An Augmented Lagrangian algorithm for nonlinear semidefinite programming applied to the covering problem

Ernesto G. Birgin[1] Walter Gómez[2] Gabriel Haeser[3] Leonardo M. Mito[3]
Daiana S. Viana[3,4]

March 26, 2019

[1] Department of Computer Science, University of São Paulo, São Paulo SP, Brazil
[2] Department of Mathematical Engineering, Universidad de La Frontera, Temuco, Chile
[3] Department of Applied Mathematics, University of São Paulo, São Paulo SP, Brazil
[4] Center of Exact and Technological Sciences, Federal University of Acre, Rio Branco AC, Brazil

In this work we present an Augmented Lagrangian algorithm for nonlinear semidefinite problems (NLSDPs), which is a natural extension of its consolidated counterpart in nonlinear programming. This method works with two levels of constraints; one that is penalized and other that is kept within the subproblems. This is done in order to allow exploiting the subproblem structure while solving it. The global convergence theory is based on recent results regarding approximate Karush-Kuhn-Tucker optimality conditions for NLSDPs, which are stronger than the usually employed Fritz John optimality conditions. Additionally, we approach the problem of covering a given object with a fixed number of balls with a minimum radius, where we exploit some convex algebraic geometry tools, such as Stengle's Positivstellensatz and its variations, which allows for a much more general model. Preliminary numerical experiments are presented.

Keywords: Augmented Lagrangian, Nonlinear Semidefinite Programming, Covering Problem, Convex Algebraic Geometry.

1 Introduction

Augmented Lagrangian methods for nonlinear semidefinite programming that consider quadratic penalty and modified barrier functions have already been introduced in [60] and [25], respectively. Convergence theories of both methods rely on constraint qualifications such as the *constant positive linear dependence* (CPLD), *Mangasarian-Fromovitz constraint qualification* (MFCQ), or non-degeneracy. The Augmented Lagrangian method introduced in the present work extends the results presented in [60, 25] by using the sequential optimality condition introduced in [4], that does not require the use of constraint qualifications.

Sequential optimality conditions [5] have played a major role in extending and unifying global convergence results for several classes of algorithms for nonlinear programming and other classes of problems. In particular, we present a more general version of the algorithm from [4] by allowing arbitrarily constrained subproblems in the Augmented Lagrangian framework. Namely, instead of penalizing all the constraints, we assume the constraints can be divided in two sets: one set of constraints that will be penalized and other that will be kept as constraints for the subproblems. Allowing constrained subproblems is an important algorithmic feature due to the ability to explore good subproblem solvers. See [1, 16]. We introduce this algorithm with a clear application in mind namely the so-called covering problem, where a new model based on convex algebraic geometry tools is introduced.

The covering problem consists of finding the centers and the minimal common radius of m balls whose union covers a given compact set. This problem has several applications, such as finding the optimal broadcast tower distribution over a region or the optimal allocation of Wi-Fi routers in a building, for example. Also, some specific cases have been solved in the last few decades, for example T. Tarnai and Z. Gáspár in 1995 presented minimal coverings of a square with up to ten circles (in [30]). Their results were generalized for rectangles by A. Heppes and H. Melissen in 1997 (see [37]), when they found the best coverings of a rectangle with up to five equal circles and also, under some conditions, with seven circles as well. Then, Melissen and Schuur in 2000 (see [46]) extended those results for six and seven circles, following their previous work [45]. In the same year, K. Nurmela and P. Östergård presented numerical approximations of circle coverings for a square with up to 30 equal circles, improving some of the previous ones until then (in [50]). Covering

a unit circle with another set of n circles has been a matter of deep study for many years. K. Bezdek is responsible for proving coverings for $n = 5$ and $n = 6$ (in [13, 14]) and the result is somewhat trivial for $n \in \{1,2,3,4,7\}$. Then, coverings of a unit circle with other eight, nine and ten circles were presented by Tóth (in [57]). Melissen and Nurmela have also provided coverings with up to 35 circles (see [44, 49]). In 2014, Gáspár, Tarnai and Hincz (see [31, 32]) published a broad paper in which they discuss recently developed techniques used to pursuit a slight variation of the covering problem, where the radius is fixed.

One of the classical algorithmic approaches, introduced in [26] and later used in other relevant works (see, for example, [61, 59]), involves approximating the compact set to be covered by a finite set of points (discretization) and then applying a point covering algorithm. Usually the whole procedure leads to a min-max-min formulation, which needs further treatment, such as smoothing. Discretizing a set is a laborous task itself and the quality of the approximation bounds the quality of the solution. Our approach is intended to be an alternative method that aims to avoid discretization, where we characterize a feasible covering by means of auxiliary semidefinite variables relying on a *Positivstellensatz* theorem.

In our model, only equality constraints will be penalized, hence avoiding eigenvalue computations in evaluating the Augmented Lagrangian function, while the (linear) semidefinite constraints will be kept within the subproblems. Thus, eigenvalues are computed in order to project onto the semidefinite cone in the application of the Spectral Projected Gradient method [19] for solving the subproblems. Numerical experiments are still very preliminary, as the complexity of the model increases rapidly with the increase of the degree of a polynomial needed in the formulation. However, we illustrate the generality of the model by solving the problem of covering a heart-shaped figure.

