Math. Program., Ser. A (2013) 141:273–293
DOI 10.1007/s10107-012-0519-x

FULL LENGTH PAPER

Sparse solutions to random standard quadratic optimization problems

Xin Chen · Jiming Peng · Shuzhong Zhang

Received: 14 October 2010 / Accepted: 14 February 2012 / Published online: 9 March 2012

Abstract 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 this paper we focus on a special scenario of the StQP where all the elements of the data matrix Q are independently identically distributed and follow a certain distribution such as uniform or exponential distribution. We show that the probability that such a random StQP has a global optimal solution with k nonzero elements decays exponentially in k. Numerical evaluation of our theoretical finding is discussed as well.

Keywords Quadratic optimization · Semidefinite optimization · Relaxation · Computational complexity · Order statistics · Probability analysis

Mathematics Subject Classification 90C20 · 90C26 · 60A99

X. Chen · J. Peng (✉)
Department of Industrial and Enterprise System Engineering,
University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
e-mail: pengj@illinois.edu

X. Chen
e-mail: xinchen@illinois.edu

S. Zhang
Industrial and Systems Engineering Program, University of Minnesota,
Minneapolis, MN 55455, USA
e-mail: zhangs@umn.edu

1 Introduction

In this paper, we consider the following quadratic optimization problem

$$\begin{aligned} \min \quad & x^T Qx \\ s.t. \quad & e^T x = 1, \quad x \geq 0, \end{aligned} \tag{1}$$

where $Q = [Q_{ij}] \in \Re^{n\times n}$ is a symmetric matrix, and $e \in \Re^n$ is the all 1-vector. Like in [4], we call the above problem the standard quadratic optimization problem (StQP). The problem appears in numerous applications such as resource allocation [21] and portfolio selection [23]. It covers also other problems such as the maximal clique problem in discrete optimization [16] and determining the co-positivity of a matrix in linear algebra [8,22,28]. It is known that the StQP is NP-hard [25]. As such, the study of StQP has attracted the attention of researchers in various fields and a host of algorithms have been proposed in the literature. For details, we refer to recent papers [8,26,30] and the references therein.

In this work we focus on a special scenario of StQPs where the matrix Q is random. There are several motivations that inspire our work. The first is the observation that some discrete optimization problems such as the knapsack problem, though NP-hard from a theoretical perspective, can be solved efficiently when the input data follow a certain distribution [2,17]. It is of interests to investigate whether we can obtain similar results for some hard continuous optimization problems. On the other hand, we note that due to the lack of testing problems, in many algorithmic works on StQPs, random instances are generated and employed to test the proposed algorithms [5,14,26,30]. Since a random matrix is usually generated by following a certain distribution, it is desirable to explore the properties of the resulting StQPs which may provide insights to the proposed algorithms and help the further development of effective algorithms for this class of problems.

Our work is also inspired by the recent progress in compressed sensing on finding the sparsest solution of a system of linear equations and inequalities [11,12,15]. Specifically, consider the following optimization problem raised in compressed sensing:

$$\begin{aligned} \min \quad & \|x\|_0 \\ s.t. \quad & Ax = b, \end{aligned} \tag{2}$$

where $\|x\|_0$ denotes the number of nonzero elements of x. The above problem is NP-hard in general [15]. An elegant approach to tackle the above problem is to replace the objective function by the L_1-norm, which leads to the following LP

$$\begin{aligned} \min \quad & \|x\|_1 \\ s.t. \quad & Ax = b. \end{aligned} \tag{3}$$

As proved in [11,12,15], under certain conditions, the optimal solution of problem (3) is also optimal to problem (2) with a high probability. In [9], the authors further discussed problem (3) with an additional requirement that x is nonnegative, which leads to the following problem

$$\begin{aligned} \min \quad & \|x\|_1 \\ s.t. \quad & Ax = b, \quad x \geq 0. \end{aligned} \tag{4}$$

To motivate the StQP model analyzed in this paper, denote a nontrivial sparsest solution of problem (4) by x^* satisfying $\rho = \|x^*\|_1 > 0$. Then x^* must be the sparsest solution of the following least squares optimization problem

$$\begin{aligned} \min \quad & \|Ax - b\|_2^2 \\ s.t. \quad & e^T x = \rho, \quad x \geq 0. \end{aligned} \tag{5}$$

Let $c = A^T b$. Due to the special constraints in the above problem, we have

$$b^T Ax = c^T x = \frac{1}{\rho} x^T (ec^T)x = \frac{1}{2\rho} x^T (ec^T + ce^T)x.$$

Ignoring the constant term in the objective, we can rewrite problem (5) as the following

$$\begin{aligned} \min \quad & x^T \left(A^T A - \frac{1}{\rho}(ec^T + ce^T) \right) x \\ s.t. \quad & e^T x = \rho, \quad x \geq 0. \end{aligned} \tag{6}$$

Note that the above problem is homogeneous in x. Without loss of generality, we can replace the constraints by $e^T x = 1, x \geq 0$ and thus problem (6) reduces to a special case of the StQP model (1). The above interesting relation and the established results on L_1 optimization indicate that for some StQPs, global optimum solutions are expected to be very sparse.

Our work is an initial attempt to establish the existence of sparse globally optimal solutions for a class of random matrices. Specifically, when the upper triangular elements of Q in problem (1) are identically independently distributed with a uniform or exponential distribution, we prove that the probability that there exists a globally optimal solution with at least k nonzero elements decays exponentially (in k).

It is worth mentioning that for problem (1), due to the specific geometric feature of the constraint set (all the extreme points are sparse), it always has sparse solutions whenever the matrix Q is negative semidefinite. For generic QPs, in [19] Hager et al. showed that if Q has $n - k$ negative eigenvalues, then there must be at least $n - k$ active constraints at an optimal solution x^*. The results in [19] were further extended by Bomze and Locatelli [7] and Hu et al. [20] where the authors showed that certain classes of nonconvex QP can be solved in polynomial time. By applying the result in [19] to the StQP, one can show that if Q has at most k positive eigenvalues, then there exists an optimal solution to the StQP model which has at most $k + 1$ nonzero elements. However, as observed in [6,24], even for generic indefinite matrix Q, there may still exist very sparse solutions to the corresponding StQP. It is also appropriate to point out that our simulation indicates that both the numbers of negative eigenvalues and positive eigenvalues are in the order of n for the random matrices Q considered here due to the semi-circular law [29].

We note that in a recent work [1], Barany et al. showed that for random 2-play/bimatrix games, there is a high probability that there is a Nash equilibrium with a support of small size (at most k). As observed in [31], a bimatrix game can be reformulated as a symmetric linear complementarity problem (LCP) which is closely related to StQP. A fully polynomial time approximation scheme was proposed in [31] to find a stationary point of the underlying QP derived from the symmetric LCP. Different from the above papers, we are mainly concerned about the global solution of the StQP.

The paper is organized as follows. In Sect. 2, we establish several probability results that are crucial in the analysis of the present work. In Sect. 3, we explore the properties of the constraint set of problem (1) and the properties of a random matrix to establish the main result in the paper. In Sect. 4, we explore the interrelation between problem (1) and its SDP relaxation. In Sect. 5 we evaluate the main theoretical result established in this work, and report some experimental results. Finally we conclude our paper with some remarks in Sect. 6.

A few sentences about notation. Throughout the paper, $Q_{\min}$ (or $Q_{\max}$) denotes the value of the minimal (or maximal) element of Q. We also refer to x^* as a sparsest globally optimal solution of problem (1) (unless else specified), i.e., it has the smallest number of positive components among all globally optimal solutions.

