JMLR: Workshop and Conference Proceedings vol 49:1–22, 2016

On the low-rank approach for semidefinite programs arising in synchronization and community detection

Afonso S. Bandeira BANDEIRA@MIT.EDU
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

Nicolas Boumal NBOUMAL@MATH.PRINCETON.EDU
Mathematics Department, Princeton University, Princeton, New Jersey, USA.

Vladislav Voroninski VVLAD@MATH.MIT.EDU
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

Abstract

To address difficult optimization problems, convex relaxations based on semidefinite programming are now common place in many fields. Although solvable in polynomial time, large semidefinite programs tend to be computationally challenging. Over a decade ago, exploiting the fact that in many applications of interest the desired solutions are low rank, Burer and Monteiro proposed a heuristic to solve such semidefinite programs by restricting the search space to low-rank matrices. The accompanying theory does not explain the extent of the empirical success. We focus on Synchronization and Community Detection problems and provide theoretical guarantees shedding light on the remarkable efficiency of this heuristic.

Keywords: Semidefinite Programming, Burer–Monteiro heuristic, SDPLR, Synchronization, Community Detection

1. Introduction

Estimation problems in many fields, including signal processing, statistics and machine learning, are formulated as intractable optimization problems. A now popular technique to attempt solving some of these problems is to replace the difficult optimization problem by a surrogate tractable convex problem and take advantage of existing machinery for convex optimization. In many instances, the surrogate problem is obtained by considering a larger, convex feasible space, and thus it is often called a *convex relaxation*. Relaxations based on semidefinite programming are among the most popular.

We will focus mostly on a certain class of problems, namely, *synchronization problems* on graphs (Singer (2011); Bandeira et al. (2015)). Synchronization problems consist in estimating n group elements $g_1, \ldots, g_n$ from information about their offsets $g_i g_j^{-1}$. A simple instance of this class is $\mathbb{Z}_2$-Synchronization (Abbe et al. (2014)), corresponding to the group on two elements. In that case, the task is to estimate n bits $x_1, \ldots, x_n \in \{\pm 1\}$ from noisy measurements of $x_i x_j$.

The problem of Community Detection in the Binary Stochastic Block Model (SBM) can also be regarded as an instance of Synchronization over $\mathbb{Z}_2$. SBM on two communities (also known as *planted partition*) is a model of a random graph $\mathcal{G}(n, p, q)$ on $n = 2m$ nodes split evenly in two communities, identified by $\{\pm 1\}$ node labels. Edges for this graph are drawn randomly and independently, where each pair of nodes in the same community gets connected with probability p and in different communities with probability q. The task is to estimate the partition, given an observation of the graph. This fascinating problem was the target of much study in the last few years, including identification of remarkable phase transitions on the values of p and q and possibility of recovering the unknown labels (Decelle et al. (2011); Mossel et al. (2014a,b); Massoulié (2014); Abbe et al. (2016); Mossel et al. (2014c)).

While computing the maximum likelihood estimator (MLE) for either of the problems above is known to be computationally intractable, several heuristics have been proposed and studied. A particularly successful approach is based on Semidefinite Programming. Following the seminal work of Goemans and Williamson (1995) in the context of the Max-Cut problem, the maximum likelihood estimation problem (originally an intractable optimization problem in n node variables) is relaxed to a semidefinite program (SDP) (a tractable problem on n^2 variables). Importantly, the solution to the SDP is only guaranteed to correspond to the solution of the original problem if it has rank 1. Enforcing this constraint explicitly would render the problem intractable. Remarkably, in some regimes, it is known that the solution of this semidefinite program is naturally of rank 1, and that furthermore this solution allows to identify the true labels (Abbe et al. (2014); Bandeira et al. (2014b); Abbe et al. (2016); Bandeira (2015b); Hajek et al. (2014)).

While SDP's are known to be solvable up to arbitrary precision in polynomial time (Nesterov (2004)), the increase in dimension due to the lifting technique (the relaxation) and the semidefiniteness constraint render solving them rather slow in practice. This is mainly because intermediate iterates involve matrix decompositions of dense, full-rank matrices of size n. On the other hand, in certain regimes, the optimal solution is expected to have low rank. Indeed, in the problems studied here, the solution being rank 1 coincides precisely with it corresponding to the desirable MLE. It is then natural to attempt to reach the solution via a sequence of similarly simple objects instead.

The most popular and successful such low-rank approach is SDPLR, as proposed in (Burer and Monteiro, 2003, 2005) (also in a particular form in (Burer et al., 2002)). In a nutshell, the idea is to restrict the search space to matrices of bounded rank. While, after adding this rank constraint, the optimization problem is no longer convex (and it is hence unclear whether it is tractable or not), this approach is empirically successful for a variety of instances (Burer and Monteiro, 2003; Journée et al., 2010; Bandeira et al., 2014b; Boumal, 2015).

The original SDPLR papers come with general theory supporting the fact that relaxing the rank up to about $\sqrt{2n}$ might work well. Yet, in practice, for well-behaved SDP's which admit a solution of rank 1, it is often seen that relaxing the rank merely to 2 works fantastically. To date, there was no satisfactory explanation for this nonconvex success. This paper provides the first such guarantee, in the context of synchronization over $\mathbb{Z}_2$ and community detection. More precisely, we show that, in certain noise regimes, there are no spurious second-order critical points for the rank-2 constrained problem, while in other (more inclusive) regimes we show that all such points need to correlate non-trivially with the ground truth. Second-order critical points (that is, points which satisfy second-order necessary optimality conditions) can be computed in our context (Boumal et al., 2016).

In this paper, we focus on a selection of well-studied problems, keeping in mind that the proposed analysis has the potential to generalize to many problems where semidefinite relaxations are successful yet demanding to solve. A more general setting will be the focus of a future publication.

It is also worth noting that semidefinite relaxations have been observed to work well on problems for which other standard heuristics seem to not perform well. One relevant example is the Multireference Alignment problem (Bandeira et al. (2014a)). Furthermore, low-rank semidefinite programming has the potential to yield a posteriori certifiably correct algorithms (Bandeira (2015a)) and is known to be robust to certain monotone adversary models (Feige and Kilian (2001); Moitra et al. (2015)).

Notation

Given a matrix M and its vector of singular values ν, we write $\|M\| = \|\nu\|_\infty$ for its operator norm (largest singular value), $\|M\|_F = \|\nu\|_2$ for its Frobenius norm and $\|M\|_* = \|\nu\|_1$ for its nuclear norm (sum of singular values). The operator $\operatorname{ddiag}\colon \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ sets all off-diagonal entries of a matrix to zero. The Hadamard or entry-wise product is written $\circ$.

2. The $\mathbb{Z}_2$ Synchronization problem

The goal of $\mathbb{Z}_2$ Synchronization is to estimate labels $z \in \{\pm 1\}^n$ from noisy pairwise measurements $Y_{ij} = z_i z_j + \sigma W_{ij}$, where $W_{i>j} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$ and $W_{ij} = W_{ji}$, $W_{ii} = 0$. In matrix notation,

$$Y = zz^T + \sigma W. \tag{1}$$

Because only zz^T is measured, the labels can only be estimated up to a global sign flip. One can also consider more realistic noise models on which Y_{ij} also takes binary values (Abbe et al. (2014)) but, for the sake of simplicity, we will consider Gaussian noise. Since $\|zz^T\| = n$ and $\|\sigma W\|$ concentrates around a constant multiple of $\sigma\sqrt{n}$ (in operator norm), a natural way to parametrize the signal-to-noise ratio is by taking $\lambda = \frac{\sqrt{n}}{\sigma}$. For this reason, some authors consider the scaled model

$$\mathcal{Y} = \frac{\lambda}{n} Y = \frac{\lambda}{n} zz^T + \frac{1}{\sqrt{n}} W. \tag{2}$$

The problems of recovering z from Y or $\mathcal{Y}$ are clearly equivalent. The former choice of notation has been used, for example, in (Bandeira et al., 2014b; Bandeira, 2015b) and has the advantage of highlighting the fact that the observation is an entry-wise noisy version of the ground truth, while the latter has been used in (Javanmard et al., 2015) and has the advantage of highlighting the spike model interpretation of the problem and making a more transparent connection with the well-studied spike model in random matrix theory and the BBP transition phenomenon (Baik et al., 2005; Féral and Péché, 2006) (this connection had already been made in (Singer, 2011)). For the sake of completeness, we will state our results in both notation choices.

