Working Paper Series

Adaptive Information Systems and Modelling in Economics and Management Science

Portfolio selection via replicator dynamics and projections of indefinite estimated covariances

Immanuel M. Bomze

Working Paper No. 75
October 2000

Working Paper Series

Portfolio selection via replicator dynamics and projections of indefinite estimated covariances

Immanuel M. Bomze

Institut f. Statistik und Decision Support Systems, Univ. Wien

This version: September 11, 2000

Abstract

Replicator dynamics are an increasingly popular device for obtaining (local) solutions of considerably high quality to so-called standard quadratic optimization problems, which consist of finding maxima of (possibly indefinite) quadratic forms over the standard simplex. In the simplest version of portfolio selection, the quadratic form is theoretically negative-semidefinite, so that any local solution automatically is a global one. However, if it comes to more realistic set-ups, then (i) no market portfolio is available, so that one ends up with an indefinite theoretical problem; (ii) estimated covariance matrices modelling risk may be indefinite also. This paper deals with both problems in a different way: (i) will be solved via escape steps to avoid low-quality local solutions while (ii) is dealt with by several projection strategies which convert the indefinite estimated covariance matrix into a positive-semidefinite one.

1 Introduction

A standard quadratic problem (QP) consists of finding (global) maximizers of a quadratic form over the standard simplex, i.e. we consider (global) optimization problems of the form

$$\mathbf{x}^\top A\mathbf{x} \to \max! \quad \text{subject to } \mathbf{x} \in \Delta\,, \tag{1}$$

where A is an arbitrary symmetric $n \times n$ matrix; a $^\top$ denotes transposition; and Δ is the standard simplex in n-dimensional Euclidean space $\mathbb{R}^n$,

$$\Delta = \{\mathbf{x} \in \mathbb{R}^n : x_i \geq 0 \text{ for all } i \in \mathcal{V},\ \mathbf{e}^\top\mathbf{x} = 1\}\,,$$

where $\mathcal{V} = \{1, \ldots, n\}$. Here $\mathbf{e} = \sum_{i\in\mathcal{V}} \mathbf{e}_i = [1, \ldots, 1]^\top$ while $\mathbf{e}_i$ denotes the i-th standard basis vector in $\mathbb{R}^n$.

Note that the maximizers of (1) remain the same if A is replaced with $A + \gamma\mathbf{e}\mathbf{e}^\top$ where γ is an arbitrary constant. So without loss of generality assume henceforth that all entries of A are non-negative with positive diagonal elements. This assumption is important for the monotonicity and convergence results in Section 2.

Furthermore, the question of finding maximizers of a general quadratic function $\mathbf{x}^\top Q\mathbf{x} + 2\mathbf{c}^\top\mathbf{x}$ over Δ can be homogenized in a similar way by considering the rank-two update $A = Q + \mathbf{e}\mathbf{c}^\top + \mathbf{c}\mathbf{e}^\top$ in (1) which has the same objective values.

Of course, quadratic optimization problems like (1) – even regarding the detection of local solutions – are NP-hard [14]. Nevertheless, there are several exact procedures which try to exploit favourable data constellations in a systematic way, and to avoid the worst-case behaviour whenever possible. One prototypical example for this type of algorithms is specified in Section 2 below. For a review on standard QPs and its applications, which also offers a justification for terminology see [4].

2 Local and global optimization of standard QPs: replicator dynamics and escape strategies

First we concentrate on the evolutionary approach to local solutions of standard QPs. To this end, consider the following dynamical system operating on Δ:

$$\dot{x}_i(t) = x_i(t)[(A\mathbf{x}(t))_i - \mathbf{x}(t)^\top A\mathbf{x}(t)]\,, \quad i \in \mathcal{V}\,, \tag{2}$$

where a dot signifies derivative w.r.t. time t, and a discrete time version

$$x_i(t+1) = x_i(t)\frac{(A\mathbf{x}(t))_i}{\mathbf{x}(t)^\top A\mathbf{x}(t)}\,, \quad i \in \mathcal{V}\,. \tag{3}$$

Both (2) and (3) arise in population genetics under the name *selection equations* where they are used to model time evolution of haploid genotypes, A being the (symmetric) fitness matrix, and $x_i(t)$ representing the relative frequency of allele i in the population. Since they serve to model replicating entities also in a much more general contex, they frequently are called *replicator dynamics* now.

