \quad \operatorname{rank}(C_1) = n - r, \tag{5.4a}$$

$$\mathcal{A}^*(g_1) + C_1 = C_2, \quad C_2 V = 0. \tag{5.4b}$$

We simplify the problem with the following proposition, proved in Appendix D.5.

[1] If $\mathcal{M}_p$ contains rank-deficient matrices, $\mathcal{M}_p/O(p)$ is not a manifold, because $\{(V, VX), V \in \mathcal{M}_p, X \in O(p)\}$ is not a submanifold of $\mathcal{M}_p^2$. We must therefore remove rank-deficient elements from $\mathcal{M}_p$.

PROPOSITION 5.8. *Without loss of generality, we can assume that*

$$X_0 = \begin{pmatrix} I_r & 0_{r,n-r} \\ 0_{n-r,r} & 0_{n-r,n-r} \end{pmatrix} \quad \textit{and} \quad V = \begin{pmatrix} 0_{r,p} \\ I_p \\ 0_{n-p-r,p} \end{pmatrix}.$$

With this assumption, the three conditions in (5.4a) are true if and only if

$$C_1 = \begin{pmatrix} 0_{r,r} & 0_{r,n-r} \\ 0_{n-r,r} & D_1 \end{pmatrix}$$

for some $D_1 \in \mathbb{S}^{(n-r)\times(n-r)}$ such that $D_1 \succ 0$. And $C_2 V = 0$ if and only if

$$C_2 = \begin{pmatrix} F_1 & 0_{r,p} & F_2 \\ 0_{p,r} & 0_{p,p} & 0_{p,n-r-p} \\ F_2^T & 0_{n-r-p,p} & F_3 \end{pmatrix}$$

for some F_1, F_2, F_3. Therefore, to ensure conditions (5.4a) and (5.4b), we must only show the existence of g_1, D_1, F_1, F_2, F_3 such that $D_1 \succ 0$ and

$$\mathcal{A}^*(g_1) = \begin{pmatrix} F_1 & 0_{r,p} & F_2 \\ 0_{p,r} & 0_{p,p} & 0_{p,n-r-p} \\ F_2^T & 0_{n-r-p,p} & F_3 \end{pmatrix} - \begin{pmatrix} 0_{r,r} & 0_{r,n-r} \\ 0_{n-r,r} & D_1 \end{pmatrix}. \tag{5.5}$$

We observe that, if these exist, $\mathcal{A}^*(g_1)$ must be of the form

$$\mathcal{A}^*(g_1) = \begin{pmatrix} G_1 & 0_{r,p} & G_2 \\ 0_{p,r} & G_3 & G_4 \\ G_2^T & G_4^T & G_5 \end{pmatrix} \tag{5.6}$$

with $G_3 \prec 0$ (since it is a minor of $-D_1$). But conversely, if there exists g_1 for which (5.6) is true, we can set

$$F_1 = G_1, \quad F_2 = G_2, \quad F_3 = G_5 + \lambda I_{n-r-p}, \quad D_1 = \begin{pmatrix} -G_3 & -G_4 \\ -G_4^T & \lambda I_{n-r-p} \end{pmatrix}$$

for some $\lambda > 0$ large enough, and (5.5) holds. (We observe that $D_1 \succ 0$ for λ large enough: all its principal minors are of the form

$$\det \begin{pmatrix} -G_3^{(sub)} & -G_4^{(sub)} \\ -G_4^{(sub)T} & \lambda I_s \end{pmatrix} = \lambda^s \det\left(-G_3^{(sub)}\right) + O(\lambda^{s-1})$$

with $-G_3^{(sub)}$ a principal submatrix of $-G_3$, whose determinant is positive because $-G_3 \succ 0$. Therefore, all principal minors of D_1 are positive if λ is large enough.)

To conclude, we must only prove the existence of g_1 for which (5.6) is true. This is a consequence of the following lemma, whose proof is in Appendix D.6 (and relies on the minimally secant property).

LEMMA 5.9. *For any $R_1 \in \mathbb{R}^{r\times p}, R_2 \in \mathbb{S}^{p\times p}$, there exist $g_1 \in \mathbb{R}^m, G_1, G_2, G_4, G_5$ such that*

$$\mathcal{A}^*(g_1) = \begin{pmatrix} G_1 & R_1 & G_2 \\ R_1^T & R_2 & G_4 \\ G_2^T & G_4^T & G_5 \end{pmatrix}.$$

∎

5.6. Construction of C: Point 3. In this subsection, we consider g_1, C_1, C_2 which satisfy properties (Optim. SDP 2), (Optim. SDP 3), (Strict complementarity), (First-order optim. 1), and (First-order optim. 2), and construct $g_1^{(mod)}$,$C_1^{(mod)}$,$C_2^{(mod)}$, also satisfying these properties, and, in addition, property (Second-order optim.):

$$\forall \dot{V} \in T_V\mathcal{M}_p, \quad \left\langle C_2^{(mod)}, \dot{V}\dot{V}^T \right\rangle \geq 0, \tag{5.7}$$

with equality if and only if $\dot{V} = VA$ for some $A \in \mathrm{Anti}(p)$.

Using Proposition 5.8 as in the previous subsection, we assume

$$X_0 = \begin{pmatrix} I_r & 0_{r,n-r} \\ 0_{n-r,r} & 0_{n-r,n-r} \end{pmatrix} \quad \text{and} \quad V = \begin{pmatrix} 0_{r,p} \\ I_p \\ 0_{n-p-r,p} \end{pmatrix}. \tag{5.8}$$

We set

$$g_1^{(mod)} = g_1, \quad C_1^{(mod)} = C_1 + t \begin{pmatrix} 0_{r+p,r+p} & 0_{r+p,n-r-p} \\ 0_{n-r-p,r+p} & I_{n-r-p} \end{pmatrix},$$
$$C_2^{(mod)} = C_2 + t \begin{pmatrix} 0_{r+p,r+p} & 0_{r+p,n-r-p} \\ 0_{n-r-p,r+p} & I_{n-r-p} \end{pmatrix}$$

for some $t \geq 0$ large. From the following proposition (proved in Appendix D.7) these definitions satisfy properties (Optim. SDP 2) to (First-order optim. 2).

PROPOSITION 5.10. *Whatever the value of $t \geq 0$, $g_1^{(mod)}$, $C_1^{(mod)}$, and $C_2^{(mod)}$ satisfy properties* (Optim. SDP 2) *to* (First-order optim. 2).

When t is large enough, it turns out that they also satisfy (5.7). This is proved in two steps, each embedded in a proposition (proofs are in Appendices D.8 and D.9): first, we observe that, to prove (5.7), one can look only at matrices $\dot{V}$ in some subspace of $T_V\mathcal{M}_p$. Then, with a compactness argument, we show that, for matrices in this subspace, (5.7) is true.

