



Mathematics of Operations Research

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

New Analysis on Sparse Solutions to Random Standard Quadratic Optimization Problems and Extensions

Xin Chen, Jiming Peng

To cite this article:

Xin Chen, Jiming Peng (2015) New Analysis on Sparse Solutions to Random Standard Quadratic Optimization Problems and Extensions. *Mathematics of Operations Research*

Published online in Articles in Advance 02 Feb 2015

<http://dx.doi.org/10.1287/moor.2014.0692>

Full terms and conditions of use: <http://pubsonline.informs.org/page/terms-and-conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2015, INFORMS

Please scroll down for article—it is on subsequent pages



INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

New Analysis on Sparse Solutions to Random Standard Quadratic Optimization Problems and Extensions

Xin Chen

Department of Industrial and Enterprise System Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois, 61801, xinchen@illinois.edu

Jiming Peng

Department of Industrial Engineering, University of Houston, Houston, Texas, 77204, jopeng@uh.edu

The standard quadratic optimization problem (StQP) refers to the problem of minimizing a quadratic form over the standard simplex. Such a problem arises from numerous applications and is known to be NP-hard. In a recent paper [Chen X, Peng J, Zhang S (2013) Sparse solutions to random standard quadratic optimization problems. *Math. Programming* 141(1–2):273–293], Chen, et al. showed that with a high probability close to 1, StQPs with random data have sparse optimal solutions when the associated data matrix is randomly generated from a certain distribution such as uniform and exponential distributions. In this paper, we present a new analysis for random StQPs combining probability inequalities derived from the first-order and second-order optimality conditions. The new analysis allows us to significantly improve the probability bounds. More important, it allows us to handle normal distributions, which is left open in Chen et al. (2013). The existence of sparse approximate solutions to convex StQPs and extensions to other classes of QPs are discussed as well.

Keywords: copositive definite matrix; standard quadratic optimization; sparse solutions; random matrices; probability analysis

MSC2000 subject classification: Primary: 90C20; secondary: 15B52

OR/MS subject classification: Primary: quadratic programming

History: Received August 1, 2012; revised July 29, 2014. Published online in *Articles in Advance*.

1. Introduction. In this paper, we consider the following quadratic programming (QP) problem:

$$\min x^T Q x \tag{1}$$

$$\text{s.t. } e^T x = 1, \quad x \geq 0, \tag{2}$$

where $Q = [Q_{ij}] \in \mathbb{R}^{n \times n}$ is a symmetric matrix, and $e \in \mathbb{R}^n$ is the all 1-vector. Like in Bomze [4], we called the above problem the standard quadratic programming problem (StQP). The problem appears in numerous applications such as resource allocation (Irabaki and Katoh [16]), portfolio selection (Markowitz [18]), and machine learning (Mukherjee et al. [20]). It also covers several other important problems such as the maximal clique problem in discrete optimization (Gibbons et al. [13]) and the determination of copositivity of a matrix in linear algebra (Bomze et al. [5]). In addition to its broad range of applications, the StQP provides a prototype for numerous classes of quadratic optimization problems. The study of StQP has caught the attention of experts in many different fields, and various algorithms have been proposed in the literature. For details, we refer to recent papers (Bomze et al. [5], Scozzari and Tardella [22], Yang and Li [23]) and the references therein.

In this work, we are mainly concerned with a special scenario of StQPs, where the matrix Q is generated from a certain distribution. The study on optimization problems with random data can be dated back to the early 1980s (Goldberg and Marchetti-Spaccamela [14]) when Goldberg and Marchetti-Spaccamela considered how to find an exact solution of the knapsack problem with random data. For more recent progress on random knapsack problems, we refer to the paper (Beier and Vöcking [3]). In recent years, the study on the so-called L_1 minimization problem with random constraints has attracted a great amount of attention from experts in various disciplines. It is shown that when the coefficient matrix is generated from a normal distribution, then with a high probability, the optimal solution of the L_1 minimization problem is the sparsest point in the constrained set (Candès and Wakin [6], Candès et al. [7], Donoho [12]). In Recht et al. [21], the authors extended the L_1 minimization to the case where the decision variable is required to be a positive semidefinite (PSD) matrix and the L_1 norm is replaced by the Nuclear norm. Other extensions can also be found in Candès et al. [8].

In our recent work (Chen et al. [9]), we noted that there exists a close relation between the StQP and the L_1 minimization. Inspired by this observation, we considered the existence of sparse solutions to StQPs with random data and proved that with a high probability, an StQP admits a very sparse optimal solution if the matrix Q is generated from a class of distributions, including the uniform or exponential distributions.

The analysis in Chen et al. [9] mainly builds on one characteristic derived from the first-order optimality condition of the global optimal solution of an StQP and some probability bounds related to order statistics for

random variables with concave cumulative distribution functions in their supports. The class of distributions, though very broad, and including notable distributions such as uniform and exponential, excludes normal distributions, which are predominately assumed in the literature on the L_1 minimization problem and its extensions.

The main purpose of this work is to extend and improve the results in Chen et al. [9]. Specifically, we develop a new analysis for StQPs with random matrices combining properties derived from the first-order and the second-order optimality conditions of the global optimal solution of an StQP. It significantly improves the results in Chen et al. [9]. More importantly, coupled with results from random matrix theory, our new analysis allows us to handle StQPs with random matrices generated from normal distributions. We also analyze an StQP with a random PSD Q and estimate the sparsity of its so-called ϵ -approximate solution.

The paper is organized as follows. In §2, we review some useful properties to be used here. In §3, we present a new analysis on sparse solutions to random StQPs. In §4, we consider the case when Q is PSD, and establish the existence of sparse approximations to convex StQPs. Finally, we conclude our paper with some remarks in §5.

Before we proceed, we list a few notations. Throughout this paper, we use e and $\exp(x)$ to denote the all-ones vector and the exponential function, respectively. The symbol E represents the all-ones square matrix with dimensions consistent with the context, and we use E_k for E when the dimension is specified as k . We refer to x^* as one of the sparsest global optimal solutions of problem (1) (unless else specified), i.e., it has the smallest number of positive components among all global optimal solutions. For a given vector $x \in \mathfrak{R}^n$, $\|x\|_0$ denotes the number of nonzero elements. We define $n!$ as the factorial of a natural number n , $C(n, k)$ the number of k -combinations from a given set with n elements, and $B(n, k)$ the Beta function with $B(n, k) = \int_0^1 u^{n-1}(1-u)^{k-1} du = (n-1)!(k-1)!/(n+k-1)!$.

2. Preliminary results. Before we present our new analysis, we first review in this section some properties established in Chen et al. [9] that will be useful here.

Let $U_r, r = 1, \dots, n$ be independent continuous random variables each with a cumulative distribution $F(\cdot)$ and a probability density function $f(\cdot)$ and $u_1 \leq u_2 \leq \dots \leq u_n$ be the order statistics of U_r 's. Let $V_r, r = 1, 2, \dots, m$ be independent continuous random variables each with a cumulative distribution $G(\cdot)$ and a probability density function $g(\cdot)$ and $v_1 \leq v_2 \leq \dots \leq v_m$ be the order statistics of V_r 's. Assume that the random vectors $[U_r]_{r=1}^n$ and $[V_s]_{s=1}^m$ are independent from each other. In addition, we make the following assumption on the distributions.

- ASSUMPTION 1. (a) $G(\cdot)$ is concave on its support.
 (b) There exists a constant $\alpha \in (0, 1]$ such that $G(x) \geq \alpha F(x)$ for any $x \in (-\infty, \infty)$.

We have the following probability bounds on the order statistics.