The discrete-time dynamics (3) is well-behaved in the sense that it leaves the simplex Δ forward-invariant if and only if

$$a_{ij} \geq 0 \quad \text{for all } (i,j) \in \mathcal{V}\times\mathcal{V} \quad \text{and} \quad a_{ii} > 0 \quad \text{for all } i \in \mathcal{V}\,. \tag{4}$$

Conditions (4) are necessary and sufficient for the properties $A\mathbf{x} \geq o$ and $\mathbf{x}^\top A\mathbf{x} > 0$ for all $\mathbf{x} \in \Delta$, which in turn guarantee that $\mathbf{x}(t) \in \Delta$ entails that $\mathbf{x}(t+1)$ is well-defined and also $\mathbf{x}(t+1) \in \Delta$. However, if (4) is violated but nevertheless for some starting point $\mathbf{x}(0) \in \Delta$, the points $\mathbf{x}(t)$ are well-defined for all $t > 0$, then neither $\mathbf{x}(t)^\top A\mathbf{x}(t)$ need to be increasing in t nor $\mathbf{x}(t)$ need to converge as $t \to \infty$, as observed in [21], for

$$A = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & 3 \\ -1 & 3 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{x}(0) = \begin{bmatrix} 2/3 \\ 0 \\ 1/3 \end{bmatrix}$$

we get a 2-periodic orbit for $\mathbf{x}(t)$ with $\mathbf{x}(0)^\top A\mathbf{x}(0) = 1/9 < 9 = \mathbf{x}(1)^\top A\mathbf{x}(1)$.

From the form of the dynamics (2) and (3) it is immediately clear that – at least under (4) – not only the whole simplex, but also its faces,

$$\Delta_S = \{\mathbf{x} \in \Delta : x_i = 0 \text{ if } i \in \mathcal{V}\setminus S\}$$

are forward invariant for every $S \subset \mathcal{V}$. Further, the *stationary points* under (2) and (3) coincide, and all local solutions of (1) are among these. Of course, there are quite many stationary points, e.g. all *vertices* $\mathbf{e}_1, \ldots, \mathbf{e}_n$ of Δ (observe that $\Delta_{\{i\}} = \{\mathbf{e}_i\}$ for all $i \in \mathcal{V}$). However, only those $\mathbf{x}$ are serious candidates for strict[1] local solutions which are *asymptotically stable*, i.e. every solution to (2) or (3) which starts close enough to $\mathbf{x}$, will converge to $\mathbf{x}$ as $t \nearrow \infty$. These results and many more connections between three different fields: optimization theory, evolutionary game theory, and qualitative theory of dynamical systems can be found, e.g., in [7].

The significance for local solution to standard QPs is best captured by the following results, the proofs of which can be found in [11],[16],[2] and [8].

Theorem 1 *Suppose that $A = A^\top$ satisfies (4). Then*

1. *the objective function $\mathbf{x}^\top A\mathbf{x}$ is (strictly) increasing over time along non-constant trajectories, i.e.*
$$\mathbf{x}(t)^\top A\mathbf{x}(t) < \mathbf{x}(t+1)^\top A\mathbf{x}(t+1) \quad \text{if } \mathbf{x}(t) \text{ is not stationary;}$$
2. *every trajectory $\mathbf{x}(t)$ converges to a stationary point;*
3. *if no principal minor of A vanishes, then with probability one any trajectory converges to a strict local solution $\overline{\mathbf{x}}$ of (1). Furthermore, if $S = \{i \in \mathcal{V} : \overline{x}_i > 0\}$, then*
   (a) *$\mathbf{y}^\top A\mathbf{y} < \overline{\mathbf{x}}^\top A\overline{\mathbf{x}}$ for all $\mathbf{y} \in \Delta_S$ with $\mathbf{y} \neq \overline{\mathbf{x}}$;*
   (b) *the relative interior of Δ_S is contained in the basin of attraction of $\overline{\mathbf{x}}$.*

Although strictly increasing objective values are guaranteed as we follow trajectories under (2) or (3), we could get stuck in an inefficient local solution $\overline{\mathbf{x}}$ of (1). A global optimization procedure therefore must incorporate a decision maker at a higher level than "blind" nature. One attempt is based on the reduction of problem dimension at the cost of generating a series of subproblems, which seems to be a promising approach in view of the NP-hardness in quadratic programming. This procedure is circumscribed by *genetic engineering via negative fitness* (G.E.N.F.) because of the following interpretation from [8].

