ON HAZAN'S ALGORITHM FOR SYMMETRIC PROGRAMMING PROBLEMS

L. FAYBUSOVICH*

Abstract. We describe the generalization of Hazan's algorithm for symmetric programming problems

Key words. Symmetric programming, Euclidean Jordan algebras, low-rank approximations to optimal solutions

AMS subject classifications. 90C25,17C99

1. Introduction. In [3] E.Hazan proposed an algorithm for solving semidefinite programming problems. Though the complexity estimates for this algorithm are not as good as for the most popular primal-dual algorithms, it nevertheless has a number of interesting properties. It provides a low-rank approximation to the optimal solution and it requires the computation of the gradient of an auxiliary convex function rather than the Hessian of a barrier function in case of interior-point algorithms. In present paper we show that these properties are preserved in a much more general case of symmetric programming problems. Moreover, we will also show that the major computational step of the algorithm (finding the maximal eigenvalue and corresponding eigenvector of a positive definite symmetric matrix) admits a natural decomposition in a general case of a direct sum of simple Euclidean Jordan algebras. For some types of irreducible blocks (e.g. ,generally speaking, infinite-dimensional spin-factors) the corresponding eigenvalue problem has a simple analytic solution. In particular, Hazan's algorithm offers a computationally attractive alternative to other methods in case of second-order cone programming (including the infinite-dimensional version considered in [2], [4]).

2. Jordan-algebraic concepts. We stick to the notation of an excellent book [1]. We do not attempt to describe the Jordan-algebraic language here but instead provide detailed references to [1]. Throughout this paper:
• V is an Euclidean Jordan algebra;
• $\mathrm{rank}(V)$ stands for the rank of V;
• $x \circ y$ is the Jordan algebraic multiplication for $x, y \in V$;
• $< x, y >= \mathrm{tr}(x \circ y)$ is the canonical scalar product in V; here tr is the trace operator on V;
• Ω is the cone of invertible squares in V;
• $\bar{\Omega}$ is the closure of Ω in V;
• An element $g \in V$ such that $g^2 = g$ and $\mathrm{tr}(g) = 1$ is called a primitive idempotent in V;
• Given $x \in V$, we denote by $L(x)$ the corresponding multiplication operator on V, i.e.

$$L(x)y = x \circ y, \quad y \in V;$$

• Given $x \in V$, we denote by $P(x)$ the so-called quadratic representation of x, i.e.

$$P(x) = 2L(x)^2 - L(x^2).$$

*The author was supported in part by National Science Foundation Grant NO. DMS07-12809. Department of Mathematics University of Notre Dame, Notre Dame, IN 46556 USA

Given $x \in V$, there exist idempotents $g_1, \cdots, g_k$ in V such that $g_i \circ g_j = 0$ for $i \neq j$ and such that $g_1 + g_2 \cdots + g_k = e$, and distinct real numbers $\lambda_1, \cdots, \lambda_k$ with the following property:

$$x = \sum_{i=1}^{k} \lambda_i g_i \tag{2.1}$$

The numbers λ_i and idempotents g_i are uniquely defined by x. (see Theorem III. 1.1 in [1]). Here e is the identity element in the Jordan algebra V. We define $\lambda_{\min}(x)$ (resp. $\lambda_{\max}(x)$) as $\min\{\lambda_i : i = 1, \ldots, k\}$ (resp. $\max\{\lambda_i : i = 1, \ldots, k\}$. Within the context of this paper the notion of rank of x is very important. By definition:

$$\mathrm{rank}(x) = \sum_{i:\lambda_i \neq 0} \mathrm{tr}(f_i). \tag{2.2}$$

Given $x \in V$, the operator $L(x)$ is symmetric with respect to the canonical scalar product. If g is an idempotent in V, it turns out that the spectrum of $L(g)$ belongs to $\{0, \frac{1}{2}, 1\}$. Following [1], we denote by $V(1,g), V(\frac{1}{2}, g), V(0,g)$ corresponding eigenspaces.

It is clear that

$$V = V(0,g) \oplus V(1,g) \oplus V(\frac{1}{2}, g) \tag{2.3}$$

and the eigenspaces are pairwise orthogonal with respect to the scalar product $<,>$. This is the so-called Peirce decomposition of V with respect to an idempotent g. However, eigenspaces have more structure (see [1], Proposition IV. 1.1). In particular, $V(0,g), V(1,g)$ are subalgebras in V.

We summarize some of the properties of algebras $V(1,g) = V(0, 1-g)$.

Proposition 1. *Let g be an idempotent in an Euclidean Jordan algebra V. Then $V(1,g)$ is an Euclidean Jordan subalgebra with identity element g. Moreover,*

$$rank(V(1,g)) = rank(g)$$

The trace operator on $V(1,g)$ coincides with the restriction of the trace operator on V. If $\tilde{\Omega}$ is the cone of invertible squares in $V(1,g)$ then $\bar{\tilde{\Omega}} = \bar{\Omega} \cap V(1,g)$.

