Journal of Global Optimization **18:** 301–320, 2000.

On Copositive Programming and Standard Quadratic Optimization Problems ⋆

IMMANUEL M. BOMZE[1], MIRJAM DÜR[2], ETIENNE DE KLERK[3], CORNELIS ROOS[3], ARIE J. QUIST[3] and TAMÁS TERLAKY[4]
[1]*ISDS, Universität Wien, Austria (e-mail: immanuel.bomze@univie.ac.at);* [2]*Department of Statistics, Vienna University of Economics, Austria;* [3]*Faculty ITS/TWI/SSOR, Delft University of Technology, The Netherlands;* [4]*Department of Computing and Software, McMaster University Hamilton, Ontario, Canada*

Abstract. A standard quadratic problem consists of finding global maximizers of a quadratic form over the standard simplex. In this paper, the usual semidefinite programming relaxation is strengthened by replacing the cone of positive semidefinite matrices by the cone of completely positive matrices (the positive semidefinite matrices which allow a factorization FF^T where F is some non-negative matrix). The dual of this cone is the cone of copositive matrices (i.e., those matrices which yield a non-negative quadratic form on the positive orthant). This conic formulation allows us to employ primal-dual affine-scaling directions. Furthermore, these approaches are combined with an evolutionary dynamics algorithm which generates primal-feasible paths along which the objective is monotonically improved until a local solution is reached. In particular, the primal-dual affine scaling directions are used to escape from local maxima encountered during the evolutionary dynamics phase.

Key words: Copositive programming, Global maximization, Positive semidefinite matrices, Standard quadratic optimization

1. Introduction

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

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

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

$$\Delta = \{x \in \mathbb{R}^n_+ : e^\top x = 1\}\,,$$

⋆ Paper presented at the GO•99 conference, Firenze, Sept. 28–Oct. 2, 1999. I.M.B. and M.D. thank M. Locatelli and F. Schoen for their hospitality and perfect organization. Large part of the work was done while T.T. was working at TU Delft. All authors are indebted to two anonymous referees for their valuable remarks.

where $e = [1, \ldots, 1]^\top$ and $\mathbb{R}^n_+$ denotes the non-negative orthant in $\mathbb{R}^n$ (of course, the region $\{y \in \mathbb{R}^n_+ : e^\top y \leqslant 1\}$ can always be represented by $\Delta \subseteq \mathbb{R}^{n+1}$, introducing a slack variable). To avoid trivial cases, we assume throughout the paper that the objective is not constant over Δ, which means that $\{A, E\}$ are linearly independent where $E = ee^\top$ is the $n \times n$ matrix consisting entirely of unit entries, so that $x^\top Ex = (e^\top x)^2 = 1$ on Δ. For a review on standard QPs and its applications, which also offers a justification for terminology see [9].

Note that the maximizers of (1) remain the same if A is replaced with $A + \gamma E$ where γ is an arbitrary constant. So without loss of generality assume henceforth that all entries of A are non-negative. Furthermore, the question of finding maximizers of a general quadratic function $x^\top Qx + 2c^\top x$ over Δ can be homogenized in a similar way by considering the rank-two update $A = Q + ec^\top + ce^\top$ in (1) which has the same objective values on Δ.

Of course, quadratic optimization problems like (1) are NP-hard [24]. Nevertheless, there are several exact procedures which try to exploit favourable data structures in a systematic way, and to avoid the worst-case behaviour whenever possible. One example for this type of algorithms is specified in this paper: the proposed procedure exploits extensively the special structure of a standard QP (e.g., that the feasible set is the standard simplex), as opposed to the general formulation of a quadratic problem.

This article deals with the application of an interior-point method to an extension of semidefinite programming called copositive programming, and is organized as follows: Section 2 contains a concise exposition of primal and dual problems in copositive programming which involves copositive rather than positive-semidefinite matrices, using an explicit characterization of the dual cone of the convex, non-polyhedral cone of all copositive matrices. We also shortly treat (the relaxation of copositive programming to) all-quadratic problems on the simplex as considered in [43]. In Section 3, this will be then specialized to be applied to standard QPs, which enjoy the property that the copositive programming relaxation becomes an exact reformulation of (1). Here the dual is in fact a univariate copositive-feasibility problem which can be seen as a straightforward generalization of an eigenvalue bound problem. Section 4 contains a short review on the replicator dynamics, which by now has become an increasingly popular local optimization procedure for standard QPs. This technique is combined with primal-dual search directions from general conic programming [30, 47], which are used to escape from inefficient local solutions returned by the replicator dynamics iteration.

2. Copositive programming problems: general setup

Consider the following primal-dual pair of linear programming problems over a pointed convex cone $\mathcal{K} \subset \mathbb{R}^d$, see, e.g. [18, 30, 39, 47]:

$$f(x) = c^\top x \to \max! \quad \text{subject to} \quad Dx = b\,,\ x \in \mathcal{K}\,, \tag{2}$$

where D is an $m \times d$ matrix with full row rank and $b \in \mathbb{R}^m$ while $c \in \mathbb{R}^d$, and

$$g(y) = b^\top y \to \min! \quad \text{subject to} \quad D^\top y - s = c\,,\ y \in \mathbb{R}^m\,,\ s \in \mathcal{K}^*\,, \tag{3}$$

where $\mathcal{K}^* = \{s \in \mathbb{R}^d : s^\top x \geqslant 0 \text{ for all } x \in \mathcal{K}\}$ is the (convex) dual cone of $\mathcal{K}$. In semidefinite programming, $d = \binom{n+1}{2}$ and $\mathcal{K}$ coincides with the cone $\mathcal{P}$ of all symmetric positive-semidefinite $n \times n$ matrices, which is self-dual $\mathcal{P} = \mathcal{P}^*$ under the usual inner product $\langle S, X\rangle = \text{trace}\,(SX)$ on the d-dimensional Euclidean space $\mathcal{S}^n$ constructed by identifying the upper triangular half of a symmetric $n \times n$ matrix with its vectorized version.

However, we need not restrict ourselves to cases of self-dual cones $\mathcal{K}$ if we can handle the dual cone $\mathcal{K}^*$, even if the geometry of $\mathcal{K}$ and $\mathcal{K}^*$ becomes more complicated. In fact, it turns out useful to study more general cases, e.g. putting $\mathcal{K}^*$ equal to the cone of copositive matrices.

Recall that a symmetric $n \times n$ matrix M is said to be *copositive* (more precisely, $\mathbb{R}^n_+$-copositive), if

$$v^\top M v \geqslant 0 \quad \text{whenever } v \geqslant o\,, \tag{4}$$

i.e., if the quadratic form generated by M takes only non-negative values on the positive orthant $\mathbb{R}^n_+$ (for lucid notation, we denote the zero vector by o while O designates a matrix of zeroes, to distinguish these entities from the number 0). The matrix M is said to be *strictly ($\mathbb{R}^n_+$-)copositive*, if the inequality in (4) is strict whenever $v \neq o$. Clearly, this cone $\mathcal{K}^*$ has non-empty interior and so does its (pre-)dual cone $\mathcal{K}$ (see Proposition 1 below) which can be described as follows, see, e.g. [20, 45]:

$$\mathcal{K} = \text{conv}\,\{xx^\top : x \in \mathbb{R}^n_+\}\,, \tag{5}$$

the convex hull of all symmetric rank-one matrices, i.e. dyadic products, generated by *non-negative* vectors. Elements of $\mathcal{K}$ are called *completely positive matrices*. Note that dropping non-negativity requirement, we again arrive at the semidefinite case. Even without constraints, checking whether or not a matrix belongs to $\mathcal{K}^*$ is co-NP-hard [38]. Some algorithms for this problem can be found, e.g. in [35, 19, 48, 5, 49, 50, 14, 6], to mention just a few. Less obvious is the primal feasibility problem (also without constraints). In fact, the authors are not aware of any finite and exact procedure to determine whether or not a given symmetric $n \times n$ matrix is completely positive if $n > 4$. See also [3, 4, 15, 34, 27, 51, 21, 31]. However, the following result may be helpful:

