



ELSEVIER

Contents lists available at ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com



Computer science

A note on probably certifiably correct algorithms



Une note sur les certificats d'algorithmes valables avec grande probabilité

Afonso S. Bandeira

Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02142, USA

ARTICLE INFO

Article history:

Received 9 September 2015

Accepted after revision 19 November 2015

Available online 4 February 2016

Presented by the Editorial Board

ABSTRACT

Many optimization problems of interest are known to be intractable, and while there are often heuristics that are known to work on typical instances, it is usually not easy to determine a posteriori whether the optimal solution was found. In this short note, we discuss algorithms that not only solve the problem on typical instances, but also provide a posteriori certificates of optimality, probably certifiably correct (PCC) algorithms. As an illustrative example, we present a fast PCC algorithm for minimum bisection under the stochastic block model and briefly discuss other examples.

© 2015 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

R É S U M É

De nombreux problèmes d'optimisation, bien qu'ils soient difficiles à traiter dans les cas les plus compliqués, peuvent souvent être résolus de manière efficace par des heuristiques lorsque les données du problème sont tirées au hasard (données typiques). Malheureusement, dans la plupart des cas, on ne sait que rarement certifier si l'heuristique produit une solution optimale au problème. Dans cette note, nous décrivons une famille d'algorithmes qui non seulement résolvent le problème sur des données typiques, mais aussi produisent un certificat d'optimalité. À titre d'illustration, nous décrivons un tel algorithme pour le problème du partitionnement de graphe dans le modèle stochastique à blocs. D'autres exemples sont aussi discutés brièvement.

© 2015 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

1. Introduction

Estimation problems in many areas are often formulated as optimization problems, maximum likelihood estimation being a prime example. Unfortunately, these optimization problems are, in many relevant instances, believed to be computationally intractable in the worst case. On the other hand, oftentimes many real-world scenarios do not resemble these worst-case instances. This motivates a line of work that attempts to propose and understand algorithms that work on only some sets of, hopefully typical, inputs. A prime example is sparse recovery, where the popular Compressed Sensing papers of Candès, Romberg, Tao, and Donoho [12,14] established that, while finding the sparsest solution to an underdetermined system is

E-mail address: bandeira@mit.edu.

computationally intractable in the worst-case, a simple efficient algorithm succeeds with high probability in many natural probabilistic models of instances.

Definition 1.1 (*Probabilistic Algorithm*). Given an optimization problem that depends on an input and a probability distribution D over the inputs, we say an algorithm is a probabilistic algorithm for this problem and distribution of instances if it finds an optimal solution with high probability,¹ with respect to D .

Although these guarantees make excellent arguments towards the efficacy of certain probabilistic algorithms, they have the drawback that, oftentimes, it is not easy to check a posteriori whether the solution computed is the optimal one. Even in the idealized scenario where the input exactly follows the distribution in the guarantee, there is a nonzero probability of the algorithm not producing the optimal solution, and it is often not possible to tell whether it did or not. To make matters worse, in practice, the exact distribution of instances rarely exactly matches the idealized one in the guarantees.

The situation is different for a certain class of algorithms, convex relaxation based ones. Some of these methods work by enlarging the feasibility set of the optimization problem to a convex set where optimizing the objective function becomes tractable. While the optimal solution is not guaranteed to be in the original feasibility set, if it is, then one is sure that it must be the optimal solution of the original problem (providing, also, an a posteriori certificate). Fortunately, this tends to be the case for many examples of problems and relaxations [6,7,9,10]. This motivates the following definition.

Definition 1.2 (*Probably Certifiably Correct (PCC) Algorithm*). Given an optimization problem that depends on an input and a probability distribution D over the inputs, we say an algorithm is a Probably Certifiably Correct (PCC) algorithm for this problem and distribution of instances if, with high probability (w.r.t. D), it finds an optimal solution and certifies to have done so. Moreover, it never incorrectly certifies a non-optimal solution.

A PCC algorithm has the advantage of being able to produce a posteriori certificates. In particular, this renders them more appealing to be used in examples where the distribution of problem instances may not coincide with the idealized ones in proved guarantees. Indeed, duality is very often used in practice to provide a posteriori certificates of quality of a solution to an optimization problem (see [13] for a particularly recent example). While this is a great argument towards the use of convex relaxation-based approaches, many such algorithms are relaxed to semidefinite programs which, while solvable (to arbitrary precision) in polynomial time, tend to be computationally costly.

