Noname manuscript No.
(will be inserted by the editor)

Testing copositivity with the help of difference-of-convex optimization

Mirjam Dür · Jean-Baptiste Hiriart-Urruty

Received: date / Accepted: date

Abstract We consider the problem of minimizing an indefinite quadratic form over the nonnegative orthant, or equivalently, the problem of deciding whether a symmetric matrix is copositive. We formulate the problem as a difference of convex functions (d.c.) problem. Using conjugate duality, we show that there is a one-to-one correspondence between their respective critical points and minima. We then apply a subgradient algorithm to approximate those critical points and obtain an efficient heuristic to verify non-copositivity of a matrix.

Keywords Copositive matrices · difference-of-convex (d.c.) · Legendre-Fenchel transforms · nonconvex duality · subgradient algorithms for d.c. functions

1 Introduction

Consider the problem of minimizing a (homogeneous) quadratic form over the nonnegative orthant:

$$(\mathrm{P}) \qquad \begin{array}{l} \min \tfrac{1}{2} x^T A x \\ \text{s.t. } x \in \mathbb{R}^n_+ . \end{array}$$

As we do not assume the matrix A to be positive semidefinite, (P) is a nonconvex constrained minimization problem. It is equivalent to the problem of determining whether A is copositive: A matrix A is called copositive, if $x^T A x \geq 0$

Dedicated to Claude Lemaréchal on the occasion of his 65th birthday and retirement.

Mirjam Dür
Department of Mathematics, University of Trier, 54286 Trier, Germany. E-mail: duer@uni-trier.de

Jean-Baptiste Hiriart-Urruty
Institut de Mathématiques, Université Paul Sabatier (Toulouse III), 118, route de Narbonne, 31062 Toulouse Cedex 4, France. E-mail: jbhu@mail.cict.fr

holds for all $x \in \mathbb{R}^n_+$. Copositivity has recently attracted quite an amount of interest in mathematical optimization, see [1,6,8] for surveys. The problem of determining copositivity, and consequently also problem (P), are known to be co-NP-complete [12].

In this note, we transfer the copositivity problem into an unconstrained problem in the same number of variables. A similar approach has been pursued by Bomze and Palagi [4], who use an exact penalization method to arrive at an objective function which is a polynomial of degree six. Our approach differs from theirs in that we formulate (P) as an unconstrained difference-of-convex (d.c. in short) optimization problem which allows us to use specific duality schemes from convex analysis and nonconvex optimization. As opposed to the objective function of Bomze and Palagi [4] which is C^∞, we derive a nonsmooth objective which is a combination of a quadratic function and the indicator function of the convex feasible set.

To be more specific, let $r > \max\{\lambda_{\max}(A), 0\}$, where $\lambda_{\max}$ denotes the maximal eigenvalue of A. Then the matrix $rI - A$ is positive definite. We now decompose the objective function as follows: For all $x \in \mathbb{R}^n$, let

$$f(x) = g(x) - h(x),$$

with

$$g(x) = \tfrac{r}{2}\|x\|^2 + \Psi_{\mathbb{R}^n_+}(x), \qquad h(x) = \tfrac{1}{2}x^T(rI - A)x,$$

where $\Psi_{\mathbb{R}^n_+}(x)$ is the indicator function of $\mathbb{R}^n_+$, that is: $\Psi_{\mathbb{R}^n_+}(x) = 0$ if $x \in \mathbb{R}^n_+$ and $\Psi_{\mathbb{R}^n_+}(x) = +\infty$ if $x \notin \mathbb{R}^n_+$. Note that both g and h are convex functions, whence $g - h$ is a description of f as a d.c. function.

On $\mathbb{R}^n_+$, the feasible set of (P), f coincides with the objective function of (P). Outside $\mathbb{R}^n_+$, $f \equiv +\infty$. Hence, (P) translates to the unconstrained d.c. minimization problem

$$\text{(P)} \qquad \min_{x \in \mathbb{R}^n} f(x) = g(x) - h(x).$$

The copositivity property of A is now that $\inf_{x \in \mathbb{R}^n} f(x) = 0$, (actually, the minimal value 0 of f on $\mathbb{R}^n$ is attained since $f(0) = 0$).

In dealing with d.c. minimization problems, there is a duality scheme specially tailored for these problems which has proved useful in various contexts, starting with calculus of mechanics, *cf.* [17,18,19]: the adjoint or dual problem associated with (P) is

$$(\text{P}^\diamond) \qquad \begin{array}{l} \min\ f^\diamond(y) := h^*(y) - g^*(y) \\ \text{s.t. } y \in \mathbb{R}^n, \end{array}$$

where φ^* stands for the Legendre-Fenchel conjugate of φ (see [7, Chapter X]). In our situation, for all $y \in \mathbb{R}^n$ we have

$$h^*(y) = \tfrac{1}{2}y^T(rI - A)^{-1}y, \qquad g^*(y) = \tfrac{1}{2r}\|y^+\|^2,$$

where y^+ means the vector $(y_1^+, \dots, y_n^+)$ with $y_i^+ = \max\{y_i, 0\}$.

