Random walk in a simplex and quadratic optimization over convex polytopes

Yu. Nesterov *

September 2003

Abstract

In this paper we develop probabilistic arguments for justifying the quality of an approximate solution for global quadratic minimization problem, obtained as a best point among all points of a uniform grid inside a polyhedral feasible set. Our main tool is a random walk inside the standard simplex, for which it is easy to find explicit probabilistic characteristics. For any integer $k \geq 1$ we can generate an approximate solution with relative accuracy $\frac{1}{k}$ provided that the quadratic objective function is non-negative in all nodes of the feasible set. The complexity of the process is polynomial in the number of nodes and in the dimension of the space of variables. We extend some of the results to problems with polynomial objective function. We conclude the paper by showing that some related problems (maximization of cubic or quartic form over the Euclidean ball, and the matrix ellipsoid problem) are NP-hard.

Keywords: Global optimization, quadratic optimization, polynomial optimization, simplex structure, random walk, polynomial-time complexity.

*Center for Operations Research and Econometrics (CORE), Catholic University of Louvain (UCL), 34 voie du Roman Pays, 1348 Louvain-la-Neuve, Belgium; e-mail: nesterov@core.ucl.ac.be.

This paper presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming.
The scientific responsibility is assumed by the author.

1 Introduction

Motivation. In nonlinear optimization the problem of minimizing a general quadratic function over the standard simplex is considered as one of the most difficult. Maybe the most famous problem of this type is the Motzkin-Straus problem [3] (see (5.1)), for which the optimal value is equal to the inverse stability number of a graph.

It follow that for such a problem we cannot expect existence of a polynomial-time procedure. Therefore it looks reasonable to try to develop a polynomial-time scheme for finding solutions with *fixed* accuracy. A first scheme of this type was proposed in [4] for accuracy $\frac{1}{2}$. Later in [1], using a semidefinite programming technique and representations of the cones of sums of squares, this result was extended to arbitrary accuracy $\frac{1}{k}$, $k \geq 1$. Of course, the complexity of the computation grows exponentially with k.

In this paper we present another justification for the estimate proposed in [1]. Our analysis is much simpler. It is based on the properties of a *random walk* in the standard simplex. In order to estimate the quality of the approximate solution we only need to compute a covariance matrix of the random process. As a result, we get a sharper estimate than in [1]. For example, we show that our approximation has relative accuracy $\frac{1}{k}$ for matrices with non-positive diagonal elements. (In Theorem 3.2 in [1] this result was established for matrices with non-positive entries.) We extend our results to quadratic minimizations problems over polytopes defined as a convex combination of several nodes, and to problems with polynomial objective function.

Contents. In Section 2 we define a random walk in a standard simplex and study its properties. In Section 3 we apply the results to justify the quality of an approximate solution, which is the best point among all points of a "uniform" grid inside the simplex. This approach is extended also to arbitrary polytopes. In Section 4 we show that the random walk technique can be applied to problems with polynomial objective functions. Finally, in Section 5, we use the Motzkin-Straus problem in order to prove that some related problems (maximization of a cubic or quartic form over Euclidean ball, and the matrix ellipsoid problem) are NP-hard.

Notation. In what follows we denote by R^n an n-dimensional real vector space comprised of column vectors. The basis vectors of this space are denoted by e_i, $i = 1, \dots, n$. For two vectors x and y from R^n $\langle x, y \rangle$ denotes the standard scalar product:

$$\langle x, y \rangle = \sum_{i=1}^{n} x^{(i)} y^{(i)}.$$

$\bar{e}_n \in R^n$ stands for the vector of all ones. Further, for any integer $k \geq 1$ we define the following simplex:

$$\Delta_n(k) = \{x \in R^n : \; x \geq 0, \; \langle \bar{e}_n, x \rangle = k\}, \quad \mathcal{I}_n(k) = \Delta_n(k) \bigcap Z^n_+,$$

where Z^n_+ denotes the set of all n-dimensional vectors with non-negative entries. Clearly,

$$\tfrac{1}{k} \Delta_n(k) = \Delta_n(1) \stackrel{\text{def}}{=} \Delta_n.$$

Finally, $\mathcal{E}[\cdot]$ stands for the expectation of some random variable.

In what follows we often work with multi-indexes. Let α be a multi-index:

$$\alpha = (\alpha^{(1)}, \ldots, \alpha^{(n)}) \in Z_+^n.$$

For $x \in R^n$ and $\alpha \in Z_+^n$ we use the following notation

$$x^\alpha = \prod_{i=1}^n (x^{(i)})^{\alpha^{(i)}}, \quad \alpha! = \prod_{i=1}^n (\alpha^{(i)})!\,.$$

In the sequel we often use the well-known Leibniz formula

$$\left(\sum_{i=1}^N p^{(i)} \right)^d = \sum_{\alpha \in \mathcal{I}_N(d)} \frac{d!}{\alpha!}\, p^\alpha, \tag{1.1}$$

which is valid for all $p \in R^N$ and $d \in Z_+$.