From Theorem 1, a strict local solution $\overline{\mathbf{x}}$ of (1) must be a global one if all $\overline{x}_i > 0$. Consequently, at an inefficient local solution necessarily $\overline{x}_i = 0$ for some i. In the usual genetic interpretation, this means that some alleles die out during the selection process, and these are therefore unfit in the environment currently prevailing. The escape step now artificially re-introduces some alleles which would have gone extinct during the natural selection process. This is done via the negative fitness approach: remove all alleles which are not unfit, i.e. all $i \in S = \{i \in \mathcal{V} : \overline{x}_i > 0\}$. Then determine fitness *minimizers* in the reduced problem, i.e. consider problem (1) with A replaced by

$$\overline{A} = [\gamma_S - a_{ij}]_{i,j \in \mathcal{V} \setminus S}$$

where $\gamma_S = \max_{i,j \in \mathcal{V} \setminus S} a_{ij}$ is the maximum fitness of all extinct alleles. After having obtained a local solution $\mathbf{y}$ of this auxiliary problem, put

$$T = \{j \in \mathcal{V} \setminus S : y_j > 0\}$$

which can be viewed as the set of "truly unfit" alleles. Now the following result can be shown [3]:

[1]A local solution $\mathbf{x}$ to (1) is said to be *strict* if there exists a neighborhood U of $\mathbf{x}$ such that $f(\mathbf{y}) < f(\mathbf{x})$ if $\mathbf{y} \in \Delta \cap U \setminus \{\mathbf{x}\}$.

Theorem 2 *Suppose* $\overline{\mathbf{x}}$ *is local solution of master problem (1) with surviving allele set* $S = \{i \in \mathcal{V} : \overline{x}_i > 0\}$. *Pick a disjoint set* T *of size* $m \geq 1$ *by 'negative genetic engineering' as above. For all* $s \in S$ *and* $t \in T$, *replace* a_{si} *with* a_{ti}, *remove all (other) unfit* $j \in T$. *Consider the reduced problem* $\mathcal{P}_{t\to s}$, *i.e. problem (1) in* $n - m$ *variables for the so obtained matrix* $A_{t\to s}$.
Then $\overline{\mathbf{x}}$ *is global solution of the master problem (1) if and only if for all* $(s,t) \in S \times T$, *the maximum of* $\mathcal{P}_{t\to s}$ *does not exceed the current best value* $\overline{\mathbf{x}}^\top A \overline{\mathbf{x}}$.
In the negative, i.e. if $\mathbf{u}^\top A_{t\to s} \mathbf{u} > \overline{\mathbf{x}}^\top A \overline{\mathbf{x}}$ *for some* $\mathbf{u} \in I\!R^{n-m}$ *in the standard simplex, and if* $j \in T$ *is chosen such that*

$$\sum_{p \notin T \cup \{s\}} a_{jp} u_p + \frac{1}{2} a_{jj} u_s \geq \sum_{p \notin T \cup \{s\}} a_{qp} u_p + \frac{1}{2} a_{qq} u_i \quad \text{for all } q \in T\,,$$

then a strictly improving feasible point $\tilde{\mathbf{x}}$ *is obtained as follows:*

$$\tilde{x}_q = \begin{cases} u_i & \text{if } q = j\,, \\ 0 & \text{if } q \in T \cup \{s\} \setminus \{j\}\,, \\ u_q & \text{if } q \in \mathcal{V} \setminus T\,. \end{cases}$$

In view of the possible combinatorial explosion in effort with increasing number of variables, this dimension reducing strategy seems to be promising: if k is the size of S, the above result yields a series of km standard QPs in $n - m$ variables rather than in n. We are now ready to describe the algorithm which stops after finitely many repetitions, since it yields strict local solutions with strictly increasing objective values (cf. Theorem 1). So suppose that A has no negative entries and that no principal minor of A vanishes (then $a_{ii} > 0$ for all $i \in \mathcal{V}$, so that (4) follows).

Algorithm

1. Start with $\mathbf{x}(0) = \frac{1}{n}\mathbf{e}$ or nearby, iterate (3) until convergence;
2. the limit $\overline{\mathbf{x}} = \lim_{t\to\infty} \mathbf{x}(t)$ is a strict local solution with probability one; call the escape procedure of Theorem 2 to improve the objective, if possible; denote the improving point $\tilde{\mathbf{x}}$;
3. repeat 1., starting with $\mathbf{x}(0) = \tilde{\mathbf{x}}$.

Of course, the escape procedure described above is not the only one which completes the above algorithmic roster. A possible alternative is a hybrid procedure which merges above-described techniques with a variant of the primal-dual interior point methods frequently used in the context of semidefinite programming (SDP). This extension is called *copositive programming* because the cone which defines the feasible set of the underlying optimization problem is no longer that of all semidefinite symmetric matrices, but rather that of all copositive matrices. For details see [6]. The close connections between copositivity and standard QPs are investigated in [5]. In section 4, we will return to the SDP framework.

