Data Randomness Makes Optimization Problems Easier to Solve?

Yinyu Ye
Department is Management Science & Engineering
and Institute of Computational & Mathematical Engineering
Stanford University

December 25, 2016

Abstract

Optimization algorithms have been recently applied to solver problems where data possess certain randomness, partly because data themselves contain randomness in a big-data environment or data are randomly sampled from their populations. It has been shown that data randomness typically makes algorithms run faster in the so-called "average behavior analysis". In this short note, we give an example to show that a general non-convex quadratically constrained quadratic optimization problem, when data are randomly generated and the variable dimension is relatively higher than the number of constraints, can be globally solved with high probability via convex optimization algorithms. The proof is based on the fact that the semidefinite relaxation of the problem with random data would likely be exact in such cases. This implies that certain randomness in the gradient vectors and/or Hessian matrices may help to solve non-convex optimization problems.

1 Introduction

Optimization algorithms have been recently applied to solver problems where data possess certain randomness, partly because data themselves contain randomness in a big-data environment or data are randomly sampled from their large populations. Besides, many optimization solver developers are also tested their products based on randomly generated data.

It has been shown that data randomness typically makes algorithms run faster in the so-called "average behavior analysis". The idea of average case analysis is to obtain rigorous probabilistic bounds on the number of iterations required by an algorithm to reach some termination criterion, when the algorithm is applied to a random instance of a problem drawn from some probability distribution. In the case of the simplex method for LP, average case analysis (see for example [1], [5], [8], [11], and lately [9]) has provided some theoretical justification for the observed practical efficiency of the method, despite its exponential worst case bound.

In the case of interior-point algorithms for LP, a "high probability" bound of $O(\sqrt{n}\ln n)$ iterations for termination (independent of data size) was proved, using a variety of algorithms applied to several different probabilistic models. Here n is the dimension or number of variables in a standard form problem, and "high probability" means that the probability that the bound holds goes to one, as $n \to \infty$; see, e.g., [13] and [4]. In [12] the authors analyzed a condition/complex number of constraint matrix A of dimension $m \times n$ for an interior-point linear programming algorithm, and showed that, if A is a standard Gaussian matrix, then the expected condition number equals $O(\min\{m \ln n, n\})$. Consequently, the interior-point algorithm terminates in strongly polynomial time in expectation.

On the other hand, specific recovering problems with random data/sampling were proved to be exact by convex optimization approaches, which include Digital Communication [10], Sensor-Network Localization [7], Phaselift Signal Recovering [3], Max-Likelihood Angular Synchronization [2], and a survey paper and references therein [6]. In these approaches, the recovering problems are relaxed to semidefinite programs, where each randomly sampled measurement becomes a constraint in the relaxation. When the number of random constraints or measurements is sufficiently large ($O(n \ln n)$) relatively to the dimension n of the variable matrix, then the relaxation contains the unique solution that needs to be recovered. However, there problems can actually be solved by more efficient deterministic targeted sampling using only $O(n)$ measurements.

In this short note, we give another evidence example to show that a general non-convex optimization problem, when data are random and the variable dimension n is relatively higher than the number of constraints m (note that this is in contrast to the covering problems mentioned earlier), can be globally solved with high probability via convex optimization algorithms. The proof is based on the fact that the convex relaxation of the problem with randomly generated data would be more likely exact when the dimension is relatively higher.

2 Quadratically Constrained Quadratic Program and its SDP Relaxation

Consider the quadratically constrained quadratic program (QCQP):

$$\begin{array}{lll} \text{minimize}_x & x^T Q x + 2c^T x & \\ \text{subject to} & x^T A_i x + 2a_i^T x & = b_i;\ \forall i = 1, ..., m, \\ & \|x\|_2 & = 1. \end{array} \tag{1}$$

where Q and A_i are general symmetric $n \times n$ matrices, c and a_i are n-dimensional vectors for $i = 1, ..., m$. This is a general non-convex optimization problem where many local minimizers exist.