2 Random walk in a simplex

Let us fix some vector $p \in \Delta_n$. Denote by $\zeta(p)$ a discrete random variable distributed as follows:

$$\text{Prob}\{\zeta(p) = i\} = p^{(i)}, \quad i = 1, \ldots, n. \tag{2.1}$$

Consider the following process:

$$\begin{aligned} x_0(p) &= 0 \in R^n, \\ x_{k+1}(p) &= x_k(p) + e_{\zeta_k(p)},\ k \geq 0, \end{aligned} \tag{2.2}$$

where all $\zeta_k(p)$ are random independent variables distributed according to (2.1).

Clearly, any realization $x_k(p)$ belongs to the simplex $\Delta_n(k)$. Hence, the process

$$y_k(p) = \tfrac{1}{k} x_k(p) \tag{2.3}$$

can be seen as a random walk in the standard simplex Δ_n. Note that the direct representation of the process $y_k(p)$ is as follows:

$$\begin{aligned} y_0(p) &= 0 \in R^n, \\ y_{k+1}(p) &= \tfrac{k}{k+1} y_k(p) + \tfrac{1}{k+1} e_{\zeta_k(p)},\ k \geq 0. \end{aligned} \tag{2.4}$$

Let us compute the main characteristics of the processes $x_k(p)$ and $y_k(p)$.

Theorem 1 *Let $k \geq 1$. Then for any $i = 1, \ldots, n$ we have*

$$\begin{aligned} \mu_k^{(i)}(p) \stackrel{\text{def}}{=} \mathcal{E}\left[x_k^{(i)}(p) \right] &= kp^{(i)}, \\ \mathcal{E}\left[\left(x_k^{(i)}(p) - \mu_k^{(i)}(p) \right)^2 \right] &= kp^{(i)}(1 - p^{(i)}). \end{aligned} \tag{2.5}$$

The covariance of two random variables $x_k^{(i)}(p)$ and $x_k^{(j)}(p)$, $i \neq j$, is as follows:

$$\mathcal{E}\left[\left(x_k^{(i)}(p) - \mu_k^{(i)}(p) \right) \cdot \left(x_k^{(j)}(p) - \mu_k^{(j)}(p) \right) \right] = -kp^{(i)} p^{(j)}. \tag{2.6}$$

Proof:
Indeed, the random variable $x_k^{(i)}$ takes discrete values $m = 0 \dots k$ with the following probabilities:

$$\text{Prob}\{x_k^{(i)} = m\} = \tfrac{k!}{m!(k-m)!} \left(p^{(i)}\right)^m \left(1 - p^{(i)}\right)^{k-m}.$$

Therefore using (1.1) with $N = 2$ and $d = k - 1$ we get

$$\begin{aligned} \mu_k^{(i)}(p) &= \sum_{m=0}^{k} \tfrac{m\,k!}{m!(k-m)!} \left(p^{(i)}\right)^m \left(1 - p^{(i)}\right)^{k-m} \\ &= kp^{(i)} \sum_{m=1}^{k} \tfrac{(k-1)!}{(m-1)!(k-m)!} \left(p^{(i)}\right)^{m-1} \left(1 - p^{(i)}\right)^{k-m} \\ &= kp^{(i)}. \end{aligned}$$

Similarly, using (1.1) for $N = 2$ twice with $d = k - 1$ and $d = k - 2$, we obtain

$$\begin{aligned} \mathcal{E}\left[\left(x_k^{(i)}(p)\right)^2\right] &= \sum_{m=0}^{k} \tfrac{m^2\,k!}{m!(k-m)!} \left(p^{(i)}\right)^m \left(1 - p^{(i)}\right)^{k-m} \\ &= kp^{(i)} \sum_{m=1}^{k} \tfrac{(1+m-1)\cdot(k-1)!}{(m-1)!(k-m)!} \left(p^{(i)}\right)^{m-1} \left(1 - p^{(i)}\right)^{k-m} \\ &= kp^{(i)} \cdot \left(1 + (k-1)p^{(i)} \sum_{m=2}^{k} \tfrac{(k-2)!}{(m-2)!(k-m)!} \left(p^{(i)}\right)^{m-2} \left(1 - p^{(i)}\right)^{k-m}\right) \\ &= kp^{(i)} + k(k-1)\left(p^{(i)}\right)^2. \end{aligned}$$

Thus we get (2.5).