THEOREM 1 (CHEN ET AL. [9]). *Let $U_r, r = 1, \dots, n$ and $V_r, r = 1, \dots, m$ be independently identically distributed random variables with cumulative distribution functions $F(\cdot)$ and $G(\cdot)$, respectively. Assume that the random vectors $[U_r]_{r=1}^n$ and $[V_s]_{s=1}^m$ are independent from each other. If $G(\cdot)$ and $F(\cdot)$ satisfy Assumption 1, then we have*

$$P\left(\sum_{r=1}^k u_r \leq kv_1\right) \leq \prod_{j=1}^k \frac{(m+n-k)(n-j+1)}{(n-k+\alpha m(k-j+1)/k)(m+n-j+1)} \quad (3)$$

$$\sum_{i=1}^m P\left(\sum_{r=1}^k u_r \leq (k+1)v_1 - v_i\right) \leq (k+1)P\left(\sum_{r=1}^k u_r \leq kv_1\right). \quad (4)$$

The following proposition presents some characteristics of a global optimal solution of the StQP (1).

PROPOSITION 1 (CHEN ET AL. [9]). *Suppose that x^* is one of the sparsest global optimal solutions of problem (1), satisfying $\|x^*\|_0 = k > 1$. Let $Q_{\mathcal{H}} \in \mathfrak{R}^{k \times k}$ be the principal submatrix of Q induced by the index set of all the nonzero elements of x^* , denoted by \mathcal{H} , and define $\lambda^* = (x^*)^T Q x^*$. Then the following conclusions hold:*

C.1 *There exists a row (or column) of $Q_{\mathcal{H}}$ such that the average of all its elements is strictly less than the minimal diagonal element of Q ;*

C.2. *The matrix $Q_{\mathcal{H}} - \lambda^* E_k$ is PSD.*

Property C.1 in the above proposition follows from the first-order optimality condition, while property C.2 is derived from the second-order optimality condition. The following sparsity results of the global optimal solution were mainly built on property C.1 and the probability bounds in Theorem 1.

THEOREM 2 (CHEN ET AL. [9]). *Assume that the upper triangular elements of Q are independent with the diagonal elements identically distributed with cdf $G(\cdot)$ and pdf $g(\cdot)$ and the strict upper triangular elements of Q identically distributed with cdf $F(\cdot)$ and pdf $f(\cdot)$. Suppose that $G(\cdot)$ and $F(\cdot)$ satisfy Assumption 1. Let x^* be a global optimal solution of problem (1). Then it holds for $k \geq 2$ that*

$$P(\|x^*\|_0 = k) \leq k\tau^{\min(k-1, \lfloor \sqrt{2n\alpha} \rfloor)} \quad (5)$$

and

$$P(\|x^*\|_0 \geq k) \leq \tau^{k-1} \left(\frac{1}{(1-\tau)^2} + \frac{k-1}{1-\tau} \right) + \frac{n(n+1)}{2} \tau^{\lfloor \sqrt{2n\alpha} \rfloor}, \quad (6)$$

where $\tau = (1/(1 + \alpha/2))^{1/2}$.

It should be pointed out that the probability bound on $P(\|x^*\|_0 = k)$ is only implicitly presented in the proof of Theorem 3.4 in Chen et al. [9]. One significant restriction of the above result is that the cdf $G(\cdot)$ is required to be concave or equivalently its pdf $g(\cdot)$ is nonincreasing in its support. Though it covers notable distributions such as uniform and exponential, it excludes normal distributions, widely used in the literature of L_1 minimization. Interestingly, as observed from the numerical experiment in Chen et al. [9], the StQPs generated from normal distributions still have very sparse optimal solutions. In what follows, we present a new analysis employing properties C.1 and C.2, which allows us to significantly improve the probability bounds in Theorem 2, and more importantly, quantify the sparsity of global optimal solutions of StQPs generated from normal distributions.

3. New analysis on sparse solutions to random StQPs. Throughout this paper, let x^* be one of the sparsest global optimal solutions of problem (1). From Proposition 1, we have that if $\|x^*\|_0 = k$, then there exists an index set \mathcal{H} with $|\mathcal{H}| = k$, which satisfies properties C.1 and C.2. To present our analysis, let $Q_{\mathcal{H}}$ be the principal submatrix induced by the index set \mathcal{H} . Define the following probability events:

$$\mathcal{H}_1^{\mathcal{H}} = \{Q_{\mathcal{H}} \text{ satisfies C.1}\}; \quad (7)$$

$$\mathcal{H}_{1,i}^{\mathcal{H}} = \left\{ \sum_{j \in \mathcal{H}} Q_{ij} \leq k \min_{j=1:n} Q_{jj} \right\}; \quad (8)$$

$$\tilde{\mathcal{H}}_{1,i}^{\mathcal{H}} = \left\{ \sum_{j \in \mathcal{H}} Q_{ij} \leq k \min_{j \in \mathcal{H} \setminus \{i\}} Q_{jj} \right\}; \quad (9)$$

$$\mathcal{H}_2^{\mathcal{H}} = \{\exists \lambda^* \in \mathfrak{R}: Q_{\mathcal{H}} - \lambda^* E_k \geq 0\}. \quad (10)$$

With these definitions, we have

$$P(\|x^*\|_0 = k) \leq C(n, k)P(\mathcal{H}_1^{\mathcal{H}} \cap \mathcal{H}_2^{\mathcal{H}}).$$

Because $\mathcal{H}_1^{\mathcal{H}} = \bigcup_{i \in \mathcal{H}} \mathcal{H}_{1,i}^{\mathcal{H}}$ and any principal submatrix of a PSD matrix is still PSD, we have that

$$\mathcal{H}_1^{\mathcal{H}} \cap \mathcal{H}_2^{\mathcal{H}} \subseteq \bigcup_{i \in \mathcal{H}} (\mathcal{H}_{1,i}^{\mathcal{H}} \cap \mathcal{H}_2^{\mathcal{H} \setminus \{i\}}).$$

Since $\mathcal{H}_{1,i}^{\mathcal{H}} \subseteq \tilde{\mathcal{H}}_{1,i}^{\mathcal{H}}$ and the two events $\tilde{\mathcal{H}}_{1,i}^{\mathcal{H}}$ and $\mathcal{H}_2^{\mathcal{H} \setminus \{i\}}$ are independent, we have for any given $i \in \mathcal{H}$,

$$P(\|x^*\|_0 = k) \leq kC(n, k)P(\tilde{\mathcal{H}}_{1,i}^{\mathcal{H}})P(\mathcal{H}_2^{\mathcal{H} \setminus \{i\}}). \quad (11)$$

In the remainder of this section, we discuss how to bound the probabilities $P(\tilde{\mathcal{H}}_{1,i}^{\mathcal{H}})$ and $P(\mathcal{H}_2^{\mathcal{H} \setminus \{i\}})$ under two different models generating the random symmetric matrix Q .

Model 1. first randomly generate Q 's lower triangular and diagonal elements independently using the same cumulative distribution $G(\cdot)$; then assign Q 's upper triangular elements with their corresponding lower triangular elements.

Model 2. first randomly generate a square matrix $M \in \mathfrak{R}^{n \times n}$ whose elements are identically independently distributed with a cumulative distribution function $G(\cdot)$; then define $Q = (M + M^T)/2$.

Bounding the probabilities $P(\tilde{\mathcal{H}}_{1,i}^{\mathcal{H}})$ and $P(\mathcal{H}_2^{\mathcal{H} \setminus \{i\}})$ is challenging for general distributions. Only under some special yet important cases, we are able to derive some probability bounds. Specifically, in §3.1, we focus on Model 1 in which G is assumed to be continuous and concave in its support, while in §3.2, we focus on Model 2 in which G is the standard normal distribution. We also provide a probability bound on $P(\|x^*\|_0 = k)$ for the scenario in which Q is generated by Model 2 with G being continuous and concave in its support by bounding $P(\tilde{\mathcal{H}}_{1,i}^{\mathcal{H}})$ only in the appendix.

