Exploiting cone approximations in an augmented Lagrangian method for conic optimization

Mituhiro Fukuda* Walter Gómez† Gabriel Haeser‡ Leonardo M. Mito§

June 2, 2024

Abstract

We propose an algorithm for general nonlinear conic programming which does not require the knowledge of the full cone, but rather a simpler, more tractable, approximation of it. We prove that the algorithm satisfies a strong global convergence property in the sense that it generates a strong sequential optimality condition. In particular, a KKT point is necessarily found when a limit point satisfies Robinson's condition. We conduct numerical experiments minimizing nonlinear functions subject to a copositive cone constraint. In order to do this, we consider a well known polyhedral approximation of this cone by means of refining the polyhedral constraints after each augmented Lagrangian iteration. We show that our strategy outperforms the standard approach of considering a close polyhedral approximation of the full copositive cone in every iteration.

Keywords: nonlinear conic programming, sequential optimality condition, augmented Lagrangian method, nonlinear copositive programming

MSC 2020: 90C30, 49M37, 90C26

1 Introduction

We study a very general class of optimization problems, sometimes referred to as *nonlinear conic programming* (NCP), which can be stated in the form:

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{Minimize}} & f(x), \\ \text{subject to} & g(x) \in \mathcal{K}, \end{array} \tag{NCP}$$

where $f\colon \mathbb{R}^n \to \mathbb{R}$ and $g\colon \mathbb{R}^n \to \mathbb{E}$ are continuously differentiable, $\mathbb{E}$ is a finite-dimensional vector space equipped with an inner product $\langle\cdot,\cdot\rangle$, and $\mathcal{K} \subset \mathbb{E}$ is a nonempty closed convex cone. We will denote the feasible set of (NCP) by Ω, which is assumed to be nonempty, and the Lagrangian function of (NCP) is defined as $L(x,\mu) = f(x) + \langle g(x), \mu\rangle$.

Despite its generality and simplicity of representation, studies on practical methods for solving (NCP) and even some of its particular cases are somewhat rare in the current literature, in comparison with the cases when $\mathcal{K}$ is polyhedral, i.e., *nonlinear programming* (NLP); or when g is affine, i.e., *conic programming* (CP). This is no surprise, in view of the fact that evaluating feasibility in (N)CP may be computationally expensive or even NP-Hard, such as in the particular case of *copositive programming* (COP). For such intricate problems, some authors have found sequences of polyhedral approximations of $\mathcal{K}$ and used them to compute bounds for the solutions of the original problems – see, for instance, [5, 7, 8] and references therein for additional information on COP. A particularly interesting algorithmic approach for solving COP problems was presented in Bundfuss and Dür [6] and also in Yıldırım [16], which consists of computing increasingly better polyhedral approximations of $\mathcal{K}$ and solving the approximate problems induced by them to generate a sequence of approximate solutions for the original problem. Both works [6, 16] present great numerical results, using distinct approximations of $\mathcal{K}$.

*Center for Mathematics, Computing and Cognition, Federal University of ABC, Santo André, São Paulo, Brazil. E-mail: mituhiro.f@ufabc.edu.br; Department of Computer Science, University of São Paulo, São Paulo, Brazil; Department of Mathematical and Computing Science, Tokyo Institute of Technology, Meguro-ku, Tokyo, Japan

†Departamento de Ingeniería Matemática, Universidad de la Frontera, Temuco, Chile. E-mail: walter.gomez@ufrontera.cl

‡Departamento de Matemática Aplicada, Universidade de São Paulo, São Paulo, Brazil. E-mail: ghaeser@ime.usp.br

§Departamento de Matemática Aplicada, Universidade de São Paulo, São Paulo, Brazil. E-mail: leonardommito@gmail.com

Recently, Andreani et al. [1] extended the so-called "sequential optimality conditions" from NLP to the NCP world. In summary, sequential optimality conditions are parametric perturbed forms of the traditional Karush-Kuhn-Tucker (KKT) conditions with three distinguishing properties:

1. They are always fulfilled at local minimizers, even for degenerate problems;
2. They are equivalent to the standard KKT condition in the presence of a constraint qualification;
3. They are naturally satisfied by all feasible accumulation points of the output sequence of at least one practical algorithm.

The interested reader may check the earlier texts on sequential conditions in NLP [2, 4] and an extension to *nonlinear semidefinite programming* (NSDP) [3] for more details. All these works use an augmented Lagrangian method to illustrate item 3 of the list above, but one of the most critical disadvantages of it in the conic context is the need of a very expensive projection onto $\mathcal{K}$ that must be computed several times per iteration. In this paper, we will study the effects of replacing $\mathcal{K}$ with increasingly better polyhedral approximations of it, and the novelty in our proposal is the possibility of improving approximation mid-execution. As far as we are concerned, there is no previous work that allows this in the literature. Our first concern is to build a solid convergence theory for our method, which is done by means of a modified variant of a sequential optimality condition called *approximate gradient projection* (AGP), first introduced in [12] for NLP and then extended to the conic framework in [1].

To illustrate the behavior of our method, we use it for solving a nonlinear COP problem via Yıldırım's approximations [16].

2 General framework

We study situations where the cone $\mathcal{K}$ can be approximated by a sequence $\{\mathcal{K}^k\}_{k\in\mathbb{N}}$ that converges to $\mathcal{K}$ in some sense, such that projecting onto each $\mathcal{K}^k$ is relatively easy.

Definition 2.1 (Continuous approximation of $\mathcal{K}$)**.** A sequence $\{\mathcal{K}^k\}_{k\in\mathbb{N}}$ of nonempty closed convex cones is a continuous approximation for $\mathcal{K}$ when:

- For every sequence $\{y^k\}_{k\in\mathbb{N}} \to y$ with $y^k \in \mathcal{K}^k$ for all $k \in \mathbb{N}$, we must have $y \in \mathcal{K}$. In other words, $\limsup_{k\in\mathbb{N}} \mathcal{K}^k \subseteq \mathcal{K}$.
- For every $y \in \mathcal{K}$, there exists a sequence $\{y^k\}_{k\in\mathbb{N}} \to y$ with $y^k \in \mathcal{K}^k$, for all $k \in \mathbb{N}$. In other words, $\mathcal{K} \subseteq \liminf_{k\in\mathbb{N}} \mathcal{K}^k$.

When Definition 2.1 holds, we denote it by $\lim_{k\in\mathbb{N}} \mathcal{K}^k = \mathcal{K}$. Let us begin with a technical lemma regarding continuous cone approximations, which is simply a generalization of some classical results on projections:

Lemma 2.1. *Let $\mathcal{K}$ and $\{\mathcal{K}^k\}_{k\in\mathbb{N}}$ be nonempty closed convex cones in $\mathbb{E}$. Then:*

1. *If $\mathcal{K} \subseteq \liminf_{k\in\mathbb{N}} \mathcal{K}^k$, then $\Pi_{\mathcal{K}^k}(z) \to z$ for every $z \in \mathcal{K}$;*
2. *If $\lim_{k\in\mathbb{N}} \mathcal{K}^k = \mathcal{K}$, then $\Pi_{\mathcal{K}^k}(z^k) \to \Pi_{\mathcal{K}}(z)$ for every converging sequence $\{z^k\}_{k\in\mathbb{N}} \to z \in \mathbb{E}$.*

Proof.