2 Some probability results

In this section we establish several probability results that will play an important role in our later analysis.

Let U_r, $r = 1, \ldots, n$ be independent continuous random variables each with a cumulative distribution $F(\cdot)$ and a probability density function $f(\cdot)$ and $u_1 \leq u_2 \leq \cdots \leq u_n$ be the order statistics of U_r's. Let V_r, $r = 1, 2, \ldots, m$ be independent continuous random variables each with a cumulative distribution $G(\cdot)$ and a probability density function $g(\cdot)$ and $v_1 \leq v_2 \leq \cdots \leq v_m$ are the order statistics of V_r's. For notational convenience, let $\bar{u}_k = \sum_{r=1}^{k} u_r$. Our goal is to develop an upper bound on the probability

$$\sum_{i=1}^{m} P(\bar{u}_k \leq (k+1)v_1 - v_i). \tag{7}$$

For simplicity of our analysis,[1] we make the following assumption on the distributions.

Assumption 2.1 (a) $G(\cdot)$ is concave on its support.
(b) There exists a constant $\alpha \in (0, 1]$ such that $G(t) \geq \alpha F(t)$ for any $x \in (-\infty, \infty)$.

Part (a) of Assumption 2.1 implies that the support of $g(\cdot)$ is an interval with a finite left end point. In addition, $g(\cdot)$ is nonincreasing in its support. Notable examples

[1] In the proof of Theorem 3.4, we will replace V_r and U_r in inequality (7) by the diagonal elements and the corresponding off-diagonal elements of Q respectively.

satisfying Assumption 2.1 part (a) include uniform and exponential distributions. When $\alpha = 1$, part (b) of Assumption 2.1 is equivalent to the statement that V_r is stochastically less than U_r. To avoid technicalities, we also make the innocuous assumption that both $f(\cdot)$ and $g(\cdot)$ are continuous.

To derive an upper bound on the probability (7), we will use the joint pdf of order statistics. Specifically, the joint pdf of $u_{n_1}, \ldots, u_{n_k}$ $(0 = n_0 < n_1 < \cdots < n_k < n_{k+1} = n + 1)$ for $-\infty = u_0 \le u_1 \le u_2 \le \cdots \le u_k \le u_{k+1} = \infty$ can be written as follows (see (2.2.2) in [13]):

$$f_{(n_1)\ldots(n_k)}(u_1, \ldots, u_k) = n! \prod_{j=1}^{k} f(u_j) \prod_{j=0}^{k} \frac{[F(u_{j+1}) - F(u_j)]^{n_{j+1}-n_j-1}}{(n_{j+1} - n_j - 1)!}.$$

We now state the major result in this section.

Theorem 2.2 *Let $U_r, r = 1, \ldots, n$ and $V_r, r = 1, \ldots, m$ be independently identically distributed (i.i.d.) 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 2.1, then we have*

$$P(\bar{u}_k \le kv_1) \le \prod_{j=1}^{k} \frac{(m+n-k)(n-j+1)}{(n-k+\frac{\alpha m(k-j+1)}{k})(m+n-j+1)}; \tag{8}$$

$$\sum_{i=2}^{m} P(\bar{u}_k \le (k+1)v_1 - v_i) \le kP(\bar{u}_k \le kv_1). \tag{9}$$

Proof For notational convenience, let us denote $\bar{n}_k = n - k$, and $F^{-1}(u) = \inf\{\xi : F(\xi) \ge u\}$. To prove relation (8), we observe that

$$\begin{aligned}
P(\bar{u}_k \le kv_1) &= \frac{n!}{\bar{n}_k!} \int_{-\infty}^{\infty} \int_{u_1}^{\infty} \cdots \int_{u_{k-1}}^{\infty} \prod_{j=1}^{k} f(u_j)[1-F(u_k)]^{\bar{n}_k}[1-G(\bar{u}_k/k)]^m d_{u_k} \cdots d_{u_1} \\
&= \frac{n!}{\bar{n}_k!} \int_{0}^{1} \int_{u_1}^{1} \cdots \int_{u_{k-1}}^{1} (1-u_k)^{\bar{n}_k} \left[1-G\left(\frac{1}{k}\sum_{j=1}^{k} F^{-1}(u_j)\right)\right]^m d_{u_k} \cdots d_{u_1} \\
&\le \frac{n!}{\bar{n}_k!} \int_{0}^{1} \int_{u_1}^{1} \cdots \int_{u_{k-1}}^{1} (1-u_k)^{\bar{n}_k} \left[1-\frac{1}{k}\sum_{j=1}^{k} G(F^{-1}(u_j))\right]^m d_{u_k} \cdots d_{u_1} \\
&\le \frac{n!}{\bar{n}_k!} \int_{0}^{1} \int_{u_1}^{1} \cdots \int_{u_{k-1}}^{1} (1-u_k)^{\bar{n}_k} \left[1-\frac{\alpha}{k}\bar{u}_k\right]^m d_{u_k} \cdots d_{u_1} \\
&\le \frac{n!}{\bar{n}_k!} \int_{0}^{1} \int_{u_1}^{1} \cdots \int_{u_{k-1}}^{1} \left[1-\frac{m\alpha\bar{u}_{k-1}}{k(m+\bar{n}_k)} - \frac{\bar{n}_k + \frac{m\alpha}{k}}{m+\bar{n}_k}u_k\right]^{m+\bar{n}_k} d_{u_k} \cdots d_{u_1}
\end{aligned}$$

$$
\begin{aligned}
&\le \frac{n!}{\bar{n}_k!}\frac{m+\bar{n}_k}{(\bar{n}_k+\frac{m\alpha}{k})(m+\bar{n}_k+1)} \\
&\quad\times \int_0^1\int_{u_1}^1\cdots\int_{u_{k-2}}^1 \left[1-\frac{m\alpha\bar{u}_{k-2}}{k(m+\bar{n}_k)}-\frac{\bar{n}_k+\frac{2m\alpha}{k}}{m+\bar{n}_k}u_{k-1}\right]^{m+\bar{n}_k+1} d_{u_{k-1}}\cdots d_{u_1} \\
&\quad\vdots \\
&\le \frac{n!}{\bar{n}_k!}\prod_{j=1}^{k}\frac{m+\bar{n}_k}{(\bar{n}_k+mj\alpha/k)(m+\bar{n}_k+j)} \\
&= \prod_{j=1}^{k}\frac{(m+n-k)(n-j+1)}{(n-k+m(k-j+1)\alpha/k)(m+n-j+1)},
\end{aligned}
$$

where the first inequality follows from the concavity assumption of $G(\cdot)$, the second inequality from Assumption 2.1 part (b), the third inequality from the inequality of weighted arithmetic and geometric means, and the last two inequalities are derived by integrating over the variables and removing the negative terms sequentially. This proves inequality (8).