PROPOSITION 5.11. *Let $\mathcal{E}_\perp$ be the orthogonal in $T_V\mathcal{M}_p$ of $\{VA, A \in \mathrm{Anti}(p)\}$. Equation* (5.7) *is true if and only if*

$$\forall \dot{V} \in \mathcal{E}_\perp - \{0\}, \quad \left\langle C_2^{(mod)}, \dot{V}\dot{V}^T \right\rangle > 0. \tag{5.9}$$

PROPOSITION 5.12. *For t large enough,* (5.9) *is true.* ◻

Appendix A. Proof of basic properties.

A.1. Proof of Proposition 2.3. The dual of problem (SDP) is

$$\text{(SDP-dual)} \qquad \begin{aligned} &\text{maximize } \langle g_1, b \rangle \\ &\text{such that } C = \mathcal{A}^*(g_1) + C_1, \\ &\qquad\qquad C_1 \succeq 0. \end{aligned}$$

If C_1, g_1 are as in the statement, they are dual feasible. Because of the complementary slackness condition $C_1X_0 = 0$, X_0 and (C_1, g_1) are primal-dual optimal. In particular, X_0 is a solution of problem (SDP).

Conversely, let us assume X_0 is a solution of problem (SDP). We temporarily admit that Slater's condition holds (that is, $\mathcal{C}$ contains a positive definite matrix). Then strong duality holds [28, p. 114] and the dual problem has at least one solution

(C_1, g_1). This pair satisfies $C = \mathcal{A}^*(g_1) + C_1$ and $C_1 \succeq 0$, because it is dual feasible. Strong duality means that

$$\langle g_1, b \rangle = \langle C, X_0 \rangle ,$$

which is equivalent to $\langle C_1, X_0 \rangle = 0$ and in turn implies $C_1 X_0 = 0$ because $C_1, X_0 \succeq 0$.

To establish Slater's condition, we assume by contradiction that it does not hold:

$$\{X \in \mathbb{S}^{n\times n}, \mathcal{A}(X) = b\} \cap \{X \in \mathbb{S}^{n\times n}, X \succ 0\} = \emptyset.$$

From a hyperplane separation theorem, there exists a nonzero $M \in \mathbb{S}^{n\times n}$, and $\mu \in \mathbb{R}$ such that

$$\forall X \in \{X \in \mathbb{S}^{n\times n}, X \succ 0\}, \quad \langle M, X\rangle > \mu, \tag{A.1a}$$

$$\text{and} \quad \forall X \in \{X \in \mathbb{S}^{n\times n}, \mathcal{A}(X) = b\}, \quad \langle M, X\rangle \le \mu. \tag{A.1b}$$

Equation (A.1a) is equivalent to $M \succeq 0$ and $\mu \le 0$. And if we fix $V \in \mathcal{M}_p$, we can see that (A.1b) is equivalent to

$$M \in \mathrm{Range}(\mathcal{A}^*) \text{ and } \left\langle M, VV^T \right\rangle \le \mu.$$

In particular, $\left\langle M, VV^T \right\rangle \le \mu \le 0$. As $M \succeq 0$, this means $MV = 0$. Denoting $g \in \mathbb{R}^m$ a vector such that $M = \mathcal{A}^*(g)$, we have $\mathcal{A}^*(g)V = 0$. Therefore, for any $\dot V \in \mathbb{R}^{n\times p}$,

$$\left\langle \mathcal{A}(V\dot V^T + \dot V V^T), g \right\rangle = \left\langle V\dot V^T + \dot V V^T, M \right\rangle = 2\left\langle \dot V, MV \right\rangle = 0,$$

which contradicts the assumption that $(\mathcal{A}, b)$ is p-regular. ∎

A.2. Proof of Proposition 2.4. We assume that strict complementary slackness holds, but X_0 is not the unique solution of problem (SDP), and we show that X_0 is not an extremal point of $\mathcal{C}$.

We observe that X_0 has maximal rank among all solutions of problem (SDP): if X_0' is another solution, as (C_1, g_1) is dual optimal, X_0' and C_1 satisfy the complementary slackness condition,

$$C_1 X_0' = 0,$$

hence $\mathrm{rank}(X_0') \le n - \mathrm{rank}(C_1) = \mathrm{rank}(X_0)$. From [18, Theorem 2.4], there exists $U_0 \in \mathbb{R}^{n\times \mathrm{rank}(X_0)}$ such that

$$X_0 = U_0 U_0^T$$

and there exists $Z \in \mathbb{S}^{\mathrm{rank}(X_0)\times\mathrm{rank}(X_0)}$ such that $\mathcal{A}(U_0 Z U_0^T) = 0$. Then $U_0(I_{\mathrm{rank}} + tZ)U_0^T$ belongs to $\mathcal{C}$ for any t close enough to 0, meaning than X_0 is not extremal. ∎

A.3. Proof of Proposition 2.5. From [8, equation 7], the gradient of the cost function of problem (Factorized SDP) at V is $2\mathrm{Proj}_V(CV)$, where $\mathrm{Proj}_V : \mathbb{R}^{n\times p} \to T_V\mathcal{M}_p$ is the orthogonal projection onto $T_V\mathcal{M}_p$. Consequently, V is a first-order critical point if and only if

$$\begin{aligned} CV \in (T_V\mathcal{M}_p)^\perp &= \left\{ \dot V \in \mathbb{R}^{n\times p}, \mathcal{A}(\dot V V^T + V\dot V^T) = 0 \right\}^\perp \\ &= \left\{ \dot V \in \mathbb{R}^{n\times p}, \forall g_2 \in \mathbb{R}^m, \left\langle \dot V V^T + V \dot V^T, \mathcal{A}^*(g_2) \right\rangle = 0 \right\}^\perp \\ &= \left(\{\mathcal{A}^*(g_2)V, g_2 \in \mathbb{R}^m\}^\perp \right)^\perp \\ &= \{\mathcal{A}^*(g_2)V, g_2 \in \mathbb{R}^m\}. \end{aligned}$$

Now, $CV = \mathcal{A}^*(g_2)V$ for some $g_2 \in \mathbb{R}^m$ if and only if $C = C_2 + \mathcal{A}^*(g_2)$, for some $g_2 \in \mathbb{R}^m, C_2 \in \mathbb{S}^{n\times n}$ such that $C_2V = 0$.

To show that, when it exists, the pair (g_2, C_2) is unique, we assume that there exists another pair (g_2', C_2') satisfying the same conditions. Then $\mathcal{A}^*(g_2 - g_2')V = (C_2' - C_2)V = 0$. The same argument as at the end of Appendix A.1 shows that $g_2 - g_2' = 0$. Therefore, $g_2 = g_2'$ and $C_2 = C_2'$. ∎

A.4. Proof of Proposition 2.6. For any $\dot V \in T_V\mathcal{M}_p$, from [8, eq. 10],

$$\mathrm{Hess} f_C(V)\cdot(\dot V, \dot V) = 2\left\langle S\dot V, \dot V\right\rangle,$$

where $S = C - \mathcal{A}^*(\mu)$ for some $\mu \in \mathbb{R}^m$ such that $2SV = \mathrm{grad} f_C(V) = 0$.