The most natural approach to recovering z is to consider the MLE, which solves

$$\min_{x \in \{\pm 1\}^n} \sum_{i,j=1}^{n} \left(Y_{ij} - x_i x_j\right)^2. \tag{3}$$

It is readily seen that solutions of this problem are the same as those of

$$\max_{x \in \{\pm 1\}^n} x^T Y x, \tag{4}$$

which is known to the *NP-hard* in general. In fact, when $Y \succeq 0$ this is known as the little Grothendieck problem (Briët et al., 2014; Bandeira et al., 2013), and, when Y corresponds to the Laplacian of a graph, it corresponds to the Max-Cut problem (Goemans and Williamson, 1995).

Following the now standard lifting technique (dating back at least to Goemans and Williamson (1995)), we set a new variable $X = xx^T$ and write the equivalent formulation

$$\begin{array}{ll} \max_X & \operatorname{Tr}(YX) \\ \text{s.t.} & X_{ii} = 1, \text{ for } 1 \le i \le n \\ & X \succeq 0 \\ & \operatorname{rank}(X) = 1. \end{array} \tag{5}$$

Problems (4) and (5) are equivalent. One arrives at a tractable SDP formulation by dropping the problematic rank constraint:

$$\begin{array}{ll} \max & \operatorname{Tr}(YX) \\ \text{s.t.} & X_{ii} = 1, \text{ for } 1 \le i \le n \\ & X \succeq 0. \end{array} \tag{6}$$

By construction, problem (6) has the same solutions if Y is replaced with $\mathcal{Y}$ (but it will have a different optimal value, since Y and $\mathcal{Y}$ are scaled versions of each other.)

Recall the model (1) (or equivalently (2)). It is known that, as long as $\sigma < \sqrt{\frac{n}{2\log n}}$, or equivalently $\lambda > \sqrt{2\log n}$, with high probability, the solution X of (6) is unique and corresponds exactly to the ground truth $X = zz^T$. We refer to this phenomenon as *exact recovery*. It is also known that on the other side of this threshold exact recovery is information-theoretically impossible, showcasing the efficiency of semidefinite relaxations for this type of task (Bandeira et al., 2014b; Bandeira, 2015b).

While exact recovery is impossible for constant λ (or, equivalently, $\sigma \sim \sqrt{n}$), simply taking the top eigenvector of $\mathcal{Y}$ is known to produce an estimator that correlates non-trivially with the ground truth, precisely when $\lambda > 1$ (Féral and Péché (2006); Singer (2011)). This phenomenon is often referred to as the BBP transition (Baik et al. (2005)). This motivates the question of whether (6) gives meaningful results for the constant λ regime (Montanari and Sen (2015); Javanmard et al. (2015); Guedon and Vershynin (2014)). In the context of community detection, this question was first addressed by Guedon and Vershynin (2014). In (Montanari and Sen, 2015), a phase transition is shown to exist at $\lambda = 1$: similarly to what happens with the top eigenvalue of $\mathcal{Y}$ (Féral and Péché (2006)). For $\lambda < 1$, the value of (6) with cost $\mathcal{Y}$ converges (in probability) to 2, and for $\lambda > 1$ its limit is strictly larger than 2. Javanmard et al. (2015) give fascinating predictions, based on non-rigorous statistical mechanics tools, for the behavior of both (6) and (4) as a function of λ.

Another, related, type of synchronization problem is *phase synchronization*, where one estimates $z \in \mathbb{C}^n$ with $|z_i| = 1$ for all i. This is a direct complex analogue of $\mathbb{Z}_2$ synchronization. The SDP relaxation works similarly in the complex field. See (Singer, 2011) for a first analysis of the SDP relaxation, see (Bandeira et al., 2014b) for a proof of tightness of the SDP relaxation in a similar regime as here and, still in the same regime, see (Boumal, 2016) for a proof that the nonconvex problem has no spurious local optima either. A key difference with the present paper is that phase synchronization is a continuous problem, whereas $\mathbb{Z}_2$ synchronization is discrete. This explains why, in the present setting, the constraint $\operatorname{rank}(X) = 1$ has to be relaxed at least to $\operatorname{rank}(X) = 2$ (to connect the search space), whereas in the former paper, the rank is not relaxed at all.

2.1. The Burer–Monteiro Approach

In transiting from (5) to (6), one effectively replaces the constraint $\operatorname{rank}(X) \le 1$ by the vacuous constraint $\operatorname{rank}(X) \le n$, thus going from a combinatorial problem to a tractable but highdimensional SDP. One of the insights of Burer and Monteiro (2003, 2005) is that relaxing the rank only partially, as $\mathrm{rank}(X) \leq p$ for variable p, gives access to a family of low-dimensional (but nonconvex) nonlinear optimization problems, which can be put to good use to understand both (5) and (6).

In general, given an SDP of the form

$$\max_X \mathrm{Tr}(YX) \quad \text{s.t.} \quad \mathcal{A}(X) = b, X \succeq 0, \tag{7}$$

where $\mathcal{A}$ is a linear operator from the symmetric matrices of size n to $\mathbb{R}^m$, the SDPLR algorithm (sometimes referred to as the Burer–Monteiro approach) consists in parameterizing X as $X = QQ^T$, where Q lives in $\mathbb{R}^{n\times p}$. In so doing, $X \succeq 0$ is naturally enforced. This yields the following nonlinear optimization problem:

$$\max_{Q\in\mathbb{R}^{n\times p}} \mathrm{Tr}(Q^TYQ) \quad \text{s.t.} \quad \mathcal{A}(QQ^T) = b, \tag{8}$$

where both the cost and the constraints are quadratic in Q. This is typically nonconvex. Both problems are equivalent, up to the additional constraint $\mathrm{rank}(X) \leq p$ forced in the nonlinear program. Of course, if the SDP admits a solution of rank at most p, then the two problems attain the same optimal value. When the search space is compact, this is known to happen as soon as $p(p+1)/2 \geq m$ (m is the number of constraints), as per general results put forth in (Shapiro, 1982; Barvinok, 1995; Pataki, 1998). Burer and Monteiro (2005) show that rank deficient local optima of the nonlinear program are globally optimal.[1] See also Lemma 10.

The latter observations motivate the algorithm SDPLR (Burer and Monteiro, 2003), where the nonlinear program is tackled in lieu of the SDP, using classical nonlinear optimization algorithms such as the augmented Lagrangian method. The above discussion suggests setting $p \approx \sqrt{2m}$. If the local method converges to a rank-deficient local optimum (which is often the case in practice but is not satisfactorily explained in theory), then we have found a global optimum. The advantage (compared to the SDP) is that the search space has lower dimension.

In practice, if we are optimistic about our chances, we can set p as low as $p = 2$ for example (Burer et al. (2002)). If the SDP has a solution of rank 1, then the corresponding nonlinear program has local optima (actually, global optima) of rank 1 as well, thus being rank deficient, which allows certifying their optimality. In practice, this is seen to work remarkably well for the problems considered in this paper (in certain noise regimes), as had already been observed in other papers as well. It is not obvious why this is so, as the rank restriction could (in general) spawn a large number of spurious local optima.

Here, for the first time, we provide a proof that, at $p = 2$ and in favorable regimes, all local optima are global optima, which explains why local methods behave appropriately. In fact, we even show that saddle points can always be escaped using solely second-order derivatives.

Local algorithms such as the augmented Lagrangian method are useful to tackle generic constrained nonlinear programs. The SDP's under scrutiny, though, have rather special structure. In particular, they are such that the search space in Q is a smooth manifold already for $p \geq 2$. As originally advocated by Journée et al. (2010), this suggests solving the nonlinear program using techniques from optimization on manifolds (Absil et al., 2008). The latter are indeed better suited to fully leverage the special geometry of these problems. See also (Wen and Yin, 2013; Boumal,

[1] This renders these methods potentially a posteriori certifiable (Bandeira (2015a)).

2015) for further numerical and theoretical investigation. See (Boumal et al., 2016) for a Riemannian version of the trust-region method which globally converges to global optima owing to the special properties of our problem, with global rates of convergence.

As described above, here we consider the rank constrained problem as follows (Burer et al. (2002)):

$$\begin{array}{ll} \max & \operatorname{Tr}(YX) \\ \text{s.t.} & X_{ii} = 1, \text{ for } 1 \le i \le n \\ & X \succeq 0 \\ & \operatorname{rank}(X) \le 2. \end{array} \tag{9}$$

By taking $X = QQ^T$ with $Q \in \mathbb{R}^{n\times 2}$, one can reformulate (9) as:

$$\begin{array}{ll} \max & \operatorname{Tr}\left(Q^T Y Q\right) \\ & \text{s.t.} \, \|Q_{i:}\|^2 = 1, \text{ for } 1 \le i \le n \\ & Q \in \mathbb{R}^{n\times 2}, \end{array} \tag{10}$$

where $Q_{i:}$ denotes the ith row of Q. Note the geometry here: the search space is a product of circles.