Let us prove equation (2.6). Note that the possible values of the random vector variable $x_k(p)$ are the multi-indexes from the set $\mathcal{I}_n(k)$, which appear with the following probabilities:

$$\text{Prob}\{x_k(p) = \alpha\} = \tfrac{k!}{\alpha!}\, p^\alpha.$$

In view of evident symmetry, it is enough to prove (2.6) for $i = 1$ and $j = 2$. Assume that $k \geq 2$. Note that

$$\mathcal{E}[x_k^{(1)}(p) \cdot x_k^{(2)}(p)] = \sum_{\substack{\alpha=(\alpha^{(1)},\alpha^{(2)},\hat{\alpha}) \\ \alpha\in\mathcal{I}_n(k)}} \tfrac{\alpha^{(1)}\alpha^{(2)}\,k!p^\alpha}{\alpha^{(1)}!\dots\alpha^{(n)}!} = \sum_{t=2}^{k} \left(\sum_{\substack{\alpha^{(1)},\alpha^{(2)}\geq 1, \\ \alpha^{(1)}+\alpha^{(2)}=t}} \tfrac{k!(p^{(1)})^{\alpha^{(1)}}(p^{(2)})^{\alpha^{(2)}}}{(\alpha^{(1)}-1)!(\alpha^{(2)}-1)!} S_{k,t} \right),$$

where the value $S_{k,t}$ can be represented as follows:

$$S_{k,t} = \sum_{\hat{\alpha}\in\mathcal{I}_{n-2}(k-t)} \tfrac{(p^{(3)})^{\hat{\alpha}^{(1)}}\dots(p^{(n)})^{\hat{\alpha}^{(n-2)}}}{\hat{\alpha}^{(1)}!\dots\hat{\alpha}^{(n-2)}!} = \tfrac{1}{(k-t)!}\left(\sum_{i=3}^{n} p^{(i)}\right)^{k-t} = \tfrac{\left(1-p^{(1)}-p^{(2)}\right)^{k-t}}{(k-t)!},$$

(we have used (1.1) with $N = n - 2$ and $d = k - t$). Denoting now

$$\alpha_1' = \alpha_1 - 1, \quad \alpha_2' = \alpha_2 - 1, \quad \alpha_3' = k - t,$$

we get

$$\begin{aligned}\mathcal{E}[x_k^{(1)}(p) \cdot x_k^{(2)}(p)] &= k(k-1)p^{(1)}p^{(2)} \sum_{\substack{\alpha'_1,\alpha'_2,\alpha'_3 \geq 0, \\ \alpha'_1+\alpha'_2+\alpha'_3 = k-2}} \frac{(k-2)!(p^{(1)})^{\alpha'_1}(p^{(2)})^{\alpha'_2}(1-p^{(1)}-p^{(2)})^{\alpha'_3}}{\alpha'_1!\alpha'_2!\alpha'_3!} \\ &= k(k-1)p^{(1)}p^{(2)},\end{aligned}$$

where for the last step we used (1.1) with $N = 3$ and $d = k - 2$.

In the case $k = 1$ the latter expression is also valid since the product $x_1^{(1)}(p) \cdot x_1^{(2)}(p)$ is always equal to zero. Now, using (2.5) we obtain (2.6). □

The following corollary describes the properties of the random process $y_k(p)$ in the standard simplex.

Corollary 1 *For any $k \geq 1$ and $i \neq j$ we have*

$$\begin{aligned}\mathcal{E}\left[\left(y_k^{(i)}(p)\right)^2\right] &= \tfrac{1}{k}p^{(i)} + \left(1 - \tfrac{1}{k}\right) \cdot \left(p^{(i)}\right)^2, \\ \mathcal{E}\left[y_k^{(i)}(p) \cdot y_k^{(j)}(p)\right] &= \left(1 - \tfrac{1}{k}\right) p^{(i)}p^{(j)}.\end{aligned} \tag{2.7}$$

Proof:
We get (2.7) directly from (2.5) and (2.6) by dividing $x_k(p)$ by k. □

3 Quadratic optimization over polytopes

Let us start from the following minimization problem:

$$\text{Find } f_* = \min_x \{ f(x) \stackrel{\text{def}}{=} \langle Qx, x \rangle : \; x \in \Delta_n \}, \tag{3.1}$$

where Q is a symmetric $n \times n$-matrix. We allow Q to be indefinite.