From the uniqueness of (C_2, g_2), we have $\mu = g_2$ and $S = C_2$. ∎

Appendix B. Discussion on Definition 2.9. Let for the time being $X_0 \in \mathbb{S}^{n\times n}, V \in \mathbb{R}^{n\times p}$ be fixed, such that $\mathrm{rank}(X_0) = r$. We assume properties 1 and 2 of Definition 2.9 are true:

$$\mathrm{rank}(V) = p \quad \text{and} \quad \mathrm{Range}(X_0) \cap \mathrm{Range}(V) = \{0\}.$$

We discuss when property 3 holds. This property is equivalent to

$$\begin{aligned} T_V\mathcal{M}_p \cap \{\dot V, \mathrm{Range}(\dot V) \subset \mathrm{Range}(X_0) + \mathrm{Range}(V)\} \\ = \{VA, A \in \mathrm{Anti}(p)\}. \end{aligned} \tag{B.1}$$

The vector spaces $\{\dot V, \mathrm{Range}(\dot V) \subset \mathrm{Range}(X_0) + \mathrm{Range}(V)\}$ and $T_V\mathcal{M}_p$ contain $\{VA, A \in \mathrm{Anti}(p)\}$ and respectively have dimensions

$$p \times \dim\left(\mathrm{Range}(X_0) + \mathrm{Range}(V)\right) = p(p+r)$$

and $np - m$.

Consequently, $(T_V\mathcal{M}_p \cap \{\dot V, \mathrm{Range}(\dot V) \subset \mathrm{Range}(X_0) + \mathrm{Range}(V)\})^\perp$ is a subset of $\{VA, A \in \mathrm{Anti}(p)\}^\perp$, with dimension at most

$$\begin{aligned} \min\Big(&\dim\{VA, A \in \mathrm{Anti}(p)\}^\perp, \\ &\dim(T_V\mathcal{M}_p)^\perp + \dim\{\dot V, \mathrm{Range}(\dot V) \subset \mathrm{Range}(X_0) + \mathrm{Range}(V)\}^\perp\Big) \\ &= \min\left(np - \frac{p(p-1)}{2}, m + np - p(p+r)\right) \\ &= \dim\{VA, A \in \mathrm{Anti}(p)\}^\perp + \min\left(0, m - \frac{p(p+1)}{2} - pr\right). \end{aligned}$$

Therefore, if $\frac{p(p+1)}{2} + pr > m$, $(T_V\mathcal{M}_p \cap \{\dot V, \mathrm{Range}(\dot V) \subset \mathrm{Range}(X_0) + \mathrm{Range}(V)\})^\perp$ is a strict subset of $\{VA, A \in \mathrm{Anti}(p)\}^\perp$, (B.1) does not hold and $\mathcal{M}_p$ cannot be X_0-minimally secant at V. On the contrary, if $\frac{p(p+1)}{2} + pr \le m$, the above upper bound on the dimension is exactly equal to the dimension if $T_V\mathcal{M}_p$ is "generic enough," hence $\mathcal{M}_p$ is X_0-minimally secant at V.

Consequently, we expect the main assumption in Theorem 3.3 (the existence of X_0, V such that $\mathcal{M}_p$ is X_0-minimally secant at V, in a setting where $\frac{p(p+1)}{2} + pr \le m$) to hold for almost all $(\mathcal{A}, b)$.

We, however, note that, although rare, there are pairs $(\mathcal{A}, b)$ for which X_0, V do not exist, hence the hypothesis cannot be trivially removed. An example is as follows.

Example B.1. We set $r = 1, p = 2$. Let $m \leq n$ be arbitrary. We consider

$$\mathcal{A} : X \in \mathbb{S}^{n\times n} \to (X_{1,1}, X_{1,n-m+2}, \dots, X_{1,n}) \in \mathbb{R}^m,$$

and $b = (1, 0, \dots, 0)$. One can check that $(\mathcal{A}, b)$ is 2-regular.

The rank-1 elements of $\mathcal{C}$ are exactly the matrices X_0 of the form

$$X_0 = \begin{pmatrix} \begin{matrix} 1 & u^T \\ u & uu^T \end{matrix} & 0_{n-m+1,m-1} \\ 0_{m-1,n-m+1} & 0_{m-1,m-1} \end{pmatrix} \quad \text{with } u \in \mathbb{R}^{(m-1)\times 1},$$

and $\mathcal{M}_2$ contains all matrices of the form $V = WX$ with $X \in O(2)$ and

$$W = \begin{pmatrix} 1 & 0 \\ w_1 & w_2 \\ \multicolumn{2}{c}{0_{m-1,2}} \end{pmatrix} \quad \text{with } w_1, w_2 \in \mathbb{R}^{(n-m)\times 1}.$$

For any rank-1 X_0 in $\mathcal{C}$ and $V \in \mathcal{M}_2$, using the above notation one can check that

$$\dot{V} = \begin{pmatrix} 0 & 0 \\ w_1 - u & \vdots \\ 0 & \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix} X$$

is in $T_V\mathcal{M}_2$, and $\text{Range}(\dot{V}) \subset \text{Range}(X_0) + \text{Range}(V)$. Nevertheless, $\dot{V} \neq VA$, for all $A \in \text{Anti}(2)$, so property 3 of Definition 2.9 does not hold (unless $w_1 = u$, in which case property 2 does not hold), even if $m \geq \frac{p(p+1)}{2} + pr = 5$.

Appendix C. Auxiliary results for the proof of Theorem 3.2.