Many probabilistic algorithms not based on convex relaxations (such as, for example, spectral methods or alternating-minimization techniques) often do not enjoy a posteriori guarantees, but tend to be considerably more efficient than the convex relaxation-based competitors. This motivates the natural question of whether it is possible to devise a posteriori certifiers for candidate solutions produced by these or other methods.

Definition 1.3 (*Probabilistic Certifier*). Given an optimization problem that depends on an input, a probability distribution D over the inputs, and a candidate solution for it, we call a Probabilistic Certifier, a method that:

- With high probability (w.r.t. D), if the candidate solution is an optimal one it outputs: The solution is optimal.² It may, with vanishing probability,³ output: The solution may be non-optimal.
- If the candidate solution is not an optimal solution, it always outputs: The solution may be non-optimal.

A particularly natural way of constructing such certifiers is to rely on convex relaxation-based PCC algorithms; given a candidate solution computed by a probabilistic algorithm, one can check whether it is an optimal solution to a certain convex relaxation. Remarkably, it is sometimes considerably faster to check whether a candidate solution is an optimal solution of a convex program than to solve the program; in many such cases, one can devise faster PCC algorithms by combining fast probabilistic algorithms with fast methods to certify that a candidate solution is an optimal solution to a convex relaxation; or even that it is the unique optimal solution (as it will be the case with Algorithm 2.1). In the next section, we use the problem of minimum bisection under the stochastic block model to illustrate these ideas.

2. A fast PCC algorithm for recovery in the Stochastic Block Model

The problem of minimum bisection on a graph is a particularly simple instance of community detection that is known to be NP-hard. Recently, there has been interest in understanding the performance of several heuristics in typical realizations of a certain random graph model that exhibits community structure, the stochastic block model: given n even and $0 \leq q < p \leq 1$, we say that a random graph G is drawn from $\mathcal{G}(n, p, q)$, the Stochastic Block Model with two communities, if G has

¹ An event happens with high probability if its probability tends to 1 as the underlying parameters tend to infinity.

² In some cases, certifiers may also certify that a solution is not only optimal, but the unique optimal solution, as it will be the case with Algorithm 2.1.

³ By vanishing probability we mean probability tending to 0 as the underlying parameters tend to infinity.

n nodes, divided in two clusters of $\frac{n}{2}$ nodes each, and for each pair of vertices i, j , (i, j) is an edge of G with probability p if i and j are in the same cluster and with probability q otherwise, independently from any other edge.

Let A denote the adjacency matrix of G . We define a signed adjacency matrix B as $B = 2A - (\mathbf{1}\mathbf{1}^T - I)$. To each partitioning of the nodes we associate a vector x with ± 1 entries corresponding to cluster memberships. The minimum bisection of G can be written as

$$\max_x x^T B x \quad \text{s.t.} \quad x_i^2 = 1 \forall_i \text{ and } \mathbf{1}^T x = 0. \tag{1}$$

Setting $p = \alpha \frac{\log n}{n}$ and $q = \beta \frac{\log n}{n}$, it is known [2,19] that the hidden partition, with high probability, coincides with the minimum bisection, and can be computed efficiently provided that

$$\sqrt{\alpha} - \sqrt{\beta} > \sqrt{2}. \tag{2}$$

On the other hand, if $\sqrt{\alpha} - \sqrt{\beta} < \sqrt{2}$, then, with high probability, the maximum likelihood estimator (which corresponds to the minimum bisection) does not coincide with the hidden partition.

Remarkably, for the stochastic block model on two communities with parameters in the regime (2), there are quasi-linear time algorithms known to be probabilistic algorithms for minimum bisection (see, for example, [3]). A convex relaxation, proposed in [2], was also recently shown to exactly compute the minimum bisection in the same regime [8,15].