Proposition 1 easily follows from the properties of Peirce decomposition on V (see section IV.2 in [1]). Notice that if c is a primitive idempotent in $V(1,g)$, then c is primitive idempotent in V. For a proof see e.g [6].

Let $g_1, \cdots, g_r$, where $r = \mathrm{rank}(V)$, be a system of primitive idempotents such that $g_i \circ g_j = 0$ for $i \neq j$ and $g_1 + \cdots g_r = e$. Such system is called a Jordan frame. Given $x \in V$, there exists a Jordan frame $g_1, \cdots, g_r$ and real numbers $\lambda_1, \cdots, \lambda_r$ (eigenvalues of x) such that

$$x = \sum_{i=1}^{r} \lambda_i g_i.$$

Such a representation is called a spectral decomposition of x.
Since primitive idempotents in $V(1,g)$ remain primitive in V, it easily follows that the rank of $x \in V(1,g)$ is the same as its rank in V.

3. Auxiliary functions. Let V be an Euclidean Jordan algebra. Consider the function

$$f_K(x) = \frac{1}{K}\ln(\sum_{i=1}^{m} e^{K(\langle a_i, x\rangle - b_i)}). \tag{3.1}$$

Here $K > 0$ is a real parameter; $a_i \in V, b_i \in \mathbf{R}, i = 1, \dots, m$.

The following Proposition is, of course, well-known, but we will prove it, since we will need the explicit expressions for the gradient and the Hessian of f_K later on.

PROPOSITION 2. *The function f_K is convex on V.*

Proof. We have:

$$Df_K(x)h = \frac{1}{\alpha(x)}\sum_{i=1}^{m}\alpha_i(x)\langle a_i, h\rangle, h \in V. \tag{3.2}$$

Here

$$\alpha_i(x) = e^{K(\langle a_i, x\rangle - b_i)}, \alpha(x) = \sum_{i=1}^{m}\alpha_i(x);$$

$$D^2 f_K(x)(h,h) = K[\frac{\sum_{i=1}^{m}\alpha_i(x)\langle a_i, h\rangle^2}{\alpha(x)} - \frac{(\sum_{i=1}^{m}\alpha_i(x)\langle a_i, h\rangle)^2}{\alpha(x)^2}]. \tag{3.3}$$

Here $Df_K(x), D^2 f_K(x)$ are the first and second Frechet derivatives of f_K, respectively. Let

$$\beta_i = \langle a_i, h\rangle e^{\frac{K(\langle a_i, x\rangle - b_i)}{2}},$$

$$\gamma_i = e^{\frac{K(\langle a_i, x\rangle - b_i)}{2}}.$$

Then Cauchy-Schwarz inequality

$$(\sum_{i=1}^{m}\beta_i\gamma_i)^2 \leq \sum_{i=1}^{m}\beta_i^2\sum_{i=1}^{m}\gamma_i^2$$

shows that $D^2 f_K(x)(h,h) \geq 0$ for all $x \in V, h \in V$. □

As usual, we introduce the gradient $\nabla f_K(x) \in V$ and the Hessian $H_{f_K}(x)$ as follows:

$$\langle \nabla f_K(c), h\rangle = Df_K(x)h,$$

$$\langle h, H_{f_K}(x)h\rangle = D^2 f_K(x)(h,h), x, h \in V. \tag{3.4}$$

In particular,

$$\nabla f_K(x) = \frac{\sum_{i=1}^{m}\alpha_i(x)a_i}{\alpha(x)}, \tag{3.5}$$

and $H_{f_K}(x)$ is a symmetric linear map from V to V. Consider the function

$$\Phi_K : \mathbf{R^m} \to \mathbf{R},$$

$$\Phi_K(y_1, \ldots, y_m) = \frac{\ln(\sum_{i=1}^{m} e^{Ky_i})}{K}. \tag{3.6}$$

It is clear that

$$f_K(x) = \Phi_K(\langle a_1, x\rangle - b_1, \ldots, \langle a_m, x\rangle - b_m).$$

The next Proposition is , of course, well-known.