C.1. Proof of Proposition 4.1. Let (V, C_2) belong to $\mathcal{E}$. We are going to exhibit a parametrization of $\mathcal{E}$ around (V, C_2). Because V has rank p, there exist a neighborhood $\mathcal{V}$ of V in $\mathbb{R}^{n\times p}$ and a smooth map $W \in \mathcal{V} \to U_W \in \mathbb{R}^{n\times(n-p)}$ such that, for any $W \in \mathcal{V}$, the columns of U_W form an orthonormal basis of $\text{Range}(W)^\perp$. We define

$$\begin{array}{cccc} \psi : & (\mathcal{M}_p^{full} \cap \mathcal{V}) \times \mathbb{S}^{n\times n} & \to & \mathbb{S}^{p\times p} \times \mathbb{R}^{(n-p)\times p}, \\ & (W, D) & \to & (W^T DW, U_W^T DW). \end{array}$$

This function is smooth. At any point (W, D), its differential with respect to D is surjective: for any $(A, B) \in \mathbb{S}^{p\times p} \times \mathbb{R}^{(n-p)\times p}$, one can check that $d_D\psi(W, D) \cdot \dot{D} = (A, B)$ if one sets

$$\dot{D} = (\, W \;\; U_W \,)^{-1T} \begin{pmatrix} A & B^T \\ B & 0_{n-p,n-p} \end{pmatrix} (\, W \;\; U_W \,)^{-1}.$$

Therefore, from [2, Proposition 3.3.3], $\psi^{-1}(0_{p,p}, 0_{n-p,p})$ is a submanifold of $\mathcal{M}_p^{full} \times \mathbb{S}^{n\times n}$, with dimension

$$\dim(\mathcal{M}_p^{full} \times \mathbb{S}^{n\times n}) - \dim(\mathbb{S}^{p\times p} \times \mathbb{R}^{(n-p)\times p}) = np - m + \frac{(n-p)(n-p+1)}{2}.$$

For any $(W, D) \in (\mathcal{M}_p^{full} \cap \mathcal{V}) \times \mathbb{S}^{n\times n}$, the following equivalences are true:

$$\begin{aligned} \Big(\psi(W, D) = (0_{p,p}, 0_{n-p,p})\Big) \quad &\Longleftrightarrow \quad \Big((\, W \;\; U_W \,)^T DW = 0_{n,p}\Big) \\ &\Longleftrightarrow \quad \Big(DW = 0_{n,p}\Big). \end{aligned}$$

Consequently, $\psi^{-1}(0_{p,p}, 0_{n-p,p})$ and $\mathcal{E}$ coincide in a neighborhood of (V, C_2), which implies that $\mathcal{E}$ is also an $(np - m + \frac{(n-p)(n-p+1)}{2})$-dimensional manifold.

Its tangent space at (V, C_2) is

$$\begin{aligned}
\mathrm{Ker}(d\psi(V, C_2)) = \Big\{ &(\dot V, \dot C_2) \in T_V\mathcal{M}_p \times \mathbb{S}^{n\times n}, \dot V^T C_2 V + V^T \dot C_2 V + V^T C_2 \dot V = 0_{p,p} \\
&\text{and } (dU_V \cdot \dot V) C_2 V + U_V^T \dot C_2 V + U_V^T C_2 \dot V = 0_{n-p,p} \Big\} \\
\overset{(C_2 V = 0)}{=} \Big\{ &(\dot V, \dot C_2) \in T_V\mathcal{M}_p \times \mathbb{S}^{n\times n}, V^T \dot C_2 V + V^T C_2 \dot V = 0_{p,p} \\
&\text{and } U_V^T \dot C_2 V + U_V^T C_2 \dot V = 0_{n-p,p} \Big\} \\
= \Big\{ &(\dot V, \dot C_2) \in T_V\mathcal{M}_p \times \mathbb{S}^{n\times n}, \left(V \; U_V \right)^T (\dot C_2 V + C_2 \dot V) = 0_{n,p} \Big\} \\
= \Big\{ &(\dot V, \dot C_2) \in T_V\mathcal{M}_p \times \mathbb{S}^{n\times n}, \dot C_2 V + C_2 \dot V = 0_{n,p} \Big\}. \qquad \square
\end{aligned}$$

C.2. Proof of Proposition 4.3. Let $U_V \in \mathbb{R}^{n\times p}, U_V^\perp \in \mathbb{R}^{n\times(n-p)}$ be matrices whose columns respectively form an orthonormal basis of $\mathrm{Range}(V)$ and of $\mathrm{Ker}(VV^T) = \mathrm{Range}(V)^\perp$. Let $G \in \mathbb{R}^{p\times p}$ be the unique matrix such that $U_V = VG$.

Because $\mathcal{C} \subset \{X \in \mathbb{S}^{n\times n}, X \succeq 0\}$, the face of $\mathcal{C}$ containing VV^T is a subset of the face of $\{X \in \mathbb{S}^{n\times n}, X \succeq 0\}$ containing VV^T, which is, from [17, section 2.4],

$$\{X \in \mathbb{S}^{n\times n}, X \succeq 0, \mathrm{Ker}(VV^T) \subset \mathrm{Ker}(X)\}.$$

Therefore, $\mathrm{Ker}(VV^T) \subset \mathrm{Ker}(X_{Face})$, which implies

$$\begin{aligned}
&X_{Face} U_V^\perp = 0_{n,n-p}, \\
&\Rightarrow \quad \left(U_V \; U_V^\perp \right)^T X_{Face} \left(U_V \; U_V^\perp \right) = \begin{pmatrix} R & 0_{p,n-p} \\ 0_{n-p,p} & 0_{n-p,n-p} \end{pmatrix} \text{ for some } R \in \mathbb{S}^{p\times p} \\
&\Rightarrow \quad X_{Face} = \left(U_V \; U_V^\perp \right) \begin{pmatrix} R & 0_{p,n-p} \\ 0_{n-p,p} & 0_{n-p,n-p} \end{pmatrix} \left(U_V \; U_V^\perp \right)^T \text{ for some } R \in \mathbb{S}^{p\times p} \\
&\Rightarrow \quad X_{Face} = VGRG^T V^T \text{ for some } R \in \mathbb{S}^{p\times p}, \\
&\Rightarrow \quad X_{Face} = VTV^T \text{ for some } T \in \mathbb{S}^{p\times p}. \qquad \square
\end{aligned}$$

Appendix D. Auxiliary results for the proof of Theorem 3.3.

D.1. Proof of Proposition 5.4. It suffices to show the following property:

$$\text{(D.1)} \qquad \forall\, k \text{ large enough}, \quad \forall X \in \mathcal{C} - B(X_0, \epsilon), \quad \langle C_k', X_0 \rangle < \langle C_k', X \rangle .$$

Indeed, in this case, for k large enough, any minimizer of $\langle C_k', . \rangle$ on the compact set $\mathcal{C} \cap \overline{B}(X_0, \epsilon)$ (there is at least one) is a minimizer of $\langle C_k', . \rangle$ on $\mathcal{C}$, and every minimizer of $\langle C_k', . \rangle$ on $\mathcal{C}$ is in $B(X_0, \epsilon)$.

We assume, by contradiction, that property (D.1) is not true. Up to replacing $(C_k')_{k\in\mathbb{N}}$ by a subsequence, we can assume that, for any $k \in \mathbb{N}$,

$$\text{(D.2)} \qquad \exists X_k' \in \mathcal{C} - B(X_0, \epsilon), \quad \langle C_k', X_0 \rangle \geq \langle C_k', X_k' \rangle .$$

For any k, let X_k' be such a matrix.