The new problem $(\mathrm{P}^\diamond)$ is again a d.c. minimization problem, now unconstrained, with a C^1 (but not C^2) objective function. Note that with the help of the Legendre-Fenchel transformation, the nonnegativity constraint in (P) has been integrated in the expression $\|y^+\|^2$ appearing in $(\mathrm{P}^\diamond)$.

The construction of $(\mathrm{P}^\diamond)$ from (P) depends the decomposition $f = g - h$ more than on f itself. This flexibility is indeed an advantage, for example adding a convex function φ to both g and h would not change the initial problem since $f = (g + \varphi) - (h + \varphi) = g - h$, but would give rise to a different adjoint problem $(\mathrm{P}^\diamond)$. Most of the arguments in this paper hold for d.c. decompositions other than the one described above, but the numerical performance of d.c. optimization algorithms will depend on the choice of the decomposition. Bomze and Locatelli [3] discuss this problem. The authors give evidence that in case of a quadratic objective x^TAx, the so called spectral d.c. decomposition is preferable. This decomposition consists of writing A as $A = A_+ - A_-$, where both A_+ and A_- are (close to) nonsingular and collect the positive (negative) eigenvalues of A, respectively. In our approach, however, we prefer to work with nonsingular decompositions, as we will frequently use the one-to-one relation between problems (P) and $(\mathrm{P}^\diamond)$.

The two problems (P) and $(\mathrm{P}^\diamond)$ are related in an "involutive" manner, *i.e.*, $(\mathrm{P}^{\diamond\diamond}) := (\mathrm{P}^\diamond)^\diamond = (\mathrm{P})$, since $\varphi^{**} := (\varphi^*)^* = \varphi$ holds for any closed convex function φ.

The next results relate infima, global minimizers and critical points in problems (P) and $(\mathrm{P}^\diamond)$. They are adaptations to our specific context of general results in d.c. dualization ([17, 19], see also [16]).

2 Comparing infima and critical points in (P) and $(\mathrm{P}^\diamond)$

In minimizing f over $\mathbb{R}^n_+$ (our problem (P)), there are only two possibilities: either f is not bounded below, *i.e.*, $\inf_{x\in\mathbb{R}^n_+} f(x) = -\infty$, or f is bounded from below. In that case, necessarily $\inf_{x\in\mathbb{R}^n_+} f(x) = 0$ and $\bar{x} = 0$ is among the global minimizers of f on $\mathbb{R}^n_+$. Like for quadratic objective functions, this peculiar property is due to the positive homogeneity of degree 2 of f: if there were $\tilde{x} \in \mathbb{R}^n_+$ with $f(\tilde{x}) < 0$, then $f(\alpha\tilde{x}) = \alpha^2 f(\tilde{x})$ for all $\alpha > 0$ would imply $\inf_{x\in\mathbb{R}^n_+} f(x) = -\infty$.

Theorem 1 *We have*

$$\inf_{x\in\mathbb{R}^n_+} x^TAx = \inf_{x\in\mathbb{R}^n} f(x) = \inf_{y\in\mathbb{R}^n} f^\diamond(y),$$

that is, the infima of (P) *and* $(\mathrm{P}^\diamond)$ *(which are either* 0 *or* $-\infty$*) are equal.*

This relationship has been proved directly in [8, Section 7.8] where it gave rise to the following characterization of copositivity: A matrix A is copositive if and only if

$$y^T(rI - A)^{-1}y \geq \frac{1}{r}\|y^+\|^2 \quad \text{for all } y \in \mathbb{R}^n.$$

We next define what we mean by a critical point of f on $\mathbb{R}^n_+$, or critical point of (P).

Definition 1 A point $\bar{x} \in \mathbb{R}^n_+$ is called a critical point of (P) if

$$-A\bar{x} \in N_{\mathbb{R}^n_+}(\bar{x}), \tag{1}$$

where $N_{\mathbb{R}^n_+}(\bar{x})$ stands for the normal cone to $\mathbb{R}^n_+$ at $\bar{x}$.

As a motivation for this definition, note hat $\nabla h(\bar{x}) = (rI - A)\bar{x}$ and that the generalized subdifferential of g at $\bar{x}$ is $\partial g(\bar{x}) = r\bar{x} + N_{\mathbb{R}^n_+}(\bar{x})$, whence writing the optimality condition $0 \in \partial f(\bar{x}) = \partial g(\bar{x}) - \nabla h(\bar{x})$ directly leads to (1).