Clearly, the problem (3.1) is non-convex. As in [1] and [4], we propose to approximate f_* by taking the minimum value of the function $f(x)$ on a "uniform" grid in Δ_n. We justify the quality of such an approximation by the random vector process (2.3). For $k \geq 1$ define

$$f_k^* = \min_\alpha \left\{ \tfrac{1}{k^2} f(\alpha) : \; \alpha \in \mathcal{I}_n(k) \right\}.$$

Theorem 2 *For any $k \geq 1$ we have*

$$0 \leq f_k^* - f_* \leq \tfrac{1}{k} \left[\max_{1 \leq i \leq n} Q^{(i,i)} - f_* \right]. \tag{3.2}$$

Proof:

Indeed, $f_* \leq f_k^*$. On the other hand, let us choose $p = x^*$, where x^* is a global solution to (3.1). Then, in view of relations (2.7), we get

$$\begin{aligned}
f_k^* &= \min_\alpha \{ f\left(\tfrac{\alpha}{k}\right) : \ \alpha \in \mathcal{I}_n(k)\} \leq \mathcal{E}[f(y_k(p))] = \mathcal{E}[\langle Q y_k(p), y_k(p)\rangle] \\
&= \sum_{i,j=1}^n Q^{(i,j)} \cdot \mathcal{E}\left[y_k^{(i)}(p) \cdot y_k^{(j)}(p) \right] \\
&= \sum_{i=1}^n Q^{(i,i)} \cdot \mathcal{E}\left[\left(y_k^{(i)}(p) \right)^2 \right] + \sum_{i \neq j} Q^{(i,j)} \cdot \mathcal{E}\left[y_k^{(i)}(p) \cdot y_k^{(j)}(p) \right] \\
&= \tfrac{1}{k} \sum_{i=1}^n Q^{(i,i)} p^{(i)} + \left(1 - \tfrac{1}{k}\right) \sum_{i,j=1}^n Q^{(i,j)} p^{(i)} p^{(j)} \\
&\leq \tfrac{1}{k} \max_{1 \leq i \leq n} Q^{(i,i)} + \left(1 - \tfrac{1}{k}\right) f_*.
\end{aligned}$$

□

Corollary 2 *If $Q^{(i,i)} \leq 0$ for all $i = 1, \ldots, n$, then $f_* \leq 0$ and the estimate f_k^* has relative accuracy $\frac{1}{k}$:*

$$f_k^* - f_* \leq \tfrac{1}{k}(-f_*). \tag{3.3}$$

Proof:
Indeed, if all diagonal entries of the matrix Q are non-positive, then $f_* \leq \max\limits_{1 \leq i \leq n} f(e_i) \leq 0$. It remains to use the estimate (3.2). □

Let us compute M_n^k, the number of integer points in the set $\mathcal{I}_n(k)$.

Lemma 1 *For any $k \geq 1$ and $n \geq 1$ we have the following relation:*

$$M_n^{k+1} = M_n^k + M_{n-1}^{k+1}. \tag{3.4}$$

Since $M_n^1 = n$, $n \geq 1$, and $M_1^k = 1$, $k \geq 1$, we obtain

$$M_n^k = \tfrac{(n+k-1)!}{k!(n-1)!}. \tag{3.5}$$

Proof:
Indeed, the integer points in the set $\mathcal{I}_n(k+1)$ can be divided into two disjoint groups:

$$\mathcal{A}_0 = \{x \in \mathcal{I}_n(k+1) : \ x^{(1)} = 0\}, \quad \mathcal{A}_1 = \{x \in \mathcal{I}_n(k+1) : \ x^{(1)} \geq 1\}.$$

Clearly $|\mathcal{A}_0| = M_{n-1}^{k+1}$ and $|\mathcal{A}_1| = M_n^k$. Thus, we get (3.4). Expression (3.5) can be easily obtained by applying recursively (3.4) to the boundary conditions. □

In other words, M_n^k is equal to the number of different possibilities to choose k elements form the set of n elements with repetitions; that is $\binom{n+k-1}{n-1}$.

Thus, if the relative accuracy is fixed, sometimes we can obtain a corresponding approximate solution in polynomial time. For example, if the conditions of Corollary 2 are satisfied, then the solution of the problem (3.1) with relative accuracy $\frac{1}{3}$ can be found in $\frac{n(n+1)(n+2)}{6}$ computations of function $f(x)$ in the nodes of the grid $\frac{1}{3}\mathcal{I}_n(3)$.