1. Let $z \in \mathcal{K}$. By the hypothesis, there exists a sequence $\{z^k\}_{k\in\mathbb{N}} \to z$ such that $z^k \in \mathcal{K}^k$ for each $k \in \mathbb{N}$. From the definition of projection, $\|\Pi_{\mathcal{K}^k}(z) - z\| \leqslant \|z^k - z\|$, whence follows the result;
2. Since the projection is nonexpansive and $0 \in \mathcal{K}^k$, $\{\Pi_{\mathcal{K}^k}(z)\}_{k\in\mathbb{N}}$ is bounded by $\|z\|$. Let $N \subseteq \mathbb{N}$ be any infinite subset of $\mathbb{N}$ such that $\{\Pi_{\mathcal{K}^k}(z)\}_{k\in N}$ converges to, say, $w \in \mathbb{E}$. Since $\limsup_{k\in\mathbb{N}} \mathcal{K}^k \subseteq \mathcal{K}$ we obtain $w \in \mathcal{K}$. Now, for every $y \in \mathcal{K}$, it follows from the previous item that
$$\|z - w\| = \lim_{k\in N} \|z - \Pi_{\mathcal{K}^k}(z)\| \leqslant \lim_{k\in N} \|z - \Pi_{\mathcal{K}^k}(y)\| = \|z - y\|,$$
which means $w = \Pi_{\mathcal{K}}(z)$, hence $\lim_{k\in\mathbb{N}} \Pi_{\mathcal{K}^k}(z) = \Pi_{\mathcal{K}}(z)$. Now,
$$\begin{aligned}\|\Pi_{\mathcal{K}^k}(z^k) - \Pi_{\mathcal{K}}(z)\| &\leqslant \|\Pi_{\mathcal{K}^k}(z^k) - \Pi_{\mathcal{K}^k}(z)\| + \|\Pi_{\mathcal{K}^k}(z) - \Pi_{\mathcal{K}}(z)\| \\ &\leqslant \|z^k - z\| + \|\Pi_{\mathcal{K}^k}(z) - \Pi_{\mathcal{K}}(z)\| \to 0,\end{aligned}$$
which completes the proof.

□

To study global convergence of algorithms that benefit from continuous approximations of $\mathcal{K}$, we propose an adapted version of the sequential optimality condition from [1], called *approximate gradient projection* (AGP):

Definition 2.2 (R-AGP)**.** Let $\{\mathcal{K}^k\}_{k\in\mathbb{N}}$ be a continuous approximation of $\mathcal{K}$, in the sense of Definition 2.1. We say a feasible point $\bar{x}$ of (NCP) satisfies the *relaxed cone AGP* (R-AGP) condition if there exist sequences $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$ and $\{\mu^k\}_{k\in\mathbb{N}}$ such that $\mu^k \in (\mathcal{K}^k)^\circ$ for every $k \in \mathbb{N}$ and

1. $\nabla L(x^k, \mu^k) := \nabla f(x^k) + Dg(x^k)^*[\mu^k] \to 0$,
2. $\langle \mu^k, \Pi_{\mathcal{K}^k}(g(x^k))\rangle \to 0$,

where $\Pi_{\mathcal{K}^k}(g(x^k))$ is the orthogonal projection of $g(x^k)$ onto $\mathcal{K}^k$.

Next, we prove that R-AGP is a necessary optimality condition. For this purpose, we prove the convergence of a variation of the external penalty algorithm mid-proof. This algorithm, even though it seems novel, is not highlighted here because R-AGP will be used later for building the convergence theory of an Augmented Lagrangian algorithm, which is where we intend to focus. For doing so, besides continuity of the cone approximation, we assume also that locally the original feasible set $g^{-1}(\mathcal{K})$ is asymptotically included in the approximate feasible set $g^{-1}(\mathcal{K}^k)$.

Theorem 2.2. *If $\bar{x}$ is a local minimizer of (NCP), $\{\mathcal{K}^k\}_{k\in\mathbb{N}}$ is a continuous approximation of $\mathcal{K}$, and there is a $\delta > 0$ such that $g^{-1}(\mathcal{K}) \cap B[\bar{x}, \delta] \subseteq \liminf_{k\in\mathbb{N}} g^{-1}(\mathcal{K}^k) \cap B[\bar{x}, \delta]$, then $\bar{x}$ satisfies R-AGP.*

Proof. Let us assume that $\delta > 0$ is small enough such that $f(\bar{x}) \leq f(x)$ for all $x \in g^{-1}(\mathcal{K}) \cap B[\bar{x}, \delta]$ and let $\{\rho_k\}_{k\in\mathbb{N}} \to \infty$. Consider the regularized penalized subproblem of (NCP):

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{Minimize}} & P_k(x) \doteq f(x) + \dfrac{\rho_k}{2}\|\Pi_{(\mathcal{K}^k)^\circ}(g(x))\|^2 + \dfrac{1}{2}\|x - \bar{x}\|^2 \\ \text{subject to} & \|x - \bar{x}\| \leqslant \delta. \end{array} \tag{RegP-k}$$

Let $\{x^k\}_{k\in\mathbb{N}}$ be a sequence of global minimizers of (RegP-k), which is bounded by the ball centered at $\bar{x}$ with radius δ, and let w be an arbitrary limit point of it. Then, by the optimality of x^k, we have $P_k(x^k) \leqslant P_k(\bar{x})$ for every $k \in \mathbb{N}$, which implies

$$\frac{\rho_k}{2}\|\Pi_{(\mathcal{K}^k)^\circ}(g(\bar{x}))\|^2 \geqslant f(x^k) - f(\bar{x}) + \frac{\rho_k}{2}\|\Pi_{(\mathcal{K}^k)^\circ}(g(x^k))\|^2 + \frac{1}{2}\|x^k - \bar{x}\|^2. \tag{1}$$

Dividing everything by $\dfrac{\rho_k}{2}$, we obtain

$$\|\Pi_{(\mathcal{K}^k)^\circ}(g(\bar{x}))\|^2 \geqslant \frac{2(f(x^k) - f(\bar{x}))}{\rho_k} + \|\Pi_{(\mathcal{K}^k)^\circ}(g(x^k))\|^2 + \frac{\|x^k - \bar{x}\|^2}{\rho_k} \geqslant \frac{2(f(x^k) - f(\bar{x}))}{\rho_k}, \tag{2}$$

for every $k \in \mathbb{N}$. Since $\Pi_{(\mathcal{K}^k)^\circ}(g(\bar{x})) = g(\bar{x}) - \Pi_{\mathcal{K}^k}(g(\bar{x})) \to 0$ (Lemma 2.1 item 1) and $\{x^k\}_{k\in\mathbb{N}}$ is bounded, and $\rho_k \to \infty$, we obtain

$$\lim_{k\to\infty} \|\Pi_{(\mathcal{K}^k)^\circ}(g(x^k))\| = 0, \tag{3}$$

which means $g(w) \in \mathcal{K}$ because

$$\|\Pi_{\mathcal{K}^\circ}(g(x^k))\| \leqslant \|g(x^k) - \Pi_{\mathcal{K}^k}(g(x^k))\| + \|\Pi_{\mathcal{K}^k}(g(x^k)) - \Pi_{\mathcal{K}}(g(x^k))\|,$$

given that both terms vanish due to (3) and Lemma 2.1 item 2, respectively.

For every $z \in g^{-1}(\mathcal{K}) \cap B[\bar{x}, \delta]$, there is a sequence $\{z^k\}_{k\in\mathbb{N}} \to z$ such that $z^k \in g^{-1}(\mathcal{K}^k) \cap B[\bar{x}, \delta]$, so

$$f(x^k) + (1/2)\|x^k - \bar{x}\|^2 \leqslant P_k(x^k) \leqslant P_k(z^k) = f(z^k) + (1/2)\|z^k - \bar{x}\|^2.$$

Taking limits, we obtain $f(w) + (1/2)\|w - \bar{x}\|^2 \leqslant f(z) + (1/2)\|z - \bar{x}\|^2$. Hence, w is a global minimizer of the following localized problem:

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{Minimize}} & f(x) + \dfrac{1}{2}\|x - \bar{x}\|^2 \\ \text{subject to} & g(x) \in \mathcal{K} \\ & \|x - \bar{x}\| \leqslant \delta, \end{array} \tag{Loc-NCP}$$

but the unique global minimizer of (Loc-NCP) is $\bar{x}$, which means $w = \bar{x}$. For k sufficiently large, we have $\|x^k - \bar{x}\| < \delta$ and, by the first-order conditions for (RegP-k), we obtain

$$\nabla f(x^k) + Dg(x^k)^*[\rho_k \Pi_{(\mathcal{K}^k)^\circ}(g(x^k))] = -(x^k - \bar{x}). \tag{4}$$

Defining $\mu^k \doteq \rho_k \Pi_{(\mathcal{K}^k)^\circ}(g(x^k))$ for every $k \in \mathbb{N}$ is enough to finish the proof, because of (4) and

$$\langle \mu^k, \Pi_{\mathcal{K}^k}(g(x^k))\rangle = \rho_k \langle \Pi_{(\mathcal{K}^k)^\circ}(g(x^k)), \Pi_{\mathcal{K}^k}(g(x^k))\rangle = 0,$$

with $x^k \to \bar{x}$.