The set $\mathbb{R}^n_+$ is a closed convex cone whose polar cone is $\mathbb{R}^n_- := -\mathbb{R}^n_+$. Hence, the normal cone $N_{\mathbb{R}^n_+}(\bar{x})$ can be given explicitly (see [7, Chapter III]):

$$\bar{y} = (\bar{y}_1, \ldots, \bar{y}_m) \in N_{\mathbb{R}^n_+}(\bar{x}) \quad \Leftrightarrow \quad \bar{y} \in \mathbb{R}^n_- \text{ and } \bar{y}_i = 0 \text{ whenever } \bar{x}_i > 0.$$

As a consequence, we have the following.

Proposition 1 *A point $\bar{x} \in \mathbb{R}^n_+$ is a critical point of* (P) *if and only if*

$$A\bar{x} \geq 0 \quad \textit{and} \quad (A\bar{x})_i = 0 \textit{ for all } i \textit{ such that } \bar{x}_i > 0. \tag{2}$$

Observe that $\bar{x} = 0$ is always a critical point of (P), and so is any local minimizer of f on $\mathbb{R}^n_+$. Due to the positive homogeneity of degree 2 of the objective function f in (P) (*i.e.*, $f(\alpha x) = \alpha^2 f(x)$ for all $\alpha > 0$ and $x \in \mathbb{R}^n$), positive homogeneity also holds true for the set of critical points in (P): if $\bar{x}$ is a critical point of (P) and $\alpha > 0$, then $\alpha\bar{x}$ is also a critical point of (P). Therefore, we have "critical rays".

In [15], the authors considered the problem of minimizing the function $\frac{1}{2}x^T Ax$ over $S = \mathbb{R}^n_+ \cap \{x \in \mathbb{R}^n \mid \|x\| = 1\}$. The resulting definition of critical point they adopted ([15, Theorem 1]),

$$A\bar{x} - (\bar{x}^T A\bar{x})\bar{x} \in \mathbb{R}^n_+$$

is less stringent than ours.

In problem ($\mathrm{P}^\diamond$), the objective function $f^\diamond$ is differentiable, so that defining critical points does not pose any difficulty.

Definition 2 A point $\bar{y} \in \mathbb{R}^n$ is a critical point of ($\mathrm{P}^\diamond$) if

$$(rI - A)\bar{y}^+ = r\bar{y}. \tag{3}$$

Indeed, $\nabla f^\diamond(\bar{y}) = \nabla h^*(\bar{y}) - \nabla g^*(\bar{y}) = (rI - A)^{-1}\bar{y} - \frac{1}{r}\bar{y}^+$, so that the relation $\nabla f^\diamond(\bar{x}) = 0$ directly leads to (3).

Observe that $\bar{y} = 0$ as well as any local minimizer of $f^\diamond$ on $\mathbb{R}^n$ are always critical points. Like for (P), the positive homogeneity of degree 2 of the objective function $f^\diamond$ in $(\mathrm{P}^\diamond)$ yields the positive homogeneity of the set of critical points of $(\mathrm{P}^\diamond)$: if $\bar{y}$ is a critical point of $(\mathrm{P}^\diamond)$, then $\alpha\bar{y}$ is a critical point of $(\mathrm{P}^\diamond)$ for any $\alpha > 0$.

There are some specific properties of the critical points in $(\mathrm{P}^\diamond)$ which can be derived from (3). Here are a few examples:

Proposition 2 *The critical points of* $(\mathrm{P}^\diamond)$ *have the following properties:*

1. *The critical points located in $\mathbb{R}^n_+$ are the points in $\mathbb{R}^n_+ \cap \operatorname{Ker} A$. Hence, if A is nonsingular, there is no critical point in $\mathbb{R}^n_+$, except for $\bar{y} = 0$.*
2. *There is no critical point in $\mathbb{R}^n_-$, except for $\bar{y} = 0$.*
3. *All the critical points belong to the cone*

$$K := \mathbb{R}^n_+ - A(\mathbb{R}^n_+). \tag{4}$$

Proof 1. The characterization (3) of critical points $\bar{y}$ located in $\mathbb{R}^n_+$ is

$$(rI - A)\bar{y} = r\bar{y},$$

that is to say $A\bar{y} = 0$.

2. If $\bar{y}$ lies in $\mathbb{R}^n_-$, then $\bar{y}^+ = 0$, so (3) reads $r\bar{y} = 0$ whence the result follows.
3. According to (3), a critical point $\bar{y}$ equals $\bar{y}^+ - \frac{1}{r}A\bar{y}^+$ which belongs to the cone $\mathbb{R}^n_+ - A(\mathbb{R}^n_+)$.