We now present our main results in the realm of $\mathbb{Z}_2$ Synchronization.

Theorem 1 *Consider model (2) (or equivalently (1)). If $\lambda > 8$ (or equivalently $\sigma < \frac{1}{8}\sqrt{n}$), then, with high probability, all second-order critical points Q of (10) correlate non-trivially with the ground truth z in the sense that, for every such λ, there exists ε such that*

$$\frac{1}{n}\left\|Q^T z\right\|_2 \ge \varepsilon. \tag{11}$$

Proof All the interesting ingredients of the proof are in Section 3. The claim will follow from Lemma 8, noting that $\left\|Q^T z\right\|_2^2 = \left\langle QQ^T, zz^T\right\rangle$, the behavior of (10) is unchanged by changing the diagonal values of $\mathcal{Y}$, and the fact that for any $\delta > 0$, with high probabily, $\frac{1}{n}\mathrm{SDP}(W - \operatorname{ddiag}(W)) \le \|W - \operatorname{ddiag}(W)\| \le (2+\delta)\sqrt{n}$ (see Lemma 16 and eq. (21)). ■

Remark 2 (Rounding) *In Theorem (1) we ask for non-trivial correlation in the form of (22), but it would also be natural to ask for an estimator $\hat{z} \in \mathbb{R}^n$ (or even $\hat{z} \in \{\pm 1\}^n$) that exhibits non trivial correlation with z. We note that if we take x and y to be the columns of Q satisfying (22), then a random linear combination $g_1 x + g_2 y$, where g_1 and g_2 are independent standard Gaussian variables, can be shown to be likely to produce meaningful correlation estimators. Indeed, $\mathbb{E}\{\|Qg\|^2\} = \|Q\|_F^2 = n$ and $\mathbb{E}\{\langle z, Qg\rangle^2\} = \|Q^T z\|_2^2$. We also direct the reader to Section 4 of (Montanari and Sen, 2015) for techniques to construct meaningful $\{\pm 1, 0\}^n$ estimators from solutions of the SDP. We note furthermore that if ε is large in (22) (as will be the case in Theorem 3, for example), then constructing such estimators becomes considerably easier.*

Theorem 3 *Consider model (2) (or equivalently (1)). If $\lambda > 16$ (or equivalently $\sigma < \frac{1}{16}\sqrt{n}$), then for any $\varepsilon > 0$, with high probability, all second-order critical points Q of (10) satisfy*

$$\frac{1}{n}\left\|Q^T z\right\|_2 \ge 1 - (1+\varepsilon)\frac{16}{\lambda} = 1 - (1+\varepsilon)\frac{16\sigma}{\sqrt{n}}. \tag{12}$$

Proof Again, all the interesting ingredients of the proof are in Section 3, as this will follow directly from Lemma 9 and the considerations in the proof of Theorem 1. ■

Theorem 4 *Consider model (2) (or equivalently (1)). There is a universal constant C such that: If $\lambda \geq Cn^{\frac{1}{3}}$ (or equivalently $\sigma \leq \frac{1}{C}n^{\frac{1}{6}}$) then, with high probability, all second-order critical points Q of (10) are optimal and correspond to the ground truth, meaning $QQ^T = zz^T$.*

Proof This follows from Theorem 15 and the high probability bounds in Lemma 16. ■

Note that there are no guarantees in general that nonlinear optimization algorithms (such as the augmented Lagrangian method) converge to local optima. Yet, we know that only local optima are stable fixed points for such methods; and that this is true regardless of initialization. Hence, one perspective is that the contribution of this paper is to show that all stable fixed points of local methods—without the need for a good initial guess—are global optima (in the given regime). As we mentioned earlier though, because in our case second-order necessary optimality conditions are sufficient for global optimality, results in (Boumal et al., 2016) show that the Riemannian trust-region method converges to global optimizers, regardless of initialization, with global rates of convergence.

2.2. Community Detection in the Binary Stochastic Block Model

The Stochastic Block Model (SBM) on two communities is a random graph model $\mathcal{G}(n, p, q)$ on $n = 2m$ nodes. The nodes are divided evenly in two communities, identified by a vector of labels $g \in \{\pm 1\}^n$ satisfying $g^T \mathbf{1} = 0$. Edges on this graph are drawn independently at random. A pair of nodes in the same community is connected by an edge with probability p; nodes in different communities with probability q. We focus on the case $p > q$. This means that the adjacency matrix A of this graph has the following distribution:

$$A_{ij} = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p, \end{cases}$$

if $g_i g_j = 1$ (thus including the diagonal), and

$$A_{ij} = \begin{cases} 1 & \text{with probability } q \\ 0 & \text{with probability } 1-q, \end{cases}$$

if $g_i g_j = -1$.

The goal of community detection is to recover the labels g from a realization $G \sim \mathcal{G}(n, p, q)$. Roughly half a decade ago, Decelle et al. (2011) conjectured a remarkable phase transition in the constant average–degree regime: if $p = \frac{a}{n}$ and $q = \frac{b}{n}$ with $a > b$ constants, Decelle et al. (2011) conjectured that as long as

$$\lambda(a, b) := \frac{a-b}{\sqrt{2(a+b)}} > 1, \tag{13}$$

it is possible to construct an estimator that, with high probability, correlates with the true partition, whereas below this threshold that would be impossible. This conjecture was proven in a series of papers (Mossel et al., 2014a,b; Massoulié, 2014).

Another natural question is to understand when it is possible to exactly recover g (or $-g$). A similar phase transition was established in (Abbe et al., 2016) and (Mossel et al., 2014c): if $p = \alpha\frac{\log n}{n}$ and $q = \beta\frac{\log n}{n}$ with $\alpha > \beta$ constants, then as long as

$$\sqrt{\alpha} - \sqrt{\beta} > \sqrt{2}, \tag{14}$$

the MLE for the partition, solution of

$$\begin{array}{ll} \max & x^T A x \\ \text{s.t.} & x \in \{\pm 1\}^n \\ & x^T \mathbf{1} = 0, \end{array} \tag{15}$$

coincides exactly with the partition, with high probability; and, moreover, below this threshold it is information-theoretically impossible to recover the partition.

Abbe et al. (2016) studied a semidefinite relaxation of (15) analogous to (6), and conjectured that its solution coincides with the ground truth (that is, the SDP achieves exact recovery) at the information-theoretical limit (14). This was confirmed independently in (Hajek et al., 2014) and (Bandeira, 2015b). The performance of the semidefinite relaxation on the constant average–degree regime has also been target of recent work (Guedon and Vershynin (2014); Montanari and Sen (2015); Javanmard et al. (2015)). Guedon and Vershynin (2014) showed that when the left hand side of (13) is a sufficiently large constant, the solution of the SDP correlates nontrivially with the partition. This was improved by Montanari and Sen (2015) who showed that, for every $\epsilon > 0$, and large enough average degree, $\lambda(a, b) > 1 + \epsilon$ suffices. Javanmard et al. (2015) give precise predictions for the behavior of the SDP in this regime, albeit using non-rigorous techniques. Interestingly, the results in (Moitra et al., 2015) suggest that there may be values of a and b satisfying (13) for which the SDP does not, with high probability, provide meaningful solutions.

We will now write A in a form close to (2) to help illustrate how community detection in the SBM is closely related to $\mathbb{Z}_2$ Synchronization. We start by noting that

$$\mathbb{E}A = \frac{p+q}{2}\mathbf{1}\mathbf{1}^T + \frac{p-q}{2}gg^T.$$

Since $\mathbf{1}$ is a fixed vector, we can consider $A^\natural$ defined as

$$A^\natural = A - \frac{p+q}{2}\mathbf{1}\mathbf{1}^T.$$

It is readily seen that, because of the balanced partition constraint ($x^T\mathbf{1} = 0$), replacing A by $A^\natural$ does not affect the solution of the MLE (15). The benefit is that now, since $A^\natural$ no longer has a bias in the direction of $\mathbf{1}\mathbf{1}^T$, we will remove the extra constraint and focus on trying to understand the efficiency of the simpler program:

$$\begin{array}{ll} \max & x^T A^\natural x \\ \text{s.t.} & x \in \{\pm 1\}^n, \end{array} \tag{16}$$

and its natural SDP relaxation

$$\begin{array}{ll} \max & \operatorname{Tr}\left(A^\natural X\right) \\ \text{s.t.} & X_{ii} = 1, \text{ for } 1 \le i \le n \\ & X \succeq 0. \end{array} \tag{17}$$

Following the Burer–Monteiro approach, we consider the natural analogue to (10):

$$\begin{array}{ll} \max & \operatorname{Tr}\left(Q^T A^{\natural} Q\right) \\ & \text{s.t. } \|Q_{i:}\|^2 = 1, \text{ for } 1 \le i \le n \\ & Q \in \mathbb{R}^{n\times 2}. \end{array} \tag{18}$$