By compactness, we can assume that $((X_k' - X_0)/||X_k' - X_0||)_{k\in\mathbb{N}}$ converges to some unit-normed limit $Z \in \mathbb{S}^{n\times n}$. From (D.2) and because $(C_k')_{k\in\mathbb{N}}$ converges to C, $\langle C, Z \rangle \leq 0$. Equivalently,

$$\text{(D.3)} \qquad \langle C, X_0 + \epsilon Z \rangle \leq \langle C, X_0 \rangle .$$

Observe that $X_0 + \epsilon Z$ belongs to $\mathcal{C}$: it is the limit of the sequence

$$\left(\left(1 - \frac{\epsilon}{||X'_k - X_0||}\right) X_0 + \frac{\epsilon}{||X'_k - X_0||} X'_k\right)_{k\in\mathbb{N}^*}.$$

Each element of this sequence belongs to $\mathcal{C}$ (X_0 and X'_k do, and $\mathcal{C}$ is convex), and $\mathcal{C}$ is closed, so the limit also belongs to $\mathcal{C}$. Consequently, (D.3) contradicts the fact that X_0 is the unique minimizer of $\langle C, .\rangle$ on $\mathcal{C}$. ∎

D.2. Proof of Lemma 5.5. As $D_k + \mathcal{A}^*(h_k) = C'_k \overset{k\to+\infty}{\to} C = C_1 + \mathcal{A}^*(g_1)$,

$$D_k - C_1 + \mathcal{A}^*(h_k - g_1) \overset{k\to+\infty}{\to} 0. \tag{D.4}$$

In particular, if $h_k \overset{k\to+\infty}{\to} g_1$, then $D_k \overset{k\to+\infty}{\to} C_1$, so we only have to show that $(h_k)_{k\in\mathbb{N}}$ converges to g_1.

By contradiction, we assume that $h_k \not\to g_1$. Up to replacing $(h_k)_{k\in\mathbb{N}}$ by a subsequence, we can assume that $(||h_k - g_1||)_{k\in\mathbb{N}}$ is lower bounded by a positive constant and that $((h_k - g_1)/||h_k - g_1||)_{k\in\mathbb{N}}$ converges to some nonzero limit g.

From (D.4),

$$\frac{C_1 - D_k}{||h_k - g_1||} \overset{k\to+\infty}{\to} \mathcal{A}^*(g). \tag{D.5}$$

From Proposition 5.4, $(X'_k)_{k\in\mathbb{N}}$ converges to X_0, so $C_1 X'_k \overset{k\to+\infty}{\to} C_1 X_0 = 0$, and because $(||h_k - g_1||)_{k\in\mathbb{N}}$ is bounded away from zero, this implies

$$\frac{C_1 X'_k}{||h_k - g_1||} \overset{k\to+\infty}{\to} 0.$$

Recalling that, from the definition of D_k, $D_k X'_k = 0$ for all k, (D.5) yields

$$\mathcal{A}^*(g) X_0 = \lim_{k\to+\infty} \left(\frac{C_1 - D_k}{||h_k - g_1||}\right) X'_k = 0.$$

Therefore, we also have $\mathcal{A}^*(g)V_0$, if we fix $V_0 \in \mathbb{R}^{n\times p}$ such that $X_0 = V_0 V_0^T$ (it is possible, as $X_0 \succeq 0$ and $\mathrm{rank}(X_0) = r \le p$). The matrix V_0 is in $\mathcal{M}_p$. Applying the same argument as at the end of Appendix A.1, we reach a contradiction. ∎

D.3. Proof of Proposition 5.6. If we compose f with a diffeomorphism along the second coordinate, we can assume that $\mathcal{M}$ is an open subset of $\mathbb{R}^d$ for some integer d (from Proposition 5.7, the critical points of the composition of a function and a diffeomorphism are exactly the image by the reciprocal diffeomorphism of the critical points of the function).

Let $\mathcal{V}$ be a neighborhood of v in $\mathcal{M}$. We define

$$\begin{array}{rccc} \chi: & E \times \mathcal{V} & \to & \mathbb{R}^d, \\ & (c', v') & \to & \nabla(f(c', .))(v'). \end{array}$$

This is a smooth map; it satisfies $\chi(c, v) = 0$ (since v is a critical point of $f(c, .)$) and its differential at (c, v) along the second coordinate is invertible (it is $\mathrm{Hess}(f(c, .))(v)$, which is positive definite by assumption). From the implicit function theorem, there exist a neighborhood $\mathcal{E}$ of c in E and a smooth function $\delta : \mathcal{E} \to \mathcal{V}$ such that $\delta(c) = v$ and $\chi(c', \delta(c')) = 0$ for any $c' \in \mathcal{E}$. We fix such $\mathcal{E}, \delta$. Then, for any $c' \in \mathcal{E}$,

$$\nabla(f(c', .))(\delta(c')) = \chi(c', \delta(c')) = 0.$$

Equivalently, $\delta(c')$ is a first-order critical point of $f(c', .)$. Additionally, the map $c' \to \mathrm{Hess}(f(c', .))(\delta(c'))$ is continuous (f and δ are smooth), and

$$\mathrm{Hess}(f(c, .))(\delta(c)) = \mathrm{Hess}(f(c, .))(v) \succ 0.$$

As a consequence, for any $c' \in \mathcal{E}$ close enough to c,

$$\mathrm{Hess}(f(c', .))(\delta(c')) \succ 0.$$

Therefore, for any c' close enough to c, $\delta(c')$, which is an element of $\mathcal{V}$, is a second-order critical point of $f(c', .)$. ◻

D.4. Proof of Proposition 5.7. Let v belong to $\mathcal{N}_1$. We have

$$\nabla(f \circ \phi)(v) = (d\phi(v))^* \nabla f(\phi(v)).$$

As $d\phi(v)^*$ is injective (it is the adjoint of a surjective map), v is a first-order critical point of $f \circ \phi$ if and only if $\phi(v)$ is a first-order critical point of f.

In this case, the Hessians of f and $f \circ \phi$ at $\phi(v)$ and v are linked by the following relation:

$$\forall x_1, x_2 \in T_v\mathcal{N}_1, \quad \mathrm{Hess}(f \circ \phi)(v) \cdot (x_1, x_2) = \mathrm{Hess} f(\phi(v)) \cdot (d\phi(v) \cdot x_1, d\phi(v) \cdot x_2).$$

As $d\phi(v)$ is surjective, $\mathrm{Hess}(f \circ \phi)(v)$ and $\mathrm{Hess} f(\phi(v))$ have the same rank, and $\mathrm{Hess}(f \circ \phi)(v)$ is positive semidefinite if and only if $\mathrm{Hess} f(\phi(v))$ is, meaning that v is a second-order critical point of $f \circ \phi$ if and only if $\phi(v)$ is a second-order critical point of f. ◻