3.1. Model 1 with concave cumulative distributions. In this subsection, we focus on the scenario in which Q is generated by Model 1 with the cumulative distribution $G(\cdot)$ being continuous and concave in its support. We first show how to bound the probability $P(\tilde{\mathcal{H}}_{1,i}^{\mathcal{H}})$, which interestingly, can be derived from the probability bound established in Theorem 1.

LEMMA 1. *Let $Q_{\mathcal{H}} \in \mathfrak{R}^{k \times k}$ be an arbitrary principal submatrix of $Q \in \mathfrak{R}^{n \times n}$, a symmetric matrix whose diagonal and lower diagonal elements are identically independently distributed with a continuous cumulative distribution function concave in its support. Then we have that*

$$P(\tilde{\mathcal{H}}_{1,i}^{\mathcal{H}}) \leq \frac{k(k-1)^{k-1}(n-k)!}{n!}. \quad (12)$$

PROOF. Recall (9) and set $U_r = Q_{ij_r}$, $r = 1, \dots, k-1$ for any j_r with $\{j_1, \dots, j_r\} = \mathcal{H} \setminus \{i\}$, and $V_r = Q_{l_r l_r}$, $r = 1, \dots, n-k+1$ for any l_r with $\{l_1, \dots, l_{n-k+1}\} = (\{1, \dots, n\} \setminus \mathcal{H}) \cup \{i\}$. Let $[u_r]_{r=1}^{k-1}$ and $[v_r]_{r=1}^{n-k+1}$ be the corresponding order statistics in the ascending order. Conditioning on the position of Q_{ii} in the order statistics of V_r , we have that for $k \geq 2$,

$$\begin{aligned} P(\tilde{\mathcal{H}}_{1,i}^{\mathcal{H}}) &= P\left(\sum_{r=1}^{k-1} u_r \leq kv_1 - Q_{ii}\right) \\ &= \sum_{j=1}^{n-k+1} P\left(\sum_{r=1}^{k-1} u_r \leq kv_1 - v_j, Q_{ii} = v_j\right) \\ &= \sum_{j=1}^{n-k+1} P(Q_{ii} = v_j) P\left(\sum_{r=1}^{k-1} u_r \leq kv_1 - v_j\right) \\ &\leq \frac{1}{n-k+1} k P\left(\sum_{r=1}^{k-1} u_r \leq kv_1\right) \\ &\leq \frac{1}{n-k+1} \frac{k(k-1)^{k-1}(n-k+1)!}{n!}, \end{aligned}$$

where the last two inequalities follow from Theorem 1. Q.E.D.

We next estimate the probability $P(\mathcal{H}_2^{\mathcal{H} \setminus \{i\}})$. For this purpose, we present the following technical result.

LEMMA 2. *For a given symmetric matrix Q , if $Q - \lambda E$ is PSD for some $\lambda \in \mathfrak{R}$, then the projection matrix of Q onto the null space of e is PSD. Moreover, it holds*

$$Q_{ij} \leq \frac{1}{2}(Q_{ii} + Q_{jj}) \leq \max(Q_{ii}, Q_{jj}), \quad \forall i, j; \quad (13)$$

and

$$Q_{ij} \leq \sqrt{Q_{ii} Q_{jj}}, \quad \forall i, j \in \mathcal{H} \quad (14)$$

when $\lambda \geq 0$.

PROOF. The first statement of the proposition follows directly from the fact that $Q - \lambda E$ is PSD, which further implies

$$\hat{Q} = \begin{pmatrix} Q_{ii} & Q_{ij} \\ Q_{ji} & Q_{jj} \end{pmatrix} - \lambda \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \succeq 0. \quad (15)$$

The positive semidefiniteness of the submatrix \hat{Q} is equivalent to the following conditions: $Q_{ii} - \lambda \geq 0$ and

$$(Q_{ij} - \lambda)^2 \leq (Q_{ii} - \lambda)(Q_{jj} - \lambda),$$

which implies that

$$Q_{ij} - \lambda \leq \sqrt{(Q_{ii} - \lambda)(Q_{jj} - \lambda)} \leq \frac{(Q_{ii} - \lambda) + (Q_{jj} - \lambda)}{2}.$$

Thus the inequality (13) holds. If $\lambda \geq 0$, then Q is the sum of two PSD matrices ($Q - \lambda E$) and λE and thus is PSD as well, which using the above argument yields the inequality (14). Q.E.D.

We now use the inequality (13) to provide a bound on the probability $P(Q - \lambda E \geq 0 \text{ for some } \lambda \in \mathfrak{R})$, which will easily apply to bound $P(\mathcal{H}_2^{\mathfrak{R}(i)})$.

THEOREM 3. *Let $Q \in \mathfrak{R}^{n \times n}$ be a symmetric matrix whose diagonal and lower diagonal elements are identically independently distributed with a continuous cumulative distribution function. Then it holds*

$$P(\exists \lambda \in \mathfrak{R}: Q - \lambda E \geq 0) \leq \frac{2^n}{(n+1)!}.$$

PROOF. From the last inequality in (13), conditioning on the event $\mathcal{A} = \{Q_{11} \geq \dots \geq Q_{nn}\}$, the event $\{\exists \lambda \in \mathfrak{R}: Q - \lambda E \geq 0\}$ implies the following event:

$$\mathcal{B} = \{Q_{ij} \leq Q_{kk}, \forall n \geq i \geq j \geq k \geq 1\}.$$

Define a series of events

$$\mathcal{B}_k = \{Q_{ik} \leq Q_{kk}, \forall n \geq i > k\}.$$

We have that

$$\mathcal{B} = \mathcal{A} \cap \bigcap_{k=1}^n \mathcal{B}_k.$$

Therefore

$$\begin{aligned} P(\exists \lambda \in \mathfrak{R}: Q - \lambda E \geq 0) &\leq P(\mathcal{B} | \mathcal{A}) \\ &= \frac{P(\mathcal{B})}{P(\mathcal{A})} \\ &= n! \int_{-\infty}^{\infty} dG(u_1) \int_{-\infty}^{u_1} dG(u_2) \dots \int_{-\infty}^{u_{n-1}} dG(u_n) \prod_{i=1}^n G(u_i)^{n-i} \\ &= n! \int_0^1 dt_1 \int_0^{t_1} dt_2 \dots \int_0^{t_{n-1}} dt_n \prod_{i=1}^n t_i^{n-i} \\ &= n! \prod_{i=1}^n \frac{1}{i(i+1)/2} \\ &= \frac{2^n}{(n+1)!}, \end{aligned}$$

where the third equality follows by a variable transformation. Q.E.D.

The above theorem can be improved for uniform/exponential distributions.