PROPOSITION 3.

$$\max\{y_i : i \in [1, m]\} \leq \Phi_K(y_1, \ldots, y_m) \leq \max\{y_i : i \in [1, m]\} + \frac{\ln m}{K}. \tag{3.7}$$

4. Feasibility problem.

Consider the following feasibility problem:

$$\langle a_i, x\rangle \leq b_i, i = 1, 2, \ldots, m, \tag{4.1}$$

$$tr(x) = 1, x \in \bar{\Omega} \subseteq V.$$

We can assume without loss of generality that $\| a_i \| = \sqrt{\langle a_i, a_i\rangle} = 1, i = 1, \ldots, m.$ The key in Hazan's approach is the following observation:

PROPOSITION 4. *Fix $\epsilon > 0$. Let x^* be an optimal solution to the optimization problem:*

$$f_K(x) \to \min, \tag{4.2}$$

$$tr(x) = 1, x \in \bar{\Omega}, \tag{4.3}$$

where $K = \frac{2 \ln m}{\epsilon}$. Let, further, $\tilde{x}$ be such that

$$f_K(\tilde{x}) \leq f_K(x^*) + \epsilon/2, tr(\tilde{x}) = 1, \tilde{x} \in \bar{\Omega}. \tag{4.4}$$

If $f_K(\tilde{x}) \leq \epsilon$, then

$$\langle a_i, \tilde{x}\rangle - b_i \leq \epsilon, i = 1, \ldots, m. \tag{4.5}$$

If $f_K(\tilde{x}) > \epsilon$, then the set (4.1) s empty.

Proof. By Proposition 3

$$\Delta(x) \leq f_K(x) \leq \Delta(x) + \frac{\ln m}{K}, \tag{4.6}$$

where

$$\Delta(x) = \max\{\langle a_i, x\rangle - b_i : i \in [1, m]\}, x \in V.$$

If $f_K(\tilde{x}) \leq \epsilon$, the lower bound in (4.6) shows that (4.5) holds true. If, however, $f_K(\tilde{x}) > \epsilon$, then $f_K(x^*) > \frac{\epsilon}{2}$ by (4.4). Suppose that x is feasible for (4.1). Then, in particular, $f_K(x) \geq f_K(x^*) > \frac{\epsilon}{2}$. But then the upper bound in (4.6) yields:

$$f_K(x) \leq \Delta(x) + \frac{\ln m}{K} \leq \frac{\epsilon}{2},$$

since $\Delta(x) \leq 0$ and $K = \frac{2 \ln m}{\epsilon}$. Thus,

$$\frac{\epsilon}{2} \geq f_K(x) > \frac{\epsilon}{2}.$$

Contradiction. Hence, $f_K(\tilde{x}) > \epsilon$ implies that (4.1) has no feasible solutions. □

5. Solving an optimization problem on the generalized simplex. In light of Proposition 4, we need to provide an efficient algorithm for solving the optimization problem (4.2),(4.3). E. Hazan uses a version of the Frank-Wolfe algorithm on the set of positive semi-definite matrices with bounded trace. Let us show that this approach works in a more general symmetric cone situation.

Consider the following optimization problem:

$$f(x) \to \min, \tag{5.1}$$

$$tr(x) = 1, x \in \bar{\Omega} \subseteq V. \tag{5.2}$$

Here f is twice continuously differentiable function on some open neighbourhood of the generalized simplex Δ defined by (5.2). Denote by $\Gamma(f)$ the following expression:

$$\Gamma(f) = \max\{\lambda_{\max}(H_f(x)) : x \in \Delta\} \tag{5.3}$$

where $\lambda_{\max}(H_f(x))$ is the maximal eigenvalue of the Hessian $H_f(x)$ of the function f evaluated at the point x. Fix $\epsilon > 0$. Let x_1 by any primitive idempotent in V. Clearly , $x_1 \in \Delta$. For $k = 1, 2, \ldots$, let s_k be a primitive idempotent in V such that

$$\langle \nabla f(x_k), s_k \rangle \leq \lambda_{\min}(\nabla f(x_k)) + \frac{\epsilon}{4}. \tag{5.4}$$

Set

$$x_{k+1} = x_k + \alpha_k(s_k - x_k), \tag{5.5}$$

where $\alpha_k = \min(1, \frac{2}{k})$. Notice that $\nabla f(x_k) \in V$ and hence $\lambda_{\min}(\nabla f(x_k)$ is defined in Section 2.

THEOREM 1. *Let x^* be an optimal solution to the problem (5.1), (5.2). Then*

$$f(x_k) \leq f(x^*) + \frac{\epsilon}{4} + \frac{4\Gamma(f)}{k}, k = 1, 2, \ldots. \tag{5.6}$$

REMARK 1. *Notice that (5.5) implies that $rank(x_{k+1}) \leq rank(x_k) + rank(s_k)$ (for the rank inequalty see Appendix). Since s_k is a primitive idempotent, we conclude that $rank(x_k) \leq k, k = 1, \ldots$.*

Proof. Our proof follows the one in [5], where the case of the cone of positive semidefinite matrices was considered. Notice that x_{k+1} is a convex combination of x_k and s_k and hence belong to Δ. By Taylor's formula:

$$f(x_{k+1}) = f(x_k) + \alpha_k \langle \nabla f(x_k), s_k - x_k \rangle + \frac{\alpha_k^2}{2} \langle H_f(x')(s_k - x_k), (s_k - x_k) \rangle \tag{5.7}$$

for some $x^{'} \in conv(x_k, x_{k+1}) \subseteq \Delta$. Now,

$$H_f(x^{'}) \leq \Gamma(f) I_V,$$

where I_V is the identity map on V. Hence,

$$f(x_{k+1}) \leq f(x_k) + \alpha_k \langle \nabla f(x_k), s_k - x_k \rangle + \frac{\alpha_k^2}{2} \Gamma(f) \parallel s_k - x_k \parallel^2 .$$