In Section 2 we give some preliminaries on sequential optimality conditions, in Section 3 we define our Augmented Lagrangian algorithm and prove its global convergence. In Section 4 we give our model to the covering problem. In Section 5 we show our numerical experiments and in Section 6 we give some conclusions.

Notation: $\mathbb{R}$ is the set of real numbers, while $\mathbb{R}_+$ is the set of non-negative real numbers. $\mathbb{S}^m$ is the set of symmetric $m \times m$ matrices over $\mathbb{R}$. $\mathbb{S}^m_+$ (correspondently, $\mathbb{S}^m_-$) is the set of positive (negative) semidefinite $m \times m$ matrices over $\mathbb{R}$, and we denote $A \in \mathbb{S}^m_+$ ($A \in \mathbb{S}^m_-$) by $A \succeq 0$ ($A \preceq 0$). $\mathrm{tr}(A)$ denotes the trace of the matrix A. $\langle A, B\rangle := \mathrm{tr}(AB)$ is the standard inner product over $\mathbb{S}^m$, and $\|A\| := \sqrt{\langle A, A\rangle}$ is the Frobenius norm on $\mathbb{S}^m$. $[a]_+ := \max\{0, a\}, a \in \mathbb{R}$.

2 Preliminaries

In this section we introduce the notation that will be adopted throughout the paper. We also formally present the nonlinear semidefinite programming problem, together with recently introduced optimality conditions. Let us consider the following problem

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Minimize}} \quad & f(x), \\ \text{subject to} \quad & G(x) \preceq 0, \end{aligned} \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $G : \mathbb{R}^n \to \mathbb{S}^m$ are continuously differentiable functions. The set $\mathscr{F}$ will denote the feasible set $\mathscr{F} := \{x \in \mathbb{R}^n \mid G(x) \preceq 0\}$. Throughout this paper $G(x) \preceq 0$ will be called a semidefinite constraint and we will consider this simplified version of a NLSDP where the equality constraints are omitted without loss of generality. Given a matrix $A \in \mathbb{S}^m$, let $A = U\Lambda U^T$ be an orthogonal diagonalization of A, where $\Lambda = \mathsf{diag}(\lambda_1^U(A), \ldots, \lambda_m^U(A))$ is a diagonal matrix and $\lambda_i^U(A)$ is the eigenvalue of A at position i when A is orthogonally diagonalized by U. We omit U when the eigenvalues considered in the diagonalization of A are in ascending order, namely, $\lambda_1(A) \leq \cdots \leq \lambda_m(A)$. The projection of A onto $\mathbb{S}^m_+$, denoted by $[A]_+$, is given by $[A]_+ := U\mathsf{diag}([\lambda_1^U(A)]_+, \ldots, [\lambda_m^U(A)]_+)U^T$, and is independent of the choice of U. The derivative operator of the mapping $G : \mathbb{R}^n \to \mathbb{S}^m$ at a point $x \in \mathbb{R}^n$ is given by $DG(x) : \mathbb{R}^n \to \mathbb{S}^m$, defined by $DG(x)h := \sum_{i=1}^{n} G_i(x)h_i$, $h \in \mathbb{R}^n$, where $G_i(x) := \dfrac{\partial G(x)}{\partial x_i} \in \mathbb{S}^m, i = 1, \ldots, m$, are the partial derivative matrices with respect to x_i. In addition, its adjoint operator is given by $DG(x)^* : \mathbb{S}^m \to \mathbb{R}^n$ with $DG(x)^*\Omega := (\langle G_1(x), \Omega\rangle, \ldots, \langle G_n(x), \Omega\rangle)^T$, where $\Omega \in \mathbb{S}^m$. We denote the Lagrangian function at $(x, \Omega) \in \mathbb{R}^n \times \mathbb{S}^m_+$ by $L(x, \Omega) := f(x) + \langle G(x), \Omega\rangle$ and the generalized Lagrangian function at $(x, \lambda, \Omega) \in \mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{S}^m_+$ by

$$L^g(x, \lambda, \Omega) := \lambda f(x) + \langle G(x), \Omega\rangle.$$

In general, we use iterative algorithms for solving the nonlinear semidefinite programming problem [25, 24, 38, 60, 42, 62]. These algorithms must be able to identify good solution candidates for (1). By "good candidates" we mean points that satisfy some necessary optimality condition. For instance, the *Fritz John* (FJ) necessary optimality condition states that whenever $\bar{x} \in \mathbb{R}^n$ is a local minimizer of (1), there exist so-called Fritz John multipliers $(\lambda, \Omega) \in \mathbb{R}_+ \times \mathbb{S}^m_+$, with $\lambda + \|\Omega\| \neq 0$, such that

$$\nabla_x L^g(\bar{x}, \lambda, \Omega) = 0 \text{ and } \langle \Omega, G(\bar{x})\rangle = 0. \tag{2}$$