THEOREM 4. *If $Q \in \mathfrak{R}^{n \times n}$ be a symmetric matrix whose diagonal and lower diagonal elements are identically independently generated from the uniform distribution (i.i.u.d.) in $(0, 1)$, then it holds*

$$P(Q \geq 0) \leq \frac{2^n}{(n+1)^n}.$$

If $Q \in \mathfrak{R}^{n \times n}$ be a symmetric matrix whose diagonal and lower diagonal elements are identically independently generated from the exponential distribution (i.i.e.d.) in $(0, \infty)$ with rate parameter 1, then it holds

$$P(Q \geq 0) \leq \frac{2^n}{(n-1)^n}.$$

PROOF. If Q is generated from i.i.u.d. in $(0, 1)$ from (14), it follows:

$$\begin{aligned} P(Q \geq 0) &\leq P(Q_{ij} \leq Q_{ii}^{1/2} Q_{jj}^{1/2}, \forall 1 \leq i < j \leq n) \\ &= \int_0^1 \dots \int_0^1 \prod_{i < j} t_i^{1/2} t_j^{1/2} dt_1 \dots dt_n \\ &= \int_0^1 \dots \int_0^1 \prod_{i=1}^n t_i^{(n-1)/2} dt_1 \dots dt_n = \frac{2^n}{(n+1)^n}. \end{aligned} \tag{16}$$

In case of exponential distribution, from the first inequality in (13), we obtain

$$\begin{aligned}
 P(Q \geq 0) &\leq P\left(Q_{ij} \leq \frac{Q_{ii} + Q_{jj}}{2}, \forall i < j\right) \\
 &= \int_0^\infty \dots \int_0^\infty \prod_{i < j} \left(1 - \exp\left(-\frac{t_i + t_j}{2}\right)\right) \exp\left(-\sum_{i=1}^n t_i\right) dt_1 \dots dt_n \\
 &= \int_0^1 \dots \int_0^1 \prod_{i < j} (1 - \sqrt{u_i u_j}) du_1 \dots du_n \\
 &\leq \int_0^1 \dots \int_0^1 \left(1 - \frac{2 \sum_{i < j} \sqrt{u_i u_j}}{n(n-1)}\right)^{n(n-1)/2} du_1 \dots du_n \\
 &\leq \int_0^1 \dots \int_0^1 \left(1 - \prod_{i=1}^n u_i^{1/n}\right)^{n(n-1)/2} du_1 \dots du_n,
 \end{aligned}$$

where the last two inequalities follow from the inequality of arithmetic and geometric means. By using the variable transformation $u_i = v_i^n$ ($i = 1, \dots, n$), we obtain

$$\begin{aligned}
 P(Q \geq 0) &\leq n^n \int_0^1 \dots \int_0^1 \left(1 - \prod_{i=1}^n v_i\right)^{n(n-1)/2} \prod_{i=1}^n v_i^{n-1} dv_1 \dots dv_n \\
 &\leq n^n \int_0^1 \dots \int_0^1 \left(1 - \prod_{i=1}^n v_i\right)^{n(n-1)/2} \prod_{i=1}^n v_i^{n-i} dv_1 \dots dv_n \\
 &= \frac{n^n}{n(n-1)/2 + 1} \int_0^1 \dots \int_0^1 \left(1 - \prod_{i=1}^{n-1} v_i\right)^{(n(n-1))/2 + 1} \prod_{i=1}^{n-1} v_i^{n-1-i} dv_1 \dots dv_{n-1} \\
 &\quad \vdots \\
 &= \frac{n^n}{\prod_{i=1}^n (n(n-1)/2 + i)} \\
 &\leq \frac{2^n}{(n-1)^n}.
 \end{aligned} \tag{17}$$

This finishes the proof of the theorem. Q.E.D.

To build some intuition about the relative magnitudes of the probability bounds in Theorems 3 and 4, we use the well-known Stirling’s approximation for large factorials, which states

$$\lim_{n \rightarrow \infty} \frac{n!}{\sqrt{2\pi n} (n/\exp(1))^n} = 1.$$

It is clear that the bounds in Theorem 4 improve upon the one in Theorem 3 roughly by an order of $\exp(n)$.

Combining Theorem 3 and Lemma 1, we can obtain the following result.

THEOREM 5. *Assume that $Q \in \mathfrak{R}^{n \times n}$ is a symmetric matrix whose diagonal and lower diagonal elements are identically independently distributed with a continuous cumulative distribution function concave in its support. Let x^* be one global optimal solution of problem (1). It holds for $k \geq 2$*

$$P(\|x^*\|_0 = k) \leq \frac{(k-1)^{k-1} 2^{k-1}}{((k-1)!)^2}. \tag{18}$$

PROOF. From (11), Theorem 4 and Lemma 1, we obtain

$$\begin{aligned}
 P(\|x^*\|_0 = k) &\leq kC(n, k)P(\mathcal{H}_{1,i}^{\mathcal{H}})P(\mathcal{H}_2^{\mathcal{H} \setminus \{i\}}) \\
 &\leq \frac{n!}{k!(n-k)!} \frac{k^2(k-1)^{k-1}(n-k)!}{n!} \frac{2^{k-1}}{k!}.
 \end{aligned}$$

Thus the theorem holds. Q.E.D.

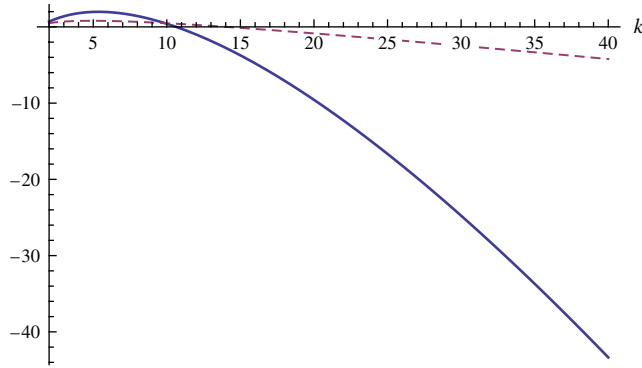


FIGURE 1. (Color online) Log of the bounds in Theorems 2 and 5.

We remark that comparing with the bound on $P(\|x^*\|_0 = k)$ in Chen et al. [9] (presented in Theorem 2), for small k , both bounds are not very meaningful. However, for reasonably large k , the new bound is tighter. Indeed, the Stirling’s approximation implies that the new bound is roughly in the order of $\exp(-O(k \ln k))$, while the one in Chen et al. [9] is $\exp(-O(k))$. To illustrate the significant improvement, we draw in Figure 1, the logarithm of the bounds for $k = 2, \dots, 40$, where the dashed curve and the thick curve are the bounds in Theorem 2 (as long as $n \geq 800$) and Theorem 5, respectively.

3.2. Model 2 with normal distribution. In this subsection, we focus on the scenario in which Q is generated by Model 2 with G being the standard normal distribution $\mathcal{N}(0, 1)$. In this case, the diagonal elements and the off-diagonal elements of Q have different variances. The matrix Q is referred to as the Gaussian Orthogonal Ensemble (GOE) (Mehta [19]) in the literature of random matrix theory.

Again, our task is to bound the probability $P(\tilde{\mathcal{H}}_{1,i}^{\mathcal{K}})$ and $P(\tilde{\mathcal{H}}_2^{\mathcal{K} \setminus \{i\}})$. Unfortunately, the analysis in the previous subsection does not carry over to the case with normal distribution. In fact, though Theorem 3 works for general distributions, Lemma 1 is only established for distributions with concave cumulative distribution functions.

To bound the probability $P(\tilde{\mathcal{H}}_{1,i}^{\mathcal{K}})$, we need the following key technical result. Its proof is quite involved and presented in the appendix.

THEOREM 6. *Let $U_i \sim N(0, 1)$ ($i = 1, \dots, k$) be independent and V_1 be the smallest order statistics of n independent standard normal random variables independent of U_i . There exists a constant $\eta > 0$ such that for any $1 \leq k \leq n$,*

$$\Pr\left(\frac{1}{k} \sum_{i=1}^k U_i \leq V_1\right) \leq \eta^k \left(\ln \frac{n+k}{k}\right)^{\lceil (k-1)/2 \rceil} B(k, n+1).$$

The above theorem allows us to bound the probability $P(\tilde{\mathcal{H}}_{1,i}^{\mathcal{K}})$.