Now we turn to the proof of relation (9). By using the formula of the joint pdf, for every $i = 2, \dots, m$, we have

$$
\begin{aligned}
&P(\bar{u}_k \le (k+1)v_1 - v_i) \\
&= \frac{n!m!}{\bar{n}_k!(i-2)!(m-i)!}\int_{-\infty}^{\infty}\int_{u_1}^{\infty}\cdots\int_{u_{k-1}}^{\infty}\int_{\frac{\bar{u}_k}{k}}^{\infty}\int_{y_1}^{(k+1)v_1-\bar{u}_k}\prod_{j=1}^{k} f(u_j)[1-F(u_k)]^{\bar{n}_k} \\
&\quad\times g(v_1)g(v_2)[1-G(v_2)]^{m-i}[G(v_2)-G(v_1)]^{i-2} d_{v_2}d_{v_1}d_{u_k}\dots d_{u_1}
\end{aligned}
$$

It follows immediately

$$
\begin{aligned}
&\sum_{i=2}^{m} P(\bar{u}_k \le (k+1)v_1 - v_i) \\
&= \frac{n!m!}{\bar{n}_k!(m-2)!}\int_{-\infty}^{\infty}\int_{u_1}^{\infty}\cdots\int_{u_{k-1}}^{\infty}\int_{\frac{\bar{u}_k}{k}}^{\infty}\int_{v_1}^{(k+1)v_1-\bar{u}_k}\prod_{j=1}^{k} f(u_j)[1-F(u_k)]^{\bar{n}_k} \\
&\quad\times g(v_1)g(v_2)[1-G(v_1)]^{m-2} d_{v_2}d_{v_1}d_{u_k}\cdots d_{u_1} \\
&= \frac{n!m!}{\bar{n}_k!(m-2)!}\int_{-\infty}^{\infty}\int_{u_1}^{\infty}\cdots\int_{u_{k-1}}^{\infty}\int_{\frac{\bar{u}_k}{k}}^{\infty}\prod_{j=1}^{k} f(u_j)[1-F(u_k)]^{\bar{n}_k} g(v_1)[1-G(v_1)]^{m-2} \\
&\quad\times[G((k+1)v_1-\bar{u}_k)-G(v_1)] d_{v_1}d_{u_k}\cdots d_{u_1}
\end{aligned}
$$

$$= \frac{n!(m-1)}{\bar{n}_k!} \int_{-\infty}^{\infty} \int_{u_1}^{\infty} \cdots \int_{u_{k-1}}^{\infty} \prod_{j=1}^{k} f(u_j)[1-F(u_k)]^{\bar{n}_k} h(\bar{u}_k, v_1, m, k) d_{u_k} \cdots d_{u_1}, \tag{10}$$

where

$$\begin{aligned} h(\bar{u}_k, v_1, m, k) &= [1-G(\bar{u}_k/k)]^m - m \\ &\quad \times \int_{\frac{\bar{u}_k}{k}}^{\infty} [1-G((k+1)v_1-\bar{u}_k)] g(v_1)[1-G(v_1)]^{m-2} d_{v_1} \\ &= -\frac{[1-G(\bar{u}_k/k)]^m}{m-1} + \frac{m(k+1)}{m-1} \\ &\quad \times \int_{\frac{\bar{u}_k}{k}}^{\infty} g((k+1)v_1-\bar{u}_k)[1-G(v_1)]^{m-1} d_{v_1} \\ &\le -\frac{[1-G(\bar{u}_k/k)]^m}{m-1} + \frac{m(k+1)}{m-1} \int_{\frac{\bar{u}_k}{k}}^{\infty} g(v_1)[1-G(v_1)]^{m-1} d_{v_1} \\ &= \frac{k}{m-1}[1-G(\bar{u}_k/k)]^m. \end{aligned} \tag{11}$$

Here the inequality follows from the non-increasing property of the probability density function $g(\cdot)$. Combining relations (10) and (11), we obtain

$$\begin{aligned} &\sum_{i=2}^{m} P(\bar{u}_k \le (k+1)v_1 - v_i) \\ &\quad \le \frac{n!k}{\bar{n}_k!} \int_{-\infty}^{\infty} \int_{u_1}^{\infty} \cdots \int_{u_{k-1}}^{\infty} \prod_{j=1}^{k} f(u_j)[1-F(u_k)]^{\bar{n}_k} [1-G(\bar{u}_k/k)]^m d_{u_k} \cdots d_{u_1} \\ &\quad = kP(\bar{u}_k \le kv_1), \end{aligned} \tag{12}$$

which further implies (9). This finishes the proof of the theorem. □

For the special case with $m = n$, define

$$\rho(n, k) = P(\bar{u}_k \le kv_1). \tag{13}$$

We next derive a simpler bound for $\rho(n,k)$. Assume that $k \le \sqrt{2n\alpha}$. From Theorem 2.2, we obtain

$$
\begin{aligned}
\rho(n,k) &\le \prod_{j=1}^{k} \frac{2n-k}{n-k+n(k-j+1)\alpha/k}\,\frac{n-j+1}{2n-j+1} \\
&\le \prod_{j=1}^{k} \frac{2n-k}{2(n-k+n(k-j+1)\alpha/k)} \\
&= \left(\prod_{j=1}^{k} \frac{(2n-k)^2}{4(n-k+n(k-j+1)\alpha/k)(n-k+nj\alpha/k)}\right)^{1/2} \\
&\le \left(\frac{(2n-k)^2}{4(n-k+n\alpha)(n-k+n\alpha/k)}\right)^{k/2} \\
&= \left(\frac{(2n-k)^2}{(2n-k)^2+(2n-k)(2n\alpha(1+1/k)-2k)+(2n\alpha-k)(2n\alpha/k-k)}\right)^{k/2} \\
&\le \left(\frac{1}{1+\alpha(1+1/k)/2}\right)^{k/2} \\
&\le \left(\frac{1}{1+\alpha/2}\right)^{k/2},
\end{aligned}
\tag{14}
$$

where the third inequality holds since for fixed n and k, $4(n-k+n(k-j+1)\alpha/k)$ $(n-k+nj\alpha/k)$ is minimized at $j=1$ or $j=k$, and the fourth inequality follows from the assumption that $k \le \lfloor\sqrt{2n\alpha}\rfloor$. Since $\rho(n,k)$ is nonincreasing function of k, we have for $k \ge \lfloor\sqrt{2n\alpha}\rfloor$,

$$
\rho(n,k) \le \left(\frac{1}{1+\alpha/2}\right)^{\lfloor\sqrt{2n\alpha}\rfloor/2}.
$$

From the above discussion, we derive the following corollary.

Corollary 2.3 *Let $U_r, r=1,\ldots,n$ and $V_r, r=1,\ldots,n$ be i.i.d. 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 2.1, then we have*

$$
\rho(n,k) = P(\bar{u}_k \le k v_1) \le \begin{cases} \left(\frac{1}{1+\alpha/2}\right)^{k/2} & \textit{if } k \le \lfloor\sqrt{2n\alpha}\rfloor \\ \left(\frac{1}{1+\alpha/2}\right)^{\lfloor\sqrt{2n\alpha}\rfloor/2} & \textit{Otherwise.} \end{cases}
\tag{15}
$$

3 Sparseness in the global solutions to random StQPs

In this section we investigate the existence of sparse solutions to StQP with random matrices. In the remaining of this paper, without loss of generality, we assume that for a given random matrix Q, the relation $Q_{11} < Q_{22} < \cdots < Q_{nn}$ holds and the probabilities involving Q are always assumed to be conditional on this relation, though we will not explicitly write down the conditional event for simplicity of discussion. If a generated random matrix Q does not satisfy such a relation, then we can perform permutations to obtain a new matrix so that the relation holds. Because we are concerned only with the sparsity of the globally optimal solutions to the underlying problem, thus our results will not be affected by the permutation process.