We further write

$$A^{\natural} = \frac{p-q}{2} g g^T + \frac{(p-q)n}{2} E + \frac{(p-q)n}{2} D, \tag{19}$$

where D is a diaognal matrix, and E is a zero-diagonal centered random matrix (we choose to normalize by the average degree $\frac{(p-q)n}{2}$ to be compatible with the notation in (Montanari and Sen, 2015), whose statements about E we will use). We note also that the diagonal matrix D does not affect the solutions of (15), (17), or (18). This means in particular that, in the constant average–degree regime ($p = \frac{a}{n}$ and $q = \frac{b}{n}$),

$$\sqrt{\frac{2}{a+b}} A^{\natural} - D = \frac{\lambda(a,b)}{n} g g^T + E,$$

for $\lambda(a,b)$ defined in (13). This illustrates the parallelism with (2). The main difficulty with this setting is that the spectral norm of E is known not to be bounded (as $n \to \infty$), with high probability (similarly to how adjacency matrices of constant average–degree Erdős-Rényi graphs typically have a fluctuation around their mean whose spectral norm is not bounded, see for example (Krivelevich and Sudakov, 2003; Montanari and Sen, 2015)). For this reason, we will focus on another quantity.

Following the notation in (Montanari and Sen, 2015), given a matrix M we define $\operatorname{SDP}(M)$ as

$$\operatorname{SDP}(M) = \begin{array}{ll} \max & \operatorname{Tr}(MX) \\ \text{s.t.} & X_{ii} = 1, \text{ for } 1 \le i \le n \\ & X \succeq 0, \end{array} \tag{20}$$

and note that, since in the search space $\|X\|_* = n$, we have

$$\operatorname{SDP}(M) \le n\|M\|. \tag{21}$$

(Use Hölder's inequality: $\langle M, X\rangle \le \|M\|\|X\|_*$ and $\|X\|_* = \operatorname{Tr}(X) = n$ since $X \succeq 0$.)

Fortunately, Lemma 7, which is crucial to establish non-trivial correlation guarantees, does not depend on $\|E\|$, which is not bounded, but rather depends only on $\frac{1}{n}\operatorname{SDP}(E)$, which is known to be bounded (Montanari and Sen, 2015) (see Lemma 17 for details). Indeed, using Lemmas 7 and 17 we immediately get the following guarantee.

Theorem 5 *Consider the Stochastic Block Model on two communities described above in the constant average–degree regime $p = \frac{a}{n}$ and $q = \frac{b}{n}$. Let d denote the average degree parameter*

$$d = \frac{a+b}{2}.$$

Then, for any $\delta > 0$ there exists d_0 such that if $d > d_0$ and $\lambda(a,b) > 8 + \delta$ (see (13)) then there exists $\varepsilon > 0$ such that: with high probability, all second-order critical points Q of (18) correlate non-trivially with the ground truth partition g in the following sense.

$$\frac{1}{n}\left\|Q^T g\right\|_2 \ge \varepsilon. \tag{22}$$

Proof The proof is analogous to Theorem 1, but based on the bound on $\mathrm{SDP(E)}$ in Lemma 17. ■

Further controlling $\|E\|$ and $\|Eg\|_\infty$ (Lemmas 18 and 19) and using Lemma 15 we can also obtain a (suboptimal) exact recovery guarantee. It is useful to define the parameter (analogous to (13)):

$$\tilde{\lambda}(p,q) := \frac{p-q}{\sqrt{2(p+q)}}\sqrt{n}.$$

Note that if $p = \frac{a}{n}$ and $q = \frac{b}{n}$ then $\tilde{\lambda}(p,q) = \lambda(a,b)$.

Theorem 6 *Consider the Stochastic Block Model on two communities described above. There exists a universal constant c such that, as long as*

$$\tilde{\lambda}(p,q) \geq cn^{1/3},$$

then, with high probability, all second-order critical points Q of (18) correspond to the ground truth partition, meaning $QQ^T = gg^T$.

Proof If $\frac{p-q}{\sqrt{2(p+q)}}\sqrt{n} \geq cn^{1/3}$, then $\sqrt{\frac{n}{2}(p+q)} \geq cn^{1/3}$ (use $p+q \geq p-q$). A fortiori, this implies that $\sqrt{\frac{n}{2}(p+q)} \gg \log n$. Together with Lemmas 18 and 19 this shows that $\|E\|$ and $\|Eg\|_\infty$ satisfy, up to constants, the same high probability bounds as obtained through Lemma 16 for $\frac{1}{\sqrt{n}}(W - \mathrm{ddiag}(W))$ and so the proof of Theorem 4 applies. ■

3. Proof of the main results

This section contains the main technical content of the paper. In particular, it provides guarantees for the Burer–Monteiro approach based solely on deterministic properties of the matrices involved. In this whole section, the cost matrix which appears in (9) and (10) is denoted by A, to mark that the analysis applies both the $\mathbb{Z}_2$-synchronization and to community detection, where the cost matrices are denoted respectively by Y and $A^\sharp$.

3.1. Partial Recovery

Our first goal is to characterize the local optima of problem (9). In particular, we mean to show that they necessarily reveal a lot of information about the ground truth (the planted solution), under some conditions on the noise. To this end, we start by deriving the first- and second-order necessary optimality conditions of (10).

Lemma 7 *If Q is a local maximum of (10) with cost matrix A, it satisfies first-order necessary optimality conditions,*

$$\left(\mathrm{ddiag}(AQQ^T) - A\right)Q = 0, \tag{23}$$

and second-order necessary optimality conditions,

$$\mathrm{ddiag}(AQQ^T) - A \circ (QQ^T) \succeq 0, \tag{24}$$

where $\mathrm{ddiag}\colon \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ *sets all off-diagonal entries to zero. Letting* $Q = [x\ y]$*, condition* (23) *also implies*

$$Ax \circ y = Ay \circ x. \tag{25}$$

(In fact, they are equivalent.)

Proof Problem (10) is a smooth optimization problem on a smooth manifold. The necessary optimality conditions correspond respectively to requiring the (projected) gradient of the cost function to vanish and the (projected) Hessian of the cost function to have appropriate curvature, see (Absil et al., 2008, eqs.(3.37,5.15)). A derivation of exactly these conditions is done in full in (Journée et al., 2010; Boumal, 2015), among others.

We now prove (23) implies (25). Note that

$$(AQQ^T)_{ii} = ([Ax\ Ay]Q^T)_{ii} = (Ax)_i x_i + (Ay)_i y_i.$$

Hence, for all $1 \leq i \leq n$, considering the first and second columns of condition (23) separately, we get

$$\begin{aligned} ((Ax)_i x_i + (Ay)_i) x_i &= (Ax)_i, \text{ and} \\ ((Ax)_i x_i + (Ay)_i) y_i &= (Ay)_i. \end{aligned}$$

Multiply the first equations by y_i and the second equations by x_i, then subtract. It follows for all i:

$$0 = (Ax)_i y_i - (Ay)_i x_i,$$

which can be compactly written as $Ax \circ y = Ay \circ x$. ■

Lemma 8 *Let* $A = zz^T + \sigma\Delta$ *for some planted signal* $z \in \{\pm 1\}^n$*, where* $\Delta = \Delta^T$*,* $\frac{1}{n}SDP(\Delta) \leq \gamma\sqrt{n}$ *and* $\sigma = c\sqrt{n}$ *for some* $\gamma, c \geq 0$*. Then, for any* Q *satisfying the second-order condition* (24) *and corresponding* $X = QQ^T$*, we have*

$$1 \geq \frac{\langle zz^T, X\rangle}{n^2} \geq \frac{1}{2} - 2\gamma c.$$

This is true in particular for all local optima, thus showing nontrivial correlation of all of those with the planted solution as soon as $\gamma c < \frac{1}{4}$*. Note also that* $\frac{1}{n}SDP(\Delta) \leq \|\Delta\|$ *(see* (21)*).*

Proof For any two positive semidefinite matrices X, Y, it holds that $\langle X, Y\rangle \geq 0$ and that $X \circ Y \succeq 0$ (Schur's product theorem). In particular, taking the inner product of the second-order optimality condition (24) with $X \circ (zz^T) \succeq 0$ on both sides yields the inequality

$$\left\langle \mathrm{ddiag}(AX), X \circ (zz^T) \right\rangle \geq \left\langle A \circ X, X \circ (zz^T) \right\rangle.$$