Note that, in the particular case of $\lambda = 1$, the Fritz John condition coincides with the *Karush-Kuhn-Tucker* (KKT) condition for NLSDPs. Although Fritz John is a useful optimality condition, there are some disadvantages in its use, for instance, the multiplier associated with the objective function may vanish. In this case, the Fritz John condition would not be very informative, as it would not depend on the objective function. The possibility of the multiplier associated with the objective function vanishing can be avoided assuming that the *Mangasarian-Fromovitz constraint qualification* (MFCQ) holds (also called Robinson's CQ in the context of NLSDPs). We say that $\bar{x} \in \mathscr{F}$ satisfies MFCQ if there exists a vector $h \in \mathbb{R}^n$ such that $G(\bar{x}) + DG(\bar{x})h$ is negative definite [21]. Hence, the FJ optimality condition under the additional assumption that MFCQ holds implies that the KKT condition is satisfied. That is, under MFCQ, every Fritz John multiplier is also a Lagrange multiplier. In fact, one can prove that FJ is equivalent to the optimality condition "KKT or not-MFCQ". Most global convergence results for NLSDP methods assume that the problem satisfies MFCQ and proves that any feasible limit point of a sequence generated by the algorithm satisfies the KKT condition. That is, feasible limit points of sequences generated by it satisfy the FJ optimality condition. In this paper we will present an Augmented Lagrangian algorithm with global convergence to an optimality condition strictly stronger than FJ.

This optimality condition is the *Approximate-Karush-Kuhn-Tucker* (AKKT) condition, which was defined for NLSDPs in [4] and is well-known in nonlinear programming [5, 53]. As the FJ condition, AKKT is an optimality condition without the need of assuming a constraint qualification. Instead of considering the punctual verification of the KKT condition at a point $\bar{x}$, one allows its approximate verification at a sequence $x^k \to \bar{x}$ with a corresponding (possibly unbounded) dual sequence $\{\Omega^k\} \subset \mathbb{S}^m_+$.

The formal definition is as follows:

Definition 2.1. *(AKKT [4]) The feasible point $\bar{x} \in \mathscr{F}$ of* (1) *satisfies the Approximate-Karush-Kuhn-Tucker (AKKT) condition if there exist sequences $x^k \to \bar{x}$ and $\{\Omega^k\} \subset \mathbb{S}^m_+$ such that*

$$\nabla f(x^k) + DG(x^k)^* \Omega^k \to 0, \tag{3}$$

$$\lambda_i^U(G(\bar{x})) < 0 \Rightarrow \lambda_i^{S_k}(\Omega^k) = 0, \text{ for all } i = 1, \ldots, m \text{ and sufficiently large } k, \tag{4}$$

where $G(\bar{x}) = U\mathsf{diag}(\lambda_1^U(G(\bar{x})), \ldots, \lambda_m^U(G(\bar{x})))U^T$, $\Omega^k = S_k\mathsf{diag}(\lambda_1^{S_k}(\Omega^k), \ldots, \lambda_m^{S_k}(\Omega^k))S_k^T$, with U and S_k orthogonal matrices such that $S_k \to U$.

The importance of the AKKT condition in the nonlinear programming literature is due to the large amount of first- and second-order algorithms that generate AKKT sequences, thus extending global convergence results under constraint qualifications strictly weaker than MFCQ [2, 3, 6, 7, 8, 9, 10, 11, 12, 18, 15, 16, 17, 28, 33, 35, 36, 43, 47, 58].

The definition of AKKT for nonlinear semidefinite programming is not straightforward from the definition known in nonlinear programming. A key difference is that the approximate constraint matrix $G(x^k)$ and the Lagrange multiplier approximation Ω^k must be approximately simultaneously diagonalizable (this is the role of $S_k \to U$ in the definition), which is necessary for pairing the corresponding eigenvalues.

In order to explain the relevance of Definition 2.1, let us say that an AKKT sequence $\{x^k\}$ is known, converging to an AKKT point $\bar{x}$, with a corresponding dual sequence $\{\Omega^k\}$, generated by an algorithm. A usual global convergence result would be to say that $\bar{x}$ is a FJ point, by dividing (3) by $\|\Omega^k\|$ and taking the limit. The corresponding FJ multiplier (λ, Ω) may be such that $\lambda = 0$ provided that $\{\Omega^k\}$ is unbounded. In order to avoid this situation and arrive at a KKT point, one usually assumes that MFCQ holds. This is equivalent to saying that no FJ multiplier with $\lambda = 0$ may exist. By employing the AKKT condition, one may allow the existence of FJ multipliers with $\lambda = 0$ (that is, MFCQ may fail) while still being able to take the limit in (3) and proving that $\bar{x}$ is a KKT point (under a constraint qualification weaker than MFCQ). This is the case, for instance, when $\{\Omega^k\}$ is bounded. Or, even in the unbounded case, one may be able to rewrite (3-4) with a different sequence $\{\tilde{\Omega}^k\}$ in place of $\{\Omega^k\}$, allowing the proof that $\bar{x}$ is a KKT point. See [4]. Note that the FJ condition is meaningful with respect to the objective function (that is, implies KKT) only when the set of Lagrange multipliers is bounded (that is, MFCQ holds), while AKKT may imply KKT even when the set of Lagrange multipliers is unbounded. Note also that boundedness or unboundedness of the Lagrange multiplier set at $\bar{x}$ do not correspond, respectively, to the boundedness or unboundedness of the dual sequence $\{\Omega^k\}$, as this sequence may be bounded even when the set of Lagrange multipliers is unbounded (see, for instance, [22, 34]).