The (convex) semidefinite programming (SDP) relaxation of problem (1) is

$$\begin{array}{lll} \text{minimize}_{X,x} & Q \cdot X + 2c^T x & \\ \text{subject to} & A_i \cdot X + 2a_i^T x & = b_i,\ \forall i = 1, ..., m, \\ & I \cdot X & = 1, \\ & X - xx^T & \succeq 0; \end{array} \tag{2}$$

and in general it is not exact. Note that the matrix inequality $X - xx^T \succeq 0$ can be equivalently written as

$$\begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix} \succeq 0.$$

The optimality conditions of the SDP relaxation are: for dual variables or multipliers y_i, $i = 1, ..., m$, and δ

$$\begin{array}{rl} (Q - \sum_i y_i A_i + \delta I)\, x & = -\left(c - \sum_i y_i a_i\right), \\ (Q - \sum_i y_i A_i + \delta I) & \succeq 0, \\ \left(X - xx^T\right)(Q - \sum_i y_i A_i + \delta I) & = 0. \end{array} \tag{3}$$

The first equation can be seen as the first-order optimality condition, the second one represents the second-order condition, and the third one presents the matrix complementarity condition. If the rank of SDP solution X of (2) is 1, then $X = xx^T$, and we would have solved the original QCQP problem (1).

If the rank of SDP solution X is not 1, then we have the following lemma, which can be directly proved from the these SDP optimality conditions.

Lemma 1. *Let the SDP optimal solution $X \neq xx^T$. Then the followings hold.*

i) Symmetric matrix $(Q - \sum_i y_i A_i + \delta I) \succeq 0$ but it is rank deficient, that is, its smallest eigenvalue equals 0.

ii) Let positive semidefinite solution matrix $X - xx^T$ have rank $r \geq 1$ and $X - xx^T = \sum_{j=1}^r \hat{x}_j \hat{x}_j^T$. Then $\hat{x}_j$ is the zero eigenvector of $(Q - \sum_i y_i A_i + \delta I)$ for all j, and

$$\hat{x}_j^T \left(c - \sum_i y_i a_i\right) = 0, \ \forall j.$$

iii) Let $(\cdot)^*$ denote the matrix pseudo inverse. Then

$$\|(Q - \sum_i y_i A_i + \delta I)^* (c - \sum_i y_i a_i)\| < 1.$$

The proof of item *i*) is from the second-order and complementarity conditions, since otherwise, $X = xx^T$. The second item proof is from multiplying $X - xx^T$ from left to both sides of the first-order optimality condition, and noting $X - xx^T \succeq 0$ and the rest of optimality conditions. The third item proof is from the fact $1 \geq I \cdot X \geq \|x\|^2$ and the first-order optimality condition.

3 QCQP with Random Data and Exactness of SDP Relaxation

Let$(Q, A_i, c, a_i),\ i = 1, ..., m$ be independently and randomly generated, for example, from the standard Gaussian. Then, we show that the SDP relaxation (2) would be exact for (1) or the SDP solution rank is 1 with high probability when $n >> m$.

We give a proof argument for case $m = 1$, in which case the problem represents a well-known and important extended trust-region subproblem, and it is known that the SDP relaxation is not exact in general; see for example [14] and [6].

Without loss of generality, consider Q and A_1 are both diagonal matrices where each diagonal entry, together with each entry of vectors c and a_1, are iid standard Gaussian. Then, if the rank of SDP solution X is not 1, from the Lemma we must have at least one entry in vector $(c - y_1 a_1)$ equals 0, since the 0-value eigenvector of positive semidefinite matrix $Q - \sum_i y_i A_i + \delta I$, say e_i, is perpendicular to vector $(c - y_1 a_1)$. Since both c and a_1 are iid standard Gaussian and there is only one multiplier y_1, exactly one entry of $(c - y_1 a_1)$ is 0 with probability one.