Since $\mathrm{diag}(X \circ (zz^T)) = \mathbf{1}$, the left hand side evaluates to $\mathrm{Tr}(AX)$, so

$$\langle A, X\rangle \geq \left\langle A, X \circ X \circ (zz^T) \right\rangle.$$

Using $A = zz^T + \sigma\Delta$ gives

$$\left\langle zz^T, X\right\rangle + \sigma\left\langle \Delta, X\right\rangle \geq \left\langle (zz^T)\circ(zz^T), X\circ X\right\rangle + \sigma\left\langle \Delta, X\circ X\circ(zz^T)\right\rangle.$$

Notice that

$$\left\langle (zz^T)\circ(zz^T), X\circ X\right\rangle = \left\langle \mathbf{1}\mathbf{1}^T, X\circ X\right\rangle = \langle X, X\rangle = \|X\|_F^2.$$

Furthermore, both X and $X\circ X\circ zz^T$ are feasible for (6). Thus, by definition (20) and property (21),

$$\langle zz^T, X\rangle \geq \|X\|_F^2 - 2\sigma\,\mathrm{SDP}(\Delta) \geq \|X\|_F^2 - 2\gamma c n^2. \tag{26}$$

Since $X \succeq 0$, $\mathrm{rank}(X) = 2$ and $\mathrm{Tr}(X) = n$, it holds that $\frac{n^2}{2} \leq \|X\|_F^2 \leq n^2$, so that

$$\langle zz^T, X\rangle \geq n^2(1/2 - 2\gamma c),$$

concluding the proof. ■

We note that, in the above lemma, if Q satisfies second-order conditions only approximately so that $\mathrm{ddiag}(AX) - A\circ C \succeq -\epsilon I_n$ with $X = QQ^\top$, then the proof still yields $\frac{\langle zz^\top, X\rangle}{n^2} \geq \frac{1}{2} - 2\gamma c - \frac{\epsilon}{n}$.

Lemma 9 *(Continued from Lemma 8.) If furthermore $\|\Delta\| \leq \gamma\sqrt{n}$ and Q also satisfies the first-order condition (23), then*

$$1 \geq \frac{\|Q^T z\|_2}{n} \geq 1 - 8\gamma c,$$

which, in particular, establishes arbitrarily strong correlation between the planted solution z and any local maximum Q, as long as γc is sufficiently small.

Proof To prove this stronger claim, we need to show that $\|X\|_F$ is arbitrarily close to n for sufficiently small γc (previously, we only had $\|X\|_F \geq n/\sqrt{2}$). Notice that the closer X is to a rank 1 matrix, the closer its Frobenius norm is to n, which is in line with our endeavor. In order to get there, we will improve on the previous lemma by incorporating the first-order optimality conditions (which we did not use in the previous lemma).

For ease of notation, we state the proof for $z = \mathbf{1}$. For the general case, apply the change of variable $A \mapsto \mathrm{diag}(z)A\mathrm{diag}(z)$, which does not require changing the assumptions about Δ.

The conditions on Q are invariant under right-orthogonal transformation. That is, we may freely replace Q by QR, where R is an arbitrary 2×2 orthogonal matrix. This allows to assume, without loss of generality, that $Q = [x\ y]$ with $\langle x, y\rangle = 0$ (simply take the thin SVD of $Q = U\Sigma V^T$ and pick $R = V$). Expand the first-order condition (25) to get

$$((\mathbf{1}\mathbf{1}^T + \sigma\Delta)x)\circ y = ((\mathbf{1}\mathbf{1}^T + \sigma\Delta)y)\circ x.$$

Thus,

$$\langle \mathbf{1}, x\rangle y - \langle \mathbf{1}, y\rangle x = \sigma(\Delta y)\circ x - \sigma(\Delta x)\circ y.$$

By taking the squared norm of both sides of the previous equation and using $\langle x, y\rangle = 0$, we have

$$\begin{aligned}
&\langle \mathbf{1}, x\rangle^2\|y\|_2^2 + \langle \mathbf{1}, y\rangle^2\|x\|_2^2 \\
&\leq \sigma^2\left(\|(\Delta y)\circ x\|_2 + \|(\Delta x)\circ y\|_2\right)^2 \\
&\leq \sigma^2\left(\|\Delta y\|_2 + \|\Delta x\|_2\right)^2 \\
&\leq \sigma^2\left(2\|\Delta\|\sqrt{n}\right)^2 \\
&\leq c^2 n(2\gamma n)^2.
\end{aligned}$$

Now, notice that $\langle \mathbf{1}, x\rangle^2 + \langle \mathbf{1}, y\rangle^2 = \langle \mathbf{1}\mathbf{1}^T, X\rangle$ and the previous lemma imply that either $\langle \mathbf{1}, x\rangle^2$ or $\langle \mathbf{1}, y\rangle^2$ (or both) is larger than or equal to $n^2(1/4 - \gamma c)$. Without loss of generality, assume it is the former:

$$\langle \mathbf{1}, x\rangle^2 \geq n^2(1/4 - \gamma c).$$

We may combine the above inequalities to get

$$n^2(1/4 - \gamma c)\|y\|_2^2 \leq \langle 1, x\rangle^2 \|y\|_2^2 \leq (2\gamma c)^2 n^3,$$

or equivalently

$$\|y\|_2^2 \leq 4(2\gamma c)^2 n + 4\gamma c\|y\|_2^2.$$

We aim to show that $\|x\|^2$ is close to n. Use $\|x\|_2^2 + \|y\|_2^2 = n$:

$$n - \|x\|_2^2 \leq (4\gamma c)^2 n + 4\gamma c n - 4\gamma c\|x\|_2^2,$$

or equivalently

$$(1 - 4\gamma c)\|x\|_2^2 \geq \left(1 - 4\gamma c - (4\gamma c)^2\right) n.$$

Assuming $1 - 4\gamma c > 0$, we may divide through:

$$\|x\|_2^2 \geq \frac{1 - 4\gamma c - (4\gamma c)^2}{1 - 4\gamma c} n.$$

As targeted, if $4\gamma c$ is small enough, the right hand side gets arbitrarily close to n. In particular, the inequality is only informative if it is nonnegative, which requires $4\gamma c \leq \frac{\sqrt{5}-1}{2}$. By orthogonality of x and y, we have

$$\|X\|_F^2 = \|x\|_2^4 + \|y\|_2^4 \geq \|x\|_2^4,$$

so that, going back to (26) and under the assumption on $4\gamma c$,

$$\left\langle \mathbf{1}\mathbf{1}^T, X\right\rangle \geq \|X\|_F^2 - 2\gamma c n^2 \geq \left(\left(\frac{1 - 4\gamma c - (4\gamma c)^2}{1 - 4\gamma c}\right)^2 - 2\gamma c\right) n^2.$$

In the considered interval, the right hand side is nonnegative if and only if $0 \leq 8\gamma c \leq 1$ (a stronger condition than before). In that interval, we may extract the square root to characterize $Q^T\mathbf{1}$:

$$n \geq \|Q^T\mathbf{1}\|_2 = \sqrt{\langle \mathbf{1}\mathbf{1}^T, X\rangle} \geq n\sqrt{\left(\frac{1 - 4\gamma c - (4\gamma c)^2}{1 - 4\gamma c}\right)^2 - 2\gamma c} \geq (1 - 8\gamma c)n.$$

The last inequality holds by concavity of the square root term in that interval. We proved the inequality holds for $8\gamma c \leq 1$. It trivially holds otherwise as well. In passing, we note that by restricting c to a smaller interval, the bound can be improved arbitrarily close to $(1 - \gamma c)n$ (because less is lost in lowerbounding the concave function by an affine function). ■

Thus, by taking $\sigma = c\sqrt{n}$ for constant $c > 0$ small enough, we have that any local maximizer Q (in fact, any point satisfying both first- and second-order necessary optimality conditions) correlates very strongly with z.

3.2. Exact recovery

To establish exact recovery, we only need to show that all second-order critical points of (10) have rank 1. Indeed, it is well-known that rank deficient second-order critical points Q are global optima.

Lemma 10 *If Q satisfies the second-order necessary optimality condition for* (10) *and it is rank deficient (i.e.,* $\operatorname{rank}(Q) = 1$*), then Q is a global optimum.*

Proof Given Q feasible for (10) with cost matrix A, observe that if

$$S = S(Q) = \operatorname{ddiag}(AQQ^T) - A \tag{27}$$

is positive semidefinite, then Q is a global optimum. Indeed, for any feasible contender $\tilde{Q}$, we have (using $\operatorname{diag}(\tilde{Q}\tilde{Q}^T) = \mathbf{1}$)

$$0 \leq \langle S, \tilde{Q}\tilde{Q}^T \rangle = \operatorname{Tr}(AQQ^T) - \operatorname{Tr}(A\tilde{Q}\tilde{Q}^T),$$

thus showing that $\operatorname{Tr}(Q^TAQ)$ is maximal over the search space. The matrix S features prominently in (Burer et al., 2002; Burer and Monteiro, 2005; Journée et al., 2010; Wen and Yin, 2013; Bandeira et al., 2014b; Boumal, 2015, 2016), understandably given its strong properties as an optimality certificate. In this paper, we rely less on it in favor of different arguments.