THEOREM 7. *If $Q \in \mathfrak{R}^{n \times n}$ is GOE, we have that for $k \geq 2$.*

$$P(\tilde{\mathcal{H}}_{1,i}^{\mathcal{K}}) \leq \left(\eta^2 \ln \frac{n+k-1}{2k-2}\right)^{k-1} B(2k-2, n-k+2).$$

PROOF. Since Q is GOE, we have that $G = (M + M^T)/2$ with each component of $M \in \mathfrak{R}^{n \times n}$ being i.i.n.d. Similar to the proof of Theorem 11, we have that

$$\tilde{\mathcal{H}}_{1,i}^{\mathcal{K}} \subseteq \mathcal{C} = \left\{ \frac{1}{k-1} \sum_{j \in \mathcal{K} \setminus \{i\}} Q_{ij} < \min_{j \notin \mathcal{K} \setminus \{i\}} Q_{ij} \right\}.$$

The term $(1/(k-1)) \sum_{j \in \mathcal{K} \setminus \{i\}} Q_{ij}$ can be treated as the average of $2(k-1)$ independent standard normal random variables and $\min_{j \notin \mathcal{K} \setminus \{i\}} Q_{ij}$ is the smallest order statistics of $n-k+1$ independent standard normal random variables. Therefore the bound on $P(\mathcal{C})$ and thus $P(\tilde{\mathcal{H}}_{1,i}^{\mathcal{K}})$ follow by using Theorem 6. Q.E.D.

The above theorem can be combined with the bound of $P(\tilde{\mathcal{H}}_2^{\mathcal{K} \setminus \{i\}})$ in Theorem 3 to provide an upper bound of the probability $P(\|x^*\|_0 = k)$. Interestingly, a much stronger bound of $P(\tilde{\mathcal{H}}_2^{\mathcal{K} \setminus \{i\}})$ is available from the random matrix theory. To present the bound, note that the GOE enjoys the following nice properties (Mehta [19]):

P.1. The ensemble is invariant under the transformation

$$Q \rightarrow WQW^T,$$

where W is any real orthogonal matrix;

- P.2. The elements Q_{ij} , $i \leq j$ are statistically independent.
 P.3. It was shown in Dean and Majumdar [11] that for the GOE,

$$P(Q \geq 0) \leq \exp^{-(n^2/4)}.$$

To apply the above properties of GOE, choose a specific orthogonal matrix W such that its last row $W_{n,:} = e^T / \sqrt{n}$. It follows immediately that the principal submatrix of WQW^T with the first $n - 1$ rows and columns, $W_{1:n-1,:}QW_{1:n-1,:}^T$, is GOE and the property P.3 implies that

$$P(W_{1:n-1,:}QW_{1:n-1,:}^T \geq 0) \leq \exp\left(-\frac{(n-1)^2}{4}\right).$$

Since

$$W_{1:n-1,:}(Q - \lambda E)W_{1:n-1,:}^T = W_{1:n-1,:}QW_{1:n-1,:}^T, \quad \forall \lambda \in \mathfrak{R},$$

we immediately obtain the following result.

THEOREM 8. *If $Q \in \mathfrak{R}^{n \times n}$ is the GOE generated from (i.i.n.d.), then it holds*

$$P(\exists \lambda \in \mathfrak{R}: Q - \lambda E \geq 0) \leq \exp\left(-\frac{(n-1)^2}{4}\right).$$

Combing Theorems 7 and 8, we have the following result.

THEOREM 9. *Assume that $Q \in \mathfrak{R}^{n \times n}$ is GOE. Let x^* be one global optimal solution of problem (1). We have that for $k \geq 2$,*

$$P(\|x^*\|_0 = k) \leq \frac{(2k-3)!}{(k-1)!(n+1)^{k-2}} \left(\eta^2 \ln \frac{n+k-1}{2k-2}\right)^{k-1} \exp\left(-\frac{(k-2)^2}{4}\right).$$

PROOF. The result follows from (11), Theorem 7, Lemma 8, and the fact that $B(n, k) = (n-1)!(k-1)!/(n+k-1)!$ and

$$\frac{n!(n-k+1)}{(n+k-1)!} \leq \frac{1}{(n+1)^{k-2}}. \quad \text{Q.E.D.}$$

The above theorem implies that one can essentially exclude the possibility of global optimal solutions with three or more positive elements for large n .

4. Sparse approximate solutions to convex StQPs and extensions. In this section, we investigate the existence of sparse approximate solutions to StQPs with a symmetric PSD matrix $Q \in \mathfrak{R}^{n \times n}$. Different from the previous section, we focus on deterministic Q . Without loss of generality, we can further assume that all the matrix elements are positive. In fact, the optimal solution set of problem (1) is the same when Q is changed by adding λE for any λ . Under such an assumption, we can decompose the matrix into the following form

$$Q = V^T V, \quad V = [v_1; v_2, \dots, v_n],$$

where the i -th column of V represent a data point in \mathfrak{R}^d for some d with $\text{rank}(Q) \leq d \leq n$, and $Q_{ij} = v_i^T v_j > 0$ for $i, j = 1, \dots, n$. It follows immediately

$$x^T Q x = \|Vx\|_2^2,$$

and the StQP reduces to a problem of finding the shortest distance between the origin in \mathfrak{R}^d and the polytope $P(V)$ defined by

$$P(V) = \{v \in \mathfrak{R}^d: v = Vx, \|x\|_1 = 1, x \geq 0\}.$$

In other words, the StQP amounts to solving the following optimization problem:

$$\min_{v \in P(V)} \|v\|. \tag{19}$$

Since $Q_{ij} = v_i^T v_j > 0$ for $i, j = 1, \dots, n$, it must hold $0 \notin P(V)$. Therefore the optimal solution to problem (19) must be achieved at a facet of the polytope $P(V)$, which implies that the optimal solution v can be represented as a convex combination of at most d points. If $d \ll n$, this indicates that the corresponding StQP has a very sparse optimal solution.

We next consider the case where d is very large and in the same order of n . We first cite a well-known result for embedding by Johnson and Lindenstrauss (Indyk and Motwani [15], Johnson and Lindenstrauss [17]).

LEMMA 3. *Given $\epsilon > 0$ and an integer n . Let k be a positive integer such that $k \geq k_0 = \Theta(\epsilon^{-2} \log(n))$. For every set P of n points in \mathbb{R}^d , there exists a mapping $h: \mathbb{R}^d \rightarrow \mathbb{R}^k$ such that*

$$(1 - \epsilon)\|u - v\|^2 \leq \|h(u) - h(v)\|^2 \leq (1 + \epsilon)\|u - v\|^2, \quad \forall u, v \in P.$$

In particular, in Achlioptas [2], the author proved the following interesting result.

THEOREM 10. *Let P be an arbitrary set of n points in \mathbb{R}^d represented by a matrix $A \in \mathbb{R}^{d \times n}$. Given $\epsilon, \beta > 0$ and let*

$$k_0 = \frac{4 + 2\beta}{\epsilon^2/2 - \epsilon^3/3} \log(n).$$

For integer $k > k_0$, let $R \in \mathbb{R}^{k \times d}$ be a random matrix with $R(i, j) = r_{ij}$, where r_{ij} are independent random variables from either one of the following two probability distributions:

$$r_{ij} = \begin{cases} 1 & \text{with probability } 1/2 \\ -1 & \dots & 1/2 \end{cases} \quad (20)$$

$$r_{ij} = \sqrt{3} \times \begin{cases} 1 & \text{with probability } 1/6 \\ 0 & \dots & 2/3. \\ -1 & \dots & 1/6 \end{cases} \quad (21)$$

Let

$$M = \frac{1}{\sqrt{k}} RA.$$

Let $h: \mathbb{R}^d \rightarrow \mathbb{R}^k$ map the i -th column of A to the i -th column of M . With probability at least $1 - n^{-\beta}$, we have

$$(1 - \epsilon)\|u - v\|^2 \leq \|h(u) - h(v)\|^2 \leq (1 + \epsilon)\|u - v\|^2, \quad \forall u, v \in P.$$

To apply the above theorem to the StQP problem where $Q = V^T V$, we can first augment the matrix V by adding one more column (the origin in \mathbb{R}^d). We then use the above-described random projection method to the augmented data set V into a low dimensional space of size $\Theta(\log n / \epsilon^2)$. Next, we solve problem (19) with the projected data set excluding the origin. By using Theorem 10, one can show that with a high probability, the solution from the projected StQP in a low dimensional space provides an $(1 + \epsilon)$ -approximation to the StQP in the original space, which indicates with probability at least $1 - n^{-\beta}$, there exists a $(1 + \epsilon)$ -approximate solution to the original StQP whose sparsity is $\Theta(\log n / \epsilon^2)$.