To conclude this section, let us present several straightforward extensions of the result. First of all, note that the optimization problem with non-homogeneous quadratic function, that is

$$\text{Find } f_* = \min_x \{ \langle Qx, x\rangle + 2\langle a, x\rangle : \; x \in \Delta_n \}, \tag{3.6}$$

can be treated by above technique since

$$\langle Qx, x\rangle + 2\langle a, x\rangle = \langle \left(Q + a\bar{e}_n^T + \bar{e}_n a^T \right) x, x\rangle \quad \forall x \in \Delta_n.$$

Finally, it appears that this approach can be extended to quadratic minimization problems over polytopes:

$$\text{Find } f_* = \min_x \{ f(x) \stackrel{\text{def}}{=} \langle \hat{Q}x, x\rangle : \; x \in \mathcal{P} \}, \tag{3.7}$$

where $\mathcal{P} = \text{Conv}\,\{u_i \in R^n, \; i = 1, \dots, N\}$. Denote by U the matrix $(u_1, \dots, u_N)$. For $k \geq 1$ define

$$\begin{aligned} f_k^* &= \min_\alpha \left\{ \tfrac{1}{k^2} f(U\alpha) : \; \alpha \in \mathcal{I}_N(k) \right\}, \\ f^* &= \max_{1 \leq i \leq N} f(u_i). \end{aligned}$$

Theorem 3 *For any $k \geq 1$ we have*

$$0 \leq f_k^* - f_* \leq \tfrac{1}{k}[f^* - f_*]. \tag{3.8}$$

If $f^* \leq 0$, then the relative accuracy of approximation f_k^* is at least $\frac{1}{k}$.

Proof:
Indeed, any $y \in \mathcal{P}$ can be represented as

$$y = Ux, \quad x \in \Delta_N.$$

Therefore the problem (3.7) is equivalent to the problem (3.1) with

$$Q = U^T \hat{Q} U.$$

Since $Q^{(i,i)} = f(u_i)$, $i = 1, \dots, N$, we get all statements of the theorem from Theorem 2 and Corollary 2. □

To conclude, let us present the rules of the random walk in the polytope $\mathcal{P}$:

$$\begin{aligned} y_0(p) &= 0 \in R^n, \\ y_{k+1}(p) &= \tfrac{k}{k+1} y_k(p) + \tfrac{1}{k+1} u_{\zeta_k(p)}, \; k \geq 0, \end{aligned} \tag{3.9}$$

where $p \in \Delta_N$ and the discrete random variables $\zeta_k(p)$ are distributed as follows:

$$\text{Prob}[\zeta_k(p) = i] = p^{(i)}, \quad i = 1, \dots, N.$$

4 Polynomials of higher degree

Note that the above approach can also be used in a more general situation. Indeed, assume we want to approximate the value

$$f_* = \min_x \{ f(x) : \; x \in \Delta_n \}, \tag{4.1}$$

where $f(x)$ is a homogeneous polynomial of degree m. Then we can use the following upper approximation:

$$f_* \approx f^*_k = \min_\alpha \{ f(\alpha) : \; \alpha \in \mathcal{I}_n(k) \} \leq \tfrac{1}{k^m} \mathcal{E}[f(x_k(p^*))],$$

where p^* is an optimal solution to (4.1). In order to compute this expectation, we need explicit formulaes for the values

$$E^\beta_k(p) \;\stackrel{\mathrm{def}}{=}\; \mathcal{E}[x^\beta_k(p)], \quad \beta \in Z^n_+, \quad p \in \Delta_n.$$

It appears that these values can be easily computed recursively.

Lemma 2 *Let $\beta \in Z^n_+$ be a multi-index and $p \in \Delta_n$. Then for any $i = 1, \ldots, n$ and $k \geq 1$ we have*

$$E^{\beta+e_i}_{k+1}(p) = (k+1) p^{(i)} \sum_{\gamma=0}^{\beta^{(i)}} \tfrac{(\beta^{(i)})\,!}{\gamma!\,(\beta^{(i)}-\gamma)!} \; E^{\beta-\gamma e_i}_k (p). \tag{4.2}$$

The boundary conditions for this recursion are given by

$$E^0_k(p) = 1, \quad k \geq 1. \tag{4.3}$$

Proof:
Clearly we can take $i = 1$. Let us partition

$$\beta = (\beta^{(1)}, \hat{\beta}), \quad \alpha = (\alpha^{(1)}, \hat{\alpha}), \text{ and } \; p = (p^{(1)}, \hat{p}).$$