If $\operatorname{rank}(Q) = 1$, then $QQ^T = qq^T$ for some $q \in \{\pm 1\}^n$. We show that if Q further satisfies the second-order condition (24), then $S(Q)$ is positive semidefinite, implying optimality of Q. For all $u \in \mathbb{R}^n$, we have (using $(QQ^T)_{ij}^2 = 1$ for all i, j)

$$\begin{aligned} u^TSu &= \left\langle \operatorname{ddiag}(AQQ^T) - A, uu^T \right\rangle \\ &= \left\langle \operatorname{ddiag}(AQQ^T) \circ (QQ^T) - A \circ (QQ^T), (uu^T) \circ (QQ^T) \right\rangle \\ &= \left\langle \operatorname{ddiag}(AQQ^T) - A \circ (QQ^T), (u \circ q)(u \circ q)^T \right\rangle \geq 0, \end{aligned}$$

where the inequality follows from (24). This holds for all u, thus $S \succeq 0$. ■

We further show in the present context another well-known result, namely that if the perturbation $\sigma\Delta$ has good properties, then the SDP has a unique solution corresponding to the ground truth. In the Wigner setting, this lemma may be improved somewhat, see (Bandeira et al., 2014b, §2.1).

Lemma 11 *Let $A = zz^T + \sigma\Delta$ for some planted solution $z \in \{\pm 1\}^n$, with $\sigma = c\sqrt{n}$, $\Delta = \Delta^T$,* $\operatorname{diag}(\Delta) = 0$*, $\|\Delta\| \leq \gamma\sqrt{n}$ and $\|\Delta z\|_\infty \leq \gamma\sqrt{n \log n}$, for some $\gamma, c \geq 0$. If*

$$\gamma c < \frac{1}{1 + \sqrt{\log n}},$$

then the unique global optimum of (9) *with cost matrix A is $X = zz^T$, so that all global optima Q of* (10) *are of the form $Q = [z\ 0]R$, where R is a 2×2 orthogonal matrix.*

Proof We show that $S = S(z) = \operatorname{ddiag}(Azz^T) - A$ (27) is positive semidefinite with rank $n - 1$, which by the argument in the proof of Lemma 10 implies optimality of zz^T. Uniqueness follows from strict complementarity ($\operatorname{rank}(zz^T) + \operatorname{rank}(S(z)) = n$). Indeed: let X be any optimum of (9); then, $\langle S, X \rangle = \left\langle A, zz^T \right\rangle - \langle A, X \rangle = 0$. Since $S \succeq 0$ and $X \succeq 0$, $\langle S, X \rangle = 0$ implies $SX = 0$.

Thus, $\mathrm{span}(X) \subset \ker S$. But $\ker S = \mathrm{span}(z)$ since the kernel has dimension 1 and $Sz = 0$. Adding that $X \succeq 0$ and $\mathrm{diag}(X) = \mathbf{1}$, it comes that $X = zz^\top$.

Using $Sz = 0$, it remains to show that for all $u \neq 0$ such that $z^T u = 0$, $u^T S u > 0$. We have:

$$\begin{aligned} u^T S u &= u^T \left(nI_n - zz^T + \sigma \left[\mathrm{ddiag}(\Delta zz^T - \Delta) \right] \right) u \\ &= n\|u\|_2^2 + \sigma \left[u^T \mathrm{ddiag}(\Delta zz^T) u - u^T \Delta u \right] \\ &\geq \|u\|_2^2 \left(n - \sigma \|\Delta z\|_\infty - \|\Delta\| \right) \\ &\geq n\|u\|_2^2 \left(1 - \gamma c \sqrt{\log n} - \gamma c \right). \end{aligned}$$

This is indeed positive under the prescribed condition. ■

We go through a few technical lemmas to establish the rank-one property of second-order critical points, under some conditions on the perturbation.

Lemma 12 *Let QQ^T be any feasible point of (9) satisfying first- and second-order necessary optimality conditions, with $Q = [x\ y] \in \mathbb{R}^{n\times 2}$. Let A_i be the ith row of A. If $\mathrm{diag}(A) \geq 0$, then*

$$\forall i, \quad \mathrm{diag}(AQQ^T)_i = \langle A_i, x\rangle x_i + \langle A_i, y\rangle y_i = \sqrt{\langle A_i, x\rangle^2 + \langle A_i, y\rangle^2} = \|e_i^T AQ\|_2.$$

Proof Row-wise, the first-order condition (23) reads

$$[\langle A_i, x\rangle, \langle A_i, y\rangle] = (\langle A_i, x\rangle x_i + \langle A_i, y\rangle y_i)\,[x_i, y_i].$$

Taking the L_2 norm of both sides and using $x_i^2 + y_i^2 = 1$, we get

$$\sqrt{\langle A_i, x\rangle^2 + \langle A_i, y\rangle^2} = |\langle A_i, x\rangle x_i + \langle A_i, y\rangle y_i|.$$

The second-order optimality condition (24) implies that (simply considering the diagonal of the positive semidefinite matrix, which must be nonnegative)

$$\mathrm{diag}(AQQ^T) \geq \mathrm{diag}(A).$$

That is,

$$\langle A_i, x\rangle x_i + \langle A_i, y\rangle y_i \geq A_{ii}.$$

Since we assume $\mathrm{diag}(A) \geq 0$, the two results combine into:

$$\langle A_i, x\rangle x_i + \langle A_i, y\rangle y_i = |\langle A_i, x\rangle x_i + \langle A_i, y\rangle y_i| = \sqrt{\langle A_i, x\rangle^2 + \langle A_i, y\rangle^2}, \tag{28}$$

concluding the proof. ■

The following lemma bounds the largest row in ΔQ. For a Wigner setting, this result is likely suboptimal, and is the bottleneck in our analysis. This is the same bottleneck that arose in both (Bandeira et al., 2014b) and (Boumal, 2016) for the phase synchronization problem.

Lemma 13 *Let QQ^T be any feasible point of (9) satisfying first- and second-order necessary optimality conditions with cost matrix $A = zz^T + \sigma\Delta$, $\sigma = c\sqrt{n}$, where $\Delta = \Delta^T$ satisfies $\mathrm{diag}(\Delta) = 0$, $\|\Delta\| \leq \gamma\sqrt{n}$ and $\|\Delta z\|_\infty \leq \gamma\sqrt{n \log n}$, for some $\gamma, c \geq 0$. Then, we have*

$$\max_i \|e_i^T \Delta Q\|_2 \leq \gamma\sqrt{n} \left(\sqrt{\log n} + 4\sqrt{\gamma c n} \right).$$

Proof Once more, we write the proof for $z = \mathbf{1}$, without loss of generality. Let $P_{\mathbf{1}} = \frac{1}{n}\mathbf{1}\mathbf{1}^\top$ be the orthogonal projector to span$(\mathbf{1})$ where $\mathbf{1} \in \mathbb{R}^n$ and, analogously, let $P_{\mathbf{1}^\perp} = I - P_{\mathbf{1}}$ be the projector to span$(\mathbf{1})^\perp$. Writing w_i for the ith column of Δ and letting $Q = [x\ y]$, we have for all i:

$$\begin{aligned}
\left\|e_i^T \Delta Q\right\|_2 &= \left\|w_i^T (P_{\mathbf{1}} + P_{\mathbf{1}^\perp}) Q\right\|_2 \\
&\leq \left\|w_i^T P_{\mathbf{1}}(Q)\|_2 + \|w_i^T P_{\mathbf{1}^\perp}(Q)\right\|_2 \\
&\leq \frac{1}{n}\left\|w_i^T \mathbf{1}\mathbf{1}^T Q\right\|_2 + \|w_i\|_2 \left\|Q - \frac{1}{n}\mathbf{1}\mathbf{1}^T Q\right\|_F \\
&\leq \frac{1}{n}\|\Delta \mathbf{1}\|_\infty \|Q^T\mathbf{1}\|_2 + \|w_i\|_2 \left\|Q - \frac{1}{n}\mathbf{1}\mathbf{1}^T Q\right\|_F .
\end{aligned}$$