We start with a simple setting in which the elements of Q of problem (1) are generated by a discrete distribution defined as follows:

$$\sum_{i=0}^{N} P(v = v_i) = \sum_{i=0}^{N} p_i = 1, \quad p_i > 0, \quad i = 0, \ldots, N. \tag{16}$$

Theorem 3.1 *Suppose that the upper triangular elements of matrix Q are generated from a discrete distribution (16) with the left end point of the support v_0. Let x^* be the sparsest globally optimal solution of problem problem (1). Then it holds*

$$P(\|x^*\|_0 = 1) \geq 1 - (1 - p_0)^n.$$

Proof Since x^* is a globally optimal solution to problem (1), we have that

$$(x^*)^T Q x^* \geq \min_{i,j} Q_{ij} \geq v_0.$$

Therefore, if there is a minimal diagonal element of Q equal to v_0, say $Q_{11} = v_0$, then $x^* = (1, 0, \ldots, 0)^T \in \Re^n$, the first unit vector, must be a globally optimal solution to problem (1). Since $P(Q_{11} = v_0) = p_0$, it follows

$$P(\|x^*\|_0 = 1) \geq P(Q_{11} = v_0) = 1 - (1 - p_0)^n.$$

This completes the proof of the theorem. □

The above theorem implies that as $n \to \infty$, with a high probability there exists an optimal solution with one positive element and the optimal objective value is given by the minimal diagonal element of Q. Interestingly, this result is not valid anymore if Q is generated from a continuous distribution as demonstrated in the following theorem.

Theorem 3.2 *Suppose that the upper triangular elements of matrix $Q \in \Re^{n \times n}$ are continuous i.i.d. random variables with CDF $G(\cdot)$ and pdf $g(\cdot)$. Then*

$$P(\|x^*\|_0 \geq 2) \geq \frac{n-1}{2n-1}.$$

Proof Consider the following 2-dimensional optimization problem

$$\begin{aligned} \min \quad & ax_1^2 + 2bx_1x_2 + cx_2^2 \\ s.t. \quad & x_1 + x_2 = 1, x_1 \geq 0, x_2 \geq 0. \end{aligned} \tag{17}$$

Let us consider a special scenario where $b < \min(a, c)$. Then the optimal solution of problem (17) is unique and given by

$$(x_1^*, x_2^*) = \left(\frac{c-b}{a+c-2b}, \frac{a-b}{a+c-2b} \right).$$

The above observation implies that if there exists an element on the first row smaller than Q_{11}, then $\|x^*\|_0 \geq 2$. Therefore,

$$\begin{aligned} P(\|x^*\|_0 \geq 2) \geq P(\exists j \neq 1 : Q_{1j} < Q_{11}) &= P(\min_{j:j\neq 1} Q_{1j} < Q_{11}) \\ &= \int_{-\infty}^{\infty} \int_{x}^{\infty} (n-1)g(t)(1-G(t))^{n-2} ng(y)(1-G(y))^{n-1} dydt \\ &= \frac{n-1}{2n-1}, \end{aligned}$$

where in the second equality $G(\cdot)$ and $g(\cdot)$ are the cdf and pdf which generate Q_{ij}, and we use the formula of joint pdf of the order statistics $\min_i Q_{ii}$ and $\min_{j:j\neq i} Q_{ij}$. □

Though the above theorem states that $P(\|x^*\|_0 \geq 2) \geq \frac{n-1}{2n-1} \to \frac{1}{2}$ as $n \to \infty$, we will show in the following that $P(\|x^*\|_0 \geq k)$ decays exponentially for a class of random matrices.

We start by exploring the optimality conditions for problem (1). Note that since the matrix Q is indefinite, there might exist multiple optimal solutions to problem (1). In particular, if there exists a globally optimal solution x^* to problem (1) with only one nonzero element ($\|x^*\|_0 = 1$), then such a solution can be identified easily by comparing all the diagonal elements of Q. Therefore, in what follows we concentrate on the cases where the sparsest globally optimal solution x^* to problem (1) has more than one nonzero elements. We have

Proposition 3.3 *Suppose that x^* is the sparsest globally optimal solution of problem (1) satisfying $\|x^*\|_0 = k > 1$. Let $\mathcal{K}$ be the index set of nonzero elements of x^* and $Q_{\mathcal{K}} \in \Re^{k\times k}$ be the corresponding principal submatrix of Q. Then the following conclusions hold:*

(I) *The matrix $Q_{\mathcal{K}} - (x^*)^T Qx^* E_k$ is positive semidefinite, where E_k is the k-by-k all ones matrix;*
(II) *There exists a row (or column) of $Q_{\mathcal{K}}$ such that the average of all its elements is strictly less than the minimal diagonal element Q_{11} of Q;*

Proof To prove the first statement, we observe that from the choice of the submatrix $Q_{\mathcal{K}}$, one can conclude that the following reduced StQP

$$\begin{aligned} \min \quad & y^T Q_{\mathcal{K}} y \\ s.t. \quad & e_k^T y = 1, \quad y \geq 0 \end{aligned} \tag{18}$$

has a positive globally optimal solution $y^* > 0$, where e_k is the all ones vector in $\Re^k$. Therefore, y^* is also the globally optimal solution of the following reduced problem

$$\begin{aligned} \min \quad & y^T Q_{\mathcal{K}} y \\ s.t. \quad & e_k^T y = 1. \end{aligned} \tag{19}$$

Using the first and second-order optimality conditions of the above problem, we obtain

$$Q_{\mathcal{K}} y^* - \lambda e_k = 0; \tag{20}$$

$$y^T Q_{\mathcal{K}} y \geq 0, \quad \forall\, e_k^T y = 0. \tag{21}$$

Here λ is the corresponding Lagrangian multiplier at the optimal solution. Because $(y^*)^T Q_{\mathcal{K}} y^* = (x^*)^T Q x^*$, from (20) we immediately have $\lambda = (x^*)^T Q x^*$. On the other hand, since for $e_k^T y = 0$, $y^T (Q_{\mathcal{K}} - \lambda E_k) y = y^T Q_{\mathcal{K}} y \geq 0$, we conclude that the matrix $Q_{\mathcal{K}} - \lambda E_k$ has at most one negative eigenvalue. Note that because

$$(Q_{\mathcal{K}} - \lambda E_k) y^* = Q_{\mathcal{K}} y^* - \lambda e_k = 0,$$

y^* is an eigenvector of the matrix $Q_{\mathcal{K}} - \lambda E_k$ corresponding to the eigenvalue 0. Suppose to the contrary that the matrix is not positive semidefinite, or in other words, there exists a vector y_- such that

$$y_-^T (Q_{\mathcal{K}} - \lambda E_k) y_- < 0.$$

From (21) we have $e_k^T y_- \neq 0$. Without loss of generality, let us assume $e_k^T y_- = 1$. Now let us consider a new vector

$$y_t = (1 - t) y^* + t y_-, \quad t \in [0, 1].$$

It is easy to see that for sufficiently small $t > 0$, the vector y_t remains still in the nonnegative orthant, and thus it is a feasible solution to the reduced problem (19). Moreover, we have

$$y_t^T (Q_{\mathcal{K}} - \lambda E_k) y_t < 0 \Longrightarrow y_t^T Q_{\mathcal{K}} y_t < \lambda,$$

which contradicts the global optimality of y^*. This proves the first statement of the proposition.

To prove the second statement of the proposition, we note that y^* is the globally optimal solution of (19) and $\lambda = (y^*)^T Q_{\mathcal{K}} y^*$. From the choice of $Q_{\mathcal{K}}$ we obtain

$$\lambda \leq Q_{11}.$$

From (20), we have

$$(y^*)^T Q_{\mathcal{K}} e_k = e_k^T Q_{\mathcal{K}} y^* = \lambda \leq Q_{11}.$$

Since $e_k^T y^* = 1$ and $y^* > 0$, the above relation implies that

$$\min(Q_{\mathcal{K}} e_k) \leq k\lambda \leq k Q_{11},$$

i.e., the smallest element of the vector $(Q_{\mathcal{K}} e_k)$ is no more than k times Q_{11}. This finishes the proof of the proposition. □

We are now ready to state the main result in this section.