□

It is not true that the continuity of g together with the fact $\{\mathcal{K}^k\}_{k\in\mathbb{N}}$ is a continuous approximation of $\mathcal{K}$ implies that $g^{-1}(\mathcal{K}) \subseteq \liminf g^{-1}(\mathcal{K}^k)$, even if all sets involved are closed convex cones. For example, take

$$g(x) \doteq \begin{pmatrix} x+1 \\ x^2 \end{pmatrix} \quad \text{and} \quad \mathcal{K} \doteq \text{cone}((1,0)) \quad \text{and} \quad \mathcal{K}^k \doteq \text{cone}((1,-1/k)),\ \forall k \in \mathbb{N}.$$

In this case, we have $\mathcal{K}^k \to \mathcal{K}$, but $g^{-1}(\mathcal{K}) = \{(1,0)\}$ and $g^{-1}(\mathcal{K}^k) = \emptyset$ for every $k \in \mathbb{N}$. In order to ensure the validity of this inclusion, we will consider from now on that $\{\mathcal{K}^k\}_{k\in\mathbb{N}}$ is an outer approximation of $\mathcal{K}$; that is, if $\mathcal{K} \subseteq \mathcal{K}^k$ for every $k \in \mathbb{N}$, then of course $g^{-1}(\mathcal{K}) \subseteq g^{-1}(\mathcal{K}^k)$ for every $k \in \mathbb{N}$, whence follows that $g^{-1}(\mathcal{K}) \subseteq \liminf\limits_{k\in\mathbb{N}} g^{-1}(\mathcal{K}^k)$.

2.1 An Augmented Lagrangian variant with projections onto $\mathcal{K}^k$

Given a positive scalar sequence $\{\rho_k\}_{k\in\mathbb{N}}$, a continuous approximation $\{\mathcal{K}^k\}_{k\in\mathbb{N}}$ of $\mathcal{K}$, a compact set $\mathcal{B}$ such that $\mathcal{B} \cap (\mathcal{K}^k)^\circ \neq \emptyset$ for every $k \in \mathbb{N}$ together with a sequence $\{\widehat{\mu}^k\}_{k\in\mathbb{N}} \subset \mathcal{B}$, let $L_{\rho_k,\widehat{\mu}^k} : \mathbb{R}^n \to \mathbb{R}$ be an *augmented Lagrangian function with a cone approximation*, defined as

$$L_{\rho_k,\widehat{\mu}^k}(x) \doteq f(x) + \frac{\rho_k}{2}\left\|\Pi_{(\mathcal{K}^k)^\circ}\left(g(x) + \frac{\widehat{\mu}^k}{\rho_k}\right)\right\|^2 - \frac{1}{2\rho_k}\left\|\widehat{\mu}^k\right\|^2. \tag{5}$$

Now, consider the Algorithm 1.

Algorithm 1 General framework: Augmented Lagrangian

Inputs: a sequence $\{\varepsilon_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ of scalars such that $\varepsilon_k \to 0$; a nonempty bounded set $\mathcal{B} \cap \mathcal{K}^\circ \neq \emptyset$; real parameters $\tau > 1$, $\sigma \in (0,1)$, and $\rho_0 > 0$; $V^{-1} \in \mathbb{E}$; and initial points $(x^{-1}, \widehat{\mu}^0) \in \mathbb{R}^n \times \mathcal{B} \cap (\mathcal{K}^0)^\circ$. For every $k \in \mathbb{N}$:

1. Compute some point x^k such that
$$\|\nabla L_{\rho_k,\widehat{\mu}^k}(x^k)\| \leq \varepsilon_k, \tag{6}$$
using an iterative method starting from x^{k-1};

2. Update the multiplier
$$\mu^k \doteq \rho_k \Pi_{(\mathcal{K}^k)^\circ}\left(g(x^k) + \frac{\widehat{\mu}^k}{\rho_k}\right), \tag{7}$$
and define $\widehat{\mu}^{k+1}$ as the projection of μ^k onto $\mathcal{B} \cap (\mathcal{K}^k)^\circ$; for the standard choice of $\mathcal{B}$ as the ball of radius R around the origin, take $\widehat{\mu}^{k+1} \doteq \dfrac{\min\{\|\mu^k\|, R\}}{\|\mu^k\|}\mu^k$;

3. Define
$$V^k \doteq \frac{\widehat{\mu}^k}{\rho_k} - \Pi_{(\mathcal{K}^k)^\circ}\left(g(x^k) + \frac{\widehat{\mu}^k}{\rho_k}\right); \tag{8}$$

4. If $\|V^k\| \leq \sigma\|V^{k-1}\|$, set $\rho_{k+1} \doteq \rho_k$. Otherwise, choose some $\rho_{k+1} \geq \tau\rho_k$.

Note that Steps 3 and 4 imply that either $\rho_k \to \infty$ or there is some $k_0 \in \mathbb{N}$ such that $\rho_k = \rho_{k_0}$ for every $k > k_0$ and $V^k \to 0$. With this in mind, we proceed by showing that Algorithm 1 generates sequences whose limit points satisfy R-AGP.

Theorem 2.3. *Let $\bar{x}$ be a feasible limit point of a sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by Algorithm 1, for any given choice of parameters $\{\widehat{\mu}^k\}_{k\in\mathbb{N}}$ and $\{\rho_k\}_{k\in\mathbb{N}}$ conforming to Steps 2 and 3. Then, $\bar{x}$ satisfies R-AGP.*

Proof. First, let us assume for simplicity that $x^k \to \bar{x}$. Let $\{\mu^k\}_{k\in\mathbb{N}}$ be as in (7) and observe that $\nabla L_{\rho_k,\widehat{\mu}^k}(x^k) = \nabla_x L(x^k,\mu^k) \to 0$. If $\rho_k \to \infty$, then

$$\langle \Pi_{\mathcal{K}^k}(g(x^k)), \mu^k\rangle = \rho_k \left\langle \Pi_{\mathcal{K}^k}(g(x^k)) - \Pi_{\mathcal{K}^k}\left(g(x^k)+\frac{\widehat{\mu}^k}{\rho_k}\right), \Pi_{(\mathcal{K}^k)^\circ}\left(g(x^k)+\frac{\widehat{\mu}^k}{\rho_k}\right)\right\rangle$$

which implies (using Cauchy-Schwarz inequality and the nonexpansiveness of the projection) that

$$\begin{aligned}|\langle \Pi_{\mathcal{K}^k}(g(x^k)), \mu^k\rangle| &\leqslant \rho_k \left\|\Pi_{\mathcal{K}^k}(g(x^k)) - \Pi_{\mathcal{K}^k}\left(g(x^k)+\frac{\widehat{\mu}^k}{\rho_k}\right)\right\| \left\|\Pi_{(\mathcal{K}^k)^\circ}\left(g(x^k)+\frac{\widehat{\mu}^k}{\rho_k}\right)\right\| \\ &\leqslant \rho_k \left\|\frac{\widehat{\mu}^k}{\rho_k}\right\| \left\|\Pi_{(\mathcal{K}^k)^\circ}\left(g(x^k)+\frac{\widehat{\mu}^k}{\rho_k}\right)\right\|\end{aligned}$$

but due to Lemma 2.1 item 2, since $\rho_k \to \infty$ we know that $\|\Pi_{(\mathcal{K}^k)^\circ}(g(x^k)+\widehat{\mu}^k/\rho_k)\| \to \|\Pi_{\mathcal{K}^\circ}(g(\bar{x}))\| = 0$, so $\langle \Pi_{\mathcal{K}^k}(g(x^k)), \mu^k\rangle \to 0$. If $\rho_k = \rho_{k_0}$ for some k_0 and every $k \geq k_0$, then using the fact

$$\mu^k = \widehat{\mu}^k - \rho_k V^k$$

we have that

$$\langle \Pi_{\mathcal{K}^k}(g(x^k)), \mu^k\rangle = \langle \Pi_{\mathcal{K}^k}(g(x^k)), \widehat{\mu}^k\rangle - \langle \Pi_{\mathcal{K}^k}(g(x^k)), \rho_k V^k\rangle,$$

but for any convergent subsequence of $\{\widehat{\mu}^k\}_{k\in\mathbb{N}}$, which we will assume to be itself, so $\{\widehat{\mu}^k\}_{k\in\mathbb{N}} \to \bar{\mu}$, we recall that $V^k \to 0$ and $\rho_k \to \rho_{k_0}$ to obtain that

$$\langle \Pi_{\mathcal{K}^k}(g(x^k)), \mu^k\rangle \to \langle \Pi_{\mathcal{K}}(g(\bar{x})), \bar{\mu}\rangle = \langle g(\bar{x}), \bar{\mu}\rangle = 0,$$

which follows from $V^k \to 0$, since it implies $\bar{\mu} = \Pi_{\mathcal{K}^\circ}(\rho_{k_0} g(\bar{x}) + \bar{\mu})$, which in turn holds if, and only if, $\bar{\mu} \in \mathcal{K}^\circ$, $g(\bar{x}) \in \mathcal{K}$, and $\langle g(\bar{x}), \bar{\mu}\rangle = 0$[1]. Thus, $\bar{x}$ satisfies R-AGP. □