By Lemma 1 (see below) $\| s_k - x_k \| \leq \sqrt{2}$. Hence,

$$f(x_{k+1}) \leq f(x_k) + \alpha_k \langle \nabla f(x_k), s_k - x_k \rangle + \alpha_k^2 \Gamma(f). \tag{5.8}$$

Convexity of f implies

$$f(x^*) - f(x_k) \geq \langle \nabla f(x_k), x^* - x_k \rangle. \tag{5.9}$$

Notice that $\lambda_{\min}(\nabla f(x_k))$ is the optimal value for the optimization problem

$$\langle \nabla f(x_k), z \rangle \to \min, z \in \Delta.$$

This is discussed in the Appendix. Combining this with (5.4), we obtain

$$\langle \nabla f(x_k), s_k \rangle \leq \langle \nabla f(x_k), x^* \rangle + \frac{\epsilon}{4}. \tag{5.10}$$

Substituting (5.10) into (5.8), we obtain:

$$f(x_{k+1}) \leq f(x_k) + \alpha_k(\langle \nabla f(x_k), x^* \rangle + \frac{\epsilon}{4}) -$$

$$\alpha_k \langle \nabla f(x_k), x_k \rangle + \alpha_k^2 \Gamma(f) \leq$$

$$f(x_k) + \alpha_k(\langle \nabla f(x_k), x^* \rangle + \frac{\epsilon}{4}) + \alpha_k[f(x^*) - f(x_k) - \langle \nabla f(x_k), x^* \rangle] + \alpha_k^2 \Gamma(f),$$

where in the last inequality we used (5.9). Hence, we obtain:

$$f(x_{k+1}) \leq (1 - \alpha_k) f(x_k) + \alpha_k (f(x^*) + \frac{\epsilon}{4}) + \alpha_k^2 \Gamma(f). \tag{5.11}$$

We will use (5.11) to prove (5.6) by induction on k. For $k = 1, 2, \alpha_k = 1$ and (5.11) yields:

$$f(x_k) \leq f(x^*) + \frac{\epsilon}{4} + \Gamma(f).$$

If (5.6) holds for some $k \geq 2$, then

$$f(x_{k+1}) \leq (1 - \frac{2}{k})(f(x^*) + \frac{\epsilon}{4} + \frac{4\Gamma(f)}{k}) + \frac{2}{k}(f(x^*) + \frac{\epsilon}{4}) + \frac{4\Gamma(f)}{k^2} =$$

$$= f(x^*) + \frac{\epsilon}{4} + 4\Gamma(f)(\frac{1}{k} - \frac{1}{k^2}) \leq f(x^*) + \frac{\epsilon}{4} + \frac{4\Gamma(f)}{k+1}.$$

□

In the proof of Theorem 1 we used the following.

Lemma 1. *Let $x, y \in \Delta$. Then $\| x - y \| \leq \sqrt{2}$.*

Proof. Let

$$x = \sum_{i=1}^{r} \lambda_i g_i, y = \sum_{i=1}^{r} \mu_i h_i$$

be spectral decompositions of x, y as discussed in Section 2. Since $x, y \in \Delta$, we have

$$\sum_{i=1}^{r} \lambda_i = \sum_{i=1}^{r} \mu_i = 1, \lambda_i \geq 0, \mu_i \geq 0$$

for $i = 1, \dots r$. Now:

$$\| x - y \| = \| \sum_{i=1}^{r} \sum_{j=1}^{r} \lambda_i \mu_j (g_i - h_j) \| \leq$$

$$\sum_{i=1}^{r} \sum_{j=1}^{r} \lambda_i \mu_j \| g_i - h_j \| \leq \max\{\| g_i - h_j \| : i = 1, \dots, r, j = 1, \dots r\}.$$

Hence, it suffices to prove that if g, h are primitive idempotents, then $\| g - h \|^2 \leq 2$. But

$$\| g - h \|^2 = \langle g, g \rangle + \langle h, h \rangle - 2\langle g, h \rangle =$$

$$tr(g^2) + tr(h^2) - 2\langle g, h \rangle = 2 - 2\langle g, h \rangle \leq 2,$$

since $\langle g, h \rangle \geq 0$. The last inequality follows, for example, from the self-duality of the cone $\bar{\Omega}$. □

Remark 2. *Notice that*

$$f(x_k) \leq f(x^*) + \frac{\epsilon}{2} \tag{5.12}$$

provided

$$k \geq \frac{16\Gamma(f)}{\epsilon}. \tag{5.13}$$

In light of this remark , we need to estimate $\Gamma(f)$ for the function f_K.