PROPOSITION 1. *Let $\mathcal{K}$ be as in (5), $d = \binom{n+1}{2}$, and denote by*

$$\mathcal{K}_+ = \{X \in \mathcal{P} : \sqrt{X} \text{ has no negative entries}\}\,. \tag{6}$$

Then

$$\mathcal{K} = \{FF^\top : F \text{ is a non-negative } n \times (d+1) \text{ matrix}\} = \operatorname{conv} \mathcal{K}_+\,. \tag{7}$$

Proof. In view of Caratheodory's theorem, the first identity (cf. Theorem 1 of [34]) is obvious by taking

$$F = [x_1, \dots, x_{d+1}] \operatorname{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_{d+1}})$$

if $X = \sum_{i=1}^{d+1} \lambda_i x_i x_i^\top$ for some $\lambda_i \geqslant 0$, all $i \in \{1, \dots, d+1\}$, and by noting that the middle set in (7) necessarily is a subset of $\mathcal{K}^{**} = \mathcal{K}$ due to the fact that $\langle M, FF^\top\rangle = \sum_{i=1}^{k} f_i^\top M f_i \geqslant 0$ if $M \in \mathcal{K}^*$ and $f_i \in \mathbb{R}^n_+$ are the columns of F. But then the inclusion $\mathcal{K}_+ \subseteq \mathcal{K}$ is also immediate. To finalize the proof, observe that $\sqrt{xx^\top} = \frac{1}{\sqrt{x^\top x}} xx^\top \in \mathcal{K}_+$ if $x \in \mathbb{R}^n_+$ implies that $\mathcal{K} \subseteq \operatorname{co} \mathcal{K}_+$. □

Unfortunately, the cone $\mathcal{K}_+$ itself is not convex and therefore strictly smaller than $\mathcal{K}$, as the following example shows:

EXAMPLE 2. *The nonsingular* 3×3 *matrix*

$$X = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 3 \end{bmatrix} = aa^\top + bb^\top + cc^\top \quad \text{with } a = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix},\; b = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix},\; c = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

belongs to $\mathcal{K}$ (cf. [3]). However, its square root is approximately

$$\sqrt{X} \approx \begin{bmatrix} 0.908 & -0.092 & 0.408 \\ -0.092 & 0.908 & 0.408 \\ 0.408 & 0.408 & 1.633 \end{bmatrix}$$

whence $X \notin \mathcal{K}_+$ although the rank-one matrices $aa^\top$, $bb^\top$ and $cc^\top$ as seen above belong to $\mathcal{K}_+$. A singular variant is obtained by $aa^\top + bb^\top$, i.e. replacing the lower right corner entry of X with 2.

As a general application of the primal-dual approach given by (2) and (3) consider the so-called all-quadratic problem on Δ which appears as a subproblem in [43]:

$$x^\top A_0 x \to \max! \quad \text{subject to } x \in \Delta\,,\; x^\top A_i x = b_i\,,\; 1 \leqslant i \leqslant m\,. \tag{8}$$

Note that also inhomogeneous quadratic constraints can be written in this form (see above), so that by introducing slacks we also can write problems with quadratic inequality constraints in the form (8). Further, additional linear constraints of the form $d^\top x = \delta$ can be written as $\langle D, X\rangle = \delta^2$ with $D = dd^\top$, given $d^\top x$ does not change sign over the feasible set (otherwise one has to subdivide this set accordingly). So the normalization condition in Δ can be written as $\langle E, X\rangle = (e^\top x)^2 = 1$. Hence with $\mathcal{K}$ as in (5), we may view the following copositive programming problem as a relaxation of (8):

$$\langle A_0, X\rangle \to \max!\quad \text{subject to}\quad \langle E, X\rangle = 1\,,\ \langle A_i, X\rangle = b_i\,,\ 1 \leqslant i \leqslant m\,,\ X \in \mathcal{K}\,. \tag{9}$$

Indeed, linearity (in fact, convexity) of the objective ensures that one solution X to the problem (9) is attained at an extreme point of the feasible set. If X happens to lie also on an extreme ray of $\mathcal{K}$, then automatically $X = xx^\top$, so that this condition can be dropped without loss of generality. In this case, the relaxation (9) becomes an exact reformulation of (8). Unfortunately, this is not always the case, as the following example shows:

EXAMPLE 3. *Consider the problem (9) to maximize* $\langle A_0, X\rangle = 2x_{11}+x_{22}$ *subject to* $\langle A_1, X\rangle = x_{11} = \frac{1}{2} = b_1$ *and, of course,* $X \in \mathcal{K}$ *as well as* $\langle E, X\rangle = 1$*. Obviously, the only solution to this problem is given by the rank-two matrix* X^* *with* $x^*_{11} = x^*_{22} = \frac{1}{2}$ *while* $x^*_{ij} = 0$ *else.*

We proceed as in the general case with the primal-dual pair (2),(3) to establish the dual problem of (9) which has $m + 1$ structural variables y_0 and $y = [y_1, \dots, y_m]^\top$, and also d slacks contained in S:

$$y_0 + b^\top y \to \min!\ \text{ subject to } y_0 E + \sum_{i=1}^{m} y_i A_i - S = A_0$$
$$\text{with } y_0 \in \mathbb{R}\,,\ y \in \mathbb{R}^m\,,\ S \in \mathcal{K}^*. \tag{10}$$

Given that we can solve the primal and dual feasibility problems with limited effort, it is possible to use the search directions for a feasible primal-dual interior point algorithm. Indeed, the following results of Nesterov and Nemirovskii [39] are valid for a general class of convex cones which include the cone $\mathcal{K}$ given by (5):

- There exist so-called self-concordant barrier functions for the cones $\mathcal{K}$ and $\mathcal{K}^*$;
- Interior point methods which converge in a polynomially bounded number of steps can be formulated using the self-concordant barriers.

Unfortunately, no polynomially computable self-concordant barriers are known for $\mathcal{K}$ and $\mathcal{K}^*$, for an elaborate discussion on this topic see [45]. However, Tunçel [47] has recently noticed that even in this case one can still formulate a class of interior point methods known as primal–dual affine scaling algorithms.

For ease of reference we now reproduce a generic roster for a primal-dual interior-point method from [22, 29, 36]. Of course, it is in general not harder to solve (8) to optimality than to resolve the feasibility questions (i.e., to check membership of $\mathcal{K}$ or $\mathcal{K}^*$) below, but there could be special instances where the procedure is still helpful.

Generic Interior-Point Primal-Dual Algorithm

1. Choose an initial point (X^0, y^0, S^0) with $X^0 \in \text{int}\,\mathcal{K}$, $y^0 = [y_0^0, \dots, y_m^0]^\top \in \mathbb{R}^{m+1}$, such that $\langle E, X^0 \rangle = 1$ and $\langle A_i, X^0 \rangle = b_i$ for all $i = 1, \dots, m$; and that $S^0 = y_0^0 E + \sum_{i=1}^m y_i^0 A_i - A^0 \in \text{int}\,\mathcal{K}^*$. Put $(X, y, S) = (X^0, y^0, S^0)$.
2. Until a stopping criterion is satisfied, repeat the following step: choose an improving feasible direction (dX, dy, dS) and step length $\alpha > 0$ such that still $X + \alpha dX \in \text{int}\,\mathcal{K}$ as well as $S + \alpha dS \in \text{int}\,\mathcal{K}^*$. Update $(X, y, S) = (X + \alpha dX, y + \alpha dy, S + \alpha dS)$.