We close this section by discussing potential extensions of the results in the previous sections to other classes of quadratic optimization problems. To start, let us consider a generic QP model defined by

$$\begin{aligned} \min \quad & x^T Q x + r^T x \\ \text{s.t.} \quad & Ax = b, \quad x \geq 0. \end{aligned} \quad (22)$$

Here $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. Let us consider the case that the feasible set, denoted by $\mathcal{F} = \{x: Ax = b, x \geq 0\}$, is a bounded polytope whose vertices belongs to the following set defined by

$$\mathcal{S} = \{s_i: i = 1, \dots, S\}.$$

From the above definition, for every $x \in \mathcal{F}$, there exists $\alpha_i \geq 0$, $i = 1, \dots, S$ such that

$$x = \sum_{i=1}^S \alpha_i s_i, \quad \sum_{i=1}^S \alpha_i = 1.$$

Therefore we can rewrite problem (22) as another StQP as follows:

$$\begin{aligned} \min \quad & \alpha^T \hat{Q} \alpha + \hat{r}^T \alpha = \alpha^T \bar{Q} \alpha \\ \text{s.t.} \quad & \sum_{i=1}^s \alpha_i = 1, \quad \alpha \geq 0, \end{aligned} \tag{23}$$

where

$$\hat{Q}_{ij} = s_i^T Q s_j, \quad \hat{r}_i = r^T s_i, \quad \bar{Q} = \hat{Q} + (\hat{r} e_s^T + e_s \hat{r}^T)/2.$$

If we assume that the matrix \bar{Q} is generated from a certain distribution, then by using the results in the previous sections, we can conclude that the global optimal solution of problem (23) is very sparse. Suppose that the global optimal solution to problem (23) has k positive elements, then the globally optimal solution to problem (22) can be written as a convex combination of k vertices of \mathcal{F} . Since $A \in \mathbb{R}^{m \times n}$, every vertex of \mathcal{F} has at most m positive elements. It follows immediately that the global optimal solution to problem (22) has at most km positive elements. Therefore, when m is a very small integer and the elements of the matrix \bar{Q} are generated from a certain distribution, we can conclude that with a high probability, the global optimal solution to problem (22) is very sparse.

A particular example that is worth mentioning is the mean-variance model introduced in Markowitz [18], where

$$\mathcal{F} = \{x: e^T x = 1, \mu^T x = \bar{\mu}, x \geq 0\}.$$

It has been observed for a long time that the optimal solution of the mean-variance model is dominated by only a few assets (Cornuejols and Tütüncü [10]), which correspond to a very sparse exact or approximate solution for the underlying QP. Note that for the mean-variance model, we have $m = 2$. Therefore the results in this paper provides a potential interesting interpretation for the observed phenomena with respect to the mean-variance model.

5. Conclusions. In this paper, we consider the so-called StQP problems that arise from various applications. Under certain probability models of the input data matrix, we show by performing probability analysis based on the first-order and second-order optimality conditions of the underling QP that there is a very high probability that the optimal solution to a random StQP has only a few nonzero elements. Our results significantly improve the corresponding ones in Chen et al. [9], employing the first-order optimality condition when the elements of the input data matrix are generated by continuous distribution functions concave in their supports. More importantly, we show that when the input data matrix is GOE, the global optimal solutions of StQP (1) essentially have at most two nonzero positive elements for large n . To prove the result, we establish a key technical result, which provides an upper bound of the probability of the event that the average of k independent standard normal random variables is no more than the smallest order statistics of n independent standard normal random variables, which we believe can be useful in other settings. We also show that if the input data matrix is a deterministic PSD matrix, with high probability a sparse approximate solution can be found.

Our results raise up several interesting research questions. First, we would like to know how tight our bounds are. Second, can we establish general and/or better results regarding the PSD of a randomly generated symmetric matrix? More specifically, can we improve the bounds in Theorems 3 and 4? Can we provide similar bounds like we do for GOE when $Q = (M + M^T)/2$ with each element of M being identically independently distributed with a general distribution? Third, our analysis for normal distribution only works for GOE. Can we provide similar bounds if Q is symmetric with identically independent distributed diagonal and lower triangle elements? Finally, we point out that though the results in this work can provide a partial interpretation regarding the mean-variance model for portfolio selection, it is far from satisfactory since we use some assumptions on the data matrix \bar{Q} , which is constructed from the original input data matrix Q and the vertices of the constraint set \mathcal{F} . It will be of interest to explore whether we can establish similar results based on characteristics of the input data matrices Q and the coefficient data matrix A directly.

Acknowledgments. The authors thank the reviewers for their constructive suggestions and Xiangyu Gao for help on LaTeX. The research of Xin Chen is partially supported by the National Science Foundation [Grants CMMI-1030923 and CMMI-1363261], and National Science Foundation of China [Grant 71228203]. The research of Jiming Peng is jointly supported by NSF [Grants DMS 09-15240 ARRA and CMMI-1131690].

Appendix A. Proof of Theorem 6. The proof is done in three steps. In the first step, we represent the probability as an integral and divide it into two parts. One part of the integral can be easily bounded from above by a function exponentially decaying in n . In the second step, we provide an upper bound for another part of the integral using a tight approximation of the complementary cumulative distribution function of the standard normal distribution. In the third step, we show that the upper bounds of the two parts of the integral have consistent orders.

Step 1. Denote $f(u) = \exp(-u^2/2)$. Let $\phi(u) = (1/\sqrt{2\pi})f(u)$, $\Phi(u) = \int_{-\infty}^u \phi(\xi)d\xi$, and $\Phi^c(u) = 1 - \Phi(u)$ be the pdf, cdf, and complementary cdf of the standard normal distribution. It is clear that

$$\Pr\left(\frac{1}{k} \sum_{i=1}^k U_i \leq V_1\right) = \frac{\sqrt{k}}{\sqrt{2\pi}} I_{n,k}(-\infty, \infty),$$

where

$$I_{n,k}(l, u) = \int_l^u f(\sqrt{k}u)\Phi^c(u)^n du.$$

To estimate $I_0(-\infty, \infty)$, we divide the integral into two parts:

$$I_{n,k}(-\infty, \infty) = I_{n,k}(-\infty, -1) + I_{n,k}(-1, \infty).$$

Since $\Phi^c(u)$ is decreasing in u ,

$$I_{n,k}(-1, \infty) \leq \Phi^c(-1)^n = \Phi(1)^n.$$

Step 2. It suffices to estimate $I_{n,k}(-\infty, -1)$. For this purpose, we use the following approximation of $\Phi^c(u)$ for $u \geq 0$ (see 7.1.13 in Abramowitz and Stegun [1]).

LEMMA 4. For $u \geq 0$,

$$\frac{2}{u + \sqrt{u^2 + 4}} \phi(u) < \Phi^c(u) \leq \frac{2}{u + \sqrt{u^2 + 8/\pi}} \phi(u).$$

Using the above lemma, we can immediately show that for $u \geq 1$,

$$\frac{1}{\rho u} f(u) < \Phi^c(u) \leq \frac{1}{\rho' u} f(u),$$

where $\rho' = \sqrt{2\pi}$ and $\rho = ((1 + \sqrt{5})/2)\sqrt{2\pi}$.