We summarize by noting that the following properties give the main reasons for using AKKT in our global convergence results:

Theorem 2.1. *([4]) Let $\bar{x}$ be a local minimizer of* (1) *. Then, $\bar{x}$ satisfies AKKT.*

Theorem 2.2. *([4]) Let $\bar{x}$ be a feasible point of* (1)*. If $\bar{x}$ satisfies AKKT, then $\bar{x}$ satisfies the FJ condition. The reciprocal is not true.*

Usually, when proving that feasible limit points of a sequence generated by an algorithm satisfy AKKT, the sequences $\{x^k\}$ and $\{\Omega^k\}$ are explicitly generated by the algorithm. Hence, one may safely use these sequences for stopping the execution of the algorithm. Given small tolerances $\varepsilon_{opt} > 0$, $\varepsilon_{feas} > 0$, $\varepsilon_{diag} > 0$ and $\varepsilon_{compl} > 0$, one may stop at iteration k when

$$\|\nabla f(x^k) + DG(x^k)^* \Omega^k\| \leq \varepsilon_{opt}, \tag{5}$$

$$\|[G(x^k)]_+\| \leq \varepsilon_{feas}, \tag{6}$$

$$\|U_k - S_k\| \leq \varepsilon_{diag}, \tag{7}$$

$$\lambda_i^{U_k}(G(x^k)) < -\varepsilon_{compl} \Rightarrow \lambda_i^{S_k}(\Omega^k) = 0, i = 1, \ldots, m. \tag{8}$$

For simplicitly, we say that x^k satisfies the ε-KKT stopping criterion when (5-8) hold with $\varepsilon_{opt} = \varepsilon_{feas} = \varepsilon_{diag} = \varepsilon_{compl} = \varepsilon \geq 0$. In this case, the point x^k together with a corresponding Ω^k, U_k and S_k will be called an ε-KKT point. A point that satisfies the stopping criterion with $\varepsilon = 0$ is precisely a KKT point, see [4]. Note that, from the compactness of the set of orthogonal matrices, we can take, if necessary, a subsequence of the orthogonal matrices sequence $\{U_k\}$ that converges to U and define it as the original sequence. Thus, the limit $\|U_k - S_k\| \to 0$ implies that $U_k \to U$ and $S_k \to U$. Since $\{\Omega^k\}$ does not need to have a convergent subsequence, the inequality given by (7) expresses only a correspondence relation for pairing the eigenvalues of $G(x^k)$ and Ω^k.

We refer the reader to [4] for a detailed measure of the strength of AKKT, in comparison with FJ, together with a characterization of the situation in which an AKKT point is in fact a KKT point. Also, we note that the developments of this paper could be done with a different optimality condition presented in [4] (Trace-AKKT), which gives a simpler stopping criterion than (5-8); however, an additional smoothness assumption on the function G would be needed, which we have decided to avoid for simplicity.

3 An Augmented Lagrangian algorithm

Let us consider the following problem

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{Minimize}} & f(x), \\ \text{subject to} & x \in \mathscr{F}_u \cap \mathscr{F}_l, \end{array} \tag{9}$$

where $\mathscr{F}_u := \{x \in \mathbb{R}^n | \; G_1(x) \preceq 0\}$ is the set of constraints in the upper-level and $\mathscr{F}_l := \{x \in \mathbb{R}^n | \; G_2(x) \preceq 0\}$ the set of constraints in the lower-level. We assume that $f : \mathbb{R}^n \to \mathbb{R}$, $G_1 : \mathbb{R}^n \to \mathbb{S}^m$ and $G_2 : \mathbb{R}^n \to \mathbb{S}^{\overline{m}}$ are all continuously differentiable functions. The choice of which constraints will go to the upper-level and which ones will stay in the lower-level can be done arbitrarily. Nevertheless, the two-level minimization technique is used, in general, when some constraints are somewhat "easier" than the others. These constraints should be kept at the lower-level, which will not be penalized. On the other hand, the difficult constraints will be maintained at the upper-level, that is, they will be penalized with the use of the Powell-Hestenes-Rockafellar (PHR) Augmented Lagrangian function for NLSDPs. Since we do not restrict ourselves to which class of constraints will be kept at the lower level and which ones will be penalized, we will be considering a large class of Augmented Lagrangian algorithms. More details of this technique applied to nonlinear optimization problems can be found in [1, 18, 16]. If the problem has, in addition, equality constraints, they may also be arbitrarily split in the upper and lower level sets and the algorithm and its global convergence proof can be carried out with the necessary modifications. In particular, this is the technique we used in our numerical experiments regarding the covering problem (more details in Section 5). Alternatively, one may incorporate an equality constraint $h(x) = 0$ as the diagonal block $\begin{pmatrix} h(x) & 0 \\ 0 & -h(x) \end{pmatrix} \preceq 0$. Note that the independency of constraint qualifications means that the optimality condition AKKT is valid regardless of the formulation, even though the sequences of multipliers may not be the same. Also, the FJ optimality condition becomes trivially satisfied at all feasible points when an equality constraint is treated as a negative semidefinite block matrix [4]. Next, we will present an Augmented Lagrangian method for NLSDPs based on the quadratic penalty function. This is an extension of the algorithm defined in [4], where only unconstrained subproblems were considered. For more details about the Augmented Lagrangian method for NLSDPs, see [1, 25, 24, 38, 60, 55, 56, 62].