□

With the help of $\bar{y}^- = (y_1^-, \ldots, y_n^-)$ where $y_i^- = \max\{-y_i, 0\}$ so that $y = y^+ - y^-$, we can rewrite characterization (3) as follows: y is a critical point in $(\mathrm{P}^\diamond)$ if and only if

$$Ay^+ = ry^-.$$

While the notion of a critical point in (P) does not depend on r (still chosen greater than $\max\{\lambda_{\max}(A), 0\}$), the critical points in $(\mathrm{P}^\diamond)$ depend a priori on r. However, the closed convex cone K defined in (4) does not depend on r. It gives a rough estimate of the location of the critical points in $(\mathrm{P}^\diamond)$ for all the admissible r. The main point in this section is the correspondence between critical points (as well as critical values) in the original problem (P) and in the adjoint problem $(\mathrm{P}^\diamond)$.

Theorem 2 *1. If $\bar{x}$ is a critical point in* (P)*, then $\bar{y} := (rI - A)\bar{x}$ is a critical point in $(\mathrm{P}^\diamond)$, and the corresponding critical values are equal: $f(\bar{x}) = f^\diamond(\bar{y})$.*

2. If $\bar{y}$ is a critical point in $(\mathrm{P}^\diamond)$, then $\bar{x} := (rI - A)^{-1}\bar{y} = \frac{1}{r}\bar{y}^+$ is a critical point in (P)*, with equality in the critical values: $f^\diamond(\bar{y}) = f(\bar{x})$.*

Proof 1. Let $\bar{x}$ be a critical point in (P). By the very first definition, we have

$$(rI - A)\bar{x} = \nabla h(\bar{x}) \in \partial g(\bar{x}) \tag{5}$$

and $f(\bar{x}) = \frac{1}{2}\bar{x}^T A\bar{x}$. Let $\bar{y} := (rI - A)\bar{x}$. By an elementary property of the Legendre-Fenchel conjugate, we have

$$\begin{aligned} \bar{y} \in \partial g(\bar{x}) \ &\Leftrightarrow\ g^*(\bar{y}) + g(\bar{x}) = \bar{x}^T\bar{y} \quad (\text{or, equivalently, } \le \bar{x}^T\bar{y}), \\ \bar{y} = \nabla h(\bar{x}) \ &\Leftrightarrow\ h^*(\bar{y}) + h(\bar{x}) = \bar{x}^T\bar{y}. \end{aligned} \tag{6}$$

Since $(g^*)^* = g$ and $(h^*)^* = h$, we deduce from (6):

$$\begin{aligned} (g^*)^*(\bar{x}) + g^*(\bar{y}) = \bar{x}^T\bar{y} \\ (h^*)^*(\bar{x}) + h^*(\bar{y}) = \bar{x}^T\bar{y} \end{aligned} \tag{7}$$

as well as

$$g(\bar{x}) - h(\bar{x}) = h^*(\bar{y}) - g^*(\bar{y}). \tag{8}$$

Equations (7) state that $\bar{x} \in \partial g^*(\bar{y}) \cap \partial h^*(\bar{y})$. Since both g^* and h^* are differentiable at $\bar{y}$, this means that $\nabla g^*(\bar{y}) = \nabla h^*(\bar{y})$ or equivalently, $0 = \nabla h^*(\bar{y}) - \nabla g^*(\bar{y}) = \nabla f^\diamond(\bar{y})$. So $\bar{y}$ is indeed a critical point of $(\mathrm{P}^\diamond)$. Equality of the critical values has been observed in (8).

2. We play the same game as above. Let $\bar{y}$ be a critical point of $(\mathrm{P}^\diamond)$. We have:

$$(rI - A)^{-1}\bar{y} = \nabla h^*(\bar{y}) = \nabla g(\bar{y}) = \tfrac{1}{r}\bar{y}^+. \tag{9}$$

Define $\bar{x} := (rI - A)^{-1}\bar{y}$. Similar to (6), we have

$$\begin{aligned} \bar{x} = \nabla h^*(\bar{y}) \ &\Leftrightarrow\ (h^*)^*(\bar{x}) + h^*(\bar{y}) = \bar{x}^T\bar{y}, \\ \bar{x} = \nabla g^*(\bar{y}) \ &\Leftrightarrow\ (g^*)^*(\bar{x}) + g^*(\bar{y}) = \bar{x}^T\bar{y}. \end{aligned} \tag{10}$$

Again, since $(h^*)^* = h$ and $(g^*)^* = g$, we derive from (10):

$$\begin{aligned} h(\bar{x}) + h^*(\bar{y}) = \bar{x}^T\bar{y} \\ g(\bar{x}) + g^*(\bar{y}) = \bar{x}^T\bar{y} \end{aligned} \tag{11}$$

and

$$h^*(\bar{y}) - g^*(\bar{y}) = g(\bar{x}) - h(\bar{x}). \tag{12}$$

What (11) expresses is:

$$\bar{y} = \nabla h(\bar{x}) \quad \text{and} \quad \bar{y} \in \partial g(\bar{x}),$$

that is to say: $\nabla h(\bar{x}) \in \partial g(\bar{x})$, a formulation equivalently saying that $\bar{x}$ is a critical point in (P). The fact that $\bar{x} = (rI - A)^{-1}\bar{y} \in \mathbb{R}^n_+$ was a by-product of (9). Equality of the critical values, *i.e.*, $f^\diamond(\bar{y}) = f(\bar{x})$ was observed in (12).