The proof above is an adaptation of the proof of [1, Theorem 4.1]. Moreover, it is easy to show that if $\bar{x}$ satisfies R-AGP and Robinson's CQ:

$$0 \in \operatorname{int}(\operatorname{Im}(Dg(\bar{x})) + \mathcal{K} - g(\bar{x})),$$

then $\bar{x}$ must also satisfy the KKT conditions. This is a consequence of the boundedness of $\{\mu^k\}_{k\in\mathbb{N}}$ in the presence of Robinson's CQ. Indeed, if $\{\mu^k\}_{k\in\mathbb{N}}$ is not bounded, then we may take a subsequence such that $\|\mu^k\| \to \infty$, so

$$\lim_{k\to\infty} \frac{\nabla f(x^k)}{\|\mu^k\|} + Dg(x^k)^* \left[\frac{\mu^k}{\|\mu^k\|}\right] = \lim_{k\to\infty} Dg(x^k)^* \left[\frac{\mu^k}{\|\mu^k\|}\right] = 0$$

but on the other hand we also have

$$\left\langle \Pi_{\mathcal{K}^k}(g(x^k)), \frac{\mu^k}{\|\mu^k\|}\right\rangle \to 0 \quad \text{and} \quad \Pi_{\mathcal{K}^k}(g(x^k)) \to \Pi_{\mathcal{K}}(g(\bar{x})) = g(\bar{x})$$

by Lemma 2.1 item 2. Define $H \doteq \operatorname{Im}(Dg(\bar{x})) + \mathcal{K} - g(\bar{x})$ and the above reasoning tells us that any limit point $\tilde{\mu} \neq 0$ of $\{\mu^k/\|\mu^k\|\}_{k\in\mathbb{N}}$ must belong to $H^\circ = \operatorname{Ker}(Dg(\bar{x})) \cap \mathcal{K}^\circ \cap \{g(\bar{x})\}^\perp$. However, since H° is a cone, this also means that $\alpha\tilde{\mu} \in H^\circ$ for every $\alpha > 0$. On the other hand, by Robinson's CQ, we also have that $\alpha\tilde{\mu} \in H$ for every $\alpha > 0$ small enough, implying that $\tilde{\mu} = 0$, which is a contradiction.

[1] In general, for any closed convex cone $K \subseteq \mathbb{E}$ and any $y, z \in \mathbb{E}$, note that $z = \Pi_{K^\circ}(y+z)$ implies that $\Pi_K(y+z) = y + z - \Pi_{K^\circ}(y+z) = y$.

3 Continuous (polyhedral) approximation of the cone of copositive matrices

Now, we illustrate a continuous approximation of $\mathcal{K}$ (Definition 2.1) by polyhedra through a concrete example.

The cone of copositive matrices and its dual, the cone of completely positive matrices, are well studied closed convex cones. It is known that many $\mathcal{NP}$-hard problems can be formulated as convex problems employing these cones [6, 7, 8, 15]. Among several inner and outer hierarchical and asymptotic approximations of these cones, we consider the polyhedral approximation proposed by Yıldırım [16], due to its simplicity and cheap orthogonal projection onto.

For integers $m \geq 1$ and $r \geq 0$, consider the discrete set of vectors in $\mathbb{R}^m$ formed by regular grids of rational points on the unit simplex:

$$\delta_r^m \doteq \bigcup_{k=0}^{r} \{z \in \Delta_m \ : \ (k+2)z \in \mathbb{N}^m\},$$

where $\Delta_m \doteq \{y \in \mathbb{R}^m \ : \ e^T y = 1\}$ is the $(m-1)$-dimensional unit simplex. Let $\ell_r^m := |\delta_r^m|$ be the number of elements in δ_r^m, which can be roughly upper bounded by $m^2(m^{r+1}-1)/(m-1)$ [16].

Let $\mathcal{S}^m$ be the set of $m \times m$ symmetric matrices. For each $r = 0, 1, \ldots$, we define the following closed convex cones, which in fact are polyhedra:

$$\mathcal{O}_r^m \doteq \{y \in \mathcal{S}^m \ : \ d^T y d \geq 0, \forall d \in \delta_r^m\}$$

and its dual

$$(\mathcal{O}_r^m)^* \doteq \left\{ \sum_{d \in \delta_r^m} \lambda_d d d^T \ : \ \lambda_d \geq 0 \right\}, \tag{9}$$

which give an outer approximation of the closed convex cone of copositive matrices

$$\mathcal{C}^m \doteq \{y \in \mathcal{S}^m \ : \ u^T y u \geq 0, \forall u \in \mathbb{R}_+^m\}$$

and an inner approximation of the closed convex cone of completely positive matrices

$$(\mathcal{C}^m)^* \doteq \left\{ \sum_{i=1}^{k} (v_i)(v_i)^T \ : \ k \geq 1, v_i \in \mathbb{R}_+^m \right\},$$

respectively.

It can be shown [16] that

$$(\mathcal{O}_0^m)^* \subseteq (\mathcal{O}_1^m)^* \subseteq \cdots \subseteq (\mathcal{C}^m)^* \subseteq \mathcal{S}_+^m \cap \mathcal{N}^m \subseteq \mathcal{S}_+^m \subseteq \mathcal{S}_+^m + \mathcal{N}^m \subseteq \mathcal{C}^m \subseteq \cdots \subseteq \mathcal{O}_1^m \subseteq \mathcal{O}_0^m,$$

$\mathrm{cl}\left(\bigcup_{r\in\mathbb{N}} (\mathcal{O}_r^m)^*\right) = (\mathcal{C}^m)^*$, and $\bigcap_{r\in\mathbb{N}} \mathcal{O}_r^m = \mathcal{C}^m$ for $\mathcal{N}^m := \{y \in \mathcal{S}^m \ : \ y_{ij} \geq 0, \forall i, j = 1, \ldots, m\}$ where $\mathcal{S}_+^m$ ($\mathcal{S}_-^m$) is the cone of $m \times m$ positive (negative) semidefinite symmetric matrices. Moreover $(\mathcal{C}^m)^* = \mathcal{S}_+^m \cap \mathcal{N}^m$, $\mathcal{C}^m = \mathcal{S}_+^m + \mathcal{N}^m$ for $m \leq 4$, and $\{\mathcal{O}_r^m\}_{r\in\mathbb{N}}$ is a continuous approximation of $\mathcal{C}^m$ (Definition 2.1).