Let us take a penalty parameter $\rho > 0$. The Augmented Lagrangian function $L_\rho : \mathbb{R}^n \times \mathbb{S}^m \to \mathbb{R}$, a natural extension of the Augmented Lagrangian function for nonlinear programming, penalizes the constraints at the upper-level and is defined below:

$$L_\rho(x, \Omega) := f(x) + \frac{1}{2\rho} \left\{ \|[\Omega + \rho G_1(x)]_+\|^2 - \|\Omega\|^2 \right\}. \tag{10}$$

The gradient with respect to x of the Augmented Lagrangian function above is given by:

$$\nabla_x L_\rho(x,\Omega) = \nabla f(x) + DG_1(x)^*\left[\Omega + \rho G_1(x)\right]_+.$$

The formal statement of the algorithm is as follows:

Algorithm 1 Augmented Lagrangian Algorithm

Let $\tau \in (0,1)$, $\gamma > 1$, $\rho_1 > 0$ and $\Omega^{\max} \in \mathbb{S}^m_+$. Take a sequence of tolerances $\{\varepsilon_k\} \subset \mathbb{R}_+$. Define $0 \preceq \bar{\Omega}^1 \preceq \Omega^{\max}$. Choose an arbitrary starting point $x^0 \in \mathbb{R}^n$. Initialize $k := 1$ and $\|V^0\| := +\infty$.

Step 1: Find an ε_k-KKT point x^k to the problem

$$\text{Minimize } L_{\rho_k}(x,\bar{\Omega}^k) \text{ subject to } x \in \mathscr{F}_l. \tag{11}$$

Step 2: Define

$$V^k := \left[\frac{\bar{\Omega}^k}{\rho_k} + G_1(x^k)\right]_+ - \frac{\bar{\Omega}^k}{\rho_k}.$$

If $\|V^k\| \leq \tau\|V^{k-1}\|$, define $\rho_{k+1} := \rho_k$, otherwise define $\rho_{k+1} := \gamma\rho_k$.
Step 3: Compute

$$\Omega^k := \left[\bar{\Omega}^k + \rho_k G_1(x^k)\right]_+,$$

and define $\bar{\Omega}^{k+1} := \mathsf{Proj}(\Omega^k)$ as the orthogonal projection of Ω^k onto $\{X \in \mathbb{S}^m | 0 \preceq X \preceq \Omega^{\max}\}$.
Step 4: Set $k := k+1$, and go back to Step 1.

Note that, in Step 3 of the algoritm above, we take the Lagrange multiplier approximation for the next subproblem as the projection of the natural update of Ω^k onto a "safeguarded box" as in [18]. This means that when the penalty parameter is too large, the method reduces to the external penalty method. The penalty parameter is updated using the joint measure of feasibility and complementarity for the conic constraint defined by V^k. Note that $V^k = 0$ implies the sequence x^k is feasible and the complementarity condition holds. Therefore, if $\|V^k\|$ is sufficiently reduced, the algorithm keeps the previous penalty parameter unchanged and increases it otherwise. In Step 1, we assume that an unspecified algorithm for solving the subproblem (11) returns an ε_k-KKT point at each iteration k, where ε_k is the tolerance for stopping the execution of the subproblem solver. Taking into account our global convergence results, a natural stopping criterion for Algorithm 1 is to stop when x^k and the corresponding dual approximations satisfy ε-KKT for the original problem (9), for some $\varepsilon > 0$.

Let us now show that the algorithm tends to find feasible points. Namely, the following result shows that Algorithm 1 finds AKKT points of an infeasibility measure of the upper-level constraints restricted to the lower-level constraints.