Proposition 5. *Under the assumptions* $\| a_i \| \leq 1, i = 1, \dots m$, *we have:*

$$\Gamma(f_K) \leq K. \tag{5.14}$$

Proof. By (3.3)

$$D^2 f_K(h, h) = \langle h, H_{f_K} h \rangle \leq K \frac{\sum_{i=1}^{m} \alpha_i(x) \langle a_i, h \rangle^2}{\alpha(x)}. \tag{5.15}$$

Let $v_1, \dots, v_N$ be an orthonormal basis in V. Then

$$Tr(H_{f_K}(x)) = \sum_{j=1}^{N} \langle v_j, H_{f_K}(x) v_j \rangle \leq$$

$$K \frac{\sum_{i=1}^{m} \alpha_i(x) \sum_{j=1}^{N} \langle a_i, v_j \rangle^2}{\alpha(x)}$$

by (5.15). By Pythagoras theorem

$$\sum_{j=1}^{N} \langle a_i, v_j \rangle^2 = \| a_i \|^2 = 1, i = 1, \ldots, m.$$

Hence,

$$Tr(H_{f_K}(x)) \le K \frac{\sum_{i=1}^{m} \alpha_i(x)}{\alpha(x)} = K, \tag{5.16}$$

since

$$\alpha(x) = \sum_{i=1}^{m} \alpha_i(x).$$

Since $H_{f_K}(x)$ is positive definite by Proposition 2, we obtain:

$$\lambda_{\max}(H_{f_K}(x)) \le Tr(H_{f_K}(x)).$$

The result follows by (5.16). □

Recall that $K = \frac{2 \ln m}{\epsilon}$. Combining Proposition 5 with (5.13), we obtain:

$$f_K(x_k) \le f_K(x^*) + \frac{\epsilon}{2}, \tag{5.17}$$

provided

$$k \ge \frac{32 \ln m}{\epsilon^2}.$$

6. Optimization problem and extensions. Consider the optimization problem

$$\langle c, x \rangle \to \min, \tag{6.1}$$

$$\langle a_i, x \rangle \le b_i, i = 1, \ldots, m,$$

$$x \in \bar{\Omega}, tr(x) = 1.$$

We can assume without loss of generality that $\| a_i \| = 1, i = 1, \ldots, m, \| c \| = 1$.

Lemma 2. *Suppose that the feasible set of (6.1) is not empty. Then the optimal value γ_{opt} of (6.1) belongs to* $[-1, 1]$.

Proof. Since

$$| \langle c, x \rangle | \le \| c \| \| x \| = \| x \|,$$

it suffices to show that $\| x \| \le 1$ for $x \in \Delta$. Let $x \in \Delta$ have a spectral decomposition

$$x = \sum_{i=1}^{r} \lambda_i g_i.$$

Then $\lambda_i \geq 0, \sum_{i=1}^r \lambda_i = 1$. We have:

$$\parallel x \parallel \leq \sum_{i=1}^{r} \lambda_i \parallel g_i \parallel .$$

But

$$\parallel g_i \parallel = \langle g_i, g_i \rangle^{1/2} = [tr(g_i^2)]^{1/2} = [tr(g_i)]^{1/2} = 1,$$

since g_i are primitive idempotents. Hence, $\parallel x \parallel \leq \sum_{i=1}^r \lambda_i = 1$. □

We can now apply the standard bisection method to feasibility problems

$$\langle c, x \rangle \leq \gamma,$$

$$\langle a_i, x \rangle \leq b_i, i = 1, \dots m,$$

$$x \in \Delta,$$

starting with $\gamma = 0$. Hence, after $\bigcirc(\frac{\ln(\frac{1}{\epsilon}) \ln m}{\epsilon^2})$ iterations described in Theorem 1 we can find ϵ-approximate solution $\tilde{x}$ to problem (6.1) such that $\langle c, \tilde{x} \rangle \leq \gamma_{opt} + \epsilon$.

At first glance, the requirement $tr(x) = 1$ is very restrictive in formulation (6.1). However, a subset in $\bar{\Omega}$ is bounded if and only if the function tr is bounded from above on it (see [1], Corollary I.1.6). Hence, any bounded set of the form

$$\langle a_i, x \rangle \leq b_i, i = 1, \dots, m, x \in \bar{\Omega},$$

can be described as

$$\langle a_i, x \rangle \leq b_i, i = 1, \dots, m, \tag{6.2}$$

$$tr(x) \leq t, x \in \bar{\Omega} \subset V$$

for some fixed $t > 0$. Making the change of variables $y = x/t$ and adding slack variable $z \geq 0$, we can rewrite (6.2) in the following equivalent form:

$$\langle a_i, y \rangle \leq t b_i, i = 1, \dots, m,$$

$$tr(y, z) = 1, (y, z) \in \bar{\Omega} \times \mathbf{R}_+ \subseteq V \times \mathbf{R}.$$

Notice that $\bar{\Omega} \times \mathbf{R}_+$ is the cone of squares in the Euclidean Jordan algebra $V \times \mathbf{R}$.