An advantage of employing these polyhedral approximations is that we can obtain an orthogonal projection onto them relatively inexpensive. Namely, for $y \in \mathcal{S}^m$, let us compute its orthogonal projection onto the polar cone of $\mathcal{K}^k := \{y \in \mathcal{S}^m \ : \ d_i^T y d_i \geq 0, \ \forall d_i \in K^k\} \supseteq \mathcal{O}_{\bar{r}}^m$ for some $\bar{r} \geq 0$ and $\emptyset \neq K^k \subseteq \delta_{\bar{r}}^m$. Since $\Pi_{(\mathcal{K}^k)^\circ}(y) = -\Pi_{(\mathcal{K}^k)^*}(-y)$, we need to solve the following optimization problem to obtain the orthogonal projection $-z$ of y:

$$\begin{cases} \min & \| -y - z\|_F^2 \\ \text{subject to} & z \in (\mathcal{K}^k)^*. \end{cases}$$

Using the definition of $\mathcal{K}^k$, the above problem is in fact a convex quadratic program with $|K^k|$ nonnegative variables:

$$\begin{cases} \min & \lambda^T R \lambda + 2 s^T \lambda \\ \text{subject to} & \lambda \in \mathbb{R}_+^{|K^k|} \end{cases} \tag{10}$$

for $R_{ij} = (d_i^T d_j)^2$ and $s_i = d_i^T y d_i$, where $d_i \in K^k$ for $(i, j = 1, 2, \ldots, |K^k|)$. This is due to the fact that if we display all vectors of K^k as rows of a larger matrix

$$D \doteq \begin{pmatrix} d_1^T \\ d_2^T \\ \vdots \\ d_{|K^k|}^T \end{pmatrix},$$

the matrix R will be the self Hadamard product of the Gram matrix DD^T, which is positive semidefinite. Thus, $R = (DD^T) \circ (DD^T)$ is also positive semidefinite.

As we can observe, other closed convex cones such as exponential, hyperbolicity [10] or any other cone can be treated in a similar manner as long as we have a continuous approximation of them by polyhedra. This will permit us to minimize nonlinear functions over these cones by our proposed algorithm.

4 Numerical Experiments

4.1 Test problems

As of our knowledge, there is no benchmark problems or reported computational results on minimizing nonlinear objective functions on difficult convex cones such as the copositive one. Therefore, we created a set of 14 (NCP) test problems in order to verify the performance of the proposed algorithm. Our test problems will have the (NCP) structure

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{Minimize}} & f(x), \\ \text{subject to} & g(x) \in \mathcal{K}, \end{array}$$

with $f(x)$ a selected nonlinear objective functions from [13, 11, 14] (see Table 1) and $g(x) = Q_0 + \sum_{i=1}^{n} x_i Q_i$ a randomly selected linear matrix conic constraint. The function $g(x)$ is taken linear just to simplify the random generation, but for the algorithm this is not a critical issue. In any case, the considered constraint $g(x) \in \mathcal{K}$ is not linear in its nature, and hard due to the difficulty already compressed in the cone $\mathcal{K}$.

We set $\mathcal{K} = \mathcal{C}^m$ and generated $g(x)$ for a given nonlinear objective function $f \colon \mathbb{R}^n \to \mathbb{R}$ by the following procedure:

- Fix a known local minimal solution $x^* \in \mathbb{R}^n$ of the objective f (see Table 1).
- Choose a randomly generated vector $\bar{x} \in \mathbb{R}^n$ where each component is from uniformly distributed samples on the interval $[10, 100]$.
- Select two positive semidefinite symmetric matrices, P_1 and P_2, whose eigenvalues are from uniformly distributed samples on the interval $[0, 1]$.
- Set the matrix variables $Q_0, Q_1, \ldots, Q_n \in \mathcal{S}^m$ as a solution of the following semidefinite program

$$\begin{cases} \min\limits_{Q_0, \ldots, Q_n \in \mathcal{S}^m} & \operatorname{tr}(Q_0 + \sum\limits_{i=1}^{n} \bar{x}_i Q_i) \\ \text{subject to} & Q_0 + \sum\limits_{i=1}^{n} x_i^* Q_i + P_1 \in \mathcal{S}_-^m \\ & Q_0 + \sum\limits_{i=1}^{n} \bar{x}_i Q_i - P_2 \in \mathcal{S}_+^m. \end{cases}$$

The above procedure guarantees that $g(x) \doteq Q_0 + \sum_{i=1}^{n} x_i Q_i$ is such that $g(x^*)$ and $g(\bar{x})$ have non-positive and non-negative eigenvalues, respectively. Therefore, $g(x^*) \notin \mathcal{K}$ and $g(\bar{x}) \in \mathcal{K}$. Cvxpy version 1.3.1 was used to solve the above semidefinite program.

Table 1: Nonlinear objective functions $f(x)$ for (NCP).

function's name	function $f(x)$	known local min. solutions
convex quadratic (cq)	$x_1^2 + x_2^2$	$(0, 0)$
fractional convex (fc)	$x_1^2/(1 + \|x_1\|) + x_2^2/(1 + \|x_2\|)$	$(0, 0)$
extended Rosenbrock (eR) [14]	$\sum_{i=1}^{n-1}(1 - x_i)^2 + 100(x_{i+1} - x_i^2)^2$	$(1, 1, \ldots, 1)$
Freudenstein and Roth (FR) [14]	$[-13 + x_1 + ((5 - x_2)x_2 - 2)x_2]^2$ $+[-29 + x_1 + ((x_2 + 1)x_2 - 14)x_2]^2$	$(5, 4)$ and $(11.41\ldots, -0.8968\ldots)$
Powell badly scaled (Pbs) [14]	$(10^4 x_1 x_2 - 1)^2 + (e^{-x_1} + e^{-x_2} - 1.0001)^2$	$(1.098\ldots \cdot 10^{-5}, 9.106\ldots)$
Beale (B) [14]	$[1.5 - x_1(1 - x_2)]^2 + [2.25 - x_1(1 - x_2^2)]^2$ $+[2.625 - x_1(1 - x_2^3)]^2$	$(3, 0.5)$
Powell singular (Ps) [14]	$(x_1 + 10x_2)^2 + \sqrt{5}(x_3 - x_4)^2 + (x_2 - 2x_3)^4$ $+10(x_1 - x_4)^4$	$(0, 0, 0, 0)$
Wood (W) [14]	$100(x_2 - x_1^2)^2 + (1 - x_1)^2 + 90(x_4 - x_3^2)^2$ $+(1 - x_3)^2 + 10(x_2 + x_4 - 2)^2 + (x_2 - x_4)^2/10$	$(1, 1, 1, 1)$
quartic polynomial (qp)	$\sum_{i=1}^{n}(x_i - 1)^2 + \left[\sum_{i=1}^{n} i(x_i - 1)\right]^2 + \left[\sum_{i=1}^{n} i(x_i - 1)\right]^4$	$(1, 1, \ldots, 1)$
Luenberger-Ye (LY) [11]	$x_1^2 - 5x_1x_2 + x_2^4 - 25x_1 - 8x_2$	$(20, 3)$
ex4_1_5 [13]	$2x_1^2 - 1.05x_1^4 + \frac{5}{30}x_1^6 - x_1x_2 + x_2^2$	$(0, 0)$ and † $(\pm 1.74755, \pm 0.97378)$
ex8_1_4 [13]	$12x_1^2 - 6.3x_1^4 + x_1^6 - 6x_1x_2 + 6x_2^2$	$(0, 0)$
ex8_1_5 [13]	$4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^6 + x_1x_2 - 4x_2^2 + 4x_2^4$	$(0, 0)$ and $(0.08984, -0.71266)$
ex8_1_6 [13]	$[0.1 + (x_1 - 4) + (x_2 - 4)^2]^{-1}$ $-[0.2 + (x_1 - 1)^2 + (x_2 - 1)^2]^{-1}$ $-[0.2 + (x_1 - 8)^2 + (x_2 - 8)^2]^{-1}$	$(1.0004, 1.0004)$ and $(3.99995, 3.99995)$

† Not reported in [13].

4.2 Details of the implementation

Algorithm 1 computes an approximate local optimal solution when $\|\nabla L_{\rho_k,\widehat{\mu}^k}(x^k)\|$ and $\|V^k\|$ are small enough since they correspond to the norms of derivatives of the safeguarded augmented Lagrangian function $L_{\rho,\widehat{\mu}}(x)$ in relation to x^k and $\widehat{\mu}^k$, respectively. Therefore, $\{x^k\}_{k\in\mathbb{N}}$ and $\{\mu^k\}_{k\in\mathbb{N}}$ computed by Algorithm 1 are sequences which define a limit point $\bar{x}$ satisfying R-AGP according to Theorem 2.3.