Theorem 3.1. *Let $\varepsilon_k \to 0^+$. Assume that $\bar{x} \in \mathbb{R}^n$ is a limit point of a sequence $\{x^k\}$ generated by Algorithm 1. Then, $\bar{x}$ is an AKKT point of the infeasibility optimization problem*

$$\underset{x\in\mathbb{R}^n}{\textit{Minimize}} \quad P(x) := \|[G_1(x)]_+\|^2 \textit{ subject to } x \in \mathscr{F}_l. \tag{12}$$

Proof. Since x^k is an ε_k-KKT sequence, from (11) we have that

$$\|\nabla_x L_{\rho_k}(x^k,\Omega^k) + DG_2(x^k)^*\underline{\Omega}^k\| \leq \varepsilon_k, \tag{13}$$

$$\|[G_2(x^k)]_+\| \leq \varepsilon_k, \tag{14}$$

$$\|U_k - S_k\| \leq \varepsilon_k, \tag{15}$$

$$\lambda_i^{U_k}(G_2(x^k)) \leq -\varepsilon_k \Rightarrow \lambda_i^{S_k}(\underline{\Omega}^k) = 0. \tag{16}$$

where U_k and S_k are orthogonal matrices that diagonalize $G_2(x^k)$ and $\underline{\Omega}^k$, respectively. From (14) we have that $\bar{x} \in \mathscr{F}_l$. Now, let us consider two cases: the first when $\{\rho_k\}$ is unbounded and the second when $\{\rho_k\}$ is bounded.

(i) If the penalty parameter $\{\rho_k\}$ is unbounded, let us define

$$\delta^k := \nabla f(x^k) + DG_1(x^k)^*[\bar{\Omega}^k + \rho_k G_1(x^k)]_+ + DG_2(x^k)^*\underline{\Omega}^k.$$

From (13) we have $\|\delta^k\| \leq \varepsilon_k$. Dividing δ^k by ρ_k,

$$\frac{\delta^k}{\rho_k} = \frac{\nabla f(x^k)}{\rho_k} + DG_1(x^k)^*\left[\frac{\bar{\Omega}^k}{\rho_k} + G_1(x^k)\right]_+ + DG_2(x^k)^*\frac{\underline{\Omega}^k}{\rho_k},$$

since $\frac{\delta^k}{\rho_k} \to 0$ and $\nabla f(x^k) \to \nabla f(\bar{x})$, we have that

$$DG_1(x^k)^* \Omega_1^k + DG_2(x^k)^* \Omega_2^k \to 0$$

where, $\Omega_1^k := \left[\frac{\bar{\Omega}^k}{\rho_k} + G_1(x^k)\right]_+$ and $\Omega_2^k := \frac{\underline{\Omega}^k}{\rho_k}$. Furthermore, from (15) and from the compactness of the orthogonal matrices set we can take a subsequence $\{U_k\}_{k\in\mathbb{N}}$ that converges to U and define it as the original sequence such that $S_k \to U$ and $U_k \to U$. Then, if $\lambda_i^U(G_2(\bar{x})) < 0$ we have that $\lambda_i^{U_k}(G_2(x^k)) < -\varepsilon_k$ for all sufficiently large k; which implies, from (16), that $\lambda_i^{S_k}(\Omega_2^k) = 0$.

(ii) If the penalty parameter $\{\rho_k\}$ is bounded, that is, for $k \geq k_0$ the penalty parameter remains unchanged and then $V^k \to 0$. The sequence $\{\bar{\Omega}^k\}$ is bounded, then, there is an $\bar{\Omega} \in \mathbb{S}_+^m$ such that

$$\bar{\Omega} = \lim \bar{\Omega}^k = \lim \left[\bar{\Omega}^k + \rho_{k_0} G_1(x^k)\right]_+ = \lim \Omega^k = \left[\bar{\Omega} + \rho_{k_0} G_1(\bar{x})\right]_+.$$

Since $\bar{\Omega} \in \mathbb{S}_+^m$, we can take a diagonalization

$$\bar{\Omega} = [\bar{\Omega} + \rho_{k_0} G_1(\bar{x})]_+ = R\mathsf{diag}([\lambda_1^R]_+, \ldots, [\lambda_m^R]_+)R^T,$$

with $RR^T = I$. Moreover,

$$\bar{\Omega} + \rho_{k_0} G_1(\bar{x}) = R\mathsf{diag}(\lambda_1^R, \ldots, \lambda_m^R)R^T.$$

In this way,

$$G_1(\bar{x}) = (1/\rho_{k_0})R\mathsf{diag}((\lambda_1^R - [\lambda_1^R]_+), \ldots, (\lambda_m^R - [\lambda_m^R]_+))R^T,$$

thus, $G_1(\bar{x}) \preceq 0$ and since $\bar{x} \in \mathscr{F}_l$ it follows that $\bar{x}$ is a global minimizer of the optimization problem (12) and, in particular, an AKKT point.

□

Let us now show that if, in addition, the limit point $\bar{x}$ is feasible for (9), then $\bar{x}$ is in addition an AKKT point for (9). In fact, $\{x^k\}$ is an associated AKKT sequence and $\{\Omega^k\}$ is its corresponding dual sequence.