Feasibility w.r.t. the equality constraints is maintained by the so-called primal-dual affine scaling (or zero-order) search direction provided by Kojima and Tunçel [30, 47]. Slightly simplified, this class of directions is the solution of the linear system in $\mathcal{S}^n \times \mathbb{R}^{m+1} \times \mathcal{S}^n$

$$\begin{aligned} \langle E, dX \rangle &= 0\,, \\ \langle A_i, dX \rangle &= 0\,, \quad i \in \{1, \dots, m\}\,, \\ E dy_0 + \textstyle\sum_{i=1}^m A_i dy_i - dS &= O\,, \\ dX + \mathcal{Q}_{\mathcal{H}} dS &= -X \end{aligned} \tag{11}$$

where $\mathcal{H}$ is an arbitrary positive-definite, symmetric linear operator on $\mathcal{S}^n$ and

$$\mathcal{Q}_{\mathcal{H}} Y = \mathcal{H} Y + \frac{\langle X, Y \rangle}{\langle X, S \rangle} X - \frac{\langle S, \mathcal{H} Y \rangle}{\langle S, \mathcal{H} S \rangle} \mathcal{H} S\,, \quad Y \in \mathcal{S}^n\,. \tag{12}$$

As usual, the remaining (strict) feasibility requirements are guaranteed by a suitable choice of the step lengths. Note that a solution to (11) always exists as also $\mathcal{Q}_{\mathcal{H}}$ is positive-definite provided that the duality gap $\langle X, S \rangle > 0$ which is guaranteed for interior point pairs $(X, S) \in \text{int}\,\mathcal{K} \times \text{int}\,\mathcal{K}^*$ (cf. Theorem 3.3 in [47]), since we assume that $\{E, A_1, \dots, A_m\}$ are linearly independent, in correspondence with the full row rank assumption on D in (2). Thus we have the same situation as in the classical SDP case for the search direction commonly used there, cf., e.g. [16].

Kojima and Tunçel prove in [30] (cf. Theorem 3.4 in [47]) that if we choose the search directions from (12), then the duality gap decreases linearly with a factor essentially being the step length, and both primal and dual objectives will be improved, unless optimality is reached. Decisive for their arguments is the positive definiteness of $\mathcal{H}$ and the property that $\mathcal{Q}_{\mathcal{H}} S = X$.

Looking at formula (12), it is evident that much would be gained if the terms containing $\mathcal{H} S$ vanished, which of course is impossible if $\mathcal{H}$ is positive-definite.

We therefore propose a positive-semidefinite variant of the above-mentioned result where $\mathcal{H}$ has a single zero eigenvalue belonging to the direction S. Note that we no longer assume that $(X, S) \in \operatorname{int} \mathcal{K} \times \operatorname{int} \mathcal{K}^*$, but only $\langle X, S\rangle > 0$. Recall that the latter relation characterizes non-optimality of pairs (X, S).

THEOREM 4. *Suppose that $\mathcal{H}$ is a positive-semidefinite, symmetric linear operator on $\mathcal{S}^n$ with*

$$\{Y \in \mathcal{S}^n : \mathcal{H}Y = O\} = \{\lambda S : \lambda \in \mathbb{R}\}.$$

Consider a pair $(X, S) \in \mathcal{K} \times \mathcal{K}^*$ with $\langle X, S\rangle > 0$ and define the symmetric linear operator $\mathcal{R}_{\mathcal{H}}$ on $\mathcal{S}^n$ by

$$\mathcal{R}_{\mathcal{H}}Y = \mathcal{H}Y + \frac{\langle X, Y\rangle}{\langle X, S\rangle}X, \quad Y \in \mathcal{S}^n. \tag{13}$$

Then $\mathcal{R}_{\mathcal{H}}$ is positive-definite and satisfies $\mathcal{R}_{\mathcal{H}}S = X$.
Furthermore, the solution (dX, dy, dS) to the system in $\mathcal{S}^n \times \mathbb{R}^{m+1} \times \mathcal{S}^n$

$$\begin{aligned} \langle E, dX\rangle &= 0, \\ \langle A_i, dX\rangle &= 0, \quad i \in \{1, \dots, m\}, \\ E dy_0 + \textstyle\sum_{i=1}^m A_i dy_i - dS &= O, \\ dX + \mathcal{R}_{\mathcal{H}} dS &= -X \end{aligned} \tag{14}$$

is unique and satisfies $\langle X + \alpha dX, S + \alpha dS\rangle = (1 - \alpha)\langle X, S\rangle$.

Proof. The first argument is quite similar to that in Theorem 3.3 of [47]. For any $Z \in \mathcal{S}^n$, consider

$$\langle Z, \mathcal{R}_{\mathcal{H}}Z\rangle = \langle Z, \mathcal{H}Z\rangle + \frac{\langle X, Z\rangle^2}{\langle X, S\rangle}$$

which is non-negative and can vanish only if both $\langle X, Z\rangle = 0$ and $\mathcal{H}Z = O$. But unless $Z = O$ this is absurd as the latter relation implies $Z = \beta S$ for some $\beta \in \mathbb{R}$ by assumption whereas $\langle X, Z\rangle = \beta\langle X, S\rangle$, and, again by assumption, $\langle X, S\rangle > 0$. Hence the operator is positive-definite. Finally, $\mathcal{R}_{\mathcal{H}}S = O + 1 \cdot X = X$. Turning to system (14), we show that the related homogeneous system in (dX, dy, dS) has only the trivial solution. Indeed, substituting for dS in the equation $dX + \mathcal{R}_{\mathcal{H}}dS = O$ yields

$$dX = -dy_0 \mathcal{R}_{\mathcal{H}}E - \sum_{i=1}^m dy_i \mathcal{R}_{\mathcal{H}}A_i,$$

and substituting then for dX in the first $m+1$ equations of (14) gives, after changing the signs, a homogeneous system in dy with coefficient matrix

$$\begin{bmatrix} \langle E, \mathcal{R}_{\mathcal{H}}E\rangle & \langle E, \mathcal{R}_{\mathcal{H}}A_1\rangle & \cdots & \langle E, \mathcal{R}_{\mathcal{H}}A_m\rangle \\ \langle A_1, \mathcal{R}_{\mathcal{H}}E\rangle & \langle A_1, \mathcal{R}_{\mathcal{H}}A_1\rangle & \cdots & \langle A_1, \mathcal{R}_{\mathcal{H}}A_m\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle A_m, \mathcal{R}_{\mathcal{H}}E\rangle & \langle A_m, \mathcal{R}_{\mathcal{H}}A_1\rangle & \cdots & \langle A_m, \mathcal{R}_{\mathcal{H}}A_m\rangle \end{bmatrix}$$

which is, due to linear independence of $\{E, A_1, \dots, A_m\}$, easily seen to be positive-definite as $\mathcal{R}_{\mathcal{H}}$ is so. Thus $dy = o$, yielding $dS = dX = O$. Hence (14) has always a unique solution. To establish reduction of the duality gap, let us first deal with the second-order term $\langle dX, dS\rangle$ which vanishes because of the feasibility conditions imposed on (dX, dy, dS) in (14): indeed,

$$\langle dX, dS\rangle = \langle dX, E dy_0 + \sum_{i=1}^{m} A_i dy_i\rangle = \sum_{i=0}^{m} 0.dy_i = 0\,.$$

Now the first-order terms $\langle X, dS\rangle + \langle dX, S\rangle = -\langle X, S\rangle$ because of

$$\begin{aligned}\langle dX, S\rangle &= -\langle X, S\rangle - \langle \mathcal{R}_{\mathcal{H}} dS, S\rangle \\ &= -\langle X, S\rangle - \langle dS, \mathcal{R}_{\mathcal{H}} S\rangle = -\langle X, S\rangle - \langle dS, X\rangle\,.\end{aligned}$$

This establishes $\langle X + \alpha dX, S + \alpha dS\rangle = (1-\alpha)\langle X, S\rangle$. □

Using similar arguments as in [30], one can also show that both primal and dual objectives are improved by the directions given by (14). We will establish this result directly in the next section for the special case we focus upon in this paper. Further observe that because X or S might be singular a zero step might occur. Extra care is needed to guarantee the existence of a positive step.