Simple schemes were sought to define $\{\varepsilon_k\}_{k\in\mathbb{N}}$ and $\{\mathcal{K}^k\}_{k\in\mathbb{N}}$ in Algorithm 1. Among others, the empirical choice $\varepsilon_k \doteq \min\{\varepsilon_0, \|V^k\|_{\max}\}$ seemed to provide fast convergence, where $\|V\|_{\max} = \max_{1\le i,j\le m} |V_{ij}|$. This strategy revealed to be superior than a conventional choice $\varepsilon_k \doteq \eta^k\varepsilon_0$ for $0 < \eta < 1$ which guarantees $\varepsilon_k \to 0$. Although our strategy may not guarantee $\varepsilon_k \to 0$, it appears to be reasonable when we expect that $\|V^k\|_{\max} \to 0$, whenever we succeed to compute x^k satisfying (6) at every iteration.

The strategy to choose $\{\mathcal{K}^k\}_{k\in\mathbb{N}}$ has more variations and after some tests, we decided that the best strategy was to fix first the positive integers $r_{\max}$, which defines $\mathcal{O}^m_{r_{\max}} \supseteq \mathcal{C}^m$ (see Section 3), and $\zeta < |\delta^m_{r_{\max}}| - |\delta^m_0|$. Then, we order all vectors in $\delta^m_{r_{\max}}$ such that the first $\ell^m_0 = |\delta^m_0|$ elements are the vectors in δ^m_0 (in any order), followed by the vectors in $\delta^m_1 \backslash \delta^m_0$ (in any order), until the vectors in $\delta^m_{r_{\max}} \backslash \delta^m_{r_{\max}-1}$ (in any order). Set $K^0 \doteq \delta^m_0$ in $\mathcal{K}^k = \{y \in \mathcal{S}^m \ : \ d_i^T y d_i \ge 0, \ \forall d_i \in K^k\}$, and $K^k \doteq K^{k-1} \cup \{\text{first } \zeta \text{ vectors (whenever possible) in } \delta^m_{r_{\max}} \text{ which are not in } K^{k-1} \text{ in the above order}\}$. That is, at every iteration of Algorithm 1, $\mathcal{K}^k$ is defined by $|K^k|$ vectors which has exactly ζ more vectors than $|K^{k-1}|$, unless $0 < |\delta^m_{r_{\max}} \backslash K^{k-1}| < \zeta$. In our numerical experiments (see Section 4.3), we considered test problems for $m = 3$ and $r_{\max} = 15$ which gives $\ell^m_{r_{\max}} = 901$, and $m = 5$ and $r_{\max} = 7$ which gives $\ell^m_{r_{\max}} = 1816$. Notice that in order for our approximation of $\mathcal{K} = \mathcal{C}^m$ to be continuous as in Definition 2.1, we would need to keep approximating the sets δ^m_r for $r > r_{max}$, however this is numerically intractable.

Another detail apparently hidden in the implementation of Algorithm 1 is the "size" of the bounded set $\mathcal{B}$ which we set as the ball of radius R, $\{x \in \mathbb{R}^n \ : \ \|x\|_\infty \le R\}$. At every iteration, we need to project μ^k onto $\mathcal{B}$ at step 2 and R should be large enough to not bound the real size of the Lagrange multiplier $\widehat{\mu}^{k+1} \in \mathcal{B} \cap (\mathcal{K}^k)^\circ$. In the numerical experiments, we set $R = 10^{12}$.

Finally, a scaling of the safeguarded augmented Lagrangian function $L_{\rho_k,\widehat{\mu}^k}(x)$ is effective, specially because some objective functions $f(x)$ are badly scaled. We divided the function $L_{\rho_k,\widehat{\mu}^k}(x)$ (5) by the average of the first five iterations of $\max\{1, \|\nabla L_{\rho_k,\widehat{\mu}^k}(x^{k-1})\|_\infty, \|\nabla f(x^{k-1})\|_\infty\}$.

4.3 Numerical results

All numerical experiments were performed on Intel Core i7-10700 (2.90GHz, 8 cores) processor with 8GB of memory running python 3.9.16. Since the function (5) is only once differentiable [9], scipy.optimize function with option "method='BFGS' " was used to solve it with "gtol=ε_k". This option was the best choice to have less failure to compute an approximate x^k satisfying (6), which seems to be the achilles heel of the algorithm. The convex quadratic problem (10) on the other hand was solved by mosek 10.0.43.

We adopted the following stopping criterion for Algorithm 1. The algorithm stops successfully if $\|\nabla L_{\rho_k,\widehat{\mu}^k}(x^k)\| \leq \varepsilon_L$, $\|V\|_{\max} \leq \varepsilon_V$, and $r = r_{\max}$. We forcefully stop the algorithm whenever we fail to compute (6) in overall of 20% of iterations, respecting a minimum of 14 iterations, since we observed that Algorithm 1 fails to converge whenever we cannot compute x^k satisfying (6) in some consecutive iterations.

The following parameter values were set for Algorithm 1: each coordenate of $x^{-1} \in \mathbb{R}^n$ randomly chosen from interval $[-100, 100]$, $V^{-1} = \dfrac{\widehat{\mu}^0}{\rho_0} - \Pi_{(\mathcal{K}^0)^\circ}\left(g(x^{-1}) + \dfrac{\widehat{\mu}^0}{\rho_0}\right)$, $\widehat{\mu}^0 = RI$, $R = 10^{12}$, $\sigma = 0.9$, $\tau = 2.0$, $\varepsilon_V = \varepsilon_L = 10^{-5}$; and $\rho_0 = 0.1$, $\varepsilon_0 = 1.0$ for $m = 3$, and $\rho_0 = 1.0$, $\varepsilon_0 = 0.1$ for $m = 5$, respectively. I is the $m \times m$ identity matrix.

Our first experiment aims to determine the parameter ζ, which is the number of vectors to add at every iteration to form K^k from K^{k-1} (see Section 3). Figures 1 and 2 show the performance profile in terms of wall-clock time when solving the 14 test problems described in Subsection 4.1 for matrix orders $m = 3$ and $m = 5$, respectively. We can conclude that, in fact, there is a clear preference for the parameter ζ and it depends on the matrix sizes in this particular setting. Therefore, we fix $\zeta = 45$ and $\zeta = 70$ for the problems with matrix orders $m = 3$ and $m = 5$, respectively, in the main part of the numerical experiments.

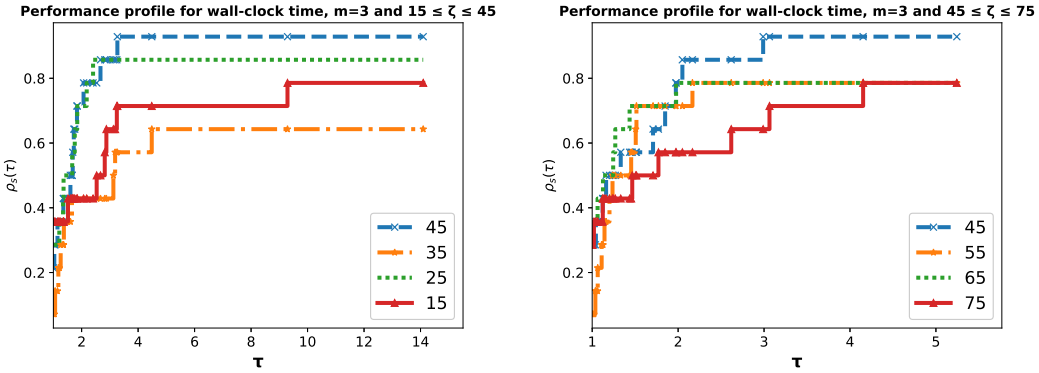

Figure 1: Performance profile for $m = 3$ when solving 14 problems by the "proposed" method for values of $\zeta = 15, 25, 35, 45, 55, 65$, and 75. It shows that $\zeta = 45$ is the best choice.

In order to measure the effectiveness of our proposal in practice, we compared our algorithm to the standard augmented Lagrangian method applied to (NCP) where we just considered $K^0 = K^1 = \cdots = \delta^m_{r_{\max}}$. That is, the $\mathcal{K}^k$ is fixed to the best approximation from the first iteration. We refer to this approach as "standard" in our numerical results in contrast to "proposed" for our proposal.