7. Appendix. Each iteration described in Theorem 1 requires solving (approximately) the optimization problem of the form

$$\langle \nabla f_K(x), z \rangle \to \min, s \in \Delta.$$

Notice that by (3.5)

$$\parallel \nabla f_K(x) \parallel \leq 1, x \in V.$$

Hence, we need to consider the following problem:

$$\langle c, x\rangle \to \min, \tag{7.1}$$

$$x \in \bar{\Omega}, tr(x) = 1,$$

with $\| c \| \leq 1$. Notice that the Lagrange dual tp (7.1) has the form:

$$\mu \to \max, \tag{7.2}$$

$$c - \mu e \in \bar{\Omega}.$$

Here e is the unit element in the Jordan algebra V. It is quite obvious that the strong duality holds. Let

$$c = \sum_{i=1}^{r} \lambda_i g_i \tag{7.3}$$

be the spectral decomposition of c. Then

$$c - \mu e = \sum_{i=1}^{r} (\lambda_i - \mu) g_i$$

and the condition $c - \mu e \in \bar{\Omega}$ is equivalent to

$$\min\{\lambda_i - \mu : i = 1, \dots, r\} \geq 0,$$

i.e. $\mu \leq \min\{\lambda_i : i = 1, \dots r\}$. Hence, the common optimal value of (7.1) and (7.2) is $\lambda_{\min}(c)$ and an optimal solution to (7.1) is g_{i*} ,where g_{i*} is the primitive idempotent in spectral decomposition (7.3) corresponding to eigenvalue $\lambda_{\min}(c)$. Let $q = 2e - c$. It is obvious that q and c have the same sets of primitive idempotents (Jordan frames) in their spectral decompositions and $\lambda_{\max}(q) = 2 - \lambda_{\min}(c)$.

Lemma 3. *We have:*

i) $\lambda_{\max}(q) \in [1, 3]$;

ii) if s is a primitive idempotent such that $\langle q, s\rangle \geq \lambda_{\max}(q) - \epsilon$,then $\langle c, s\rangle \leq \lambda_{\min}(c) + \epsilon$.

Proof. Let (7.3) be the spectral decomposition of c. Then

$$\| c \| = \langle c, c\rangle^{1/2} = (\sum_{i=1}^{r} \lambda_i^2)^{1/2}.$$

Since $\| c \| \leq 1$, we have $| \lambda_i | \leq 1, i = 1, \dots, r$. Furthermore,

$$q = \sum_{i=1}^{r} (2 - \lambda_i) g_i.$$

We obviously have:

$$1 \leq 2 - | \lambda_i | \leq | 2 - \lambda_i | = 2 - \lambda_i \leq 2 + | \lambda | \leq 3,$$

which proves i). If $\langle q, s\rangle \geq \lambda_{\max}(q) - \epsilon$, then

$$\langle c, s\rangle = 2 - \langle q, s\rangle \leq 2 - \lambda_{\max}(q) + \epsilon = \lambda_{\min}(c) + \epsilon.$$

□

Each iteration described in Theorem 1 requires solving the problem of the type (7.1) approximately. In light of Lemma 3 and preceding discussion, it suffices to solve the following problem: given $q \in \Omega$ such that $\lambda_{\max}(q) \in [1, 3]$ and $\epsilon > 0$, find a primitive idempotent $s \in V$ satisfying the following property:

$$\langle q, s\rangle \geq \lambda_{\max}(q) - \epsilon.$$

Let

$$V = V_1 \oplus V_2 \oplus \ldots \oplus V_l$$

be the decomposition of V into the direct sum of pairwise orthogonal simple algebras. Such a decomposition exists and, moreover, all Euclidean simple algebras are completely classified (see[1],Chapter 5). If

$$q = \sum_{i=1}^{l} q_i, q_i \in V_i, i = 1, \ldots, l,$$

than it is clear that

$$\lambda_{\max}(q) = \max\{\lambda_{\max}(q_i) : i = 1, \ldots, l\}.$$

Hence, solving (7.1) is reduced to finding $\lambda_{\max}(q_i)$ for each i. Thus, it suffices to consider the problem of finding $\lambda_{\max}(q)$ and the corresponding primitive idempotent for the case, where V is simple.