D.5. Proof of Proposition 5.8. Let us consider for a moment an arbitrary invertible matrix $G \in \mathbb{R}^{n \times n}$, and define

$$\tilde{\mathcal{A}} : X \in \mathbb{S}^{n \times n} \to \mathcal{A}(GXG^T) \in \mathbb{R}^m \quad \text{and} \quad \tilde{b} = b,$$
$$\tilde{X}_0 = G^{-1} X_0 (G^T)^{-1} \quad \text{and} \quad \tilde{V} = G^{-1} V.$$

We denote $\tilde{\mathcal{M}}_p$ the set of feasible points for problem (Factorized SDP) where $\mathcal{A}$ and b have been replaced with $\tilde{\mathcal{A}}$ and $\tilde{b}$:

$$\begin{aligned} \tilde{\mathcal{M}}_p &= \{W \in \mathbb{R}^{n \times p}, \tilde{\mathcal{A}}(WW^T) = \tilde{b}\} \\ &= \{G^{-1}W, W \in \mathcal{M}_p\}. \end{aligned}$$

The pair $(\tilde{\mathcal{A}}, \tilde{b})$ is p-regular (because $(\mathcal{A}, b)$ is) and one can check that $\tilde{\mathcal{M}}_p$ is $\tilde{X}_0$-minimally secant at $\tilde{V}$ (because $\mathcal{M}_p$ is X_0-minimally secant at V).

Now imagine we can construct $\tilde{C}_1, \tilde{g}_1, \tilde{C}_2, \tilde{g}_2$ satisfying properties (Optim. SDP 2) to (Second-order optim.) (with $\tilde{X}_0, \tilde{V}, \tilde{\mathcal{A}}, \tilde{\mathcal{M}}_p$ in place of their non-tilde versions). Then, if we define

$$C_1 = (G^T)^{-1} \tilde{C}_1 G^{-1}, \quad C_2 = (G^T)^{-1} \tilde{C}_2 G^{-1}, \quad g_1 = \tilde{g}_1, \quad g_2 = \tilde{g}_2,$$

we see that these objects satisfy conditions (Optim. SDP 2) to (Second-order optim.) (with the non-tilde versions). Therefore, if we are able to construct $\tilde{C}_1, \tilde{g}_1, \tilde{C}_2, \tilde{g}_2$ satisfying conditions (Optim. SDP 2) to (Second-order optim.), it proves the existence of C_1, g_1, C_2, g_2 satisfying these same conditions.

To conclude, it suffices to show that if we properly define G, then

$$\tilde{X}_0 = \begin{pmatrix} I_r & 0_{r,n-r} \\ 0_{n-r,r} & 0_{n-r,n-r} \end{pmatrix} \quad \text{and} \quad \tilde{V} = \begin{pmatrix} 0_{r,p} \\ I_p \\ 0_{n-p-r,p} \end{pmatrix}. \tag{D.6}$$

Let $U_0 \in \mathbb{R}^{n\times r}$ be such that $X_0 = U_0 U_0^T$ (it exists: X_0 is semidefinite positive and has rank r). We define

$$G = \begin{pmatrix} U_0 & V & W \end{pmatrix} \in \mathbb{R}^{n\times n},$$

where $W \in \mathbb{R}^{n\times(n-r-p)}$ is any matrix that makes G invertible (it exists, as the columns of U_0 and V are linearly independent, from properties 1 and 2 of Definition 2.9). Equation (D.6) holds. ∎

D.6. Proof of Lemma 5.9. We define $L : \mathbb{S}^{n\times n} \to \mathbb{R}^{r\times p} \times \mathbb{S}^{p\times p}$ the linear map such that, for any $R_1, R_2, G_1, G_2, G_4, G_5$,

$$L\left(\begin{pmatrix} G_1 & R_1 & G_2 \\ R_1^T & R_2 & G_4 \\ G_2^T & G_4^T & G_5 \end{pmatrix}\right) = (R_1, R_2).$$

Proving the lemma amounts to showing that $L \circ \mathcal{A}^*$ is surjective. Equivalently, it suffices to show that the dual map $\mathcal{A} \circ L^*$ is injective. Let (R_1, R_2) be in its kernel:

$$\mathcal{A}\left(\begin{pmatrix} 0 & R_1/2 & 0 \\ R_1^T/2 & R_2 & 0 \\ 0 & 0 & 0 \end{pmatrix}\right) = 0.$$

We recall that we have assumed, following Proposition 5.8,

$$X_0 = \begin{pmatrix} I_r & 0_{r,n-r} \\ 0_{n-r,r} & 0_{n-r,n-r} \end{pmatrix} \quad \text{and} \quad V = \begin{pmatrix} 0_{r,p} \\ I_p \\ 0_{n-p-r,p} \end{pmatrix}.$$

Therefore, if we set

$$\dot{V} = \begin{pmatrix} R_1/2 \\ R_2/2 \\ 0_{n-p-r,p} \end{pmatrix},$$

we have $\text{Range}(\dot{V}) \subset \text{Range}(X_0) + \text{Range}(V)$ and $\mathcal{A}(V\dot{V}^T + \dot{V}V^T) = 0$, hence $\dot{V} \in T_V\mathcal{M}_p$. From Property 3 of Definition 2.9, there exists $A \in \text{Anti}(p)$ such that $\dot{V} = VA$. As a consequence, $R_1 = 0_{r,p}$ and R_2 is both symmetric and antisymmetric, hence $R_2 = 0_{p,p}$. This proves that $\text{Ker}(\mathcal{A} \circ L^*) = \{(0_{r,p}, 0_{p,p})\}$, which is what we needed. ∎

D.7. Proof of Proposition 5.10. Properties (Optim. SDP 2), (Optim. SDP 3), (First-order optim. 1), and (First-order optim. 2) are a direct consequence of (5.8) and of the fact that g_1, C_1, C_2 satisfy these same properties.

For property (Strict complementarity), we have $\text{rank}(C_1^{(mod)}) \geq \text{rank}(C_1) = n - \text{rank}(X_0)$, since adding a semidefinite positive matrix to another one cannot decrease the rank. Additionally, as $C_1^{(mod)} X_0 = 0$ (property (Optim. SDP 3)), we also have $\text{rank}(C_1^{(mod)}) \leq n - \text{rank}(X_0)$ and, therefore, $\text{rank}(C_1^{(mod)}) = n - \text{rank}(X_0)$. ∎

D.8. Proof of Proposition 5.11. Equation (5.7) naturally implies (5.9). Let us assume that (5.9) is true and show the converse.