□

In the scheme below we summarize the correspondence between critical points in (P) and ($\mathrm{P}^\diamond$) via the nonsingular linear mapping $rI - A$:

(P)		($\mathrm{P}^\diamond$)
critical point $\bar{x}$	$\xrightarrow{rI-A}$	critical point $\bar{y} = (rI - A)\bar{x}$
critical point $\bar{x} = (rI - A)^{-1}\bar{y}$	$\xleftarrow{(rI-A)^{-1}}$	critical point $\bar{y}$

3 Comparing global and local minimizers in (P) and ($\mathrm{P}^\diamond$)

We have shown in Section 2 that the infima in (P) and ($\mathrm{P}^\diamond$) are the same. Now, using the same techniques from convex analysis as in Section 2, we show that there is also a one-to-one correspondence between global minimizers in (P) and ($\mathrm{P}^\diamond$).

Theorem 3 *Let $\bar{x}$ be a global minimizer in* (P). *Then*

(i) $\bar{x}^T A\bar{x} = 0$,
(ii) $\bar{y} := (rI - A)\bar{x}$ *is a global minimizer in* ($\mathrm{P}^\diamond$),
(iii) $\bar{y}^T(rI - A)^{-1}\bar{y} = \frac{1}{r}\|\bar{y}^+\|^2$.

Conversely, let $\bar{y}$ be a global minimizer in ($\mathrm{P}^\diamond$). *Then*

(i') $\bar{y}^T(rI - A)^{-1}\bar{y} = \frac{1}{r}\|\bar{y}^+\|^2$
(ii') $\bar{x} := (rI - A)^{-1}\bar{y} = \frac{1}{r}\bar{y}^+$ *is a global minimizer in* (P).

Proof Let $\bar{x}$ be a global minimizer in (P). Then necessarily $\bar{x}^T A\bar{x} = 2f(\bar{x}) = 0$. With $\bar{y} := (rI - A)\bar{x}$, we have $f^\diamond(\bar{y}) = 0$ (*cf.* Theorem 1), and it remains to prove that $f^\diamond(\bar{y}) = h^*(\bar{y}) - g^*(\bar{y}) = 0$. By linearity of the constraints, no qualifications are needed to ensure that $\bar{x}$ is a critical point. Then Theorem 2 gives the desired results (ii) and (iii). The proofs of (i') and (ii') are in the same vein. □

What we have proved for global minimizers also holds true for local minimizers: the proofs have to be slightly adapted. We present below the version "from ($\mathrm{P}^\diamond$) to (P)", which is the case of more interest.

Theorem 4 *If $\bar{y}$ is a local minimizer in* ($\mathrm{P}^\diamond$), *then* $\bar{x} = (rI - A)^{-1}\bar{y} = \frac{1}{r}\bar{y}^+$ *is a local minimizer in* (P).

Proof Let $\bar{y}$ be a local minimizer in ($\mathrm{P}^\diamond$). Then there exists a neighborhood V of $\bar{y}$ such that

$$h^*(y) - g^*(y) = f^\diamond(y) \geq f^\diamond(\bar{y}) = h^*(\bar{y}) - g^*(\bar{y}) \quad \text{for all } y \in V. \tag{13}$$

Since $\nabla h : x \in \mathbb{R}^n \mapsto \nabla h(x) = (rI - A)x$ is a bijective linear mapping, there exists a neighborhood N of $\bar{x} = (rI - A)^{-1}\bar{y}$ such that $\nabla h(N) \subset V$. So choose $x \in N$. Then $y = \nabla h(x)$ and x satisfy

$$h^*(y) + h(x) = x^T y$$

(this is the characterization of $\nabla h(x)$ in terms of h^*), and

$$g^*(y) + g(x) \geq x^T y$$

by the definition of the conjugate fuction. Because $y = \nabla h(x)$ lies in V, these two inequalities combined with (13) imply

$$g(x) - h(x) \geq h^*(y) - g^*(y) \geq f^\diamond(\bar{y}). \tag{14}$$

We know from Theorem 2 that $f^\diamond(\bar{y}) = f(\bar{x})$. So, we have proved with (14) that

$$f(x) \geq f(\bar{x}) \quad \text{for all } x \text{ in the neighborhood } N \text{of } \bar{x},$$

which completes the proof. □

4 DC Algorithm

Any local optimization procedure can be used to compute critical points of (P) and (P$^\diamond$). In our context, it seems suitable to adapt the DCA algorithm of [14,13,9]. This method works as follows: given a starting point $x^0 \in \mathbb{R}^n$, we construct sequences $\{x^k\}$ and $\{y^k\}$ with

$$y^k \in \partial h(x^k) \quad \text{and} \quad x^{k+1} \in \partial g^*(y^k)$$

such that the sequences $(g - h)(x^k)$ and $(h^* - g^*)(y^k)$ are decreasing and such that the limit points of $\{x^k\}$ (resp. $\{y^k\}$) is a critical point of $(g - h)$ (resp. $(h^* - g^*)$).