3 Portfolio selection as a standard QP

The familiar mean/variance portfolio selection problem (see, e.g. [17, 18]) can be formalized as follows: suppose there are n securities to invest in, at an amount expressed in relative shares

$x_i \geq 0$ of an investor's budget. Thus, the budget constraint reads $\mathbf{e}^\top \mathbf{x} = 1$, and the set of all feasible portfolio (investment plans) is given by Δ. Now, given the expected return m_i of security i during the forthcoming period, and an $n \times n$ covariance matrix V across all securities, the investor faces a multiobjective problem to maximize expected return $\mathbf{m}^\top \mathbf{x}$ and simultaneously minimize the risk $\mathbf{x}^\top V \mathbf{x}$ associated by her decision $\mathbf{x}$. One of the most popular approaches to such type of problems in general applications is that the user prespecifies a parameter β which in her eyes balances the benefit of high return and low risk, i.e. consider the parametric QP

$$f_\beta(\mathbf{x}) = \mathbf{m}^\top \mathbf{x} - \beta \mathbf{x}^\top V \mathbf{x} = \frac{1}{2} \mathbf{x}^\top (\mathbf{m}\mathbf{e}^\top + \mathbf{e}\mathbf{m}^\top - 2\beta V)\mathbf{x} \to \max! \quad \text{subject to } \mathbf{x} \in \Delta \,. \tag{5}$$

Note that for fixed β, this is again a standard QP (cf. Section 1). Moreover in theory the matrix V is, as an exact covariance matrix, positive semidefinite (although it could be singular in many applications [18]), so that (5) is a convex problem. Hence in the algorithm of Section 2, the escape step would be superfluous, so there were no repetitions (note that the transformation proposed in Section 1 could destroy positive semidefiniteness on the whole of $\mathbb{R}^n$, but the objective function would still be convex over Δ). Anyhow, the question remains how to choose β. In finance applications, the notion of market portfolio is used to determine a reasonable value for this parameter. This emerges more or less from an exogenous artefact, namely by introduction of a completely risk-free asset which is used to scale return versus risk. For details see, e.g., Chapter 3 of [15].

Following a result of Best and Ding [1] who consider the problem

$$\max_{\beta > 0} \max_{\mathbf{x} \in \Delta} \frac{1}{\beta} f_\beta(\mathbf{x}) \,, \tag{6}$$

also a purely endogenous derivation of market portfolio seems to be possible: they show how optimal solutions $(\beta^*, \mathbf{x}^*)$ for (6) emerge from a single standard QP (1) with, e.g. $A = 2\mathbf{m}\mathbf{m}^\top - V$. Hopefully some empirical results with this approach on the Austrian stock market can be reported soon.

Of course, there are many approaches to solve the parametric QP (5), e.g. in [18, 1]. Given we know the solutions as specified in [1]

$$\mathbf{x}^*(\beta) = \mathbf{g}_i + \beta \mathbf{h}_i \,, \quad \text{if } \beta_{i-1} < \beta < \beta_i \,, \; i \in \{1, \ldots, t\}$$

with $0 = \beta_0 < \beta_1 < \ldots < \beta_t = \infty$ and known vectors $\mathbf{g}_i$ and $\mathbf{h}_i$, it is natural to ask which of the investor's utility structure would yield the same outcome [9, 12]. In the present context, this question (referring to a nonparametric function class giving the same solution) can be rephrased as a typical application of hypersensitivity analysis in connection with composite quadratic programming, which will be addressed in a forthcoming paper [10].

4 Orthoprojection on the semidefinite cone

In practice, securities are often highly correlated, and in time-series analysis one frequently encounters the situation that some of the most reliable estimators $\tilde{V}$ of the unknown covariance matrix V lack semidefiniteness properties [20], pp.134ff. More recently, also the econometric community has been aware of this phenomenon [19], which of course is a nuisance but nevertheless can be handled by the procedures proposed in Section 2, with the expectation to receive a good local solution. However, since we know that indefiniteness is an artefact generated by the estimation procedure, it seems to be natural to correct $\tilde{V}$ for semi-definiteness, arriving thus at some alternative estimator $\widehat{V}$. Observe that under mild conditions, $\tilde{V}$ will be close to the true covariance matrix V, so that we should try to keep $\widehat{V}$ as close to $\tilde{V}$ as possible. Since the two competing estimators $\tilde{V}$ and $\widehat{V}$ have quadratic forms of different curvature, it is clear that the distance between the solutions (the maximizers) may be large even if the maximum objective values are close to each other. This is of particular interest when it comes to the usual sensitivity questions.