Theorem 3.4 *Assume that the diagonal elements of the symmetric matrix Q are independently identically distributed with cdf $G(\cdot)$ and pdf $g(\cdot)$ and its strict upper triangular elements of Q are independently identically distributed with cdf $F(\cdot)$ and pdf $f(\cdot)$. Suppose that $G(\cdot)$ and $F(\cdot)$ satisfy Assumption 2.1. Let x^* be the sparsest globally optimal solution of problem (1). Then it holds*

$$P(\|x^*\|_0 \geq k) \leq \tau^{k-1}\left(\frac{1}{(1-\tau)^2} + \frac{k-1}{1-\tau}\right) + \frac{n(n+1)}{2}\tau^{\lfloor\sqrt{2n\alpha}\rfloor}, \tag{22}$$

where $\tau = (1+\alpha/2)^{-\frac{1}{2}}$ and α is a constant depending on the distribution functions F and G.

Proof Denote $Q_{i,:}$ the ith row of the matrix Q and $\bar{Q}_{i,:}$ the sorted sequence in increasing order consisting of all the elements in Q_{ij} for $i \neq j$. Define the following probability events

$$\mathcal{H}_i^k = \left\{ Q_{ii} + \sum_{j=1}^{k-1} \bar{Q}_{ij} \leq k Q_{11} \right\}; \quad i = 1, \ldots, n; \quad \mathcal{H}^k = \bigcup_{i=1}^{n} \mathcal{H}_i^k. \tag{23}$$

If the sparsest globally optimal solution to problem (1) has exactly k positive elements with $k > 1$, it follows from the second conclusion of Proposition 3.3 that there exists a row of the submatrix $Q_{\mathcal{K}}$ such that the average of its elements is strictly less than Q_{11}, where $\mathcal{K}$ is the index set of nonzero elements of x^*. This implies that there exists a row, say i, such that

$$Q_{ii} + \sum_{i=1}^{k-1} \bar{Q}_{ij} \leq k Q_{11}.$$

Therefore,

$$P(\|x^*\|_0 \geq k) \leq P\left(\bigcup_{i=k}^{n} \mathcal{H}^i\right).$$

From (12) and (14) we get

$$\begin{aligned}
P(\mathcal{H}^k) &= P\left(\bigcup_{i=1}^{n} \mathcal{H}_i^k\right) \\
&\leq \sum_{i=1}^{n} P(\mathcal{H}_i^k) \\
&= \sum_{i=1}^{n} P\left(\sum_{j=1}^{k-1} \bar{Q}_{ij} \leq kQ_{11} - Q_{ii}\right) \qquad (24) \\
&= \sum_{i=1}^{n} P\left(\sum_{j=1}^{k-1} u_j \leq kv_1 - v_i\right) \qquad (25) \\
&\leq k\rho(n, k-1),
\end{aligned}$$

where u and v are defined in Sect. 2 in the special case $m = n$. Finally, from Corollary 2.3 we obtain

$$\begin{aligned}
P\left(\bigcup_{i=k}^{n} \mathcal{H}^i\right) &\leq \sum_{i=k}^{n} P(\mathcal{H}^i) \\
&\leq \sum_{i=k}^{\lfloor\sqrt{2n\alpha}\rfloor} i\tau^{i-1} + \sum_{i=\lfloor\sqrt{2n\alpha}\rfloor+1}^{n} i\tau^{\lfloor\sqrt{2n\alpha}\rfloor} \\
&\leq \sum_{i=k}^{\infty} i\tau^{i-1} + \left(\frac{n(n+1)}{2} - \frac{\sqrt{2n\alpha}(\sqrt{2n\alpha}+1)}{2}\right)\tau^{\lfloor\sqrt{2n\alpha}\rfloor} \\
&\leq \tau^{k-1}\sum_{i=0}^{\infty}(i+k)\tau^i + \frac{n(n+1)}{2}\tau^{\lfloor\sqrt{2n\alpha}\rfloor} \\
&\leq \tau^{k-1}\left(\frac{1}{(1-\tau)^2} + \frac{k-1}{1-\tau}\right) + \frac{n(n+1)}{2}\tau^{\lfloor\sqrt{2n\alpha}\rfloor},
\end{aligned}$$

where the last inequality follows from the following relation

$$\sum_{i=0}^{\infty}(i+1)t^i = \left(\sum_{i=0}^{\infty} t^{i+1}\right)' = \frac{1}{(1-t)^2}, \quad \forall t \in [0, 1).$$

This finishes the proof of the theorem. □

The upper bound is valid for symmetric matrices whose upper triangular elements are independently generated by the distributions satisfying Assumption 2.1. Another common way of generating a symmetric random matrix is to generate all the elements of a matrix, say $\tilde{Q}$, independently and set $Q = \frac{\tilde{Q}+\tilde{Q}^T}{2}$. In this scenario, similar results can be established under Assumption 2.1.

Theorem 3.4 implies that the probability $P(\|x^*\|_0 \geq k)$ is bounded above by a function which decays exponentially in k for reasonably large k. However, for small k, the right hand side bound in relation (22) might turn out to be larger than 1. More accurate estimates can be obtained by invoking a more careful analysis. In the "Appendix", we present such a detailed analysis to estimate the probability $P(\|x^*\|_0 = 2)$.

4 The SDP relaxation of StQP

In Sect. 3, we have proved that there is a high probability that problem (1) has a sparse solution. We provide some computational results to validate our theoretical conclusion in Sect. 5. To find a globally optimal solution of problem (1), we propose to solve the SDP relaxation of problem (1) derived as follows. For a given feasible solution x of problem (1), denote $X = xx^T$. It is clear that all the elements of X are nonnegative and the summation of all the elements of X should be equal to 1. Moreover, X is positive semidefinite matrix of rank 1. By relaxing the rank constraint, we end up with the following SDP

$$\begin{aligned} \min\ & \operatorname{Tr}(QX) \\ s.t.\ & \operatorname{Tr}(EX) = 1, \\ & X \succeq 0,\ X_{ij} \geq 0, \quad i, j = 1, \ldots, n. \end{aligned} \tag{26}$$

Let X^* denote an optimal solution of the above SDP. It is easy to verify that if X^* is of rank one, then one can construct a globally optimal solution to problem (1) as follows:

$$x^* = (x_1^*, x_2^*, \ldots, x_n^*)^T, \quad x_i^* = X_{ii}^{1/2}.$$

Next we try to characterize the conditions under which we can extract the optimal solution to problem (1) from a solution X^* to problem (26). The following result is a slight refinement of Theorem 1 in [24].

Theorem 4.1 *Suppose that X^* be an optimal solution of problem (26). Let $\mathcal{K}$ be the index set induced by the nonzero diagonal elements of X^* and $X^*_{\mathcal{K}}$ be the principal submatrix of X^* induced by the index set $\mathcal{K}$. If all the elements in the principal submatrix $X^*_{\mathcal{K}}$ are positive, then we can extract an optimal solution to problem (1) from X^*.*

Proof By restricting the reduced SDP relaxation induced by the index set $\mathcal{K}$ if necessary, we can assume without loss of generality that all the elements of X^* are positive.