Then

$$\begin{aligned}
E^{\beta+e_1}_{k+1}(p) \quad &= \quad \sum_{\alpha \in \mathcal{I}_n(k+1)} \tfrac{(k+1)!}{\alpha!} \, p^\alpha \cdot \alpha^{\beta+e_1} \\
&= \quad (k+1)! \sum_{\alpha^{(1)}=0}^{k+1} \left[\tfrac{(\alpha^{(1)})^{\beta^{(1)}+1} (p^{(1)})^{\alpha^{(1)}}}{(\alpha^{(1)})!} \sum_{\hat{\alpha} \in \mathcal{I}_{n-1}(k+1-\alpha^{(1)})} \tfrac{\hat{p}^{\hat{\alpha}}}{\hat{\alpha}!} \cdot \hat{\alpha}^{\hat{\beta}} \right] \\
&= \quad (k+1)! \, p^{(1)} \sum_{\alpha^{(1)}=1}^{k+1} \left[\tfrac{(\alpha^{(1)})^{\beta^{(1)}} (p^{(1)})^{\alpha^{(1)}-1}}{(\alpha^{(1)}-1)!} \sum_{\hat{\alpha} \in \mathcal{I}_{n-1}(k+1-\alpha^{(1)})} \tfrac{\hat{p}^{\hat{\alpha}}}{\hat{\alpha}!} \cdot \hat{\alpha}^{\hat{\beta}} \right] \\
(\tau \equiv \alpha^{(1)} - 1) \quad &= \quad (k+1) p^{(1)} \sum_{\tau=0}^{k} \left[\tfrac{k!\,(1+\tau)^{\beta^{(1)}} (p^{(1)})^\tau}{(\tau)!} \sum_{\hat{\alpha} \in \mathcal{I}_{n-1}(k-\tau)} \tfrac{\hat{p}^{\hat{\alpha}}}{\hat{\alpha}!} \cdot \hat{\alpha}^{\hat{\beta}} \right].
\end{aligned}$$

It remains to note that

$$\sum_{\tau=0}^{k}\left[\frac{(1+\tau)^{\beta^{(1)}}\,k!\,(p^{(1)})^{\tau}}{(\tau)!}\sum_{\hat\alpha\in\mathcal{I}_{n-1}(k-\tau)}\frac{\hat p^{\hat\alpha}}{\hat\alpha!}\cdot\hat\alpha^{\hat\beta}\right]$$

$$=\sum_{\tau=0}^{k}\left[\left(\sum_{\gamma=0}^{\beta^{(1)}}\frac{(\beta^{(1)})!}{\gamma!(\beta^{(1)}-\gamma)!}\,\tau^{\beta^{(1)}-\gamma}\right)\frac{k!(p^{(1)})^{\tau}}{(\tau)!}\sum_{\hat\alpha\in\mathcal{I}_{n-1}(k-\tau)}\frac{\hat p^{\hat\alpha}}{\hat\alpha!}\cdot\hat\alpha^{\hat\beta}\right]$$

$$=\sum_{\gamma=0}^{\beta^{(1)}}\frac{(\beta^{(1)})!}{\gamma!(\beta^{(1)}-\gamma)!}\left[\sum_{\tau=0}^{k}\frac{k!(p^{(1)})^{\tau}}{(\tau)!}\,\tau^{\beta^{(1)}-\gamma}\sum_{\hat\alpha\in\mathcal{I}_{n-1}(k-\tau)}\frac{\hat p^{\hat\alpha}}{\hat\alpha!}\cdot\hat\alpha^{\hat\beta}\right]$$

$$=\sum_{\gamma=0}^{\beta^{(1)}}\frac{(\beta^{(1)})!}{\gamma!(\beta^{(1)}-\gamma)!}\,E_k^{\beta-\gamma e_1}(p).$$

The proof of the boundary conditions is straightforward. □

Let us present explicit formulae for the expectations $E_k^{\beta}(p)$ with $\beta\in\mathcal{I}_n(m)$, $1\le m\le 4$.

$m=1$	$E_k^{e_1}(p) = kp^{(1)}$
$m=2$	$E_k^{2e_1}(p) = kp^{(1)}+\frac{k!}{(k-2)!}(p^{(1)})^2$ $E_k^{e_1+e_2}(p) = \frac{k!}{(k-2)!}p^{(1)}p^{(2)}$
$m=3$	$E_k^{3e_1}(p) = kp^{(1)}+3\frac{k!}{(k-2)!}(p^{(1)})^2+\frac{k!}{(k-3)!}(p^{(1)})^3$ $E_k^{2e_1+e_2}(p) = \frac{k!}{(k-2)!}p^{(1)}p^{(2)}+\frac{k!}{(k-3)!}(p^{(1)})^2p^{(2)}$ $E_k^{e_1+e_2+e_3}(p) = \frac{k!}{(k-3)!}p^{(1)}p^{(2)}p^{(3)}$
$m=4$	$E_k^{4e_1}(p) = kp^{(1)}+7\frac{k!}{(k-2)!}(p^{(1)})^2+6\frac{k!}{(k-3)!}(p^{(1)})^3+\frac{k!}{(k-4)!}(p^{(1)})^4$ $E_k^{3e_1+e_2}(p) = p^{(1)}p^{(2)}\left[\frac{k!}{(k-2)!}+3\frac{k!}{(k-3)!}p^{(1)}+\frac{k!}{(k-4)!}(p^{(1)})^2\right]$ $E_k^{2e_1+2e_2}(p) = p^{(1)}p^{(2)}\left[\frac{k!}{(k-2)!}+\frac{k!}{(k-3)!}(p^{(1)}+p^{(2)})+\frac{k!}{(k-4)!}p^{(1)}p^{(2)}\right]$ $E_k^{2e_1+e_2+e_3}(p) = p^{(1)}p^{(2)}p^{(3)}\left[\frac{k!}{(k-3)!}+\frac{k!}{(k-4)!}p^{(1)}\right]$ $E_k^{e_1+e_2+e_3+e_4}(p) = \frac{k!}{(k-4)!}p^{(1)}p^{(2)}p^{(3)}p^{(4)}$