Table 2 details the numerical results for 14 problems when $m = 3$ and $m = 5$. The numbers of third and fourth columns are in boldface when they meet the stopping criterion as well as the column for r. The values of this column should be equal to 15 (when $m = 3$) and 7 (when $m = 5$), which correspond to the $r_{\max}$ fixed beforehand. Observe that in problems such as "fc", "proposed" stops without attaining the pre-defined approximation for $\mathcal{K}$, where r should be 15 and 7 (instead of 14 and 5) due to excessive failure (4 out of 17 iterations and 10 out of 14 iterations) to satisfy (6), respectively. The column "fails" indicates the number of times (6) is not met for the current ε_k among all iterations at "it."

We also computed $\min_{y \in \mathcal{S}^m_+ + \mathcal{N}^n} \|g(x^k) - y\|_F$ that gave values between 3.6e-15 to 6.9e-02, omitted from the table. If this value is close to zero, we can understand that $g(x^k) \in \mathcal{K} = \mathcal{C}^m$, since $\mathcal{S}^m_+ + \mathcal{N}^m \subseteq \mathcal{C}^m$, with equality holding for $m \leq 4$ (see Section 3). For instance for problem "Pbs" with $m = 3$ and "proposed", we obtained this smallest value, indicating that $g(x^k) \in \mathcal{C}^3$, while for problem "LY" with $m = 5$, both

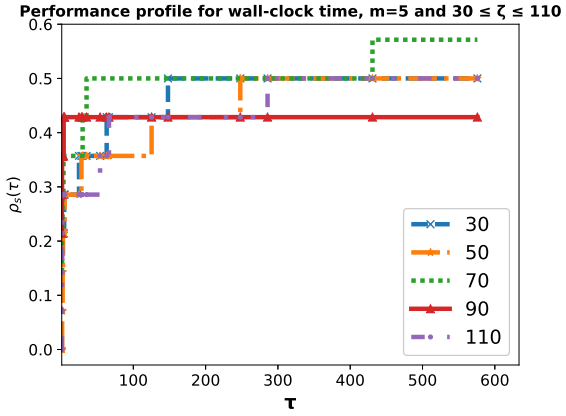

Figure 2: Performance profile for $m = 5$ when solving 14 problems by the "proposed" method for values of $\zeta = 30, 50, 70, 90$, and 110. It shows that $\zeta = 70$ is preferred.

for "proposed" and "stardard", we obtained this largest value indicating that $g(x^k) \notin \mathcal{S}^5_+ + \mathcal{N}^5$, but not certain whether $g(x^k) \in \mathcal{C}^5$.

In general, the gradual polyhedral approximation of the closed convex cone $\mathcal{K} = \mathcal{C}^m$, which we are proposing, seems superior than considering a standard augmented Lagrangian method with fixed approximation $\mathcal{K}^k$ from the beginning. This can be easily concluded from Figure 3, which show the performance profile for the wall-clock time in Table 2, when $m = 3$ and $m = 5$, respectively. Observe that some instances could not be solved at all by neither of methods, specially for the case $m = 5$, showing that some nonlinear functions can be challenging for these type of algorithms.

In Table 2, the last column gives the average wall-clock time per iteration, which starts at iteration 0. These values are only reference values since each iteration requires different amount of time, and more time is required when fail occurs due to increasing BFGS iterations to minimize (5). As we can observe, in general, "standard" requires more time per iteration than "proposed", because it needs to solve a larger problem with $K^0 = \delta^m_{\max}$ from the first iterations, that is, larger problems to minimize in the projection (10). In the "proposed" method, K^0 is set to δ^m_0 and gradually increased. There are few exceptions, "B" for $m = 5$, "Ps" for $m = 3$, and "W" for $m = 3$, even removing the cases when "proposed" has more failed iterations than "standard". We believe that these are due to the averaging of the computational time since all these cases have a higher number of iterations than other cases.

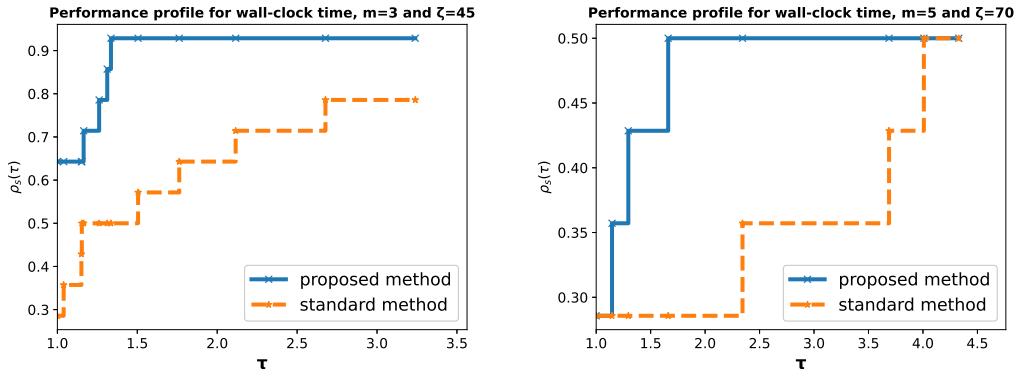

Figure 3: Performance profile for $m = 3$ (left) and $m = 5$ (right) when solving 14 problems by the "proposed" method and the "standard" augmented Lagrangian method.

In conclusion, the "proposed" method can save computation time per iteration in the first iterations and a careful (and maybe conservative) update of the method seems to avoid the failure of satisfying $\|\nabla L_{\rho_k,\widehat{\mu}^k}(x^k)\| \leq \varepsilon_k$ (6) in later iterations as it happens for the "standard" method.

Table 2: Numerical results for 14 problems when solving by the proposed method ("proposed") and the "standard" augmented Lagrangian method with order of matrices $m = 3$ and $m = 5$. It is considered solved whenever $\|\nabla L_{\rho_k,\widehat{\mu}^k}(x^k)\| \leq \varepsilon_L$, $\|V^k\|_{\max} \leq \varepsilon_V$ and $r = r_{\max}$ (which are in bold); $r_{\max} = 15$ for $m = 3$ and $r_{\max} = 7$ for $m = 5$; "fails" means # of iterations (6) was not satisfied.