Under such an assumption, X^* must be the optimal solution of the following simplified SDP

$$\begin{aligned} \min \quad & \mathrm{Tr}(QX) \\ s.t. \quad & \mathrm{Tr}(EX) = 1, \\ & X \succeq 0. \end{aligned} \tag{27}$$

Let us consider the dual of the above problem (27) defined by

$$\begin{aligned} \max \quad & \lambda \\ s.t. \quad & \lambda E + Z = Q, Z \succeq 0. \end{aligned} \tag{28}$$

Let (λ^*, Z^*) be the optimal solution to the dual problem. We have

$$Q - \lambda^* E = Z^*, \quad Z^* X^* = 0, \quad X^*, Z^* \succeq 0. \tag{29}$$

Now let us consider the first eigenvector x^* of X^* corresponding to its largest eigenvalue. Since $X^* > 0$, it holds $x^* > 0$ (see Theorem 1.3 in Chapter 2 of [3]). By scaling if necessary, we can further assume that $\|x^*\|_1 = 1$. From (29) we can conclude

$$Z^* x^* = 0.$$

It follows

$$Q - \lambda^* E \succeq 0, \quad Qx^* - \lambda^* e = 0, \tag{30}$$

which is precisely the first-order and second-order optimality conditions for problem (1). In other words, x^* is the globally optimal solution of the following problem

$$\begin{aligned} \min \quad & x^T (Q - \lambda^* E) x \\ s.t. \quad & e^T x = 1, x \geq 0. \end{aligned} \tag{31}$$

Now recall the observation [6] that the optimal solution to the StQP (1) remains invariant if we replace the matrix Q by another matrix $Q + \alpha E$ for any $\alpha \in \Re$. Therefore, x^* must be the globally optimal solution to the original StQP. □

We remark that as shown in [3], for $n \leq 4$, the cone of completely positive matrices (the dual cone of copositive matrices) is identical to the cone of doubly nonnegative matrices (see also [10]). This implies there is no gap between the optimal values at the optimal solutions of problem (1) and its SDP relaxation (26). When $n \leq 3$, in [24] the authors discussed (see Theorem 2 of [24]) how to obtain an optimal solution of problem (1) from the optimal solution of problem (26). Such a result can be easily extended to the case $n = 4$. However, it is unclear how to recover the globally optimal solution to problem (1) from a solution of problem (26) for $n \geq 5$.

Table 1 Sparsity of optimal solutions to random StQPs

n	RAND: $\|x^*\|_0 = i$					RANDN: $\|x^*\|_0 = i$				
	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$
3	574	370	56	0	0	526	415	59	0	0
4	541	396	63	0	0	545	397	57	1	0
6	511	386	96	7	0	498	430	70	2	0
10	510	400	85	4	1	402	521	71	5	1
20	509	401	85	5	0	424	507	68	1	0
40	513	417	67	3	0	434	500	64	2	0
50	492	437	69	2	0	432	521	46	1	0

5 Numerical evaluation of main results

To gain some understanding of the strength of our theoretical results, we generate test problems using uniform distributions. For a given problem size n, we use the Matlab function RAND to generate a vector in the space of $\Re^{n(n+1)/2}$, which gives us the upper triangular elements of Q. We then define $Q_{ij} = Q_{ji}$ for $i > j$ to obtain the lower triangular elements of Q. After Q is generated, we construct a globally optimal solution from the solution of the relaxed problem (26). Interestingly, except for a couple cases, the global optimality of all the constructed solutions have been numerically verified by comparing against the solution of the relaxed problem (26).

Our experiments are done on a Pentium D with 3.6GHz CPU and 2 GB memory. We use the latest version of CVX [18] and SDPT3 [27] under Matlab R2006a to solve our problem. We report our numerical results in Table 1 for problems with sizes $n = 3, 4, 6, 10, 20, 40, 50$. For every fixed size, we test 1,000 random instances. In Table 1, the columns $i = k, k = 1, \dots, 5$ record the number of random instances among those 1,000 generated random instances whose optimal solution to problem (1) satisfies the condition $\|x^*\|_0 = i$. We note that the sparsest optimal solutions for all those generated random instances have no more than five nonzero elements and our numerical study indicates that the actual relative frequency $P(\|x^*\|_0 = k)$ for $k \geq 3$ can be significantly smaller than the upper bound we provide in Theorem 3.4. We also generate test problems with Normal distributions using the Matlab function RANDN and report the result under the column RANDN. Though we face some technical challenge to provide an upper bound on probability $P(\|x^*\|_0 = k)$ under Normal distribution, our numerical results indicate that sparse optimal solutions still exist for broader settings beyond the ones we proved.

6 Conclusions

In this paper, we consider the so-called standard quadratic optimization problems that arise from various applications. We concentrated on a special scenario where the input data matrix is random whose elements follow *i.i.d.* By employing probability analysis, optimality conditions of the underling QP and the combinatorial structure of the data matrix, we show that there is a very high probability that the optimal solution to a random StQP has only a few nonzero elements.

To gain some basic understanding of our theoretical findings, we propose to solve an SDP relaxation of the StQP. We characterize the conditions under which a solution to the original StQP can be recovered from a solution of the relaxed SDP.

There are several different ways to extend our results. First in our framework we assume all the elements of the input matrix satisfy Assumption 2.1, which requires the concavity of the cdf of the diagonal elements. This assumption can be slightly relaxed. More specifically, if the supports of $F(\cdot)$ and $G(\cdot)$ are lower bounded, without loss of generality, say by zero, we can assume that

Assumption 6.1 (a) $g(v') \leq \beta(k)g(v), \forall\, v' \leq (k+1)v$.
(b) $G(\frac{1}{k}\sum_{j=1}^{k} F^{-1}(u_j)) \geq \alpha(k)\frac{1}{k}\sum_{j=1}^{k} u_j$.

Under this assumption, we can show that

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

which differs slightly from inequality (8) by changing α to $\alpha(k)$ and multiplying the right hand side of (8) by $\beta(k)$. Imposing certain conditions on $\alpha(k)$ and $\beta(k)$, say for example assume $\alpha(k) \leq 1$, $\lim_{k\to\infty} k\alpha(k) = \infty$ and $\beta(k)$ does not increase exponentially fast, we can still establish an exponential decay bound similar to inequality (15).

It will be interesting to investigate whether we can establish similar results for other distributions.

Secondly, we are currently exploring more general optimization problem structures which may allow the existence of sparse optimal solutions.

Thirdly, our numerical study indicates that there is still a big gap between our theoretical estimation and the actual probabilities. In fact, our probability bound may be even greater than 1 for small k. In “Appendix A”, we prove that $P(\|x^*\|_0 = 2) \leq \frac{7}{12}$ by more careful analysis (As a comparison, note, by Theorem 3.2, that $P(\|x^*\|_0 \geq 2) \geq \frac{n-1}{2n-1}$). A similar approach may provide better bound for larger k. However, the analysis is quite challenging because we have to deal with complicated dependency of random events. It is our hope that refined optimality conditions can be utilized to narrow the gap. For instance, in addition to parts (I) and (II) of Proposition 3.3, we can show that the principal submatrix of Q associated with the indices of positive elements in a sparsest global optimal solution contains at least $2k-2$ off-diagonal elements that are strictly less than $Q^d_{\min}$, where k is the cardinality of a sparsest global optimal solution (see “Appendix B”).

Finally, we have observed that for random data, with a very high probability that the SDP relaxation of the StQP has a rank-1 solution. It is of great interest to develop a new theoretical analysis to explain the observation. Further study is needed to address these issues.