Therefore, it is reasonable to use some projection method which transforms an indefinite matrix $\tilde{V}$ into a positive-semidefinite one $\widehat{V}$, and this is the content of the subsequent sections.

Let $\mathcal{M}_n$ denote the $d = \binom{n+1}{2}$-dimensional space of all symmetric $n \times n$ matrices. Euclidean geometry on $\mathcal{M}_n$ can be described either via vectorization of the (upper triangular part of the) matrices, thus transferring everything to $\mathbb{R}^d$ with the usual geometry, or via the trace operation,

$$\langle A, B\rangle = \text{trace}\,(AB) = [\text{vec}(A)]^\top[\text{vec}(B)]\,,\ A, B \in \mathcal{M}_n\,,$$

where $^\top$ denotes transposition. Note that if vectorize the upper triangular part instead of the full matrices, we still would get $\langle A, B\rangle = [\text{vec}(A)]^\top L[\text{vec}(B)]$ for a positive definite diagonal $d \times d$ matrix L which does not depend on (A, B), so essentially everything remains unchanged.

In this section we calculate explicitly the orthoprojection of an (arbitrary) indefinite symmetric $n \times n$ matrix $V \in \mathcal{M}_n$ (we drop the tilde for notational convenience, in the contex above $\tilde{V}$ corresponds to V, and $\widehat{V}$ to $\Pi(V)$ defined in Theorem 3 below, respectively) onto the cone of positive semidefinite matrices

$$\mathcal{P}_n = \{Q \in \mathcal{M}_n : O \preceq Q\}\,,$$

where $A \preceq B$ denotes the Löwner ordering on $\mathcal{M}_n$, i.e. signifies positive semidefiniteness of $B - A$. It is well known that w.r.t. $\langle .,.\rangle$, the cone $\mathcal{P}_n$ is self-dual, i.e. $\langle A, Q\rangle \geq 0$ holds for all $Q \in \mathcal{P}_n$ if and only if $A \in \mathcal{P}_n$.

In the sequel, a vector $\mathbf{z}$ is called *normalized* if it satisfies $\|\mathbf{z}\| = \sqrt{\mathbf{z}^\top\mathbf{z}} = 1$.

Theorem 3 *Consider an (arbitrary) symmetric $n \times n$ matrix $V \in \mathcal{M}_n$ and denote by U_V the orthogonal $n \times n$ matrix containing the normalized eigenvectors of V as its columns while $L_V = \text{diag}\big(\lambda_1(V), \ldots, \lambda_n(V)\big)$ contains the eigenvalues in an order compatible with U_V. Let $\alpha^+ = \max\{0, \alpha\}$ for any $\alpha \in \mathbb{R}$, and denote by $L_V^+ = \text{diag}\big([\lambda_1(V)]^+, \ldots, [\lambda_n(V)]^+\big)$. Then the map*

$$\begin{aligned} \Pi : V \;&\mapsto U_V L_V^+ U_V^\top \\ \mathcal{M}_n \;&\to \mathcal{P}_n \end{aligned} \tag{7}$$

is well-defined on $\mathcal{M}_n$ and satisfies $\Pi(V) \in \mathcal{P}_n$ as well as $V\Pi(V) = \Pi(V)V$. Moreover, Π is the orthoprojector onto $\mathcal{P}_n$ w.r.t. $\langle .,.\rangle$, and thus we have for $\|B\| = \sqrt{\langle B, B\rangle}$

$$\|V - \Pi(V)\| \leq \|V - Q\| \quad \textit{for all } Q \in \mathcal{P}_n\,,$$

with equality only if $Q = \Pi(V)$.

Proof. It is evident that Π is well defined since $\Pi(V)$ does not depend on the ordering of eigenvalues and -vectors. Further, L_V^+ is positive semidefinite and thus $\Pi(V) = U_V L_V^+ U_V^\top \in \mathcal{P}_n$. Let $B = V - \Pi(V)$ be the projection error. Since $\alpha^+ \geq \alpha$, we have $-B = U_V(L_V^+ - L_V)U_V^\top \in \mathcal{P}_n$ and therefore $-\langle B, Q\rangle \geq 0$ for all $Q \in \mathcal{P}_n$ by self-duality of $\mathcal{P}_n$. Also, we have $(\alpha - \alpha^+)\alpha^+ = 0$ for all α and hence $\langle B, \Pi(V)\rangle = 0$, so that we arrive at $\langle B, \Pi(V) - Q\rangle = -\langle B, Q\rangle \geq 0$ for all $Q \in \mathcal{P}_n$. The rest of the argumentation is standard: $\|V - Q\|^2 = \|B + \Pi(V) - Q\|^2 = \|B\|^2 + \|\Pi(V) - Q\|^2 + 2\langle B, \Pi(V) - Q\rangle \geq \|B\|^2$ with equality only if $Q = \Pi(V)$. □