Since, by Lemma 9, $\|Q^T\mathbf{1}\|_2 \geq n\,(1 - 8\gamma c)$, we may further bound the last term using

$$\begin{aligned}
\left\|Q - \frac{1}{n}\mathbf{1}\mathbf{1}^T Q\right\|_F^2 &= \|Q\|_F^2 + \frac{1}{n^2}\|\mathbf{1}\mathbf{1}^T Q\|_F^2 - \frac{2}{n}\|Q^T\mathbf{1}\|_2^2 \\
&= n - \frac{1}{n}\|Q^T\mathbf{1}\|_2^2 \\
&\leq n(1 - (1 - 8\gamma c)^2) \leq 16\gamma c n,
\end{aligned}$$

where we used $\|Q\|_F^2 = n$. Combining and using $\|Q^T\mathbf{1}\|_2 \leq n$, we get for all i that

$$\begin{aligned}
\left\|e_i^T \Delta Q\right\|_2 &\leq \gamma\sqrt{n\log n} + \|\Delta\|\sqrt{16\gamma c n} \\
&\leq \gamma\sqrt{n}\left(\sqrt{\log n} + 4\sqrt{\gamma c n}\right).
\end{aligned}$$

This concludes the proof. ■

The next lemma is central to our endeavor: it identifies a regime in which all second-order critical points have rank 1. The first inequality used in this proof is an important step inspired by the proof of (Wen and Yin, 2013, Thm. 3). Crucially, it is because we inject both first- and second-order optimality conditions (as opposed to only first-order conditions in (Wen and Yin, 2013)) that we are able to make a substantial statement via this rank-control argument. Indeed, even for the noiseless case, Theorem 3 in (Wen and Yin, 2013) does not lead to sufficiency of $p = 2$ for SDPLR.

Lemma 14 *Under the assumptions of Lemma 13 and using the same notation, if*

$$\gamma c < \frac{1}{9 + \sqrt{\log n} + 4\sqrt{\gamma c n}},$$

then $\operatorname{rank}(Q) = 1$.

Proof From the first-order condition (23), we may control the rank of Q as

$$\begin{aligned}
\operatorname{rank}(Q) &\leq \operatorname{null}\left(\operatorname{ddiag}(AQQ^T) - A\right) \\
&= n - \operatorname{rank}\left(\operatorname{ddiag}(AQQ^T) - A\right) \\
&= n - \operatorname{rank}\left(\operatorname{ddiag}(AQQ^T) - \sigma\Delta - zz^T\right) \\
&\leq n + 1 - \operatorname{rank}\left(\operatorname{ddiag}(AQQ^T) - \sigma\Delta\right).
\end{aligned}$$

To ensure $\operatorname{rank}(Q) = 1$, it remains to force $\operatorname{rank}(\operatorname{ddiag}(AQQ^T) - \sigma\Delta) = n$. Since the first matrix is diagonal, this is the case in particular if

$$\min_i \operatorname{diag}(AQQ^T)_i > \sigma\|\Delta\|.$$

This can be controlled by Lemmas 12 and 13:

$$\begin{aligned}\min_i \|e_i^T AQ\|_2 - \sigma\|\Delta\| &= \min_i \|e_i^T zz^T Q + \sigma \cdot e_i^T \Delta Q\|_2 - \sigma\|\Delta\| \\ &\geq \|Q^T z\|_2 - \sigma \cdot \max_i \|e_i^T \Delta Q\|_2 - \gamma cn \\ &\geq n - 9\gamma cn - \gamma cn\left(\sqrt{\log n} + 4\sqrt{\gamma cn}\right).\end{aligned}$$

Forcing the latter to be positive (as with the condition in this lemma's statement) is sufficient to imply $\operatorname{rank}(Q) = 1$. ■

Theorem 15 *Let $A = zz^T + \sigma\Delta$ for some planted solution $z \in \{\pm 1\}^n$, with $\sigma = c\sqrt{n}$, $\Delta = \Delta^T$, $\operatorname{diag}(\Delta) = 0$, $\|\Delta\| \leq \gamma\sqrt{n}$ and $\|\Delta z\|_\infty \leq \gamma\sqrt{n \log n}$, for some $\gamma, c \geq 0$. If*

$$\gamma c < \frac{1}{9 + \sqrt{\log n} + 4\sqrt{\gamma cn}},$$

then all second-order critical points Q of (10) with cost matrix A are global optima of rank 1 such that $QQ^T = zz^T$. There exists a constant k such that, if $\gamma c \leq kn^{-1/3}$, then the theorem applies.

Proof By Lemma 14, all second-order critical points Q have rank 1. By Lemma 10, such Q's are thus globally optimal. By Lemma 11 (whose conditions are satisfied a fortiori), they all satisfy $QQ^T = zz^T$. ■

Acknowledgments

ASB was supported by NSF grant DMS-1317308. ASB acknowledges Wotao Yin for pointing the author to the relevant reference (Wen and Yin, 2013) which helped motivate the start of the investigation in this paper. NB, formerly hosted by the SIERRA group at Inria and ENS, was supported by the "Fonds Spéciaux de Recherche" (FSR) at UCLouvain and by the Chaire Havas "Chaire Economie et gestion des nouvelles données", the ERC Starting Grant SIPA and a Research in Paris grant in Paris. VV acknowledges support from the Office of Naval Research.

References

E. Abbe, A. S. Bandeira, A. Bracher, and A. Singer. Decoding binary node labels from censored edge measurements: Phase transition and efficient recovery. *Network Science and Engineering, IEEE Transactions on*, 1(1):10–22, 2014.

E. Abbe, A.S. Bandeira, and G. Hall. Exact recovery in the stochastic block model. *Information Theory, IEEE Transactions on*, 62(1):471–487, 2016.

P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, NJ, 2008. ISBN 978-0-691-13298-3.

J. Baik, G. Ben-Arous, and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability*, 33(5):1643–1697, 2005.

A. S. Bandeira and R. v. Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *Annals of Probability, to appear*, 2015.

A. S. Bandeira, M. Charikar, A. Singer, and A. Zhu. Multireference alignment using semidefinite programming. *5th Innovations in Theoretical Computer Science (ITCS 2014)*, 2014a.

A.S. Bandeira. A note on probably certifiably correct algorithms. *Available at arXiv:1509.00824 [math.OC]*, 2015a.

A.S. Bandeira. Random Laplacian matrices and convex relaxations. *arXiv preprint arXiv:1504.03987*, 2015b.

A.S. Bandeira, C. Kennedy, and A. Singer. Approximating the little Grothendieck problem over the orthogonal and unitary groups. *arXiv preprint arXiv:1308.5207*, 2013.

A.S. Bandeira, N. Boumal, and A. Singer. Tightness of the maximum likelihood semidefinite relaxation for angular synchronization. *arXiv preprint arXiv:1411.3272*, 2014b.

A.S. Bandeira, Y. Chen, and A. Singer. Non-unique games over compact groups and orientation estimation in cryo-EM. *arXiv preprint arXiv:1505.03840*, 2015.

A.I. Barvinok. Problems of distance geometry and convex properties of quadratic maps. *Discrete & Computational Geometry*, 13(1):189–202, 1995. doi: 10.1007/BF02574037.

N. Boumal. A Riemannian low-rank method for optimization over semidefinite matrices with block-diagonal constraints. *arXiv preprint arXiv:1506.00575*, 2015.

N. Boumal. Nonconvex phase synchronization. *arXiv preprint arXiv:1601.06114*, 2016.

N. Boumal, P.-A. Absil, and C. Cartis. Global rates of convergence for nonconvex optimization on manifolds. *arXiv preprint arXiv:1605.08101*, 2016.

J. Briët, F.M. de Oliveira Filho, and F. Vallentin. Grothendieck inequalities for semidefinite programs with rank constraint. *Theory of Computing*, 10(4):77–105, 2014. doi: 10.4086/toc.2014.v010a004.

S. Burer and R.D.C. Monteiro. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, 95(2):329–357, 2003. doi: 10.1007/s10107-002-0352-8.

S. Burer and R.D.C. Monteiro. Local minima and convergence in low-rank semidefinite programming. *Mathematical Programming*, 103(3):427–444, 2005.

S. Burer, R.D.C. Monteiro, and Y. Zhang. Rank-two relaxation heuristics for Max-Cut and other binary quadratic programs. *SIAM Journal on Optimization*, 12(2):503–521, 2002.

A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E*, 84, December 2011.

U. Feige and J. Kilian. Heuristics for semirandom graph problems. *Journal of Computer and System Sciences*, 63(4):639 – 671, 2001.

D. Féral and S. Péché. The largest eigenvalue of rank one deformation of large wigner matrices. *Communications in Mathematical Physics*, 272(1):185–228, 2006.

M.X. Goemans and D.P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42(6): 1115–1145, 1995. doi: 10.1145/227683.227684.