REMARKS

In semidefinite programming (SDP) where $\mathcal{K} = \mathcal{K}^* = \mathcal{P}$, there are many possible choices of the operator $\mathcal{R}_{\mathcal{H}}$; only one choice is known to allow convergent algorithms to an optimal solution, namely the Nesterov-Todd primal–dual affine-scaling direction, where $\mathcal{R}_{\mathcal{H}} Y = TYT$ with

$$T = \sqrt{X}\left(\sqrt{\sqrt{X} S \sqrt{X}}\right)^{-1}\sqrt{X}.$$

Note that $\mathcal{R}_{\mathcal{H}} S = X$ and that $\mathcal{R}_{\mathcal{H}}$ is a positive definite linear operator. Also note that this choice of $\mathcal{R}_{\mathcal{H}}$ is not possible for copositive programming, since $\sqrt{X} S \sqrt{X}$ is not positive-definite for all copositive matrices S. For SDP, the primal-dual algorithm using this search direction is globally convergent and polynomial for a suitable choice of the step length [28].

Another choice of primal-dual scaling direction is the so-called (primal) HKM affine-scaling direction, where $\mathcal{R}_{\mathcal{H}} Y = \frac{1}{2}\left(XYS^{-1} + S^{-1}YX\right)$. As mentioned, this search direction is not globally convergent to an optimal solution for any choice of step length. In particular, it can converge to a non-optimal point [37]. Moreover, it cannot be used for copositive programming because a copositive matrix S can be singular despite $\langle X, S\rangle > 0$ for all $X \in \mathcal{K} \setminus \{O\}$, e.g. $S = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$, which is linearly independent from E.

Finally, a primal-dual affine scaling direction for SDP which is also defined for copositive programming is the so-called dual HKM direction, which is given as the solution of

$$\begin{aligned} \langle E, dX \rangle &= 0\,, \\ \langle A_i, dX \rangle &= 0\,, \quad i \in \{1, \ldots, m\}\,, \\ E dy_0 + \textstyle\sum_{i=1}^m A_i dy_i - dS &= O\,, \\ dS + \tfrac{1}{2}\left(X^{-1}(dX)S + S(dX)X^{-1}\right) &= -S. \end{aligned} \tag{15}$$

As with the (primal) HKM direction, no primal-dual SDP algorithm using this search direction is globally convergent [37].

The preceding observations prove two things:

– Using only primal-dual affine scaling directions in interior point methods for conic programming does not necessarily lead to a globally convergent algorithm;
– One cannot guarantee a fixed feasible step length for all primal-dual affine scaling directions (even in the SDP case); in other words, ‘jamming’ can occur.

Therefore we will discuss a hybrid algorithm in Section 4 which uses primal-dual affine-scaling steps only as an escape strategy.

The reason why we use affine-scaling directions, as opposed to path following directions, is that we have no characterization of the central path for our problem. We could imitate the dual path-following HKM direction of SDP by using the above system (15) and adding a term μX^{-1} on the right hand side, where μ now imitates the centrality parameter. Note that this will not work for the other directions, since the inverse of S does not necessarily exist, and the terms involving μ there also involve S^{-1}.

3. Standard quadratic optimization and copositive programming

First note that the standard QP (1) is a special case of the all-quadratic problem (8) with no quadratic constraints and $A_0 = A$. Hence in this case we arrive at the copositive programming problem (9) with a single constraint:

$$\langle A, X \rangle \to \max! \quad \text{subject to} \quad \langle E, X \rangle = 1\,,\ X \in \mathcal{K}\,, \tag{16}$$

so that the dual has only one structural variable $y = y_0$:*

$$y \to \min! \quad \text{subject to} \quad yE - S = A \text{ with } y \in \mathbb{R}\,,\ S \in \mathcal{K}^*\,. \tag{17}$$

This amounts to search for the smallest y such that $yE - A$ is copositive. In this sense, the dual problem (17) is related to the question of eigenvalue bounds [26] (replace E with the identity matrix I and ‘copositive’ with ‘semidefinite’).

* From this point on y will denote a scalar variable.

Further observe that in this case, (16) is no relaxation but indeed an exact reformulation of the standard QP (1): indeed, the objective function in (16) is linear so that a solution of this problem is attained at an extremal point of the feasible set. Now the next result shows that these extremal points are exactly $X = xx^\top$, the rank-one matrices based on vectors $x \in \Delta$, so that from extreme solutions of (16) we can easily construct a solution of the original standard QP (1) with the same objective value $\langle A, X\rangle = x^\top Ax$.

LEMMA 5. *The extremal points of the feasible set of (16) are exactly the rank-one matrices $X = xx^\top$ with $x \in \Delta$.*

Proof. Of course all $X = xx^\top$ with $x \in \Delta$ belong to $\mathcal{M} = \{X \in \mathcal{K} : \langle E, X\rangle = 1\}$. Now suppose that for a vector $x \in \Delta$, we have $xx^\top = (1-\lambda)U + \lambda Z$ for some $Z, U \in \mathcal{M}$ and some λ with $0 < \lambda < 1$. Choose an orthogonal basis $\{x_1, x_2, \dots, x_n\}$ of $\mathbb{R}^n$ with $x = x_n$. Then since Z and U also are positive semidefinite, we get from

$$0 = (x_i^\top x)^2 = (1-\lambda)x_i^\top Ux_i + \lambda x_i^\top Zx_i$$

that $x_i^\top Zx_i = x_i^\top Ux_i = 0$ for all $i < n$ and therefore both Z and U have rank one. As both belong to $\mathcal{K}$, we thus obtain $Z = zz^\top$ and $U = uu^\top$ for some $z, u \in \mathbb{R}^n_+$. But then we obtain $x_i^\top z = x_i^\top u = 0$ for all $i < n$, so that Z and U must be positive multiples of $xx^\top$. The requirement $\langle E, Z\rangle = \langle E, U\rangle = 1$ shows that $Z = U = xx^\top$.

To show the converse, suppose that X is an extremal point of $\mathcal{M} \subset \mathcal{K}$. Then $X = \sum_{i=1}^{d+1} \lambda_i x_i x_i^\top$ with $x_i \in \mathbb{R}^n_+ \setminus \{o\}$ and $\lambda_i \geqslant 0$ for all i as well as $\sum_{i=1}^{d+1} \lambda_i = 1$. Since $X \in \mathcal{M}$, we get

$$1 = \langle E, X\rangle = \sum_{i=1}^{d+1} \lambda_i (e^\top x_i)^2 \tag{18}$$

where $e^\top x_i > 0$ for all i. Now put $u_i = (e^\top x_i)^{-1} x_i \in \Delta$, so that $U_i = u_i u_i^\top \in \mathcal{M}$ for all i. Hence

$$X = \sum_{i=1}^{d+1} \lambda_i (e^\top x_i)^2 U_i$$

is, due to (18), a convex combination of matrices U_i in $\mathcal{M}$, whence by the extremality assumption $X = U_1$ is of the form stated. □

In principle, the roster of the algorithm of Section 2 applies, but the update equations (14) now reduces to

$$\begin{aligned} \langle E, dX\rangle &= 0 \\ E dy - dS &= O \\ dX + \mathcal{R}_{\mathcal{H}} dS &= -X \end{aligned} \tag{19}$$

which for the dual part means simply that we have to continue the line search for the generalized eigenvalue bound of A as in (17). Of course, a similar reduction applies to the Kojima-Tunçel search directions from (11), where $\mathcal{Q}_{\mathcal{H}}$ replaces $\mathcal{R}_{\mathcal{H}}$.