5 An SDP projection approach

Rather than concentrating on Euclidean geometry on $\mathcal{M}_n$, one may prefer picking that $Q \in \mathcal{P}_n$ which minimizes the difference in (estimated) variances of random variates $y = z^\top x$ for arbitrary coefficients z_i which are, without loss of generality, assumed to give a normalized vector $\mathbf{z} = [z_1, \ldots, z_n]^\top \in \mathbb{R}^n$. Taking the least favorable case into account, this means to minimize $\max_{\|\mathbf{z}\|=1} |\mathbf{z}^\top(V - Q)\mathbf{z}|$. This quantity coincides with the operator norm $\|B\|_{\text{op}} = \max_{\|\mathbf{z}\|=1} \|B\mathbf{z}\|$ if $B = V - Q$, and there are several ways to express it:

Lemma 4 *Let $B \in \mathcal{M}_n$ be an arbitrary symmetric $n \times n$ matrix and define $|B| = \sqrt{B^2} = \sqrt{B^\top B}$ with $\sqrt{Q}$ denoting the symmetric square root factorization of a positive semidefinite $Q \in \mathcal{P}_n$. Then*

$$\max_{\|\mathbf{z}\|=1} |\mathbf{z}^\top B\mathbf{z}| = \|B\|_{\text{op}} = \lambda_{\max}(|B|) = \sqrt{\lambda_{\max}(B^\top B)}\,, \tag{8}$$

and this number concides with the solution t^* of the problem

$$\min\{t \in \mathbb{R} : B \preceq tI_n \text{ and } -B \preceq tI_n\}\,. \tag{9}$$

Proof. Let us start establishing (8). First note that for any normalized $\mathbf{z} \in \mathbb{R}^n$ we get $|\mathbf{z}^\top B\mathbf{z}| \leq \|\mathbf{z}\|\,\|B\mathbf{z}\| \leq \|\mathbf{z}\|\,\|B\|_{\text{op}} = \|B\|_{\text{op}}$. Next, if a normalized $\mathbf{w} \in \mathbb{R}^n$ is chosen such that $\|B\mathbf{w}\| = \|B\|_{\text{op}}$, we obtain $\|B\|_{\text{op}}{}^2 = \|B\mathbf{w}\|^2 = \mathbf{w}^\top(B^\top B)\mathbf{w} \leq \lambda_{\max}(B^\top B) = \lambda_{\max}(|B|)^2$, where the last equality follows from the spectral mapping theorem (observe that $|B|$ and $B^\top B$ commute by construction). Finally, choose $\mathbf{u}$ to be a normalized eigenvector corresponding to $\lambda_{\max}(B^\top B)$. Then, using again the spectral mapping theorem (of course, also B and $|B|$ commute because, using the notation of Theorem 3, $|B| = U_B|L_B|U_B^\top$ where $|L_B| = \text{diag}\big(|\lambda_1(B)|, \ldots, |\lambda_n(B)|\big)$, the latter $|\lambda|$ denoting the familiar modulus of a real number λ),

$$\lambda_{\max}(|B|)^2 = \mathbf{u}^\top(B^\top B)\mathbf{u} = \pm\lambda_{\max}(|B|)\mathbf{u}^\top B\mathbf{u} \leq \lambda_{\max}(|B|) \max_{\|\mathbf{z}\|=1} |\mathbf{z}^\top B\mathbf{z}|\,.$$

Hence we close the circle of inequalities between all quantities occurring in (8), so that the identities of them are established. It remains to show the assertion on the problem (9). To this end, observe that any t which is feasible for this problem, satisfies $(tI_n - B) \in \mathcal{P}_n$ and also $(tI_n + B) \in \mathcal{P}_n$. Since these two matrices commute, we get also

$$t^2 I_n - B = (tI_n - B)(tI_n + B) = \sqrt{tI_n + B}\,(tI_n - B)\sqrt{tI_n + B} \in \mathcal{P}_n\,,$$

or, equivalently, $B^2 \preceq t^2 I_n$, which entails $\lambda_{\max}(B^2) \leq t^2$ which also holds for the optimal solution $t = t^*$. Hence $t^* \geq \|B\|_{\text{op}}$ via (8). On the other hand, $t = \lambda_{\max}(|B|)$ is feasible for (9), since, by construction, $\pm B \preceq |B| \preceq \lambda_{\max}(|B|)I_n$. Hence the result follows. □