Let $\dot{V}$ be in $T_V\mathcal{M}_p$. We must show that $\langle C_2^{(mod)}\dot{V}\dot{V}^T\rangle \geq 0$, with equality if and only if $\dot{V} = VA$ for some $A \in \mathrm{Anti}(p)$. We write

$$\dot{V} = \dot{W} + VA \text{ for some } \dot{W} \in \mathcal{E}_\perp, A \in \mathrm{Anti}(p).$$

Using at the last line the fact that $C_2^{(mod)}V = 0$ (property (First-order optim. 2)), we see that

$$\begin{aligned}\left\langle C_2^{(mod)}, \dot{V}\dot{V}^T\right\rangle &= \left\langle C_2^{(mod)}, \dot{W}\dot{W}^T\right\rangle + \left\langle C_2^{(mod)}, VA\dot{W}^T\right\rangle \\ &\quad + \left\langle C_2^{(mod)}, \dot{W}A^TV^T\right\rangle + \left\langle C_2^{(mod)}, VAA^TV^T\right\rangle \\ &= \left\langle C_2^{(mod)}, \dot{W}\dot{W}^T\right\rangle + 2\left\langle C_2^{(mod)}V, \dot{W}A^T\right\rangle + \left\langle C_2^{(mod)}V, VAA^T\right\rangle \\ &= \left\langle C_2^{(mod)}, \dot{W}\dot{W}^T\right\rangle.\end{aligned}$$

Therefore, from (5.9), $\langle C_2^{(mod)}\dot{V}\dot{V}^T\rangle \geq 0$, with equality if and only if $\dot{W} = 0$, that is, $\dot{V} \in \{VA, A \in \mathrm{Anti}(p)\}$. ∎

D.9. Proof of Proposition 5.12. First, we observe that for any $\dot{V} \in \mathcal{E}_\perp - \{0\}$,

$$\left\langle \begin{pmatrix} 0_{r+p,r+p} & 0_{r+p,n-r-p} \\ 0_{n-r-p,r+p} & I_{n-r-p}\end{pmatrix}, \dot{V}\dot{V}^T\right\rangle \geq 0$$

because it is the scalar product of two semidefinite positive matrices. It is zero if and only if the last $n-r-p$ rows of $\dot{V}$ are zero, that is,

$$\mathrm{Range}(\dot{V}) \subset \mathrm{Range}(X_0) + \mathrm{Range}(V).$$

Because $\mathcal{M}_p$ is X_0-minimally secant at V, this is possible only if $\dot{V} = VA$ for some $A \in \mathrm{Anti}(p)$, which contradicts the fact that $\dot{V}$ is in $\mathcal{E}_\perp - \{0\}$. Therefore,

$$\left\langle \begin{pmatrix} 0_{r+p,r+p} & 0_{r+p,n-r-p} \\ 0_{n-r-p,r+p} & I_{n-r-p}\end{pmatrix}, \dot{V}\dot{V}^T\right\rangle > 0.$$

We set $\mathcal{B}_\perp = \{\dot{V} \in \mathcal{E}_\perp, ||\dot{V}||_F = 1\}$. From the previous remark and because $\mathcal{B}_\perp$ is compact, there exists $\epsilon > 0$ such that

$$\forall \dot{V} \in \mathcal{B}_\perp, \quad \left\langle \begin{pmatrix} 0_{r+p,r+p} & 0_{r+p,n-r-p} \\ 0_{n-r-p,r+p} & I_{n-r-p}\end{pmatrix}, \dot{V}\dot{V}^T\right\rangle \geq \epsilon.$$

We define $\gamma = \inf_{\dot{V}\in\mathcal{B}_\perp}\langle C_2\dot{V}\dot{V}^T\rangle$. For any t such that $\gamma + t\epsilon > 0$, it holds that

$$\forall \dot{V} \in \mathcal{B}_\perp, \quad \left\langle C_2^{(mod)}, \dot{V}\dot{V}^T\right\rangle \geq \gamma + t\epsilon > 0.$$

In this case, by homogeneity, (5.9) also holds. ∎

Appendix E. Proof of Corollary 3.5. The first part of the corollary is a direct consequence of Theorem 3.2, so we focus on the second one. Let p, n be such that $p(p+1)/2 + p \leq n$. It suffices to check the three hypotheses of Theorem 3.3. The first two are classical, setting

$$X_0 = U_0U_0^T \qquad \text{with} \qquad U_0 = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^{n\times 1}.$$

Let us construct $V \in \mathcal{M}_p$ such that $\mathcal{M}_p$ is X_0-minimally secant at V. In Definition 2.9, the only delicate part is property 3. We do not have a better method than checking it by doing direct computation. Hence, we must choose V as simple as possible, so that the equations defining $T_V\mathcal{M}_p$ and $\{\dot{V} \in \mathbb{R}^{n\times p}, \mathrm{Range}(\dot{V}) \subset \mathrm{Range}(U_0) + \mathrm{Range}(V)\}$ are relatively easy to manipulate. This matrix must satisfy two constraints: it has to be in $\mathcal{M}_p$ (that is, all its rows must have norm 1) and it must have at least $\frac{p(p+1)}{2} + p$ different lines (otherwise, one can check that the aforementioned equations are degenerate).

The simplest matrix V that satisfies these constraints is arguably the following one: for any $i \le p$ and $j \in \{i+1, \dots, p\}$, we respectively set the ith, $(p+i)$th, and $(2p + \phi(i,j))$th lines of V as

$$V_{i,:} = e_i, \quad V_{p+i,:} = -e_i, \quad V_{2p+\phi(i,j),:} = \frac{e_i + e_j}{\sqrt{2}}, \tag{E.1}$$

where $(e_1, \dots, e_p)$ is the canonical basis of $\mathbb{R}^{1\times p}$ and $\phi : \{i, j \text{ s.t. } 1 \le i < j \le p\} \to \{1, \dots, \frac{p(p-1)}{2}\}$ is an arbitrary bijection. For the last $n - (\frac{p(p+1)}{2} + p)$ lines, we choose any unit-normed elements of $\mathbb{R}^{1\times p}$.

This definition ensures that V has rank p (it contains I_p as a submatrix). Moreover, $\begin{pmatrix} U_0 & V \end{pmatrix}$ has rank $p+1$ (its $p+1$ first lines form an invertible matrix). Therefore, properties 1 and 2 of Definition 2.9 hold.

We check property 3. Let $\dot{V} \in T_V\mathcal{M}_p$ be such that

$$\mathrm{Range}(\dot{V}) \subset \mathrm{Range}(U_0) + \mathrm{Range}(V).$$

Then there exists $(R, A) \in \mathbb{R}^{1\times p} \times \mathbb{R}^{p\times p}$ such that $\dot{V} = U_0 R + VA$. We fix such R, A, and show that $R = 0_{1,p}$ and A is antisymmetric.