The condition $y^k \in \partial h(x^k)$ is equivalent to $x^k \in \partial h^*(y^k)$, which in turn means by definition of the subdifferential that

$$h^*(y) - (x^k)^T y \geq h^*(y^k) - (x^k)^T y^k \quad \text{for all } y \in \operatorname{dom} h^*,$$

or equivalently, that y^k can be derived as

$$y^k \in \operatorname{Argmin}\{h^*(y) - (x^k)^T y : y \in \operatorname{dom} h^*\}. \tag{15}$$

The latter is equivalent to saying that y^k is a solution of the problem

$$(\mathrm{D}_k) \qquad \inf_{y \in \operatorname{dom} h^*} \{h^*(y) - g^*(y^{k-1}) - (x^k)^T (y - y^{k-1})\}.$$

which can be interpreted as a linearized version of (P$^\diamond$), where we have linearized g^* at the point y^{k-1}. Observe that in doing so, we have obtained a convex problem (D_k). Since $x^k \in \partial g^*(y^{k-1})$ by construction, we moreover have that $\inf(\mathrm{P}^\diamond) \leq \inf(\mathrm{D}_k)$.

Analogously, the condition $x^{k+1} \in \partial g^*(y^k)$ says that x^{k+1} can be obtained as

$$x^{k+1} \in \operatorname{Argmin}\{g(x) - (y^k)^T x : x \in \mathbb{R}^n\}. \tag{16}$$

which in turn is equivalent to saying that x^{k+1} is a solution to the problem

$$(\mathrm{P}_k) \qquad \inf_{x\in\mathbb{R}^n} \{g(x) - h(x^k) - (y^k)^T(x - x^k)\}.$$

which just as above can be interpreted as a linearized version of (P), where we have linearized h at the point x^k. Since $y^k \in \partial h(x^k)$, we get $\inf(\mathrm{P}) \leq \inf(\mathrm{P}_k)$.

Note that to actually compute the sequences $\{x^k\}$ and $\{y^k\}$, we can use (15) and (16). In our particular setting, (15) reads as:

$$y^k \in \operatorname{Argmin}\{\tfrac{1}{2} y^T (rI - A)^{-1} y - (x^k)^T y : y \in \mathbb{R}^n\}.$$

Since this is a differentiable problem, the solution is obtained by setting the gradient to zero, which gives

$$y^k = (rI - A)x^k.$$

Similarly, in our context (16) translates to

$$x^{k+1} \in \operatorname{Argmin}\{\tfrac{r}{2}\|x\|^2 - (y^k)^T x : x \in \mathbb{R}^n_+\}.$$

Here, the objective function is separable, so we can consider each component function separately. For each of them, this amounts to minimizing a strictly convex quadratic function subject to a sign constraint, and the solution is easily seen to be

$$x_i^{k+1} = \begin{cases} \frac{1}{r} y_i^k & \text{if } y_i^k \geq 0, \\ 0 & \text{else,} \end{cases} \qquad \text{i.e.,} \qquad x^{k+1} = \tfrac{1}{r}(y^k)^+.$$

Combined, this yields the recursions:

$$\begin{aligned} &x^0 \text{ given,} && x^{k+1} = \tfrac{1}{r}\Big[(rI - A)x^k\Big]^+ = \Big[(I - \tfrac{1}{r}A)x^k\Big]^+ \\ &y^0 = (rI - A)x^0, && y^{k+1} = (rI - A)\tfrac{1}{r}(y^k)^+ = (I - \tfrac{1}{r}A)\big[y^k\big]^+. \end{aligned}$$

A detailed convergence theorem for this method is proved in [13]. We present here a reduced version of this theorem which, however, is sufficient for our purposes:

Theorem 5 *Suppose that the sequences $\{x^k\}$ and $\{y^k\}$ are defined as described above. Then we have*

$$(h^* - g^*)(y^{k+1}) \leq (g - h)(x^{k+1}) \leq (h^* - g^*)(y^k) \leq (g - h)(x^k).$$

If $(g - h)(x^{k+1}) = (g - h)(x^k)$, then x^k, x^{k+1} are critical points of $g - h$. Similarly, if $(h^* - g^*)(y^{k+1}) = (h^* - g^*)(y^k)$, then y^k, y^{k+1} are critical points of $h^* - g^*$.