Now let us calculate the update steps explicitly, in order to avoid unnecessary numerical complications. Remember that we have still freedom in choosing the positive-semidefinite operator $\mathcal{H}$ as long as S gives the unique direction to the zero eigenvalue of $\mathcal{H}$ (note that by assumption on linear independence of $\{A, E\}$, the matrix $S = yE - A$ never can vanish regardless whether it belongs to $\mathcal{K}^*$ or not). For instance, we may assume that the orthoprojection of E onto the orthogonal complement of S in $\mathcal{S}^n$ is also an eigenvector of $\mathcal{H}$ with a suitably chosen eigenvalue $\lambda > 0$. As $\mathcal{H}S = O$, this is equivalent to imposing

$$\mathcal{H}E = \lambda(E - \frac{\langle E, S\rangle}{\langle S, S\rangle}S) . \tag{20}$$

THEOREM 6. *Put $S = yE - A$, assume $\langle X, S\rangle > 0$ and denote by (dX, dy, dS) the solution of (19). Then (20) implies*

$$\begin{aligned} dy &= -[\lambda(n^2 - \tfrac{\langle E,S\rangle^2}{\langle S,S\rangle}) + \tfrac{1}{\langle X,S\rangle}]^{-1} , \\ dS &= Edy , \\ dX &= -[1 + \tfrac{dy}{\langle X,S\rangle}]X + \lambda dy(\tfrac{\langle A,S\rangle}{\langle S,S\rangle}E - \tfrac{\langle E,S\rangle}{\langle S,S\rangle}A) . \end{aligned} \tag{21}$$

Proof. From $\langle E, dX\rangle = 0$ we get $0 = \langle X, E\rangle + \langle \mathcal{R}_{\mathcal{H}}dS, E\rangle$. Inserting $dS = Edy$ we further obtain $dy\langle \mathcal{R}_{\mathcal{H}}E, E\rangle = -1$. Now $\mathcal{R}_{\mathcal{H}}E = \mathcal{H}E + \frac{1}{\langle X,S\rangle}X$ and relation (20) yields the result for dy, observing that $\langle E, E\rangle = n^2$.
Similarly, we derive $dX = -X - dy\mathcal{R}_{\mathcal{H}}E = -X - \lambda dyE + \lambda dy\frac{\langle E,S\rangle}{\langle S,S\rangle}S - \frac{dy}{\langle X,S\rangle}X$, and the proof is complete. □

For further formulation, it may be convenient to write $X+\alpha dX = (1-\alpha)X+\alpha Y$ with

$$Y = dy\left[\lambda\left(\frac{\langle A, S\rangle}{\langle S, S\rangle}E - \frac{\langle E, S\rangle}{\langle S, S\rangle}A\right) - \frac{1}{\langle X, S\rangle}X\right] . \tag{22}$$

We now directly show that both objectives are indeed improved by the chosen directions.

THEOREM 7. *Assume that $(X, S) \in \mathcal{K} \times \mathcal{K}^*$ with $\langle X, S\rangle > 0$. If the improving feasible direction (dX, dy, dS) is chosen as in Theorem 6, then for $\alpha > 0$ both primal and dual objective function improve strictly, i.e.*

$$\langle A, X + \alpha dX\rangle > \langle A, X\rangle \qquad \textit{and} \qquad y + \alpha dy < y.$$

Proof. First, dy is strictly negative, since $n^2 - \frac{\langle E,S\rangle^2}{\langle S,S\rangle} = \langle E, E\rangle - \frac{\langle E,S\rangle^2}{\langle S,S\rangle} > 0$ by the Cauchy-Schwarz inequality (note that also $\{E, S\}$ are linearly independent).

To see the strict monotonicity in the primal objective function, compare the reduction of the duality gap with the improvement in the dual objective. From Theorem 4, we know that the reduction of the duality gap is $\alpha\langle X, S\rangle$. Therefore, to show that also the primal objective contributes to this reduction, we have to show that $-\alpha dy < \alpha\langle X, S\rangle$. But

$$-dy = \frac{\langle X, S\rangle}{\langle X, S\rangle\lambda(n^2 - \frac{\langle E,S\rangle^2}{\langle S,S\rangle}) + 1} < \langle X, S\rangle,$$

since the denominator of the above fraction is a positive number bigger than 1. □

For the sake of completeness, we now also provide explicit update formulae for the original Kojima/Tunçel search direction, i.e. for the solutions to the system

$$\begin{aligned} \langle E, dX\rangle &= 0 \\ Edy - dS &= O \\ dX + \mathcal{Q}_{\mathcal{H}} dS &= -X \end{aligned} \tag{23}$$

with $\mathcal{H}$ now again positive-definite but otherwise arbitrary, and $\mathcal{Q}_{\mathcal{H}}$ from (12). Of course, for concrete implementation it remains to specify the values $\mathcal{H}E$ and $\mathcal{H}S = y\mathcal{H}E - \mathcal{H}A$.

THEOREM 8. *Assume that* $(X, S) \in \mathcal{K} \times \mathcal{K}^*$ *with* $\langle X, S\rangle > 0$. *Put* $S = yE - A$ *and denote by* (dX, dy, dS) *the solution of (23). Then*

$$\begin{aligned} dy &= -[\langle \mathcal{H}E, E\rangle - \frac{\langle \mathcal{H}E,S\rangle^2}{\langle \mathcal{H}S,S\rangle} + \frac{1}{\langle X,S\rangle}]^{-1}, \\ dS &= Edy, \\ dX &= -[1 + \frac{dy}{\langle X,S\rangle}]X + dy(\frac{\langle \mathcal{H}E,S\rangle}{\langle \mathcal{H}S,S\rangle}\mathcal{H}S - \mathcal{H}E). \end{aligned} \tag{24}$$

Furthermore, both primal and dual objectives are strictly improved if $\alpha > 0$.

Proof. The arguments are very similar to that of Theorems 6 and 7, and therefore omitted. □

4. A hybrid method: replicator dynamics and primal-dual escape steps

To find local solutions to the standard QP (1), we propose to use replicator dynamics. For the reader's convenience, we here provide a short overview, and refer for more detail to [7, 11, 12]. Consider the following dynamical system operating on Δ:

$$\dot{x}_i(t) = x_i(t)[(Ax(t))_i - x(t)^\top Ax(t)], \quad i \in \{1, \dots, n\}, \tag{25}$$

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

$$x_i(t+1) = x_i(t)\frac{(Ax(t))_i}{x(t)^\top Ax(t)}\,, \quad i \in \{1, \dots, n\}\,. \tag{26}$$

Note that $x(0) \in \mathbb{R}^n_+$ implies $x(t) \in \mathbb{R}^n_+$ for all $t \geqslant 0$ since A is nonnegative by assumption.

The *stationary points* under (25) and (26) coincide, and all local solutions of (1) are among these. Of course, there are quite many stationary points, e.g. all vertices of Δ. However, it can be shown [7] that $\overline{x}$ is a strict local solution if and only if $\overline{x}$ is *asymptotically stable*, i.e. every solution to (25) or (26) which starts close enough to $\overline{x}$, will converge to $\overline{x}$ as $t \nearrow \infty$.

Both (25) and (26) 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. The Fundamental Theorem of Selection states that average fitness, i.e. the objective function $x(t)^\top Ax(t)$ is (strictly) increasing over time along trajectories [13, 23], and moreover every trajectory $x(t)$ converges to a stationary point [23, 33]. Furthermore, one can prove [7, 12] the following facts: if no principal minor of $A = A^\top$ vanishes, then with probability one any trajectory converges to a strict local solution $\overline{x}$ of (1); further, if $\sigma = \{i \in \{1, \dots, n\} : \overline{x}_i > 0\}$, then $y^\top Ay < \overline{x}^\top A\overline{x}$ for all $y \in \Delta_\sigma$ with $y \neq \overline{x}$; and Δ°_σ is contained in the basin of attraction of $\overline{x}$, where for a subset $\sigma \subseteq \{1, \dots, n\}$, we shall denote the face of Δ corresponding to σ by

$$\Delta_\sigma = \{x \in \Delta : x_i = 0 \text{ if } i \notin \sigma\}$$

and its relative interior by

$$\Delta^\circ_\sigma = \{x \in \Delta_\sigma : x_i > 0 \text{ if } i \in \sigma\}\,.$$

The dynamical systems (25) and (26) are frequently called *replicator dynamics*, and are well suited for implementation in practical applications, see [8, 11, 42]. This is reflected also in theory by the result that (25) is most efficiently approaching fixed points in the sense that it is a Shahshahani gradient system [46]. The discrete time version (26) also corresponds to a particular instance of an algorithm widely popular in computer vision. These *relaxation labeling processes* are closely related to artificial neural network learning systems, and have found applications in a variety of practical tasks, e.g. to solve certain labeling problems arising in the 3-D interpretation of ambiguous line drawings [25, 41, 44]. Furthermore, the dynamics (26) belongs to a class of dynamical systems investigated in [1, 2], which has proven to be useful in the speech recognition domain [32].