1. V is a spin algebra , i.e. $V = \mathbf{R} \times W$, where $(W, \langle , \rangle)$ is an Euclidean vector space (possibly infinite-dimensional [2],[4]). Let $v = (s, x) \in V$. Then $tr(v) = 2s$. the spectral decomposition has the form:

$$v = \lambda_1 g_1 + \lambda_2 g_2,$$

$$\lambda_1 = s + \parallel x \parallel, \lambda_2 = s - \parallel x \parallel,$$

$$g_1 = \frac{1}{2}(1, \frac{x}{\parallel x \parallel}), g_2 = \frac{1}{2}(1, -\frac{x}{\parallel x \parallel}),$$

assuming $x \neq 0$. Hence, $\lambda_{\max}(v) = s + \parallel x \parallel$ and the maximum is attained at g_1. The problem (7.1) admits an analytic solution.

2. $V = Herm(n, \mathbf{R}), V = Herm(n, \mathbf{C})$, i.e. V is the algebra of n by n real (respectively, complex) symmetric (respectively Hermitian) matrices. This case is analyzed in detail in [5], where it is proposed to use the standard power method.

3. $V = Herm(n, \mathbf{H})$, i.e. the algebra of n by n Hermitian matrices with quaternion entries. As in [6], we consider a realization of V by Hermitian matrices with complex entries. Namely,

$$V = \{C \in Herm(2n, \mathbf{C} : JC = \bar{C}J\}.$$

Here

$$J = \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix},$$

where I_n is n by n identity matrix. Let $q \in V$. We can consider the usual power iterations

$$\xi_k = \frac{q^k \xi_0}{\| q^k \xi_0 \|_{2n}},$$

where $\xi_0 \in \mathbf{C^{2n}}$ is an arbitrary nonzero vector. Notice that for $\xi \in \mathbf{C^{2n}}, \| \xi \|_{2n} = \sqrt{\xi^* \xi}$, where $\xi^* = \bar{\xi}^T$. Then we can expect that $\rho_k = \xi_k^* q \xi_k$ will converge to $\lambda_{\max}(q)$ and ξ_k will converge to an eigenvector $\eta, \| \eta \|_{2n} = 1$ of q with the eigenvalue $\lambda_{\max}(q)$. In this case , the corresponding primitive idempotent will have the form:

$$\varphi(\eta) = \eta\eta^* + J\bar{\eta}(J\bar{\eta}^*).$$

For details, see [6]. In particular, it is shown in [6] that

$$\langle q, \varphi(\eta) \rangle = \eta^* q \eta = \lambda_{\max}(q).$$

It is worthwhile to mention that the multiplicity of eigenvalue $\lambda_{\max}(q)$ of a Hermitian matrix in V, will be at least 2. Thus, it is important that the complexity analysis of power method in [5] allows multiple maximal eigenvalue. Notice that

$$\rho_k = \frac{\xi_0^* q^{2k+1} \xi_0}{\xi_0 q^{2k} \xi_0} = \frac{\langle q^{2k+1}, \varphi(\xi_0) \rangle}{\langle q^{2k}, \varphi(\xi_0) \rangle}.$$

Let

$$q = \sum_{i=1}^{n} \lambda_i g_i$$

be the spectral decomposition of q as an element of Jordan algebra V. Let, further,

$$\varphi(\xi_0) = \sum_{i=1}^{n} w_i g_i + \sum_{1 \le i < j \le n} \varphi_{ij}$$

be the Peirce decomposition of $\varphi(\xi_0)$ with respect to the Jordan frame $g_1, \dots g_n$. (see [1] Theorem IV.2.1).Then

$$\rho_k = \frac{\sum_{i=1}^{n} w_i \lambda_i^{2k+1}}{\sum_{i=1}^{n} w_i \lambda_i^{2k}},$$

where $w_i = \langle \varphi(\xi_0), g_i \rangle \ge 0, i = 1, \dots, n$. Compare this expression with (5.15) in [5]. It is therefore clear that the complexity analysis in [5] of power iterations is applicable in our situation. Finally, we notice that for any $q \in \Omega$ we have $P(q^k) = P(q)^k$ and $\lambda_{\max}(P(q)) = \lambda_{\max}(q)^2$. Here P is the quadratic representation of V (see [1], p. 32). Thus, for general simple algebra V one can apply power iterations to $P(q)$ (which is a linear map of V into itself) to find $\lambda_{\max}(q)$. While in general it would dramatically increase the dimension of the vector space, it definitely can be applied to the exceptional simple algebra $Herm(3, \mathbf{O})$ whose real dimension is just 27.