From the above expressions we observe a decrease in the quality of the approximation f_k^* as m increases. However, there exists one particular case in which the situation is under control.

Lemma 3 *Let* $f(x) = \sum\limits_{\alpha\in\mathcal{A}} f_\alpha x^\alpha$ *with* $\mathcal{A} \subseteq \{\alpha \in \mathcal{I}_n(m) :\ \alpha^{(i)} \leq 1,\ i = 1,\dots,n\}$*. Then in the problem (4.1)* $f_* \leq 0$ *and for* $k > m \geq 2$ *we have*

$$0 \leq f_k^* - f_* \leq \left(1 - \tfrac{k!}{(k-m)!\,k^m}\right) \cdot (-f_*) \leq \tfrac{m(m-1)}{2k}(-f_*). \tag{4.4}$$

Proof:
In view of Lemma 2 we have

$$\mathcal{E}[x_k^\alpha(p)] = \tfrac{k!}{(k-m)!}\, p^\alpha.$$

Therefore, taking p equal to the optimal solution of the problem (4.1), we get

$$f_k^* \leq \mathcal{E}[f(\tfrac{1}{k}x_k(p))] = \sum_{\alpha\in\mathcal{A}} f_\alpha \mathcal{E}[\tfrac{1}{k^m}x_k^\alpha(p)] = \tfrac{k!}{(k-m)!k^m}\sum_{\alpha\in\mathcal{A}} f_\alpha p^\alpha = \tfrac{k!}{(k-m)!k^m} f_*,$$

and that is the first inequality in (4.4). Let us estimate the value

$$\delta \overset{\text{def}}{=} 1 - \tfrac{k!}{(k-m)!k^m}$$

from above. Note that for $m = 2$ the last inequality in (4.4) is trivial. Let us assume that $m \geq 3$.

Since function $\psi(\tau) = \ln(1 - \frac{\tau}{k})$ is concave and $\psi(0) = 0$, we have

$$\begin{aligned}
\ln(1-\delta) &= \sum_{i=1}^{m-1} \ln\left(1 - \tfrac{i}{k}\right) = \sum_{i=1}^{m-1} \psi(i) \\
&\geq \sum_{i=1}^{m-1} \left[(1 - \tfrac{i}{m})\psi(0) + \tfrac{i}{m}\psi(m)\right] \\
&= \tfrac{m-1}{2}\psi(m) = \tfrac{m-1}{2}\ln(1 - \tfrac{m}{k}).
\end{aligned}$$

Thus, since $m \geq 3$, we conclude that

$$\delta \leq 1 - \left(1 - \tfrac{m}{k}\right)^{\frac{m-1}{2}} \leq \tfrac{m(m-1)}{2k}.$$

□

5 Maximization over the Euclidean ball

Let $G = (V, E)$ be a graph with the set of nodes $V = \{1 \dots n\}$ and the set of undirected arcs $E = \{(i_k, j_k),\ k = 1,\dots,m\}$. Denote by A its adjacency matrix. Then the Motzkin-Straus problem can be posed as follows:

$$\text{Find } f_* = \min_x \{\langle (I_n + A)x, x\rangle :\ x \in \Delta_n\}, \tag{5.1}$$

where I_n is the identity $n \times n$-matrix. It can be proved [3] that $f_* = \frac{1}{\alpha(G)}$, where $\alpha(G)$ is the stability number of the graph G. In what follows we denote by $\| \cdot \|$ the standard Euclidean norm of corresponding vector.

Theorem 4 *After an appropriate change of variables, problem (5.1) can be posed in any of the following settings.*

1. Quartic maximization *over the Euclidean ball:*

$$\max_{u \in R^n} \left\{ \sum_{k=1}^{m} \langle A_k u, u \rangle^2 : \; \|u\| = 1 \right\}. \tag{5.2}$$

2. Cubic maximization *over the Euclidean ball:*

$$\max_{v \in R^{m+n}} \left\{ g(v) : \; \|v\| = 1 \right\}, \tag{5.3}$$

where $g(v)$ is a form of degree three.