Furthermore, the entry-index of the 0-entry in $(c - y_1 a_1)$ is identical to the one where the entry of positive semidefinite diagonal matrix $Q - y_1 A_1 + \delta I$ is 0, or the entry of diagonal matrix $Q - y_1 A_1$ who has the lowest eigenvalue. Given each entry of Q and A_1 are iid standard Gaussian and they are independent of c and a_1, this matching occurs with probability $1/n$. That is, with probability $1 - 1/n$, the SDP relaxation solution X is rank-one, which solves the original QCQP problem exactly.

In summary, we have

Theorem 2. *Consider the non-convex QCQP problem (1) where $m = 1$. Let entries of (Q, A_1, c, a_1) be independently and randomly generated from the standard Gaussian. Then, with probability at least $1 - 1/n$, the SDP relaxation (2) optimal solution is rank-one, that is, the SDP relaxation solves the original problem exactly.*

A possible application of the theorem is to randomly perturb the linear term or gradient of the quadratic functions a little to achieve a rank-one SDP solution, which solves the original problem approximately. For example, consider the case $0 = c = a_i$ for all i in the original QCQP problem. Then, the SDP relaxation solution is not rank-one generically. But we may make a small random perturbation to c and a_i to solve the problem approximately. This implies that certain randomness in the gradient vectors and/or Hessian matrices may help to solve non-convex optimization problems approximately.

References

[1] I. Adler and N. Megiddo (1985), A simplex algorithm whose average number of steps is bounded between two quadratic functions of the smaller dimension. Journal of the Association of Computing Machinery 32, 891-895.

[2] A.S. Bandeira, N. Boumal, and A. Singer (2016), Tightness of the maximum likelihood semidefinite relaxation for angular synchronization. https://arxiv.org/pdf/1411.3272v3.pdf

[3] E.J. Candes, S. Thomas, and V. Voroninski (2011), PhaseLift: Exact and Stable Signal Recovery from Magnitude Measurements via Convex Programming. https://arxiv.org/abs/1109.4499

[4] K.M. Anstreicher, J. Ji, F. Potra, and Y. Ye (1999), Probabilistic analysis of an infeasible primal-dual algorithm for linear programming. Math. Oper. Res. 24, 176–192

[5] K.H. Borgwardt (1987), The Simplex Method – A Probabilistic Approach. (Springer-Verlag, New York).

[6] Z. Luo, W. Ma, A.M. So, Y. Ye, S. Zhang (2010), Semidefinite relaxation of quadratic optimization problems. IEEE Signal Processing Magazine 27 (3) 20.

[7] D. Shamsi, N. Taheri , Z. Zhu, and Y. Ye (2010), Conditions for Correct Sensor Network Localization Using SDP Relaxation. https://arxiv.org/abs/1010.2262 and Discrete Geometry and Optimization, Volume 69 of the series Fields Institute Communications, pp 279-301 (2013)

[8] S. Smale (1983), On the average number of steps of the simplex method of linear programming. Mathematical Programming 27, 241-262.

[9] D.A. Spielman and S. Teng (2001), Smoothed Analysis of Algorithms: Why the Simplex Algorithm Usually Takes Polynomial Time. https://arxiv.org/abs/cs/0111050

[10] A.M. So (2010) Probabilistic analysis of the semidefinite relaxation detector in digital communications Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms, Pages 698-711

[11] M.J. Todd (1986), Polynomial expected behavior of a pivoting algorithm for linear complementarity and linear programming problems. Mathematical Programming 35, 173-192.

[12] M.J. Todd, L. Tunel, and Y. Ye (2001), Characterizations, bounds, and probabilistic analysis of two complexity measures for linear programming problems -Mathematical Programming 90, 59-69.

[13] Y. Ye (1994): Toward probabilistic analysis of interior-point algorithms for linear programming. Math. Oper. Res. 19, 38–52

[14] Y. Ye and S. Zhang (2003), New Results on Quadratic Minimization. SIAM J. Optim., 14(1), 245267.