Although strictly increasing objective values are guaranteed as we follow trajectories under (25) or (26), we could get stuck in an inefficient local solution $\overline{x}$ of (1). From the preceding results, then necessarily $\overline{x}_i = 0$ for some i. One possibility to escape from $\overline{x}$ is by the G.E.N.F. approach [12]. An alternative is to merge the replicator dynamics method with the usual interior-point steps borrowed from semidefinite programming, and this will be described in the sequel. But given any escape procedure, we are now ready to describe the principal algorithm for solving (1) globally. Note that this procedure stops after finitely many repetitions, since it yields strict local solutions (in every relative interior Δ_σ° there is at most one of these) with strictly increasing objective values:

Replicator Dynamics Algorithm

1. Start with $x(0) = \frac{1}{n}e$ or nearby, iterate (26) until convergence; the limit $\overline{x} = \lim_{t\to\infty} x(t)$ is a strict local solution with probability one (provided no principal minor of A vanishes);
2. call an escape procedure to improve the objective, if this is still possible; denote the improving point $\widetilde{x}$;
3. repeat step 1., starting with $x(0) = \widetilde{x}$.

Now we are ready to present a combination of the above procedure and the interior-point method yielding improving direction, in the hope that this way it will be possible to escape from inefficient local solutions.

A Hybrid Algorithm

1. Initialization: choose $x(0) = \frac{1}{n}e$ or nearby. Put $y_0 = \max_{i,j} a_{ij}$. Then $y_0 E - A \in \mathcal{K}^*$.
2. Replicator dynamics for fast primal updates: starting from $x(0)$ iterate (26) until convergence; the limit $\overline{x} = \lim_{t\to\infty} x(t)$ is a strict local solution with probability one; put $\overline{y} = \overline{x}^\top A\overline{x}$.
3. Dual update: check copositivity of $\overline{y}E - A$ via shortcuts (cf. Fig. 1 in [6], for a special case see [10]).
   In the affirmative, $\overline{x}$ is the global solution of (1), since the duality gap is zero (cf. also Theorem 7 in [7]); stop.
   If however a point $\widetilde{x} \in \Delta$ is found such that $\widetilde{x}^\top(\overline{y}E - A)\widetilde{x} < 0$, then $\widetilde{x}$ improves the objective; repeat step 2 starting with this point.
   Else (no decision), keep the old value of y_0, and proceed to step 4.
4. Step back from the boundary: Choose $\rho > 0$ so small, that the point $x = (1-\rho)\overline{x} + \frac{\rho}{n}e \in \Delta$ and the matrix $X = (1-\rho)xx^\top + \frac{\rho}{n}I$ satisfies (with $\widehat{x}$ the previous iterate, so that $\overline{x}^\top A\overline{x} - \widehat{x}^\top A\widehat{x}$ is the previously obtained improvement)

$$\langle A, X\rangle - \widehat{x}^\top A\widehat{x} \geqslant \frac{1}{2}[\overline{x}^\top A\overline{x} - \widehat{x}^\top A\widehat{x}] > 0\,.$$

   This condition is a quadratic inequality for ρ which ensures that not more than half of the previously obtained improvement is lost. Note that by construction,

X is both positive-definite and has a non-negative square root:

$$\sqrt{X} = \sqrt{\frac{\rho}{n}} I + \frac{\sqrt{\frac{\rho}{n} + (1-\rho)x^\top x} - \sqrt{\frac{\rho}{n}}}{x^\top x} x x^\top .$$

5. Since X is both positive-definite and has a non-negative square root, one can choose Y as in (22) and $\alpha > 0$ sufficiently small such that $\widetilde{X} = (1-\alpha)X + \alpha Y$ shares the same properties. This is possible because the mapping $X \mapsto \sqrt{X}$ is Hölder continuous around X. Hence primal feasibility of $\widetilde{X}$ is maintained (cf. Proposition 1), and we get an explicit positive (square root) factorization of $\widetilde{X} = FF^\top$ with $F = [f_1, \dots, f_{d+1}]$ where $f_i \in \mathbb{R}^n \setminus \{o\}$ are all non-negative (a square root factorization of $\tilde{X}$ can be computed by computing the spectral decomposition of $\tilde{X}$ first, and subsequently replacing the eigenvalues by their square roots to get the spectral decomposition of the square root of $\tilde{X}$). Thus $\widetilde{X} = \sum_{i=1}^{d+1} \lambda_i x_i x_i^T$ with $\lambda_i \geqslant 0$ and $x_i = \frac{1}{e^\top f_i} f_i \in \Delta$ for all $i \in \{1, \dots, d+1\}$.
6. Primal update: Now $\overline{x}^\top A \overline{x} \approx \langle A, X \rangle < \langle A, \widetilde{X} \rangle = \sum_{i=1}^{d+1} \lambda_i x_i^\top A x_i$. If possible, choose $\widetilde{x}$ such that

$$\widetilde{x}^\top A \widetilde{x} = \max_{1 \leqslant i \leqslant d+1} x_i^\top A x_i > \overline{x}^\top A \overline{x} .$$

This will always be possible if $\langle A, \widetilde{X} \rangle > \overline{x}^\top A \overline{x}$ but otherwise may fail sometimes.
Repeat from step 2, starting with $x(0) = \widetilde{x}$.

The following small example illustrates the ideas behind the hybrid algorithm. In particular, the example is meant to illustrate how the escape strategy in steps 4 through 6 works.

EXAMPLE 9. *Let* $A = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$ *and suppose we arrived via replicator dynamics starting at* $\widehat{x} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$ *already at the (suboptimal) local solution* $\overline{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. *Then* $\overline{X} = \overline{x}\,\overline{x}^T = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ *and* $\overline{y} = \overline{x}^\top A \overline{x} = 2$ *with improvement*

$$\overline{x}^\top A \overline{x} - \widehat{x}^\top A \widehat{x} = \frac{1}{4} .$$

As $\overline{x}$ *is not the global solution, the matrix* $\overline{S} = \overline{y} E - A = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix}$ *is not copositive. Following step 3 of the hybrid algorithm, we return to the old* $y_0 = \max_{i,j} a_{ij} = 3$ *and arrive at the dually feasible (i.e., copositive) matrix* $S = y_0 E - A = \begin{bmatrix} 1 & 2 \\ 2 & 0 \end{bmatrix}$

which incidentally coincides with the optimal $S^* = y^*E - A$ *(see below). Note that although neither* $\overline{X}$ *nor* S *is interior, we have* $\langle \overline{X}, S\rangle = 1 > 0$.