Acknowledgments We would like to thank many colleagues for their generous support in the preparation of the paper. In particular, we thank Prof. N. Balakrishnan, Prof. P. Glynn, Prof. H. N. Nagaraja and Prof. Z. J. Zhang for useful discussions with respect to order statistics. We also thank the two anonymous referees for their many useful suggestions that lead to substantial improvement of the paper. The research of Xin Chen is supported by NSF grant CMMI-1030923. The research of the Jiming Peng is jointly supported by AFOSR grant FA9550-09-1-0098, NSF grants DMS 09-15240 ARRA and CMMI-1131690.

Appendix A

In this appendix, we derive a sharper bound for $P(\|x^*\|_0 = 2)$ when $F(\cdot) = G(\cdot)$ and $G(\cdot)$ is concave. Recall the notation $\mathcal{H}_i^k$ defined in the proof in Theorem 3.4. We have

$$P(\|x^*\|_0 = 2) \le P\left(\bigcup_{i=1}^n \mathcal{H}_i^2\right) = P(\mathcal{H}_1^2) + P\left(\bigcup_{i=2}^n \mathcal{H}_i^2 \backslash \mathcal{H}_1^2\right)$$
$$\le P(\mathcal{H}_1^2) + \sum_{i=2}^n P(\mathcal{H}_i^2 \backslash \mathcal{H}_1^2).$$

Recall the proof of Theorem 3.2, we have

$$P(\mathcal{H}_1^2) = \frac{n-1}{2n-1}.$$

On the other hand, for $i \ge 2$, on the event $\mathcal{H}_i^2 \backslash \mathcal{H}_1^2$ we have $\min_{j \ne 1} Q_{j1} > Q_{11}$ and $\min_{j \ne i,1} Q_{ij} \le 2Q_{11} - Q_{ii}$. Thus, we have that

$$\sum_{i=2}^n P(\mathcal{H}_i^2 \backslash \mathcal{H}_1^2)$$
$$= (n-1)(n-2) \int_{-\infty}^{\infty} dt \int_t^{\infty} dz f(t)(1-F(t))^{n-3} f(z)(1-F(z))^{n-2} \int_t^z dy_1$$
$$\times \left\{ \int_{y_1}^{2y_1 - t} dy_2 \sum_{i=2}^n \frac{n!}{(i-2)!(n-i)!} g(y_1) g(y_2) [1-G(y_2)]^{n-i} [G(y_2) - G(y_1)]^{i-2} \right\}$$
$$= (n-1)(n-2) \int_{-\infty}^{\infty} dt \int_t^{\infty} dz f(t)(1-F(t))^{n-3} f(z)(1-F(z))^{n-2} \int_t^z dy_1$$
$$\times \left\{ n(n-1) \int_{y_1}^{2y_1 - t} dy_2 g(y_1) g(y_2) [1-G(y_1)]^{n-2} \right\}$$

$$= n(n-1)^2(n-2)\int_{-\infty}^{\infty} dt \int_{t}^{\infty} dy_1 \int_{y_1}^{\infty} dz f(t)(1-F(t))^{n-3} f(z)(1-F(z))^{n-2}$$

$$\times \left\{ n(n-1) \int_{y_1}^{2y_1-t} dy_2 g(y_1)g(y_2)[1-G(y_1)]^{n-2} \right\}$$

$$= n(n-1)(n-2)\int_{-\infty}^{\infty} dt \int_{t}^{\infty} dy_1 f(t)(1-F(t))^{n-3} f(z)(1-F(y_1))^{n-1}$$

$$\times \left\{ [G(2y_1-t)-G(y_1)]g(y_1)[1-G(y_1)]^{n-2} \right\}$$

$$= n(n-1)(n-2)\int_{0}^{1} dt \int_{t}^{1} dy_1 (1-t)^{n-3}(1-F(G^{-1}(y_1)))^{n-1}$$

$$\times \left\{ [G(2G^{-1}(y_1)-F^{-1}(t))-y_1][1-y_1]^{n-2} \right\}.$$

Since we assume $G(\cdot)$ is concave and $F(\cdot) = G(\cdot)$, we have the following

$$\sum_{i=2}^{n} P(\mathcal{H}_i^2 \backslash \mathcal{H}_1^2)$$

$$= \frac{n!}{(n-3)!}\int_{0}^{1} dt \int_{t}^{1} dy_1 (1-t)^{n-3}(1-y_1)^{2n-3}[G(2G^{-1}(y_1)-G^{-1}(t))-y_1]$$

$$\leq \frac{n!}{(n-3)!}\int_{0}^{1} dt \int_{t}^{1} dy_1 (1-t)^{n-3}(1-y_1)^{2n-3}(y_1-t)$$

$$= \frac{1}{12}\frac{n(n-2)}{(n-1/2)(n-1)},$$

where the first inequality holds since from the concavity of $G(\cdot)$,

$$G(G^{-1}(y_1)) \geq \frac{1}{2}[G(2G^{-1}(y_1)-G^{-1}(t))+G(G^{-1}(t))].$$

It follows immediately

$$P(\{\|x^*\|_0 = 2\}) \leq P(\mathcal{H}_1^2) + \sum_{i=2}^{n} P(\{\mathcal{H}_i^2\}\backslash\{\mathcal{H}_1^2\})$$

$$\leq \frac{n-1}{2n-1} + \frac{1}{12}\frac{n(n-2)}{(n-1/2)(n-1)}$$

$$\leq \frac{7}{12}.$$

Appendix B

We now show that the principal submatrix of Q associated with the indices of positive elements in a sparsest global optimal solution contains at least $2k-2$ off-diagonal elements that are strictly less than Q_{11}, where k is the cardinality of a sparsest global optimal solution x^*. Let $\mathcal{K}$ be the index set of nonzero elements and $Q_{\mathcal{K}}$ be the corresponding principal submatrix of Q. Assume to the contrary that $Q_{\mathcal{K}}$ contains less than $2k-2$ off-diagonal elements that are strictly less than Q_{11}. We now construct an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with the node set $\mathcal{V} = \mathcal{K}$ and the edge set $\mathcal{E} = \{(i, j) : Q_{ij} < Q_{11}, i, j \in \mathcal{K}\}$. Since $Q_{\mathcal{K}}$ is assumed to contain less than $2k-2$ off-diagonal elements that are strictly less than Q_{11}, the graph $\mathcal{G}$ has less than $2k-2$ edges, which immediately implies that $\mathcal{G}$ contains at least two independent subgraphs. Let $\mathcal{I}_1$ be one subgraph and $\mathcal{I}_2 = \mathcal{V} \backslash \mathcal{I}_1$. It is clear that $Q_{ij} \geq Q_{11}$ for $i \in \mathcal{I}_1$ and $j \in \mathcal{I}_2$ (or simply $(i, j) \in \mathcal{I}_{12}$) and

$$
\begin{aligned}
\lambda = (y^*)^T Q_{\mathcal{K}} y^* &= (y^*_{\mathcal{I}_1})^T Q_{\mathcal{I}_1} y^*_{\mathcal{I}_1} + 2(y^*_{\mathcal{I}_1})^T Q_{\mathcal{I}_1\mathcal{I}_2} y^*_{\mathcal{I}_2} + (y^*_{\mathcal{I}_2})^T Q_{\mathcal{I}_2} y^*_{\mathcal{I}_2} \\
&\geq (y^*_{\mathcal{I}_1})^T Q_{\mathcal{I}_1} y^*_{\mathcal{I}_1} + 2\|y^*_{\mathcal{I}_1}\|_1 \|y^*_{\mathcal{I}_2}\|_1 Q_{11} + (y^*_{\mathcal{I}_2})^T Q_{\mathcal{I}_2} y^*_{\mathcal{I}_2} \\
&> (y^*_{\mathcal{I}_1})^T Q_{\mathcal{I}_1} y^*_{\mathcal{I}_1} + 2\|y^*_{\mathcal{I}_1}\|_1 (1 - \|y^*_{\mathcal{I}_1}\|_1)\lambda + (y^*_{\mathcal{I}_2})^T Q_{\mathcal{I}_2} y^*_{\mathcal{I}_2} \\
&= (y^*_{\mathcal{I}_1})^T Q_{\mathcal{I}_1} y^*_{\mathcal{I}_1} + 2\|y^*_{\mathcal{I}_2}\|_1 (1 - \|y^*_{\mathcal{I}_2}\|_1)\lambda + (y^*_{\mathcal{I}_2})^T Q_{\mathcal{I}_2} y^*_{\mathcal{I}_2},
\end{aligned}
$$