5 A heuristic for testing non-copositivity

The DC algorithm described in the last section inspires the following easy heuristic for testing whether a matrix is not copositive.

Algorithm 1: Heuristic to detect whether a matrix A is not copositive.

Input: symmetric matrix A; parameter $r > \max\{\lambda_{\max}(A), 0\}$
Output: a certificate that A is not copositive
or statement "this instance is undecidable by the heuristic"

while *stopping criterion is not fulfilled* **do**
    generate a starting point x^0;
    $k \leftarrow 0$;
    repeat
        $x^{k+1} = \left[(I - \frac{1}{r}A)x^k\right]^+$;
        if $(x^{k+1})^T A(x^{k+1}) < 0$ **then**
            return *"A is not copositive"*
        end
        $k \leftarrow k + 1$;
    until $x^{k+1} = x^k$ *or stopping criterion is fulfilled*;
end
return *"this instance is undecidable by the heuristic"*

As a stopping criterion, one may choose for example that a prescribed number of iterations has been performed. Note that this algorithm has the ability to do many multistarts, and can be parallelized without any further effort.

5.1 Examples

We first study the behavior of our heuristic for a copositive input matrix. In particular, it is interesting to see the role of the parameter r (recall that by assumption $r > \max\{\lambda_{\max}(A), 0\}$). Consider the so-called Horn-matrix

$$H = \begin{pmatrix} 1 & -1 & 1 & 1 & -1 \\ -1 & 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 & 1 \end{pmatrix}$$

which is known to be copositive (see e.g. [10]). It has $\lambda_{\max}(H) = 3.2361$, so we need to choose r bigger than this value. Applying our heuristic to H will obviously never yield a certificate for non-copositivity.

Running the heuristic for different values of r and 1000 starting points each gives the results displayed in Table 1. In this table, the columns min/max/average

r	iterations average	min	max	% failure
20	219.91	2	362	0.12
10	102.99	2	171	0.28
6	54.50	2	87	1.20
4	28.74	2	55	2.93
3.5	21.77	2	45	0.46
3.2365	18.09	2	42	0.53

Table 1 Results for the Horn-matrix: Running the heuristic for different values of r and 1000 starting points each.

iterations gives the respective number of iterations needed by the heuristic until a critical point was reached. In column "% failure" we list the percentage of instances where the heuristic did not yield a critical point within 1000 iterations.

Observe that convergence seems to be faster for values of r that are closer to $\lambda_{\max}(H)$.

As a second example, consider the behavior of the heuristic when applied to noncopositive matrices. To construct such matrices, recall that the clique number ω of a graph can be written as:

$$\omega = \min\{\alpha \in \mathbb{R} : [\alpha(E - A) - E] \text{ is copositive } \},$$

where E is the all-ones matrix and A is the adjacency matrix of the graph. Now consider the 5-cycle on a graph with five vertices whose adjacency matrix is given by

$$C = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Its clique number is $\omega = 2$. Therefore, matrices of the form

$$B_\alpha := \alpha(E - C) - E$$

are noncopositive for $\alpha < 2$ and copositive for $\alpha \geq 2$. Note that $\alpha = 2$ is the smallest value for which this is true, and that we have $B_2 = 2(E - C) - E = E - 2C = H$, the Horn-matrix from above.

Below is a summary of the behavior of the heuristic. We choose different values of α, and for the value of r we choose $r = c\lambda_{\max}(B_\alpha)$ with $c \in \{1.1, 2, 10\}$. Observe that as $\alpha \to 2$ from below, the matrix B_α approaches the boundary of the copositive cone, since for $\alpha = 2 = \omega$, the matrix B_ω lies on the boundary of that cone.

α	c	noncopositive %	critical point %	av.	min	max	failure %
0.5	1.1	66.3	33.7	2	2	2	0
	2	67.2	32.8	2	2	2	0
	10	66.2	33.8	2	2	2	0
1	1.1	52.7	47.3	2.782	2	3	0
	2	52.0	48.0	2.79	2	3	0
	10	52.5	47.5	2.78	2	3	0
1.5	1.1	9	91	306.45	2	5181	0
	2	8.4	91.6	355.23	2	5181	0
	10	10	90	280.99	3	5179	0
1.7	1.1	6.3	0	–	–	–	93.7
	2	5.3	0	–	–	–	94.7
	10	5.5	0	–	–	–	94.5
1.9	1.1	2.5	1	2	2	2	97.4
	2	2.4	1	2	2	2	97.5
	10	2.9	0	–	–	–	97.1

Table 2 Results for the matrix $B_\alpha := \alpha(E - C) - E$.