The convex relaxation in [2,8,15] is obtained by writing (1) in terms of a new variable $X = xx^T$. More precisely, (1) is equivalent to

$$\max_X \text{Tr}(BX) \quad \text{s.t.} \quad X_{ii} = 1 \forall_i, X \succeq 0, \text{rank}(X) = 1, \text{ and } \text{Tr}(X\mathbf{1}\mathbf{1}^T) = 0. \tag{3}$$

The semidefinite programming relaxation considered is obtained by removing the last two constraints.

$$\max_X \text{Tr}(BX) \quad \text{s.t.} \quad X_{ii} = 1 \forall_i \text{ and } X \succeq 0. \tag{4}$$

The argument in [8,15] use the fact that the optimal value of dual program given by

$$\min_D \text{Tr}(D) \quad \text{s.t.} \quad D - B \succeq 0 \text{ and } D \text{ is diagonal} \tag{5}$$

is known to match the optimal value of (4). More precisely, given the hidden partition $x_{\text{h}} \in \{\pm 1\}^n$, Abbe et al. [2] propose $D_{\text{h}} := D_{\text{diag}(x_{\text{h}})B\text{diag}(x_{\text{h}})}$ as a candidate solution for the dual,⁴ where $\text{diag}(x_{\text{h}})$ is a diagonal matrix whose diagonal is given by x_{h} and $D_{\text{h}} = D_{\text{diag}(x_{\text{h}})B\text{diag}(x_{\text{h}})}$ is a diagonal matrix whose diagonal is given by

$$[D_{\text{h}}]_{ii} = [D_{\text{diag}(x_{\text{h}})B\text{diag}(x_{\text{h}})}]_{ii} = \sum_{j=1}^n [\text{diag}(x_{\text{h}})B\text{diag}(x_{\text{h}})]_{ij} = (x_{\text{h}})_i \sum_{j=1}^n B_{ij} (x_{\text{h}})_j.$$

More recently, [8,15] showed that, in the parameter regime given by (2) and with high probability, this dual candidate solution is indeed a feasible solution to (5) whose value matches the value of the $x_{\text{h}}^T B x_{\text{h}}$. This implies that $x_{\text{h}} x_{\text{h}}^T$ is an optimal solution of (4). Moreover, since [8,15] show that

$$D_{\text{h}} - B \succeq 0 \quad \text{and} \quad \lambda_2(D_{\text{h}} - B) > 0, \tag{6}$$

where λ_2 denotes the second smallest eigenvalue, the argument can be easily strengthened (using complementary slackness) to show that $x_{\text{h}} x_{\text{h}}^T$ is the unique optimal solution (see [8,15] for details).

A particularly enlightening way of showing that (6) certifies that the partitioning given by x_{h} is the unique solution to the minimum bisection problem is to note that, for any candidate bisection $x \in \{\pm 1\}^n$,

$$x_{\text{h}}^T B x_{\text{h}} - x^T B x = x^T [D_{\text{h}} - B]^T x + \sum_{i=1}^n [D_{\text{h}}]_{ii} (1 - x_i^2) = x^T [D_{\text{h}} - B] x. \tag{7}$$

Since $D_{\text{h}} - B$ is known to satisfy (6), then $x^T [D_{\text{h}} - B] x \geq 0$. Moreover, since $[D_{\text{h}} - B] x_{\text{h}} = 0$, if x corresponds to another bisection (meaning that $x \neq x_{\text{h}}$ and $x \neq -x_{\text{h}}$) then $x^T [D_{\text{h}} - B] x > 0$, implying that $x_{\text{h}}^T B x_{\text{h}} - x^T B x > 0$.

Remark 1. A particularly fruitful interpretation of (7) is to think of it as a sum-of-squares certificate. More precisely, since $D_{\text{h}} - B$ is positive semidefinite it has a Cholesky decomposition $D_{\text{h}} - B = V V^T$ which means that, for $x \in \{\pm 1\}^n$,

⁴ Interestingly, in this case, the equality constraints and complementary slackness conditions are enough to pin-point a single possible candidate for a dual certificate, and only the semidefinite constraint needs to be checked; this is the case for semidefinite programs satisfying certain properties, see [5] for more details.

$$x_{\square}^{\top} B x_{\square} - x^{\top} B x = x^{\top} V V^{\top} x = \left\| V^{\top} x \right\|^2 = \sum_{j=1}^n \left(\sum_{i=1}^n V_{ij} x_i \right)^2.$$

By writing $x_{\square}^{\top} B x_{\square} - x^{\top} B x$ has a sum of squares, we certify that x_{\square} is an optimal solution. It turns out that certificates of this type always exist, potentially having to include polynomials of larger degree, and that, with a fixed bound on the degree of the polynomials involved, these certificates can be found with semidefinite programming. For more on the sum-of-squares technique see [20,11] and reference therein.

This suggests the following PCC algorithm for minimum bisection in the Stochastic Block Model.

Algorithm 2.1. Given a graph G , use the quasi-linear time algorithm in [3] to produce a candidate bisection x_* . Set $D_* := D_{\text{diag}(x_*)} B \text{diag}(x_*)$.