In the Remark after Theorem 1 we used the rank inequality. We now prove it.
THEOREM 2. *Let* $x, y \in \bar{\Omega}$. *Then*

$$rank(x+y) \leq rank(x) + rank(y).$$

Proof. Since rank is obviously an an additive function with respect to decomposition of V into direct sum of irreducible subalgebras, we can assume without loss of generality that V is simple. Consider first the case $x+y \in \Omega$. Denote by A the linear operator

$$P(x+y)^{-1/2} = P((x+y)^{-1/2}).$$

Here P is the quadratic representation of V. Then we have

$$A(x) + A(y) = e.$$

Since A is the automorphism of the cone Ω, it preserve the rank (see [1], Proposition IV.3.1). Thus, it suffices to consider the case

$$x + y = e.$$

If

$$x = \sum_{i=1}^{r} \lambda_i g_i$$

is the spectral decomposition of x, then

$$y = \sum_{i=1}^{r} (1-\lambda_i) g_i$$

is the spectral decomposition of y. Since $x, y \in \bar{\Omega}$, we have $0 \leq \lambda_i \leq 1$ for all i. Let

$$\Delta_0 = \{i \in [1,r] : \lambda_i = 0\}, \Delta_1 = \{i \in [1,r] : \lambda_i = 1\}, \Delta_2 = \{i \in [1,r] : 0 < \lambda_i < 1\}.$$

Then $rank(x) = card\Delta_1 + card\Delta_2, rank(y) = card\Delta_0 + card\Delta_2$. Hence,

$$rank(x) + rank(y) = card\Delta_0 + card\Delta_1 + 2card\Delta_2$$

$$\geq card\Delta_0 + card\Delta_1 + card\Delta_2 = r = rank(e) = rank(x+y).$$

Consider now the case $rank(x+y) < rank(V)$. Let

$$x + y = \sum_{i=1}^{r} \lambda_i g_i$$

be the spectral decomposition with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_s > \lambda_{s+1} = \ldots = \lambda_r = 0$. Consider the idempotent

$$g = \sum_{i=s+1}^{r} g_i.$$

Then $x+y$ belongs to subalgebra

$$V(g,0) = \{z \in V : z \circ g = 0\}.$$

But then both x and y belong to this subalgebra. Indeed,

$$x \circ g + y \circ g = (x+y) \circ g = 0.$$

Hence, $tr(x \circ g) + tr(y \circ g) = 0$. Or

$$\langle x, g \rangle + \langle y, g \rangle = 0.$$

Since $\langle x, g \rangle \geq 0, \langle y, g \rangle \geq 0$, it implies

$$\langle x, g \rangle = \langle y, g \rangle = 0.$$

But then $x \circ g = y \circ g = 0$. See Proposition 2 in [6]. Notice that $x+y$ is invertible in $V(g,0)$. Indeed, $e-g$ is the unit element in $V(g,0)$ and

$$(x+y) \circ (\sum_{i=1}^{s} \frac{1}{\lambda_i} g_i) = e - g$$

. Hence, we reduced our problem to the previous case (notice that the rank with respect to subalgebra $V(g,0)$ coincides with the rank in V. See e.g. [6]). □

8. Concluding remarks. In present paper we have generalized Hazan's algorithm to the case of a general symmetric programming problem with a bounded feasible domain. We have shown that the major property of the algorithm which detects a low rank approximation to an optimal solution is preserved if the concept of rank of a symmetric matrix is substituted (generalized) by the concept of rank of an element of an Euclidean Jordan algebra. That gives an additional evidence that the concept of Jordan-algebraic rank play an important role in understanding the approximate solutions of general symmetric programming problems(see in this respect also [7]). In addition, we have described in detail how to use the "block diagonal" structure in decomposing Frank-Wolfe algorithm for general symmetric cones. In particular, in case of second-order cone programming problems (even with infinite-dimensional blocks) the corresponding eigenvalue problem admits a simple analytic solution.

REFERENCES

[1] J. Faraut and A. Koranyi, *Analysis on Symmetric Cones*,Clarendon Press, Oxford, 1994,382p.

[2] L. Faybusovich and T. Tsuchiya, *Primal-dual algorithms and infinite-dimensional Jordan algebras of finite rank*, Mathematical Programming, vol.97 (2003), no. 3, pp. 471-493.

[3] E.Hazan, *Sparse approximate solutions to semidefinite programs* , In Proceedings of the 8th Latin American conference on Theoretical Informatics (LNCS 4957),pp. 306-316. Springer-Verlag,Berlin, 2008.

[4] L. Faybusovich, T. Mouktonglang and T. Tsuchiya, *Implementation of infinite - dimensional interior - point method for solving multi - criteria linear - quadratic control problem*, Optim. Methods Softw. 21(2006) no. 2, 315 - 341.

[5] B. Gartner and J. Matousek, *Approximation Algorithms and Semidefinite Programming*, Springer-Verlag, 2010, 251p.

[6] L. Faybusovich ,*Jordan-algebraic approach to convexity theorems for quadratic mappings*, SIAM J. Optim. 17(2006),no 2, 558-576.

[7] L.Faybusovich, *Jordan-algebraic aspects of optimization:randomization*, Optim. Methods Softw. 25(2010) no 4-6, 763-779.