Theorem 3.2. *Let $\varepsilon_k \to 0^+$. Assume that $\bar{x} \in \mathscr{F}_u \cap \mathscr{F}_l$ is a feasible limit point of a sequence $\{x^k\}$ generated by Algorithm 1. Then, $\bar{x}$ is an AKKT point of problem* (9)*.*

Proof. Let $\bar{x} \in \mathscr{F}_u \cap \mathscr{F}_l$ be a limit point of a sequence $\{x^k\}$ generated by Algorithm 1. Let us assume without loss of generality that $x^k \to \bar{x}$. From Step 1 of the algorithm we have that

$$\lim_{k\to\infty} \nabla f(x^k) + DG_1(x^k)^* \Omega^k + DG_2(x^k)^* \underline{\Omega}^k = 0,$$

where

$$\Omega^k := \left[\bar{\Omega}^k + \rho_k G_1(x^k)\right]_+.$$

Now, let us prove the complementarity condition in AKKT, both for the upper-level and the lower-level constraints. For this, we will use appropriate orthogonal matrices S_k, R_k, U_k, Z_k, R and U which diagonalize $\underline{\Omega}^k$, Ω^k, $G_2(x^k)$, $G_1(x^k)$, $G_1(\bar{x})$ and $G_2(\bar{x})$ respectively such that $S_k \to U$ and $R_k \to R$. Note that, similarly to what was done in Theorem 3.1, x^k is an ε_k-KKT point for problem (11). Then, by Step 2 of Algorithm 1 we get that

$$\|\nabla_x L_{\rho_k}(x^k, \Omega^k) + DG_2(x^k)^* \underline{\Omega}^k\| \leq \varepsilon_k, \tag{17}$$

$$\|[G_2(x^k)]_+\| \leq \varepsilon_k, \tag{18}$$

$$\|U_k - S_k\| \leq \varepsilon_k, \tag{19}$$

$$\lambda_i^{U_k}(G_2(x^k)) \leq -\varepsilon_k \Rightarrow \lambda_i^{S_k}(\underline{\Omega}^k) = 0. \tag{20}$$

(i) If $\rho_k \to +\infty$, for the upper-level constrants, since the sequence $\{\bar{\Omega}^k\}$ is bounded, we have that $\frac{\bar{\Omega}^k}{\rho_k} + G_1(x^k) \to G_1(\bar{x})$.

Let us take a diagonalization

$$\frac{\bar{\Omega}^k}{\rho_k} + G_1(x^k) = R_k\mathsf{diag}(\lambda_1^k, \ldots, \lambda_m^k)R_k^T.$$

Taking a subsequence if necessary, let us take a diagonalization

$$G_1(\bar{x}) = R\mathsf{diag}(\lambda_1,\ldots,\lambda_m)R^T,$$

where $R_k \to R$, $\lambda_i^k \to \lambda_i$ for all i. Then,

$$\Omega^k = [\bar{\Omega}^k + \rho_k G_1(x^k)]_+ = R_k\mathsf{diag}(\rho_k[\lambda_1^k]_+,\ldots,\rho_k[\lambda_m^k]_+)R_k^T.$$

Now, assume that $\lambda_i^R(G_1(\bar{x})) = \lambda_i < 0$. Then, $\lambda_i^k < 0$ for all sufficiently large k; which implies that $\lambda_i^{R_k}(\Omega^k) = \rho_k[\lambda_i^k]_+ = 0$. Furthermore, from (19) and from the compactness of the orthogonal matrices set we can take a subsequence $\{U_k\}_{k\in\mathbb{N}}$ that converges to U and define it as the original sequence such that $S_k \to U$ and $U_k \to U$. For the lower-level constraints note that, if $\lambda_i^U(G_2(\bar{x})) < 0$ we have $\lambda_i^{U_k}(G_2(x^k)) < -\varepsilon_k$ for k large enough and then by (20) it follows that $\lambda_i^{S_k}(\underline{\Omega}^k) = 0$.

(ii) If $\{\rho_k\}$ is bounded, for the upper-level constraints, analogously to what was done in Theorem 3.1, for $k \geq k_0$ we have that $V^k \to 0$. Thus,

$$\bar{\Omega} = \lim\bar{\Omega}^k = \lim\left[\bar{\Omega}^k + \rho_{k_0}G_1(x^k)\right]_+ = \lim\Omega^k = \left[\bar{\Omega} + \rho_{k_0}G_1(\bar{x})\right]_+.$$

Writing the orthogonal decomposition of the matrix Ω^k we have

$$\Omega^k = [\bar{\Omega}^k + \rho_{k_0}G_1(x^k)]_+ = R_k\mathsf{diag}([\lambda_1^{R_k}]_+,\ldots,[\lambda_m^{R_k}]_+)R_k^T,$$

and taking a subsequence if necessary

$$\bar{\Omega} = [\bar{\Omega} + \rho_{k_0}G_1(\bar{x})]_+ = R\mathsf{diag}([\lambda_1^R]_+,\ldots,[\lambda_m^R]_+)R^T,$$

where $R_k \to R$, $\lambda_i^{R_k} \to \lambda_i^R$ for all i . Moreover,

$$\bar{\Omega} + \rho_{k_0}G_1(\bar{x}) = R\mathsf{diag}(\lambda_1^R,\ldots,\lambda_m^R)R^T.$$