- If $x_*^{\top} \mathbf{1} = 0$ and

$$\lambda_2(D_* - B) > 0, \tag{8}$$
 output: x_* is the minimum bisection.
- If not, output: Not sure whether x_* is the minimum bisection.

Note that, since $(D_* - B)x_* = 0$, (8) automatically implies condition (6).

The following follows immediately from the results in [3] and [8,15].

Proposition 2.2. *Algorithm 2.1 is a Probably Certifiably Correct Algorithm for minimum bisection under the stochastic block model in the regime of parameters given by (2).*

2.1. Randomized certificates

While Algorithm 2.1 is considerably faster than solving the semidefinite program (4) it still requires one to check that an $n \times n$ matrix has a positive second-smallest eigenvalue, which we do not know how to do in quasi-linear time. A potentially faster alternative would be to use a randomized power-method-like algorithm, such as randomized Lanczos method [18], to estimate the second smallest eigenvalue of $D_* - B$. Note that since $B = 2A - (\mathbf{1}\mathbf{1}^{\top} - I)$, where A is a sparse matrix, matrix-vector multiplies with $D_* - B$ can be computed in quasi-linear time. While the use of such randomized methods would not provide a probabilistic certificate, it could potentially provide a randomized certificate that has a small probability (with respect to a source of randomness independent to D) of “certifying” an incorrect solution. However, since it would rely on independent randomness, the process would be able to be repeated to achieve an arbitrarily small probability of providing false certificates.⁵

3. Other examples of PCC algorithms and future directions

One of the most appealing characteristics of Algorithm 2.1 above is that it can be easily generalized to many other settings. In fact, given an optimization problem and a distribution D over the instances, if there is a fast probabilistic algorithm and a convex relaxation that is known to be tight (meaning that its optimal solution is feasible in the original problem), one can make use of both algorithms, similarly to Algorithm 2.1, and devise a fast PCC algorithm: by first running the fast probabilistic algorithm and then checking whether the candidate solution is the optimal solution to the convex relaxation. Unfortunately, it is not clear, in general, whether one can check optimality in the convex relaxation considerably faster than simply solving it. On the other hand, many proofs of tightness of convex relaxations also provide a candidate dual solution and, in many instances, checking whether this dual solution is indeed a dual certificate is significantly faster.

For some problems, such as multisection in the stochastic block model with multiple communities, both fast probabilistic algorithms [3] and tightness guarantees for convex relaxations have already been established [16,4,21] suggesting that this framework could be easily applied there. Other problems for which convex relaxations are known to be tight include: Synchronization over \mathbb{Z}_2 [1] and $SO(2)$ [9], sparse PCA [6], k -medians and k -means clustering [7,17], sensor network localization [22], and many others. We suspect that this framework may be useful in devising fast PCC algorithms for a large class of problems, perhaps including many of the described above.⁶ Moreover, even when probabilistic algorithms are not available, fast certifiers may be useful to test the performance of heuristics in real world problems.

⁵ It is conceivable that the analysis of the typical behavior of the second eigenvalue of $D_{\square} - B$ in [8,15] and the guarantees for the performance of randomized Lanczos method [18] can be used to devise a quasi-linear time randomized procedure that can serve as a randomized certificate for minimum bisection in the stochastic block model, as described above. However, such construction falls outside of the scope of this short note, and so it is left for future research.

⁶ An alternative strategy to developing fast PCC algorithms is to exploit particular structural properties of the problem to devise fast methods to solve the corresponding convex relaxations.

Acknowledgements

The author thanks Dustin G. Mixon, Soledad Villar, Nicolas Boumal, and Amit Singer for interesting discussions and valuable comments on an earlier version of this manuscript. The author also acknowledges Dustin G. Mixon for suggesting the term *Probably Certifiably Correct* and Philippe Rigollet for help with the translation of the title and abstract to French.

A.S. Bandeira was partially supported by AFOSR Grant No. FA9550-12-1-0317 and NSF grant DMS-1317308. Part of this work was done while the author was at Princeton University.