Suppose for the moment we ignored step 4 above and tried to proceed directly in forming the matrix $\check{Y}$ *(the sign* ˇ *shall emphasize that this trial is preliminary) for the escape step along (21) and (22). The key quantities in (21) and (22) are* $\langle E, S\rangle = 5$; $\langle A, S\rangle = 6$; *and* $\langle S, S\rangle = 9$. *Hence*

$$\frac{\langle A, S\rangle}{\langle S, S\rangle}E - \frac{\langle E, S\rangle}{\langle S, S\rangle}A = \frac{1}{9}\begin{bmatrix} -4 & 1 \\ 1 & -9 \end{bmatrix} \tag{27}$$

and, furthermore, $\check{dy} = -[\lambda(4 - \frac{25}{9}) + \frac{1}{\langle \overline{X}, S\rangle}]^{-1} = -[\frac{11}{9}\lambda + 1]^{-1}$. *Therefore*

$$\check{Y} = -\check{dy}\begin{bmatrix} \frac{4}{9}\lambda + 1 & -\frac{\lambda}{9} \\ -\frac{\lambda}{9} & \lambda \end{bmatrix}$$

which is positive-definite but has negative off-diagonal entries so that $\check{X} = (1 - \alpha)\overline{X} + \alpha\check{Y}$ *is infeasible for all positive* λ. *This shows that the step back from the boundary,* i.e. *step 4 in the hybrid algorithm is really necessary. Note that ignoring primal feasibility in this respect, one could investigate instead whether the vector* $\check{X}e = (1-\alpha)\overline{x} + \alpha\check{Y}e$ *belongs to* Δ *(i.e., has no negative coordinate, as automatically* $e^\top\check{X}e = \langle E, \check{X}\rangle = 1$*), and improves the objective. With regard to the latter aim, it is desirable to take* α *as large as possible (recall that* $\overline{x}$ *is locally optimal and the objective is quadratic so that the improvement will be largest for the largest possible distance – if there is one at all). In our case, this means considering as a candidate for an improving point in* Δ *the vector*

$$\check{Y}e = \begin{bmatrix} \frac{3\lambda+9}{11\lambda+9} \\ \frac{8\lambda}{11\lambda+9} \end{bmatrix} \to \begin{bmatrix} \frac{3}{11} \\ \frac{8}{11} \end{bmatrix} \quad as \quad \lambda \to \infty\,,$$

and indeed we manage to escape because we get an improvement if λ *is chosen large enough, as* $(\check{Y}e)^\top A(\check{Y}e) \to \frac{258}{121} > 2$ *for* $\lambda \to \infty$.

But let us return to the hybrid algorithm as proposed above: choose, e.g., $\rho = 0.2$ *in step 4, so that*

$$x = (1-\rho)\overline{x} + \frac{\rho}{n}e = \begin{bmatrix} 0.9 \\ 0.1 \end{bmatrix} \quad and \quad xx^\top = \begin{bmatrix} 0.81 & 0.09 \\ 0.09 & 0.01 \end{bmatrix}$$

as well as

$$X = (1-\rho)xx^\top + \frac{\rho}{n}I_n = \begin{bmatrix} 0.748 & 0.072 \\ 0.072 & 0.108 \end{bmatrix} \quad with \quad \langle X, S\rangle = 1.036\,,$$

so that the duality gap will be slightly increased, as expected. Note that S *and the update part (27) of* Y *remain the same as the dual variable* y *does not change, and observe that, as required in step 4, the choice of* $\rho = 0.2$ *satisfies*

$$\langle A, X\rangle - \widehat{x}^\top A\widehat{x} = 1.964 - 1.75 > \frac{1}{8} = \frac{1}{2}[\overline{x}^\top A\overline{x} - \widehat{x}^\top A\widehat{x}]\,.$$

Now

$$Y = -dy\left[\frac{1}{\langle X, S\rangle}X + \frac{\lambda}{9}\begin{bmatrix} 4 & -1 \\ -1 & 9 \end{bmatrix}\right].$$

Motivated by the trial with $\check{X}$ above, we choose λ large enough to enable an escape, e.g. $\lambda = 9/\langle X, S\rangle \approx 8$, and arrive at the positive-definite matrix

$$Y = \frac{1}{12}\begin{bmatrix} 4.748 & -0.928 \\ -0.928 & 9.108 \end{bmatrix}.$$

Hence the primal feasibility requirement $\widetilde{X} = (1-\alpha)X + \alpha Y \in \mathcal{K}$ will be met if and only if all entries of the latter matrix are non-negative, which means $\alpha \leqslant \frac{0.864}{1.792} \approx 0.482$. A typical choice of α in step 5 would be $\alpha = \frac{0.482}{2}$ (cf. [47]), but for simplicity we choose here $\alpha = \frac{1}{3}$. Then

$$\widetilde{X} \approx \begin{bmatrix} 0.6306 & 0.0222 \\ 0.0222 & 0.3250 \end{bmatrix} \in \operatorname{int}\mathcal{K} \quad \textit{with} \quad \sqrt{\widetilde{X}} \approx \begin{bmatrix} 0.7939 & 0.0163 \\ 0.0163 & 0.5699 \end{bmatrix}$$

and $\langle \widetilde{X}, A\rangle = 2.2806 > 2 = \langle \overline{X}, A\rangle$. In step 6 of the hybrid algorithm we therefore obtain

$$\widetilde{x} \approx [0.0278,\ 0.9722]^\top,$$

by normalizing the last column of $\sqrt{\widetilde{X}}$, with objective value $\widetilde{x}^\top A\widetilde{x} \approx 2.8912 > 2$. The last steps in the hybrid algorithm are as follows: use the improving point $\widetilde{x}$ as the starting vector for the replicator dynamics iteration, which finally leads to the global solution $x^* = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. For the final check for optimality we now calculate $X^* = x^*(x^*)^\top = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$; $y^* = \widehat{x}^\top A\widehat{x} = 3$; and $S^* = y^*E - A = S$ as specified above, with $\langle X^*, S^*\rangle = 0$.

While exploiting sparsity is problematic in general for primal-dual interior point methods for semidefinite programs (cf. also [16, 17]), this is not the case here since we only consider a special problem, namely (16). In this case the dual problem has only one variable, and the Newton system may be efficiently formed and solved. A similar approach is described in [26] where a primal-dual interior point method is formulated to find the smallest eigenvalue of a symmetric matrix using a semidefinite program. In this case the dual looks just like (17), if one replaces E by the identity matrix and $\mathcal{K}^*$ by $\mathcal{P}$. For such problems, sparsity issues are not problematic. We even know the solution of the Newton system explicitly via (22) and Theorem 8 where we can exploit sparsity of A for the matrix calculations explicitly.

All the steps in the algorithm can be implemented efficiently, except the line search in step 5 to find $\tilde{X}$, where we may have to calculate several square root factorizations. This is the reason why we recommend not to iterate step 5 but rather use this as an escape procedure. The workhorse of the optimization process is more or less the (locally optimizing) replicator dynamics iteration. It remains to be seen if this can be done in a practical way.

While convergence to a local solution is guaranteed with probability one (referring to the choice of a starting point) by virtue of the replicator dynamics under mild conditions [12], it should be plausible from NP-hardness of problem (1) that one cannot hope for a general convergence result of the whole hybrid algorithm. Rather, we suggest to use the affine-scaling steps as an escape procedure which hopefully enables us to find an improving feasible point if the local solution found by replicator dynamics turns out to be inefficient.

5. Conclusions

The problem of maximizing a quadratic form over the simplex has an exact reformulation as a copositive programming problem, i.e. a conic programming problem over the cone of copositive matrices. The advantage of such a reformulation is that successful ideas from the theory of interior point methods can thus be applied to nonconvex quadratic optimization. In particular, primal-dual affine scaling directions can be used in escape strategies if inefficient local solutions are obtained from local optimization procedures like replicator dynamics.