If we now, given an indefinite $V \in \mathcal{M}_n$, search for a positive semidefinite $Q \in \mathcal{P}_n$ such that $\lambda_{\max}(|V - Q|)$ is minimized, we arrive at the following SDP

$$\min\{t \in \mathbb{R} : -tI_n \preceq (V - Q) \preceq tI_n\} . \tag{10}$$

For easy recognition, let us reformulate (10) into standard form: put

$$Y = \begin{bmatrix} tI_n + Q - V & | & O \\ O & | & tI_n - Q + V \end{bmatrix} \quad \text{and} \quad X = \begin{bmatrix} t & | & o^\top & | & o^\top \\ o & | & Q & | & O \\ o & | & O & | & Y \end{bmatrix} .$$

Then t is feasible for problem (9) if and only if $Y \in \mathcal{P}_{2n}$; further, this relation holds with $Q \in \mathcal{P}_n$ if and only if $X \in \mathcal{P}_{3n+1}$ (observe that automatically $t \geq 0$ holds for every (9)-feasible t). Hence problem (10) can be reformulated into

$$\min\{\langle C, X\rangle : \langle A_\nu, X\rangle = b_\nu, \nu \in N , \text{ and } X \in \mathcal{P}_{3n+1}\} \tag{11}$$

with $C = \mathbf{e}_1\mathbf{e}_1^\top \in \mathcal{M}_{3n+1}$, where the matrices $A_\nu \in \mathcal{M}_{3n+1}$ and the numbers b_ν serve to express the block-diagonal structure of X and Y, as well as the relations between X, Y, Q, and t. Numerically, (11) is a sparse SDP for which recently quite efficient methods have been proposed (see, e.g. [13] and the references therein). Returning to the original SDP (10), we see that the variables are (t, Q), and the optimal solution (t^*, Q^*) represents both the projection Q^* of V onto $\mathcal{P}_n$, and the distance from V to Q^*, in that $t^* = \|V - Q^*\|_{\text{op}}$, cf. Lemma 4.

We now prove that the orthoprojection $\Pi(V)$ introduced in the previous section is part, in the sense that $Q^* = \Pi(V)$, of an optimal solution (t^*, Q^*) to the SDP (10).

Theorem 5 *Consider an indefinite symmetric $n \times n$ matrix $V \in \mathcal{M}_n$. With the terminology of Theorem 3, let $\lambda_i = \lambda_i(V)$ and denote by $\lambda^-(V) = \max\{|\lambda_i| : \lambda_i < 0\}$. Let (t^*, Q^*) be an optimal solution to the SDP (10). Then $t^* = \lambda^-(V)$ and also $\big(t^*, \Pi(V)\big)$ is an optimal solution to (10).*

Proof. Let $Q = \Pi(V) \in \mathcal{P}_n$. Then (t, Q) is feasible for the SDP (10) if and only if $-tI_n \preceq V - \Pi(V) \preceq tI_n$. Now $V - \Pi(V) = U_V \text{diag}\big(\lambda_1 - \lambda_1^+, \ldots, \lambda_n - \lambda_n^+\big) U_V^\top$ so that $-tI_n \preceq V - \Pi(V) \preceq tI_n$ if and only if $|\lambda_i - \lambda_i^+| \leq t$ holds for all i, which is the same as $t \geq \lambda^-(V)$. Hence also $\big(\lambda^-(V), \Pi(V)\big)$ is (10)-feasible, which entails $t^* \leq \lambda^-(V)$. On the other hand, let $\mathbf{v}$ be a normalized eigenvector of V for the eigenvalue $-\lambda^-(V)$. Since the optimal solution satisfies $Q^* \in \mathcal{P}_n$, we obtain

$$\mathbf{v}^\top (Q^* - V)\mathbf{v} \geq 0 + \lambda^-(V) = \lambda^-(V)$$

and hence $t^* = \|Q^* - V\|_{\text{op}} \geq \lambda^-(V)$ via Lemma 4, which together with the previously established inequality yields $t^* = \lambda^-(V)$. Of course, this optimal objective value can also be realized with the feasible choice $Q = \Pi(V)$. $\square$

Example 1. Let $V = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \in \mathcal{M}_2 \setminus \mathcal{P}_2$. Then $t^* = \lambda^-(V) = 1$ and $\Pi(V) = \frac{1}{2} \begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix} \in \mathcal{P}_2$, whereas also $Q^* = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \in \mathcal{P}_2$ yields an optimal solution (t^*, Q^*) to (10). Indeed, $Q^* - V = I_2$ so that $-t^* I_2 \preceq V - Q^* \preceq t^* I_2$ holds. This example shows that in general, the SDP (10) has more than one solution, in sharp contrast to the uniqueness result of Theorem 3. A more detailed investigation reveals that the set of all optimal solutions to (10) coincides with $\{\gamma \mathbf{e}\mathbf{e}^\top : 1 \leq \gamma \leq 2\}$.