problem	strategy	$\|\nabla L_{\rho_k,\widehat{\mu}^k}(x^k)\|$	$\|V^k\|_{\max}$	r	it.	fails	wall time (s)	time/it. (s)
cq ($m = 3$)	proposed	**4.197e-06**	**3.590e-07**	**15**	30	2	148.17	4.78
	standard	**4.203e-06**	**3.928e-06**	**15**	29	1	170.45	5.68
cq ($m = 5$)	proposed	**6.079e-07**	**3.702e-06**	**7**	31	5	1621.38	50.67
	standard	**6.836e-07**	**3.141e-06**	**7**	26	1	1251.94	46.37
fc ($m = 3$)	proposed	1.272e-02	4.045e-03	14	17	4	94.74	5.26
	standard	1.225e-02	4.155e-03	**15**	17	4	221.83	12.32
fc ($m = 5$)	proposed	4.005e-02	4.167e-04	5	14	10	170.68	11.38
	standard	9.934e-03	5.303e-02	**7**	14	9	2920.29	194.69
eR ($m = 3$)	proposed	**1.432e-07**	**1.107e-08**	**15**	45	1	235.02	5.11
$n = 5$	standard	**3.528e-08**	**9.324e-11**	**15**	43	0	201.86	4.59
eR ($m = 5$)	proposed	**3.068e-07**	**7.386e-09**	**7**	36	1	2763.04	74.68
$n = 5$	standard	**9.626e-06**	**2.278e-07**	**7**	35	0	1663.28	46.20
FR ($m = 3$)	proposed	**9.131e-06**	**7.198e-07**	**15**	37	0	56.19	1.48
	standard	**1.323e-08**	**9.427e-09**	**15**	37	0	150.44	3.96
FR ($m = 5$)	proposed	**3.925e-06**	**4.340e-07**	**7**	33	1	258.59	7.61
	standard	**4.008e-06**	**1.848e-10**	**7**	33	0	1036.54	30.49
Pbs ($m = 3$)	proposed	**2.467e-15**	**2.014e-10**	**15**	20	0	16.04	0.76
	standard	**1.779e-10**	**3.332e-10**	**15**	1	0	28.27	14.14
Pbs ($m = 5$)	proposed	**9.794e-06**	**1.352e-06**	**7**	48	2	1369.89	27.96
	standard	1.429e-05	**5.431e-06**	**7**	60	12	4753.46	77.93
B ($m = 3$)	proposed	**5.611e-08**	**8.549e-08**	**15**	79	0	228.36	2.85
	standard	**2.299e-09**	**8.696e-07**	**15**	79	0	237.34	2.97
B ($m = 5$)	proposed	1.129e-03	2.099e-03	**7**	84	17	8708.77	102.46
	standard	2.757e-03	2.714e-04	**7**	82	17	8407.61	101.30
Ps ($m = 3$)	proposed	**7.449e-06**	**3.722e-06**	**15**	43	5	617.06	14.02
	standard	1.964e-04	**2.181e-06**	**15**	55	11	693.93	12.39
Ps ($m = 5$)	proposed	**8.866e-06**	1.422e-05	**7**	37	8	2787.54	73.36
	standard	1.111e-05	1.584e-05	**7**	40	8	4178.89	101.92
W ($m = 3$)	proposed	**9.936e-06**	**5.948e-06**	**15**	45	1	345.71	7.52
	standard	**8.095e-06**	**8.633e-06**	**15**	43	2	274.10	6.23
W ($m = 5$)	proposed	2.418e-04	1.629e-04	**7**	49	10	6986.63	139.73
	standard	5.245e-05	8.211e-05	**7**	50	10	7367.50	144.46
qp ($m = 3$)	proposed	**9.756e-06**	**9.944e-06**	**15**	34	1	134.30	3.84
$n = 5$	standard	1.210e-05	**1.464e-06**	**15**	47	10	608.66	12.68
qp ($m = 5$)	proposed	**2.375e-07**	**9.955e-07**	**7**	34	1	1098.84	31.40
$n = 5$	standard	**1.564e-06**	**9.912e-09**	**7**	34	0	959.36	27.41
LY ($m = 3$)	proposed	**9.260e-07**	**4.750e-06**	**15**	37	3	206.93	5.45
	standard	**1.892e-06**	**2.072e-06**	**15**	40	5	437.75	10.68
LY ($m = 5$)	proposed	3.497e-04	1.375e-05	**7**	39	8	3110.35	77.76
	standard	2.615e-05	**3.067e-06**	**7**	39	8	3712.36	92.81
ex4_1_5 ($m = 3$)	proposed	**4.855e-06**	**1.588e-06**	**15**	34	4	246.38	7.04
	standard	**8.450e-07**	**7.721e-06**	**15**	33	1	187.79	5.52
ex4_1_5 ($m = 5$)	proposed	**7.883e-07**	**4.096e-06**	**7**	26	0	394.90	14.63
	standard	**5.657e-06**	**8.653e-07**	**7**	25	1	1456.68	56.03
ex8_1_4 ($m = 3$)	proposed	**5.602e-06**	**2.281e-06**	**15**	46	6	367.15	7.81
	standard	**7.500e-06**	**5.508e-06**	**15**	50	10	552.55	10.83
ex8_1_4 ($m = 5$)	proposed	8.105e-04	1.535e-04	**7**	35	7	2393.85	66.50
	standard	1.346e-03	2.285e-05	**7**	35	7	3237.19	89.92
ex8_1_5 ($m = 3$)	proposed	**4.909e-06**	**2.280e-07**	**15**	31	1	151.47	4.73
	standard	**3.230e-10**	**5.899e-06**	**15**	29	0	174.64	5.82
ex8_1_5 ($m = 5$)	†proposed	**4.751e-11**	**1.929e-08**	**7**	26	1	570.25	21.12
	standard	**3.946e-09**	**1.736e-06**	**7**	21	0	1336.28	60.74
ex8_1_6 ($m = 3$)	proposed	**2.048e-08**	**7.997e-08**	**15**	20	2	80.90	3.85
	standard	**7.126e-09**	**5.140e-11**	**15**	18	0	60.59	3.19
ex8_1_6 ($m = 5$)	proposed	**3.858e-08**	**7.980e-09**	6	23	5	861.55	35.90
	standard	**2.909e-08**	**1.500e-09**	**7**	15	1	974.65	60.92

† indicates mosek could not solve (10) properly in some iterations, but it did not affect the final results.

5 Concluding Remarks

The general optimization problem of minimizing a nonlinear function subject to nonlinear conic constraints has received increasing attention in the recent years. General algorithms for dealing with such generality are still under development. The linear case is of particular interest considering that in many applications the nonlinearities of the problem may be concentrated to lie in the cone itself. However, when it is not clear how to project onto the cone, a practical implementation is usually out of hand. In contrast, most algorithms require that one is able to deal with the full cone in every iteration of the method. In this paper we proposed an augmented Lagrangian algorithm that considers the possibility of iteratively approximating the cone in each iteration, without hindering well established global convergence results. Numerical experiments are conducted with the copositive cone and its polyhedral outer approximation where we demonstrate in a small collection of problems that our strategy is superior than the alternative one of considering the full cone in each iteration.

Acknowledgements

This work was partially supported by grants 2018/24293-0, 2020/04585-7, 2020/07421-5, and 2023/08706-1 from the São Paulo Research Foundation (FAPESP) and grants 302000/2022-4 and 407147/2023-3 from CNPq.

References

[1] R. Andreani, W. Gómez, G. Haeser, and L.M. Mito, *On optimality conditions for nonlinear conic programming*, Mathematics of Operations Research 47 (2022), pp. 2160–2185.

[2] R. Andreani, G. Haeser, and J.M. Martínez, *On sequencial optimality conditions for smooth constrained optimization*, Optimization 60 (2011), pp. 627–641.

[3] R. Andreani, G. Haeser, and D.S. Viana, *Optimality conditions and global convergence for nonlinear semidefinite programming*, Mathematical Programming 180 (2020), pp. 203–235.

[4] R. Andreani, J.M. Martínez, and B.F. Svaiter, *A new sequencial optimality condition for constrained optimization and algorithmic consequences*, SIAM Journal of Optimization 20 (2010), pp. 3533–3554.

[5] A. Berman, M. Dür, and N. Shaked-Monderer, *Open problems in the theory of completely positive and copositive matrices*, Electronic Journal of Linear Algebra 29 (2015), pp. 45–58.

[6] S. Bundfuss and M. Dür, *An adaptive linear approximation algorithm for copositive programs*, SIAM Journal on Optimization 20 (2009), pp. 30–53.

[7] M. Dür, *Copositive programming – a survey*, M. Diehl, F. Glineur, E. Jarlebring, and W. Michiels, eds. Springer, Berlin, Heidelberg, 2010, pp. 3–20.

[8] M. Dür and F. Rendl, *Conic optimization: A survey with special focus on copositive optimization and binary quadratic problems*, EURO Journal on Computational Optimization 9 (2021), 100021.

[9] G. Haeser, and D.O. dos Santos, *A simple proof of existence of Lagrange multipliers*, preprint (2024), arXiv:2402.05335v2, 9 Feb 2024.

[10] B.F. Lourenço, V. Roshchina, and J. Saunderson, *Hyperbolicity cones are amenable*, Mathematical Programming 204 (2024), pp. 753–764.

[11] D.G. Luenberger and Y. Ye, *Linear and Nonlinear Programming, 3rd ed.*, Springer, New York, 2007.

[12] J.M. Martínez and B.F. Svaiter, *A practical optimality condition without constraint qualifications for nonlinear programming*, Journal of Optimization Theory and Applications 118 (2003), pp. 117–133.

[13] *MINLPLib (A Library of Mixed-Integer and Continuous Nonlinear Programming Instances).* https://www.minlplib.org/. Accessed December 26, 2023.

[14] J.J. Moré, B.S. Garbow, and K.E. Hillstrom, *Testing unconstrained optimization software*, ACM Transactions on Mathematical Software 7 (1981), pp. 17–41.

[15] N. Shaked-Monderer, and A. Berman, *Copositive and Completely Positive Matrices*, World Scientific, Singapore, 2021.

[16] E.A. Yıldırım, *On the accuracy of uniform polyhedral approximations of the copositive cone*, Optimization Methods and Software 27 (2012), pp. 155–173.