References

1. Baum, L.E. and Eagon, J.A. (1967), An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology. *Bull. Amer. Math. Soc.* 73: 360–363.
2. Baum, L.E. and Sell, G.R. (1968), Growth transformations for functions on manifolds. *Pacif. J. Math.* 27: 211–227.
3. Berman, A. (1988), Complete positivity. *Linear Algebra Appl.* 107: 57–63.
4. Berman, A. and Hershkowitz, D. (1987), Combinatorial results on completely positive matrices. *Linear Algebra Appl.* 95: 111–125.
5. Bomze, I.M. (1987), Remarks on the recursive structure of copositivity. *J.Inf.Optimiz.Sci.* 8: 243–260.
6. Bomze, I.M. (1996), Block pivoting and shortcut strategies for detecting copositivity. *Linear Algebra Appl.* 248: 161–184.
7. Bomze, I.M. (1997), Evolution towards the maximum clique. *J. Global Optimiz.* 10: 143–164.
8. Bomze, I.M. (1997), Global escape strategies for maximizing quadratic forms over a simplex. *J. Global Optimiz.* 11: 325–338.
9. Bomze, I.M. (1998), On standard quadratic optimization problems. *J. Global Optimiz.* 13: 369–387.
10. Bomze, I.M. (2000), Linear-time detection of copositivity for tridiagonal matrices and extension to block-tridiagonality. *SIAM J. Matrix Anal. Appl.*. 21: 840–848.
11. Bomze, I.M., Pelillo, M. and Giacomini, R. (1997), Evolutionary approach to the maximum clique problem: empirical evidence on a larger scale. In: Bomze, I.M., Csendes, T., Horst, R. and Pardalos, P.M. (eds.), *Developments in Global Optimization*, Kluwer, Dordrecht, pp. 95–108.

12. Bomze, I.M. and Stix, V. (1999), Genetical engineering via negative fitness: evolutionary dynamics for global optimization. *Annals of O.R.* 89: 279–318.
13. Crow, J.F., Kimura, M. (1970), *An Introduction to Population Genetics Theory,* Harper & Row, New York.
14. Danninger, G. (1990), A recursive algorithm to detect (strict) copositivity of a symmetric matrix. In: Rieder, U., Gessner, P., Peyerimhoff, A. and Radermacher, F.J., (eds.), *Methods of Operations Research* 62: 45–52. Hain, Meisenheim.
15. Drew, J.H., Johnson, C.R. and Loewy, R. (1994), Completely positive matrices associated with M-matrices. *Linear and Multilinear Algebra* 37: 303–310.
16. Fujisawa, K., Kojima, M. and Nakata, K. (1997), Exploiting sparsity in primal-dual interior-point methods for semidefinite programming. *Math. Prog.* 79: 235–253.
17. Fukuda, M., Kojima, M., Murota, K. and Nakata, K. (1999), Exploiting Sparsity in Semidefinite Programming via Matrix Completion I: General Framework, Research Report B-358, Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, Tokyo 152-8552, Japan.
18. Goemans, M.X. (1997), Semidefinite programming in combinatorial optimization. *Math. Prog.* 79: 143–161.
19. Hadeler, K.P. (1983), On copositive matrices. *Linear Algebra Appl.* 49: 79–89.
20. Hall, M., Jr. and M. Newman, (1963), Copositive and completely positive quadratic forms. *Proc. Cambridge Philos. Soc.* 59: 329–339.
21. Hannah, J. and Laffey, T.J. (1983), Nonnegative factorization of completely positive matrices. *Linear Algebra Appl.* 55: 1–9.
22. Helmberg, C., Rendl, F., Vanderbei, R.J. and Wolkowicz, H. (1996), An interior-point method for semidefinite programming. *SIAM J. Optimiz.* 6: 342–361.
23. Hofbauer, J. and Sigmund, K. (1988), *The Theory of Evolution and Dynamical Systems*. Cambridge University Press, Cambridge, UK.
24. Horst, R., Pardalos, P.M. and Thoai, N.V. (1995), *Introduction to Global Optimization,* Kluwer, Dordrecht.
25. Hummel, R.A. and Zucker, S.W. (1983), On the foundations of relaxation labeling processes. *IEEE Trans. Pattern Anal. Machine Intell.* 5: 267–287.
26. Jansen, B., Roos, C. and Terlaky, T. (1995), Interior point methods: a decade after Karmarkar. A survey, with application to the smallest eigenvalue problem. *Statistica Neerlandica* 50: 146–170.
27. Kaykobad, M. (1987), On nonnegative factorization matrices. *Linear Algebra Appl.* 96: 27–33.
28. Klerk, E. de, Roos, C. and Terlaky, T. (1998), Polynomial primal–dual affine scaling algorithms in semidefinite programming. *J. Combin. Optimiz.* 2: 51–69.
29. Kojima, M., Shindoh, S. and Hara, S. (1997), Interior-point methods for the monotone semidefinite linear complementarity problem in symmetric matrices. *SIAM J. Optimiz.* 7: 86–125.
30. Kojima, M. and Tunçel, L. (1998), Monotonicity of primal-dual interior-point algorithms for semidefinite programming problems. *Optimiz. Methods Softw.* 10: 275–296.
31. Lau, C.M. and Markham, T.L., (1978), Square triangular factorizations of completely positive matrices. *J. Ind. Math. Soc.* 28: 15–24.
32. Levinson, S.E., Rabiner, L.R. and Sondhi, M.M. (1983), An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition. *Bell Syst. Tech. J.* 62: 1035–1074.
33. Lyubich, Y., Maistrowskii, G. D. and Ol'khovskii, Yu.G. (1980), Selection-induced convergence to equilibrium in a single-locus autosomal population. *Problems of Information Transmission* 16: 66–75.
34. Markham, T. L. (1971), Factorizations of completely positive matrices. *Proc. Cambridge Philos. Soc.* 69: 53–58.

35. Martin, D.H. and Jacobson, D.H. (1981), Copositive matrices and definiteness of quadratic forms subject to homogeneous linear inequality constraints, *Linear Algebra Appl.* 35: 227–258.
36. Monteiro, R.D.C. (1997), Primal-dual path-following algorithms for semidefinite programming. *SIAM J. Optimiz.* 7: 663–678.
37. Muramatsu, M. and Vanderbei, R.J. (1997), Primal-dual affine-scaling algorithm fails for semidefinite programming. Technical Report SOR 97-04, Princeton University, NJ 08544.
38. Murty, K.G. and Kabadi, S.N. (1987), Some NP-complete problems in quadratic and linear programming. *Math. Prog.* 39: 117–129.
39. Nesterov, Y.E. and Nemirovskii, A.S. (1994), *Interior point methods in convex programming: theory and applications.* SIAM, Philadelphia, PA.
40. Nowak, I. (1999), A new semidefinite programming bound for indefinite quadratic forms over a simplex. *J. Global Optimiz.* 14: 357–364.
41. Pelillo, M. (1994), On the dynamics of relaxation labeling processes. *Proc. IEEE Int. Conf. Neural Networks*, Orlando, FL, pp. 1006–1011.
42. Pelillo, M. (1995), Relaxation labeling networks for the maximum clique problem. *J. Artif. Neural Networks* 2: 313–327.
43. Raber, U. (1999), *Nonconvex all-quadratic global optimization problems: solution methods, application and related topics.* Ph.D. dissertation, University of Trier.
44. Rosenfeld, A., Hummel, R.A. and Zucker, S.W. (1976), Scene labeling by relaxation operations. *IEEE Trans. Syst. Man Cybern.* 6: 420–433.
45. Quist, A.J., de Klerk, E., Roos, C. and Terlaky, T. (1998), Copositive relaxation for general quadratic programming. *Optimiz. Methods Softw.* 9: 185–209.
46. Sigmund, K. (1987), Game dynamics, mixed strategies, and gradient systems. *Theor. Pop. Biol.* 32: 114–126.
47. Tunçel, L. (1999), Generalization of primal-dual interior-point methods to convex optimization problems in conic form. Technical Report CORR 99-35, University of Waterloo.
48. Väliaho, H. (1986), Criteria for copositive matrices. *Linear Algebra Appl.* 81: 19–34.
49. Väliaho, H. (1988), Testing the definiteness of matrices on polyhedral cones. *Linear Algebra Appl.* 101: 135–165.
50. Väliaho, H. (1989), Quadratic-programming criteria for copositive matrices. *Linear Algebra Appl.* 119: 163–182.
51. Xiang, S.H. and Xiang, S.W. (1998), Notes on completely positive matrices. *Linear Algebra Appl.* 271: 273–282.