Thus, if one wants to go a small step into the interior of $\mathcal{P}_2$, i.e. considers $(1+\varepsilon)\gamma\mathbf{e}\mathbf{e}^\top - \varepsilon V$ instead of the projection $\gamma\mathbf{e}\mathbf{e}^\top$ for some small $\varepsilon > 0$, one obtains $\mathbf{x} = [\frac{1}{2}, \frac{1}{2}]^\top$ as the (only) solution whereas $\mathbf{x}^\top V\mathbf{x}$ itself has two local minimizers over Δ, namely $\mathbf{e}_1$ and $\mathbf{e}_2$, with the same objective value. Observe that we flip from maximizing in the original standard QP (1) to minimizing, since the indefinite covariance matrix V to be projected enters the objective (5) with a negative sign.

References

[1] M. J. Best, B. Ding (1997), Global and local quadratic minimization. *J. Global Optimiz.* **10**, 77–90.

[2] I.M. Bomze (1997), Evolution towards the maximum clique. *J. Global Optimiz.* **10**, 143–164.

[3] — (1997), Global escape strategies for maximizing quadratic forms over a simplex. *J. Global Optimiz.* **11**, 325–338.

[4] — (1998), On standard quadratic optimization problems. *J. Global Optimiz.* **13**, 369–387.

[5] — (2000), Copositivity aspects of standard quadratic optimization problems. In: E. Dockner, R. Hartl, M. Luptacik, G. Sorger, eds., *Dynamics, optimization and economic analysis,* 1–11. Physica, Heidelberg.

[6] —, M. Dür, E. de Klerk, A. Quist, C. Roos, T. Terlaky (2000), On copositive programming and standard quadratic optimization problems, to appear in: Journal of Global Optimization.

[7] —, M. Pelillo, R. Giacomini (1997), Evolutionary approach to the maximum clique problem: empirical evidence on a larger scale. In: I.M. Bomze, T. Csendes, R. Horst, P.M. Pardalos, eds., *Developments in global optimization,* 95–108. Kluwer, Dordrecht.

[8] — , V. Stix (1999), Genetical engineering via negative fitness: evolutionary dynamics for global optimization. *Annals of O.R.* **89**, 279–318.

[9] G. Chamberlain (1983), A characterization of the distributions that imply mean-variance utility functions. *J. Econ. Theo.* **29**, 185–201.

[10] L. Churilov, I. M. Bomze, M. Sniedovich (1999), Hypersensitivity analysis for portfolio selection, submitted to Annals of OR.

[11] J.F. Crow, M. Kimura (1970), *An introduction to population genetics theory,* Harper & Row, New York.

[12] L. Epstein (1985), Decreasing risk aversion and mean-variance analysis. *Econometrica* **53**, 945–962.

[13] K. Fujisawa, M. Kojima, K. Nakata (1997), Exploiting sparsity in primal-dual interior-point methods for semidefinite programming. *Math. Prog.* **79**, 235–253.

[14] R. Horst, P. M. Pardalos, N. V. Thoai (1995), *Introduction to Global Optimization,* Kluwer, Dordrecht.

[15] C.-F. Huang, R. H. Litzenberger (1988), *Foundations for financial economics.* 6th print., North-Holland, Amsterdam.

[16] Y. Lyubich, G. D. Maistrowskii, Yu. G. Ol'khovskii (1980), Selection-induced convergence to equilibrium in a single-locus autosomal population. *Problems of Information Transmission* **16**, 66–75.

[17] H. M. Markowitz (1952), Portfolio selection. *J. of Finance* **7**, 77–91.

[18] H. M. Markowitz (1995), The general mean-variance portfolio selection problem. In: S. D. Howison, F. P. Kelly, and P. Wilmott, eds., *Mathematical models in finance*, pp. 93–99. Chapman & Hall, London.

[19] W. K. Newey and K. D. West (1987), A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55**, 703–708.

[20] B. M. Pötscher and I. R. Prucha (1997), *Dynamic nonlinear econometric models – asymptotic theory.* Springer, Berlin.

[21] C. Schnitzer (2000), *Mathematical Models in Population Genetics: Selection Dynamics for Three Alleles* (in german). Master Thesis, Univ. of Vienna.