References

- [1] E. Abbe, A.S. Bandeira, A. Bracher, A. Singer, Decoding binary node labels from censored edge measurements: phase transition and efficient recovery, *IEEE Trans. Netw. Sci. Eng.* 1 (1) (2014) 10–22.
- [2] E. Abbe, A.S. Bandeira, G. Hall, Exact recovery in the stochastic block model, *IEEE Trans. Inf. Theory* 62 (1) (2016) 471–487.
- [3] E. Abbe, C. Sandon, Community detection in general stochastic block models: fundamental limits and efficient algorithms for recovery, in: FOCS, 2015, 2015 IEEE 56th Annual Symposium on Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on Foundations of Computer Science (FOCS), 2015, pp. 670–688, <http://dx.doi.org/10.1109/FOCS.2015.47>.
- [4] N. Agarwal, A.S. Bandeira, K. Koiliaris, A. Kolla, Multisection in the stochastic block model using semidefinite programming, available online at [arXiv:1507.02323](http://arxiv.org/abs/1507.02323) [cs.DS], 2015.
- [5] F. Alizadeh, J.-P. Haeberly, M.L. Overton, Complementarity and nondegeneracy in semidefinite programming, *Math. Program.* 77 (1) (1997) 111–128.
- [6] Arash A. Amini, Martin J. Wainwright, High-dimensional analysis of semidefinite relaxations for sparse principal components, *Ann. Stat.* 37 (5B) (2009) 2877–2921.
- [7] P. Awasthi, A.S. Bandeira, M. Charikar, R. Krishnaswamy, S. Villar, R. Ward, Relax, no need to round: integrality of clustering formulations, in: 6th Innovations in Theoretical Computer Science (ITCS 2015), 2015.
- [8] A.S. Bandeira, Random Laplacian matrices and convex relaxations, available online at [arXiv:1504.03987](http://arxiv.org/abs/1504.03987) [math.PR], 2015.
- [9] A.S. Bandeira, N. Boumal, A. Singer, Tightness of the maximum likelihood semidefinite relaxation for angular synchronization, available online at [arXiv:1411.3272](http://arxiv.org/abs/1411.3272) [math.OC], 2014.
- [10] A.S. Bandeira, Y. Khoo, A. Singer, Open problem: tightness of maximum likelihood semidefinite relaxations, in: Proceedings of the 27th Conference on Learning Theory, in: JMLR W&CP, vol. 35, 2014, pp. 1265–1267.
- [11] B. Barak, D. Steurer, Sum-of-squares proofs and the quest toward optimal algorithms, in: Survey, ICM 2014, 2014.
- [12] E.J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, *IEEE Trans. Inf. Theory* 52 (2006) 489–509.
- [13] L. Carlone, D.M. Rosen, G.C. Calafiore, J.J. Leonard, F. Dellaert, Lagrangian duality in 3d slam: verification techniques and optimal solutions, in: Int. Conf. on Intelligent Robots and Systems (IROS), 2015.
- [14] D.L. Donoho, Compressed sensing, *IEEE Trans. Inf. Theory* 52 (2006) 1289–1306.
- [15] B. Hajek, Y. Wu, J. Xu, Achieving exact cluster recovery threshold via semidefinite programming, available online at [arXiv:1412.6156](http://arxiv.org/abs/1412.6156), 2014.
- [16] B. Hajek, Y. Wu, J. Xu, Achieving exact cluster recovery threshold via semidefinite programming: extensions, available online at [arXiv:1502.07738](http://arxiv.org/abs/1502.07738), 2015.
- [17] T. Iguchi, D.G. Mixon, J. Peterson, S. Villar, On the tightness of an SDP relaxation of k-means, available online at [arXiv:1505.04778](http://arxiv.org/abs/1505.04778) [cs.IT], 2015.
- [18] J. Kuczynski, H. Wozniakowski, Estimating the largest eigenvalue by the power and Lanczos algorithms with a random start, *SIAM J. Matrix Anal. Appl.* 13 (4) (1992) 1094–1122.
- [19] E. Mossel, J. Neeman, A. Sly, Consistency thresholds for the planted bisection model, available online at [arXiv:1407.1591v2](http://arxiv.org/abs/1407.1591v2) [math.PR], July 2014.
- [20] P.A. Parrilo, Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization, PhD thesis, 2000.
- [21] W. Perry, A.S. Wein, A semidefinite program for unbalanced multisection in the stochastic block model, available online at [arXiv:1507.05605](http://arxiv.org/abs/1507.05605) [cs.DS], 2015.
- [22] A.M.-C. So, Y. Ye, Theory of semidefinite programming for sensor network localization, *Math. Program., Ser. B* 109 (2–3) (2007) 367–384.