where the second inequality follows from the fact $\lambda < Q_{11}$, and the last equality from the relation $\|y^*_{\mathcal{I}_1}\|_1 + \|y^*_{\mathcal{I}_2}\|_1 = 1$. Therefore, we have

$$
\begin{aligned}
(y^*_{\mathcal{I}_1})^T Q_{\mathcal{I}_1} y^*_{\mathcal{I}_1} + (y^*_{\mathcal{I}_2})^T Q_{\mathcal{I}_2} y^*_{\mathcal{I}_2} &< \lambda - \lambda(\|y^*_{\mathcal{I}_1}\|_1 - \|y^*_{\mathcal{I}_1}\|_1^2 + \|y^*_{\mathcal{I}_2}\|_1 - \|y^*_{\mathcal{I}_2}\|_1^2) \\
&= \lambda(\|y^*_{\mathcal{I}_1}\|_1^2 + \|y^*_{\mathcal{I}_2}\|_1^2).
\end{aligned}
$$

The above inequality implies that one of the following two inequalities must hold

$$
\frac{(y^*_{\mathcal{I}_1})^T Q_{\mathcal{I}_1} y^*_{\mathcal{I}_1}}{\|y^*_{\mathcal{I}_1}\|_1^2} < \lambda, \quad \frac{(y^*_{\mathcal{I}_2})^T Q_{\mathcal{I}_2} y^*_{\mathcal{I}_2}}{\|y^*_{\mathcal{I}_2}\|_1^2} < \lambda.
$$

This, however, contradicts to the global optimality of x^*. Therefore, the principal submatrix of Q associated with the indices of positive elements in a sparsest global optimal solution contains at least $2k-2$ off-diagonal elements that are strictly less than Q_{11}.

References

1. Barany, I., Vempala, S., Vetta, A.: Nash equilibria in random games. Random Struct. Algorithms **31**(4), 391–405 (2007)
2. Beier, R., Vöcking, B.: Random knapsack in expected polynomial time. In: Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing, pp. 232–241 (2003)
3. Berman, A., Plemmons, R.J.: Nonnegative Matrices in the Mathematical Sciences. SIAM Classics in Applied Mathematics, Philadelphia (1994)

4. Bomze, I.M.: On standard quadratic optimization problems. J. Global Optim. **13**, 369–387 (1998)
5. Bomze, I.M., D, M., de Klerk, E., Roos, C., Quist, A.J., Terlaky, T.: On copositive programming and standard quadratic optimization problems. J. Global Optim. **18**, 301–320 (2000)
6. Bomze, I.M., de Klerk, E.: Solving standard quadratic optimization problems via linear, semidefinite and copositive programming. J. Global Optim. **24**, 163–185 (2002)
7. Bomze, I.M., Locatelli, M.: Separable standard quadratic optimization problems. Optim. Lett. (2011). doi:10.1007/s11590-011-0309-z
8. Bomze, I.M., Locatelli, M., Tardella, F.: New and old bounds for standard quadratic optimization: dominance, equivalence and incomparability. Math. Program. **115**(1), 31–64 (2008)
9. Bruckstein, A.M., Elad, M., Zibulevsky, M.: A non-negative and sparse enough solution of an underdetermined linear system of equations is unique. In: IEEE Xplore, 3rd International Symposium on Communications, Control and Signal Processing, ISCCSP, pp. 762–767 (2008)
10. Burer, S., Anstreicher, K.M., Duer, M.: The difference between 5×5 doubly nonnegative and completely positive matrices. Linear Algebra Appl. **431**, 1539–1552 (2009)
11. Candès, E., Wakin, M.: An introduction to compressive sampling. IEEE Signal Process. Mag. **25**(2), 21–30 (2008)
12. Candès, E., Romberg, J., Tao, T.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inf. Theory **52**(2), 489–509 (2006)
13. David, H.A., Nagaraja, H.N.: Order Statistics. Wiley, NJ (2003)
14. de Klerk, E., Pasechnik, D.V.: Approximation of the stability number of a graph via copositive programming. SIAM J. Optim. **12**, 875–892 (2002)
15. Donoho, D.: Compressed sensing. IEEE Trans. Inf. Theory **52**(4), 1289–1306 (2006)
16. Gibbons, L.E., Hearn, D.W., Pardalos, P., Ramana, M.V.: Continuous characterizations of the maximal clique problem. Math. Oper. Res. **22**, 754–768 (1997)
17. Goldberg, A.V., Marchetti-Spaccamela, A.: On finding the exact solution of a zero-one knapsack problem. In: Proceedings of the Sixteenth Annual ACM Symposium on Theory of Computing, pp. 359–368 (1984)
18. Grant, M., Boyd, S., Ye, Y.: CVX Users' Guide (2007). http://www.stanford.edu/~boyd/cvx/cvx_usrguide.pdf.
19. Hager, W.W., Pardalos, P.M., Roussos, I.M., Sahinoglou, H.D.: Active constraints, indefinite quadratic test problems, and complexity. J. Opt. Theor. Appl. **68**, 499–511 (1991)
20. Hu, J., Mitchell, J.E., Pang, J.S.: An LPCC approach to nonconvex quadratic programs. Math. Program. Ser. A. doi:10.1007/s10107-010-0426-y
21. Irabaki, T., Katoh, N.: Resource Allocation Problems: Algorithmic Approaches. MIT Press, Cambridge (1988)
22. Kaplan, W.: A test for copositive matrices. Linear Algebra Appl. **313**, 203–206 (2000)
23. Markowitz, H.M.: Portfolio selection. J. Finance **7**, 77–91 (1952)
24. Mukherjee, L., Singh, V., Peng, J., Hinrichs, C.: Learning Kernels for variants of normalized cuts: convex relaxations and applications. In: Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR). San Francisco (2010)
25. Murty, K.G., Kabadi, S.N.: Some NP-complete problems in quadratic and nonlinear programming. Math. Program. **39**, 117–129 (1987)
26. Scozzari, A., Tardella, F.: A clique algorithm for standard quadratic programming. Discrete Appl. Math. **156**, 2439–2448 (2008)
27. Toh, K.C., Tütüncü, R.H., Todd, M.: On the implementation and usage of SDPT3 (2006). http://www.math.nus.edu.sg/~mattohkc/guide4-0-draft.pdf
28. Väliaho, H.: Quadratic programming criteria for copositive matrices. Linear Algebra Appl. **119**, 163–182 (1989)
29. Wigner, E.P.: On the distribution of the roots of certain symmetric matrices. Ann. Math. **67**, 325–327 (1958)
30. Yang, S., Li, X.: Algorithms for determining the copositivity of a given matrix. Linear Algebra Appl. **430**, 609–618 (2009)
31. Zhu, Z., Dang, C., Ye, Y.: A FPTAS for Computing a Symmetric Leontief Competitive Economy Equilibrium. Math. Program. (to appear)