We now derive an upper bound of $I_{n,k}(-\infty, -1)$ using Lemma 4.

$$\begin{aligned} I_{n,k}(-\infty, -1) &= \frac{\sqrt{k}}{\sqrt{2\pi}} \int_{-\infty}^{-1} f(\sqrt{k}u)\Phi^c(u)^n du \\ &= \frac{\sqrt{k}}{\sqrt{2\pi}} \int_1^{\infty} f(\sqrt{k}u)(1 - \Phi^c(u))^n du \\ &\leq \int_1^{\infty} f(\sqrt{k}u) \left(1 - \frac{1}{\rho u} f(u)\right)^n du \\ &= \int_{v_0}^{\infty} f(\sqrt{k}v)(1 - f(v))^n (\rho u(v))^k \frac{v}{u(v) + 1/u(v)} dv, \end{aligned}$$

where the second equality follows from the symmetry of the normal distribution, the first inequality from Lemma 4, the last equality is derived by a change of variable $v = \sqrt{2(u^2/2 + \ln(\rho u))}$ and $v_0 = \sqrt{1 + 2\ln(\rho)}$. Since $u \leq v$, we have from the above formula that

$$I_{n,k}(-\infty, -1) \leq \rho^k J_{n,k},$$

where

$$J_{n,k} = \int_{v_0}^{\infty} f(\sqrt{k}v)(1 - f(v))^n v^k dv.$$

Making a change of variable $z = f(v)$, we have that

$$J_{n,k} = 2^{(k-1)/2} \int_0^{z_0} z^{k-1} (1-z)^n (-\ln z)^{(k-1)/2} dz,$$

where $z_0 = f(v_0) < 1$.

To provide an upper bound for $J_{n,k}$, we analyze the derivatives of the beta function. Define for a natural number i ,

$$\tau_i = (-1)^i \frac{d^i B(k, n+1)}{dk^i}.$$

Note that

$$\tau_i = \int_0^1 z^{k-1} (1-z)^n (-\ln z)^i dz.$$

Since $B(k, n+1) = \Gamma(k)\Gamma(n+1)/\Gamma(k+n+1)$ with the gamma function

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt,$$

we have that

$$\tau_1 = -B(k, n+1)(\psi_0(k) - \psi_0(k+n+1)),$$

where

$$\psi_i(z) = \frac{d^{i+1} \ln \Gamma(z)}{dz^{i+1}} = (-1)^{i+1} i! \sum_{l=0}^{\infty} \frac{1}{(z+l)^{i+1}}$$

is the polygamma function (for the definitions and relationships of the special functions, see Chapter 6 in Abramowitz and Stegun [1]). Therefore

$$\tau_i = \sum_{j=0}^{i-1} C(i-1, j) (-1)^{i-j} (\psi_{i-1-j}(k) - \psi_{i-1-j}(k+n+1)) \tau_j.$$

We now provide an upper bound on τ_i . From the above recursion, we have that for any $i-j > 1$ and $i \leq k/2$,

$$\begin{aligned} C_{i-1}^j (-1)^{i-j} (\psi_{i-1-j}(k) - \psi_{i-1-j}(k+n+1)) &= \frac{(i-1)!}{j!} \sum_{l=0}^n \frac{1}{(k+l)^{i-j}} \\ &\leq \frac{(i-1)!}{j!} \frac{1}{k^{i-j-2}} \sum_{l=0}^{\infty} \frac{1}{(k+l)(k+l-1)} \\ &= \frac{(i-1)!}{j!} \frac{1}{k^{i-j-2}} \frac{1}{k-1} \\ &\leq \left(\frac{1}{2}\right)^{i-j-1}. \end{aligned}$$

Thus we have

$$\begin{aligned} \tau_i &\leq \tau_{i-1} \sum_{l=0}^n \frac{1}{k+l} + \sum_{j=0}^{i-2} \tau_j \left(\frac{1}{2}\right)^{i-j-1} \\ &\leq \max_{j=0, \dots, i-1} \tau_j \left(1 + \sum_{l=0}^n \frac{1}{k+l}\right) \\ &\leq \max_{j=0, \dots, i-1} \tau_j \left(\ln(n+k) - \ln(k) + \frac{1}{k} + 1\right) \\ &\leq \left(\ln\left(\frac{\exp(2)(n+k)}{k}\right)\right)^i \tau_0 \\ &= \left(\ln\left(\frac{\exp(2)(n+k)}{k}\right)\right)^i B(k, n+1). \end{aligned}$$

We now use the above result on τ_i to find an upper bound for $J_{n,k}$ as follows:

$$\begin{aligned} J_{n,k} &= 2^{(k-1)/2} \int_0^{z_0} z^{k-1} (1-z)^n (-\ln z)^{(k-1)/2} dz \\ &\leq 2^{(k-1)/2} \int_0^{z_0} z^{k-1} (1-z)^n [(-\ln z)^{\lceil (k-1)/2 \rceil} + (-\ln z)^{\lfloor (k-1)/2 \rfloor}] / 2 dz \\ &\leq 2^{(k-1)/2} \int_0^1 z^{k-1} (1-z)^n [(-\ln z)^{\lceil (k-1)/2 \rceil} + (-\ln z)^{\lfloor (k-1)/2 \rfloor}] / 2 dz \\ &\leq \frac{\sqrt{k}}{\sqrt{2\pi}} 2^{(k-1)/2} \left[\left(\ln\left(\frac{\exp(2)(n+k)}{k}\right)\right)^{\lceil (k-1)/2 \rceil} + \left(\ln\left(\frac{\exp(2)(n+k)}{k}\right)\right)^{\lfloor (k-1)/2 \rfloor} \right] \cdot \frac{1}{2} \cdot B(k, n+1) \\ &\leq \frac{\sqrt{k}}{\sqrt{2\pi}} 2^{(k-1)/2} \left(\ln\left(\frac{\exp(2)(n+k)}{k}\right)\right)^{\lceil (k-1)/2 \rceil} B(k, n+1), \end{aligned}$$

where the first inequality follows from the convexity of x^n in n , and the last inequality holds since the \ln term is greater than 1.

Step 3. The above analysis implies that

$$\Pr\left(\frac{1}{k} \sum_{i=1}^k U_i \leq V_1\right) \leq \rho^k J_{n,k} + \Phi(1)^n.$$

When k is fixed as a constant, the first term in the above inequality dominates the second term. However, when k is large, their relationship is less clear. We now show that there exists a constant $\gamma \geq 1$ such that $\Phi(1)^n$ is dominated by $\gamma^k B(k, n+1)$.