For any $i = 1, \dots, p$, because $\dot{V}$ is in $T_V\mathcal{M}_p$,

$$\begin{aligned}
&\left(\mathrm{diag}(V\dot{V}^T + \dot{V}V^T)_i = 0 \text{ and } \mathrm{diag}(V\dot{V}^T + \dot{V}V^T)_{p+i} = 0\right)\\
&\overset{\text{(E.1)}}{\iff} \quad \left(\dot{V}_{i,i} = 0 \text{ and } \dot{V}_{p+i,i} = 0\right)\\
&\iff \quad ((U_0R + VA)_{i,i} = 0 \text{ and } (U_0R + VA)_{p+i,i} = 0)\\
&\overset{\text{(E.1)}}{\iff} \quad (R_{1,i} + A_{i,i} = 0 \text{ and } R_{1,i} - A_{i,i} = 0)\\
&\iff \quad (R_{1,i} = 0 \text{ and } A_{i,i} = 0)\,.
\end{aligned}$$

Consequently, $R = 0_{1,p}$ and $\mathrm{diag}(A) = 0$. Similarly, for any $1 \le i < j \le p$,

$$\begin{aligned}
&\left(\mathrm{diag}(V\dot{V}^T + \dot{V}V^T)_{2p+\phi(i,j)} = 0\right)\\
&\overset{\text{(E.1)}}{\iff} \quad \left(\dot{V}_{2p+\phi(i,j),i} + \dot{V}_{2p+\phi(i,j),j} = 0\right)\\
&\iff \quad ((U_0R + VA)_{2p+\phi(i,j),i} + (U_0R + VA)_{2p+\phi(i,j),j} = 0)\\
&\overset{\text{(E.1)}}{\iff} \quad \left(R_{1,i} + \frac{A_{i,i} + A_{i,j}}{\sqrt{2}} + R_{1,j} + \frac{A_{j,i} + A_{j,j}}{\sqrt{2}} = 0\right)\\
&\iff \quad (A_{i,j} + A_{j,i} = 0)\,.
\end{aligned}$$

The matrix A is therefore antisymmetric. ∎

Acknowledgment. We thank the reviewers for their careful reading and suggestions, which allowed us to significantly simplify our results and proofs.

REFERENCES

[1] E. Abbe, A. S. Bandeira, and G. Hall, *Exact recovery in the stochastic block model*, Trans. Inform. Theory, 62 (2016), pp. 471–487.

[2] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, Princeton, NJ, 2009.

[3] A. S. Bandeira, N. Boumal, and V. Voroninski, *On the low-rank approach for semidefinite programs arising in synchronization and community detection*, in Proceedings of the Conference on Computational Learning Theory, 2016.

[4] S. Bhojanapalli, N. Boumal, P. Jain, and P. Netrapalli, *Smoothed analysis for low-rank solutions to semidefinite programs in quadratic penalty form*, in Proceedings of the 31st Conference on Learning Theory, 2018, pp. 3243–3270.

[5] B. Borchers and J. Young, *Implementation of a primal–dual method for SDP on a shared memory parallel architecture*, Comput. Optim. Appl., 37 (2007), pp. 355–369.

[6] N. Boumal, *A Riemannian Low-rank Method for Optimization over Semidefinite Matrices with Block-diagonal Constraints*, Tech. report, http://arxiv.org/abs/1506.00575, 2015.

[7] N. Boumal, P.-A. Absil, and C. Cartis, *Global rates of convergence for nonconvex optimization on manifolds*, IMA J. Numer. Anal., 39 (2019), pp. 1–33.

[8] N. Boumal, V. Voroninski, and A. S. Bandeira, *Deterministic Guarantees for Burer-Monteiro Factorizations of Smooth Semidefinite Programs*, preprint, https://arxiv.org/abs/1804.02008, 2018.

[9] S. Burer and R. D. C. Monteiro, *A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization*, Math. Program., 95 (2003), pp. 329–357.

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

[11] C. Delorme and S. Poljak, *Laplacian eigenvalues and the maximum cut problem*, Math. Program., 62 (1993), pp. 557–574.

[12] R. Ge, J. D. Lee, and T. Ma, *Matrix Completion has no Spurious Local Minimum*, in Advances in Neural Information Processing Systems 29, Curran Associates, 2016.

[13] M. X. Goemans and D. P. Williamson, *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming*, J. ACM, 42 (1995), pp. 1115–1145.

[14] M. Jaggi, *Revisiting Frank-Wolfe: Projection-free sparse convex optimization*, in Proceedings of the 30th International Conference on Machine Learning, 2013, pp. 427–435.

[15] M. Journée, F. Bach, P.-A. Absil, and R. Sepulchre, *Low-rank optimization on the cone of positive semidefinite matrices*, SIAM J. Optim., 20 (2010), pp. 2327–2351.

[16] S. Laue, *A hybrid algorithm for convex semidefinite optimization*, in Proceedings of the 29th International Conference on Machine Learning, 2012, pp. 177–184.

[17] M. Laurent and F. Rendl, *Semidefinite Programming and Integer Programming*, Hand. Oper. Res. Management Sci., 12 (2005), pp. 393–514.

[18] A. Lemon, A. M.-C. So, and Y. Ye, *Low-rank Semidefinite Programming: Theory and Applications*, Now Publishers, 2016.

[19] Q. Li, Z. Zhu, and G. Tang, *The non-convex geometry of low-rank matrix optimization*, Inform. Inference, 8 (2019), pp. 51–96.

[20] Y. Nesterov, *Smooth minimization of non-smooth functions*, Math. Program., 103 (2005), pp. 127–152.

[21] G. Pataki, *On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues*, Math. Oper. Res., 23 (1998), pp. 339–358.

[22] S. Poljak and F. Rendl, *Nonpolyhedral relaxations of graph-bisection problems*, SIAM J. Optim., 5 (1995), pp. 467–487.

[23] T. Pumir, S. Jelassi, and N. Boumal, *Smoothed Analysis of the Low-rank Approach for Smooth Semidefinite Programs*, in Advances in Neural Information Processing Systems, 2018.

[24] D. M. Rosen, L. Carlone, A. S. Bandeira, and J. J. Leonard, *SE-Sync: A certifiably correct algorithm for synchronization over the special Euclidean group*, to appear in Internat. J. Robotics Res., 38 (2019).

[25] J. Sun, Q. Qu, and J. Wright, *A geometric analysis of phase retrieval*, Found. Comput. Math., 18 (2018), pp. 1131–1198.

[26] I. Waldspurger, A. d'Aspremont, and S. Mallat, *Phase recovery, maxcut and complex semidefinite programming*, Math. Program., 149 (2015), pp. 47–81.

[27] I. Waldspurger and A. Waters, *Rank Optimality for the Burer-Monteiro Factorization,* preprint, https://arxiv.org/abs/1812.03046, 2018.

[28] H. Wolkowicz, R. Saigal, and L. Vandenberghe, *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*, Internat. Ser. Oper. Res. Management Sci. 27, Springer, New York, 2012.

[29] A. Yurtsever, M. Udell, J. A. Tropp, and V. Cevher, *Sketchy decisions: Convex low-rank matrix optimization with optimal storage*, in Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, PMLR 54, 2017, pp. 1188–1196.