3. Matrix ellipsoid *problem:*

$$\max_{r} \left\{ r : \; I + \sum_{k=1}^{m} w^{(k)} A_k \succeq 0 \quad \forall w \in R^m, \; \|w\| \leq r \right\}. \tag{5.4}$$

Proof:
Denote $Q = \bar{e}_n \bar{e}_n^T - I_n - A$. Then for any $x \in \Delta_n$ we have

$$f(x) \stackrel{\text{def}}{=} \langle Qx, x \rangle = 1 - \langle (I_n + A)x, x \rangle.$$

Thus, $f^* \equiv \max_x \{ \langle Qx, x \rangle : \; x \in \Delta_n \} = 1 - f_*$. Note that

$$f(x) = 2 \sum_{(i,j) \in E} x^{(i)} x^{(j)}.$$

Changing variables by setting $x^{(i)} = \left(u^{(i)} \right)^2$, $i = 1, \ldots, n$, we conclude that f^* can be found from the problem of type (5.2) with

$$A_k = \frac{e_{i_k} e_{j_k}^T + e_{j_k} e_{i_k}^T}{\sqrt{2}}, \quad (i_k, j_k) \in E, \; k = 1, \ldots, m.$$

Further, the problem (5.4) can be transformed as follows:

$$\max_{r,w} \left\{ r : \; I + \sum_{k=1}^{m} w^{(k)} A_k \succeq 0 \quad \forall w \in R^m, \; \|w\| \leq r \right\}$$

$$= \max_{r,w,u} \left\{ r : \; \|u\|^2 \geq \sum_{k=1}^{m} w^{(k)} \langle A_k u, u \rangle \quad \forall w \in R^m, \; \|w\| \leq r, \; \forall u \in R^n \right\}$$

$$= \max_{r,u} \left\{ r : \; \|u\|^2 \geq r \left[\sum_{k=1}^{m} \langle A_k u, u \rangle^2 \right]^{1/2} \quad \forall u \in R^n \right\}.$$

It is clear that the latter problem is of the form (5.2).

Finally, let us show that the problem (5.2) can be rewritten in the form (5.3). Indeed,

$$\begin{aligned} \omega &\stackrel{\text{def}}{=} \left(\max_{u \in R^n} \left\{ \sum_{k=1}^{m} \langle A_k u, u \rangle^2 : \ \|u\| = 1 \right\} \right)^{1/2} \\ &= \max_{u \in R^n, w \in R^m} \left\{ \sum_{k=1}^{m} w^{(k)} \langle A_k u, u \rangle : \ \|u\| = 1, \ \|w\| = 1 \right\}. \end{aligned}$$

Denote $v = (u, w)$ and $g(v) = \sum\limits_{k=1}^{m} w^{(k)} \langle A_k u, u \rangle$. For positive β and γ define

$$\omega(\beta, \gamma) = \max_{u \in R^n, w \in R^m} \{ g(v) : \ \|u\| = \beta, \ \|w\| = \gamma \} = \beta^2 \gamma \omega.$$

Then

$$\begin{aligned} \max_{v \in R^{n+m}} \{ g(v) : \|v\| = 1 \} &= \max_{\beta,\gamma} \{ \omega(\beta, \gamma) : \beta^2 + \gamma^2 = 1 \} \\ &= \omega \cdot \max_{\beta,\gamma} \{ \beta^2 \gamma : \beta^2 + \gamma^2 = 1 \} = \tfrac{2\omega}{3\sqrt{3}}. \end{aligned}$$

Thus, ω can be computed by maximizing $g(v)$ over Euclidean ball. □

In our knowledge, the first result on NP-hardness of the matrix ellipsoid problem (5.4) was presented in [2], where the author used another argumentation.

Acknowledgement. The author is very thankful to Laurence Wolsey for his useful comments on the paper.

References

[1] I.M. Bomze and E. de Klerk. Solving standard quadratic optimization problems via linear, semidefinite, and copositive programming. J. Global Optim. 24 (2002), 163-185.

[2] T. Margalit. “Robust linear programming”, M.Sc. Thesis, Faculty of Industrial Engineering and Management, Technion, 1996.

[3] T.S. Motzkin and E.G. Straus. Maxima for graphs and a new proof of a theorem of Túran, Canadian J. Math., 17: 533-540, 1965.

[4] Yu. Nesterov. Global quadratic optimization on the sets with simplex structure. CORE Discussion Paper #9915, CORE, March 1999.