For this purpose, we use the Stirling's formula $n! = \sqrt{2\pi n}(n/e)^n e^{\lambda_n}$ with $1/(12n+1) < \lambda_n < 1/(12n)$ (see 6.1.37 in Abramowitz and Stegun [1]). We have that for $k > 1$,

$$\begin{aligned} \ln B(k, n+1) &= \ln \frac{(k-1)!n!}{(n+k)!} \\ &\geq \ln \sqrt{\frac{2\pi(k-1)n}{(n+k)}} + (k-1)\ln(k-1) + n\ln n - (n+k)\ln(n+k) \\ &= \ln \sqrt{\frac{2\pi n}{(n+k)(k-1)}} + k\ln(1-1/k) + k[-\ln(1+n/k) - (n/k)\ln(1+k/n)] \\ &\geq \ln \sqrt{\frac{2\pi}{2k}} + k\ln(1-1/k) + n\ln \Phi(1) \\ &\quad - k \max_{0 \leq \alpha \leq 1} \{\ln(1+1/\alpha) + \ln[(1+\alpha)\Phi(1)]/\alpha\} \\ &= \ln \sqrt{\frac{2\pi}{2k}} + k\ln(1-1/k) - k\ln \delta + n\ln \Phi(1), \end{aligned}$$

where the first inequality follows from the Stirling's formula, and

$$\delta = \max_{\alpha \in [0, 1]} (1+1/\alpha)[(1+\alpha)\Phi(1)]^{1/\alpha}.$$

Therefore, for $k > 1$,

$$\Phi(1)^n \leq \sqrt{\frac{2k}{2\pi}} (1-1/k)^{-k} \delta^k B(k, n+1),$$

and hence

$$\Pr\left(\frac{1}{k} \sum_{i=1}^k U_i \leq V_1\right) \leq \frac{\sqrt{k}}{\sqrt{2\pi}} \left(2^{(k-1)/2} \rho^k \left(\ln\left(\frac{\exp(2)(n+k)}{k}\right)\right)^{[(k-1)/2]} + \sqrt{2}(1-1/k)^{-k} \delta^k\right) B(k, n+1).$$

Using Mathematica, we observe that $\delta \leq 6.4$. It is clear that there exists a constant $\eta > 0$ such that for any $k > 1$,

$$P\left(\frac{1}{k} \sum_{i=1}^k U_i \leq V_1\right) \leq \eta^k \left(\ln\left(\frac{\exp(2)(n+k)}{k}\right)\right)^{[(k-1)/2]} B(k, n+1).$$

Apparently, the $\exp(2)$ term inside \ln can be combined with the η term. For $k = 1$, we have that

$$P\left(\frac{1}{k} \sum_{i=1}^k U_i \leq V_1\right) = \frac{1}{n+1} = B(k, n+1),$$

and thus Theorem 6 holds.

Appendix B. Model 2 with concave cumulative distribution function. When Q is generated by Model 2 with a continuous cumulative distribution function concave in its support, it is much more complicated to provide an upper bound for $P(Q - \lambda E \geq 0 \text{ for some } \lambda \in \mathfrak{R})$. However, Theorem 1 alone allows to provide an upper bound employing the approach in Chen et al. [9].

THEOREM 11. *Assume that $Q = (M + M^T)/2$ in which all elements of $M \in \mathfrak{R}^{n \times n}$ are identically independently distributed with a continuous cumulative distribution function concave in its support. We have that for $k \geq 2$,*

$$P(\|x^*\|_0 = k) \leq \frac{k(2k-2)^{2k-2}}{\prod_{j=1}^{k-2} (n+1+j)}.$$

PROOF. Define

$$\mathcal{C} = \left\{ \frac{1}{k-1} \sum_{j \in \mathcal{K} \setminus \{i\}} Q_{ij} < \min_{j \in \mathcal{K} \setminus \{i\}} Q_{jj} \right\}.$$

Note that $\tilde{\mathcal{H}}_1^{\mathcal{K}} \subseteq \mathcal{C}$. The term $(1/(k-1)) \sum_{j \in \mathcal{K} \setminus \{i\}} Q_{ij}$ can be treated as the average of $2(k-1)$ independent random variables and $\min_{j \in \mathcal{K} \setminus \{i\}} Q_{jj}$ is the smallest order statistics of $n-k+1$ independent random variables, each of which has the same continuous cumulative distribution function concave in its support. Therefore

$$\begin{aligned} P(\|x^*\|_0 = k) &\leq kC(n, k)P(\tilde{\mathcal{H}}_{1,i}^{\mathcal{K}}) \\ &\leq kC(n, k)P(\mathcal{C}) \\ &\leq k(2k-2)^{2k-2} \frac{n!(n-k+1)!}{k!(n-k)!(n+k-1)!} \\ &\leq \frac{k(2k-2)^{2k-2}}{\prod_{j=1}^{k-2} (n+1+j)}, \end{aligned}$$

which gives the bound in the theorem. Q.E.D.

Observe that for any fixed $k > 2$, the above bound goes to zero as n goes to infinity. Unfortunately, when k is large, the above probability bound is not as good as the bound in Theorem 5.

References

- [1] Abramowitz M, Stegun IA (1972) *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables* (Dover, New York).
- [2] Achlioptas D (2001) Database-friendly random projections. *Proc. 20th Annual ACM SIGACT-SIGMOD-SIGART Sympos. Principles of Database Systems* (ACM, New York), 274–281.
- [3] Beier R, Vöcking B (2003) Random knapsack in expected polynomial time. *Proc. Thirty-Fifth Annual ACM Sympos. Theory Comput.* (ACM, New York), 232–241.
- [4] Bomze IM (1998) On standard quadratic optimization problems. *J. Global Optim.* 13:369–387.
- [5] Bomze IM, Locatelli M, Tardella F (2008) New and old bounds for standard quadratic optimization: Dominance, equivalence and incomparability. *Math. Programming* 1(115):31–64.
- [6] Candès E, Wakin M (2008) An introduction to compressive sampling. *IEEE Signal Processing Magazine* 25(2):21–30.
- [7] Candès E, Romberg J, Tao T (2006) Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory* 52(2):489–509.
- [8] Candès E, Li XD, Ma Y, Wright J (2011) Robust principal component analysis. *J. ACM* 58(1):1–37.
- [9] Chen X, Peng J, Zhang S (2013) Sparse solutions to random standard quadratic optimization problems. *Math. Programming* 141(1–2):273–293.
- [10] Cornuéjols G, Tütüncü R (2007) *Optimization Methods in Finance* (Cambridge University Press, Cambridge, UK).
- [11] Dean DS, Majumdar SN (2008) Extreme value statistics of eigenvalues of Gaussian random matrices. *Physical Rev. E* 77(4):1–12.
- [12] Donoho D (2006) Compressed sensing. *IEEE Trans. Inform. Theory* 52(4):1289–1306.
- [13] Gibbons LE, Hearn DW, Pardalos P, Ramana MV (1997) Continuous characterizations of the maximal clique problem. *Math. Oper. Res.* 22(3):754–768.
- [14] Goldberg AV, Marchetti-Spaccamela A (1984) On finding the exact solution of a zero-one knapsack problem. *Proc. Sixteenth Annual ACM Sympos. Theory Comput.* (ACM, New York), 359–368.
- [15] Indyk P, Motwani R (1998) Approximate nearest neighbors: Towards removing the curse of dimensionality. *Proc. 30th Annu. ACM Sympos. Theory Comput.* (ACM, New York), 604–613.
- [16] Irabaki T, Katoh N (1988) *Resource Allocation Problems: Algorithmic Approaches* (MIT Press, Cambridge, MA).
- [17] Johnson WB, Lindenstrauss J (1984) Extensions of Lipschitz mapping into Hilbert space. *Contemporary Math.* 26:189–206.
- [18] Markowitz HM (1952) Portfolio selection. *J. Finance* 7:77–91.
- [19] Mehta ML (1991) *Random Matrices* (Academic Press, San Diego).
- [20] Mukherjee L, Singh V, Peng J, Hinrichs C (2010) Learning kernels for variants of normalized cuts: Convex relaxations and applications. *Proc. Conf. Comput. Vision and Pattern Recognition (CVPR), San Francisco* (—IEEE, Piscataway, NJ).
- [21] Recht B, Fazel M, Parrilo PA (2010) Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.* 52(3):471–501.
- [22] Scozzari A, Tardella F (2008) A clique algorithm for standard quadratic programming. *Discrete Appl. Math.* 156:2439–2448.
- [23] Yang S, Li X (2009) Algorithms for determining the copositivity of a given matrix. *Linear Algebra Appl.* 609–618.