O. Guedon and R. Vershynin. Community detection in sparse networks via Grothendieck's inequality. *Available online at arXiv:1411.4686 [math.ST]*, 2014.

B. Hajek, Y. Wu, and J. Xu. Achieving exact cluster recovery threshold via semidefinite programming. *Available online at arXiv:1412.6156*, 2014.

A. Javanmard, A. Montanari, and F. Ricci-Tersenghi. Phase transitions in semidefinite relaxations. *arXiv preprint arXiv:1511.08769*, 2015.

M. Journée, F. Bach, P.-A. Absil, and R. Sepulchre. Low-rank optimization on the cone of positive semidefinite matrices. *SIAM Journal on Optimization*, 20(5):2327–2351, 2010. doi: 10.1137/080731359.

M. Krivelevich and B. Sudakov. The largest eigenvalue of sparse random graphs. *Combinatorics, Probability and Computing*, 12:61–72, 2003.

L. Massoulié. Community detection thresholds and the weak ramanujan property. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, STOC '14, pages 694–703, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2710-7. doi: 10.1145/2591796.2591857. URL http://doi.acm.org/10.1145/2591796.2591857.

A. Moitra, W. Perry, and A. S. Wein. How robust are reconstruction thresholds for community detection? *Available online at arXiv:1511.01473 [cs.DS]*, 2015.

A. Montanari and S. Sen. Semidefinite programs on sparse random graphs. *Available online at arXiv:1504.05910 [cs.DM]*, 2015.

E. Mossel, J. Neeman, and A. Sly. Stochastic block models and reconstruction. *Probability Theory and Related Fields (to appear)*, 2014a.

E. Mossel, J. Neeman, and A. Sly. A proof of the block model threshold conjecture. *Available online at arXiv:1311.4115 [math.PR]*, January 2014b.

E. Mossel, J. Neeman, and A. Sly. Consistency thresholds for the planted bisection model. *Available online at arXiv:1407.1591v2 [math.PR]*, July 2014c.

Y. Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87 of *Applied optimization*. Springer, 2004. ISBN 978-1-4020-7553-7.

G. Pataki. On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues. *Mathematics of operations research*, 23(2):339–358, 1998. doi: 10.1287/moor.23.2.339.

A. Shapiro. Rank-reducibility of a symmetric matrix and sampling theory of minimum trace factor analysis. *Psychometrika*, 47(2):187–199, 1982.

A. Singer. Angular synchronization by eigenvectors and semidefinite programming. *Applied and Computational Harmonic Analysis*, 30(1):20–36, 2011. doi: 10.1016/j.acha.2010.02.001.

Z. Wen and W. Yin. A feasible method for optimization with orthogonality constraints. *Mathematical Programming*, 142(1–2):397–434, 2013. doi: 10.1007/s10107-012-0584-1.

Appendix A. Some technical steps

A.1. Other needed Lemmas

Lemma 16 *Let W be a symmetric Wigner matrix whose entries are independent standard gaussian random variables and $z \in \{\pm 1\}$ a fixed vector. $W - \operatorname{ddiag}(W)$ is the same matrix with the diagonal elements replaced by zeros. Then, the following holds for any $t \geq 0$:*

$$\operatorname{Prob}\left(\|W - \operatorname{ddiag}(W)\| \geq 2\sqrt{n} + t\right) \leq \exp\left(-\frac{t^2}{4}\right), \tag{29}$$

and

$$\operatorname{Prob}\left(\|(W - \operatorname{ddiag}(W))z\|\infty \geq \sqrt{2\log n + t}\right) \leq \exp\left(-\frac{t}{2}\right). \tag{30}$$

Proof (29) follows from combining $\mathbb{E}\|W - \operatorname{ddiag}(W)\| \leq \mathbb{E}\|W\|$ (which follows from Jensen's inequality), the well known fact that $\mathbb{E}\|W\| \leq 2\sqrt{n}$, and Gaussian Concentration. (30) follows from standard tail bounds on gaussian random variables together with a simple union bound. ■

Recall the definition of E in (19), E is symmetric, its entries are independently distributed, has diagonal zero and the distribution of the entries is: for $i \neq j$ and $g_i = g_j$:

$$\sqrt{\frac{(p+q)n}{2}} E_{ij} = \begin{cases} 1-p & \text{with probability } p \\ -p & \text{with probability } 1-p, \end{cases} \tag{31}$$

and, if $g_i \neq g_j$:

$$\sqrt{\frac{(p+q)n}{2}} E_{ij} = \begin{cases} 1-q & \text{with probability } q \\ -q & \text{with probability } 1-q. \end{cases} \tag{32}$$

Lemma 17 *There exists a constant C such that, with high probability*

$$\operatorname{SDP}(E) \leq 2 + C\frac{\log d}{d^{1/10}},$$

where $d = \frac{(p+q)n}{2} = \frac{a+b}{2}$ is the average degree parameter.

Proof See Lemma H.2. and equation (259) in (Montanari and Sen, 2015). ■

Lemma 18 *Consider E as defined above. With high probability, there exists a constant C such that*

$$\|E\| \leq 3 + C\sqrt{\frac{\log n}{\frac{n}{2}(p+q)}}.$$

Proof Set $\tilde{E} = \sqrt{\frac{(p+q)n}{2}}E$. Since the entries of $\tilde{E}$ are independent, we can use Corollary 3.12 in (Bandeira and v. Handel, 2015) (with, say, $\varepsilon = 3$) to obtain

$$\text{Prob}\left(\|\tilde{E}\| \geq 3\tilde{\sigma} + t \geq\right) \leq n\exp\left(-\frac{t^2}{c\tilde{\sigma_*}^2}\right),$$

for any $t > 0$ and a universal constant c. Here

$$\tilde{\sigma} = \max_i \sqrt{\sum_j \mathbb{E}E_{ij}^2} \leq \sqrt{\frac{n}{2}p(1-p) + \frac{n}{2}p(1-p)},$$

and

$$\tilde{\sigma_*} = \max_{ij} \|E_{ij}\| \leq 1.$$

This means that, with high probability, there exists a constant C such that

$$\|\tilde{E}\| \leq 3\sqrt{\frac{n}{2}p(1-p) + \frac{n}{2}q(1-q)} + C\sqrt{\log n} \leq 3\sqrt{\frac{(p+q)n}{2}} + C\sqrt{\log n},$$

concluding the proof. ■

Lemma 19 *Consider E as defined above. With high probability, there exists a constant C such that*

$$\|Eg\|_\infty \leq C\sqrt{\log n} + C\frac{\log n}{\sqrt{\frac{n}{2}(p+q)}}$$

Proof Set $\tilde{E} = \sqrt{\frac{(p+q)n}{2}}E$. For each $i \in [n]$, we have

$$(\tilde{E}g)_i = \sum_{j=1}^{\frac{n}{2}-1} \xi_j - \sum_{j=1}^{\frac{n}{2}} \xi'_j,$$

where ξ_j are independent random variables taking the value $1-p$ with probability p and the value $-p$ with probability $1-p$; ξ'_j are independent random variables (and independent to the random variables $\xi'_{j'}$) taking the value $1-q$ with probability q and the value $-q$ with probability $1-q$. It is easy to see that

$$\sum_{j=1}^{\frac{n}{2}-1} \mathbb{E}\xi_j^2 - \sum_{j=1}^{\frac{n}{2}} \mathbb{E}\left(\xi'_j\right)^2 = \frac{1}{2}(\frac{n}{2}-1)p(1-p) + \frac{n}{2}q(1-q) \leq \frac{n}{2}(p(1-p) + q(1-q)),$$

and that the summands are almost surely bounded by 1. This means that we can use Bernstein's inequality to get

$$\operatorname{Prob}\left((\tilde{E}g)_i > t\right) \le \exp\left(-\frac{\frac{1}{2}t^2}{\frac{n}{2}(p(1-p)+q(1-q))+\frac{1}{3}t}\right).$$

A union bound gives

$$\operatorname{Prob}\left(\|\tilde{E}g\|_\infty > t\right) \le 2n\exp\left(-\frac{\frac{1}{2}t^2}{\frac{n}{2}(p(1-p)+q(1-q))+\frac{1}{3}t}\right),$$

which means that, with high probability,

$$\|\tilde{E}g\|_\infty \lesssim \log n + \sqrt{\frac{n}{2}(p(1-p)+q(1-q))}\sqrt{\log n} \le \log n + \sqrt{\frac{n}{2}(p+q)}\sqrt{\log n},$$

concluding the proof. ■

Remark 20 *It is worth noting that the quantities $\|E\|$ and $\|Eg\|_\infty$ are tightly connected to the control of the spectrum of Γ_{SBM} in (Bandeira, 2015b, Definition 4.8).*