In Table 2, we list the results for various values of α and r for 1000 starting points each. The column "noncopositive" gives the percentage of runs that noncopositivity was detected. The next columns list the percentage of runs where the method converged to a critical point. We also list in which iteration convergence occurred. Finally, the right-most column indicates the percentage of runs where the method failed (*i.e.*, neither detected noncopositivity nor converged to a critical point).

It can be seen that as $\alpha \to 2$, the heuristic performs poorer. As discussed above, this is due to the fact that in this case B_α approaches the boundary of the copositive cone. This behavior corresponds to what has been observed in other methods, see e.g. [5].

Observe that in this setting, the choice of the parameter r does not seem to have any impact.

In all cases, if the heuristic detected noncopositivity, it was in the very first iteration, which means the sampled starting point has already been a certificate for noncopositivity.

Acknowledgements

This research was initiated during a research visit of Mirjam Dür to Université Paul Sabatier in Toulouse. She would like to thank the members of the optimization group for their warm hospitality and inspiring discussions. Her work was partially supported by the Netherlands Organisation for Scientific Research (NWO) through Vici grant no.639.033.907. Both authors wish to thank the referees for highly detailed and valuable comments.

References

1. I. M. Bomze, Copositive optimization — recent developments and applications, *European Journal of Operational Research* 216 (2012), 509–520.
2. I.M. Bomze and G. Eichfelder, Copositivity detection by difference-of-convex decomposition and ω-subdivision. *Mathematical Programming*, in print. Preprint available online at: `http://www.optimization-online.org/DB_HTML/2010/01/2523.html`
3. I.M. Bomze and M. Locatelli, Undominated d.c.decompositions of quadratic functions and applications to branch-and-bound approaches *Computational Optimization and Applications* 28 (2004), 227–245.
4. I.M. Bomze and L. Palagi, Quartic Formulation of Standard Quadratic Optimization Problems. *Journal of Global Optimization* 32 (2005), 181–205.
5. S. Bundfuss and M. Dür, Algorithmic Copositivity Detection by Simplicial Partition. *Linear Algebra and its Applications* 428 (2008), 1511–1523.
6. M. Dür, Copositive Programming – a Survey. In: M. Diehl, F. Glineur, E. Jarlebring, W. Michiels (Eds.), *Recent Advances in Optimization and its Applications in Engineering*, Springer 2010, pp. 3–20.
7. J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms I and II.* Grundlehren der Mathematischen Wissenschaften 305, Springer-Verlag Berlin (1993).
8. J.-B. Hiriart-Urruty and A. Seeger, A variational approach to copositive matrices. *SIAM Review* 52 (2010), 593-629.
9. Le Thi Hoai An and Pham Dinh Tao, Solving a class of linearly constrained indefinite quadratic problems by D.C. algorithms. *Journal of Global Optimization* 11 (1997), 253–285.
10. M. Hall Jr. and M. Newman, Copositive and Completely Positive Quadratic Forms, *Proceedings of the Cambridge Philosophical Society* **59** (1963), 329–339.
11. M. Mongeau and M. Torki, Computing eigenelements of real symmetric matrices via optimization. *Computational Optimization and Applications* 29 (2004), 203–226.
12. K.G. Murty and S.N. Kabadi, Some NP-complete problems in quadratic and nonlinear programming. *Mathematical Programming* 39 (1987), 117–129.
13. Pham Dinh Tao and Souad El Bernoussi, Duality in D.C. (difference of convex functions) optimization Subgradient methods. In: K.H. Hoffmann, J.-B. Hiriart-Urruty, C. Lemaréchal, and J. Zowe (Eds.): Trends in mathematical optimization (Irsee, 1986). *International Series of Numerical Mathematics* 84 (1988), 277–293.
14. Pham Dinh Tao and Souad El Bernoussi, Algorithms for solving a class of nonconvex optimization problems. Methods of subgradients. In: J.-B. Hiriart-Urruty (Ed.): *FERMAT Days '85: Mathematics for Optimization*, North-Holland Mathematical Studies 129 (1986), 249–271.
15. A. Seeger and M. Torki, Local minima of quadratic forms on convex cones. *Journal of Global Optimization* 44 (2009), 1–28.
16. I. Singer, A Fenchel-Rockafellar type duality theorem for maximization. *Bulletin of the Australian Mathematical Society* 20 (1979), 193–198.
17. J.F. Toland, A duality principle for non-convex optimization and the calculus of variations. *Archive for Rational Mechanics and Analysis* 71 (1979), 41–61.
18. J.F. Toland, On the stability of rotating heavy chains. *Journal of Differential Equations* 32 (1979), 15–31.
19. J.F. Toland, Duality in non-convex optimization. *Journal of Mathematical Analysis and Applications* 66 (1978), 399–415.
20. Wen-Yu Sun, R. Sampaio, and M. Candido, Proximal point algorithm for minimization of D.C. functions. *Journal of Computational Mathematics* 21 (2003), 451–462.