In this way,

$$G_1(\bar{x}) = (1/\rho_{k_0})R\mathsf{diag}((\lambda_1^R - [\lambda_1^R]_+),\ldots,(\lambda_m^R - [\lambda_m^R]_+))R^T,$$

and then, $\lambda_i^R(G_1(\bar{x})) = \dfrac{\lambda_i^R - [\lambda_i^R]_+}{\rho_{k_0}}$. If $\lambda_i^R(G_1(\bar{x})) < 0$, we have that $\lambda_i^R < [\lambda_i^R]_+$; hence, $[\lambda_i^R]_+ = 0$. Then, $\lambda_i^{R_k} < 0$ for all sufficiently large k; which implies that $\lambda_i^{R_k}(\Omega^k) = [\lambda_i^{R_k}]_+ = 0$. Furthermore, from (19) and from the compactness of the orthogonal matrices set we can take a subsequence $\{U_k\}_{k\in\mathbb{N}}$ that converges to U and define it as the original sequence such that $S_k \to U$ and $U_k \to U$. Now, for the lower-level constraints note that, if $\lambda_i^U(G_2(\bar{x})) < 0$ then $\lambda_i^{U_k}(G_2(x^k)) < -\varepsilon_k$ for k large enough; which implies by (20), $\lambda_i^{S_k}(\underline{\Omega}^k) = 0$.

□

Note that for proving the first theorem, concerning feasibility of limit points, one could relax the condition in (13) by replacing $\{\varepsilon_k\}$ by a bounded sequence $\{\tilde{\varepsilon}_k\}$, not necessarily converging to zero, while keeping $\varepsilon_k \to 0^+$ in conditions (14-16). This allows including, without further analysis, an adaptive computation of $\tilde{\varepsilon}_k := \varphi(\|[G_1(x^{k-1})]_+\|)$, for some $\varphi(t) \to 0^+$ as $t \to 0^+$.

4 The covering problem

In this section we present a new approach to the so-called covering problem. It consists of computing the smallest collection of m closed balls with common radius, such that their union contains a given compact set.

The problem will be an NLSDP of the following form:

$$\begin{array}{ll} \text{Minimize} & f(r,c_1,\ldots,c_m,S_1,\ldots,S_p), \\ \text{subject to} & H(r,c_1,\ldots,c_m,S_1,\ldots,S_p) = 0, \\ & r \geq 0, \quad S_i \succeq 0, \quad i = 1,\ldots,p, \end{array} \tag{21}$$

where $r \in \mathbb{R}, c_j \in \mathbb{R}^n, j = 1,\ldots,m$, $S_i \in \mathbb{S}^N, i = 1,\ldots,p$ and f and H are continuously differentiable functions, where H takes values on $\mathbb{R}^W$.

The nonlinear equality constraints defined by H will be penalized (upper-level constraints) while the non-negative constraints $r \geq 0$ and the conic ones $S_i \succeq 0, i = 1, \ldots, p$, will be kept as constraints in the subproblems (lower-level constraints). Hence, since the projection of $(r, S_1, \ldots, S_p)$ onto the lower-level feasible set can be computed from a spectral decomposition of each variable, namely, $([r]_+, [S_1]_+, \ldots, [S_p]_+)$, we will solve each subproblem by the Spectral Projected Gradient (SPG) method [19, 20]. In the nonlinear programming case, an Augmented Lagrangian method where subproblems are solved by the SPG method was previously considered in [27]. We start by describing our model of the covering problem, by means of convex algebraic geometry tools, which is interesting in its own in the sense that it allows dealing with this difficult problem in its full generality, without resorting to discretizations.

Let $V \subset \mathbb{R}^n$ be a compact set to be covered with m balls. Let $c_1, \ldots, c_m \in \mathbb{R}^n$ be the centers and $\sqrt{r}$ the common radius of the balls, to be determined, where $r \geqslant 0$. We use the notation $\mathbb{R}[\mathbf{x}] := \mathbb{R}[x_1, \ldots, x_n]$ for the ring of real polynomials in n variables. The i-th closed ball will be denoted by

$$B_i[c_i, r] := \{x \in \mathbb{R}^n \mid p_i(x, c_i, r) \leqslant 0\},$$

where $p_i(x, c_i, r) := \|x - c_i\|_2^2 - r \in \mathbb{R}[\mathbf{x}]$, for $i = 1, \ldots, m$, and their union will be denoted by $\mathscr{B}[c_1, \ldots, c_m, r]$. More specifically, we intend to compute a solution to the following problem:

$$\begin{array}{ll} \underset{c_i \in \mathbb{R}^n, r \in \mathbb{R}}{\text{Minimize}} & r, \\ \text{subject to} & V \subseteq \mathscr{B}[c_1, \ldots, c_m, r], \\ & r \geqslant 0. \end{array} \tag{22}$$

In order to rewrite this problem in a tractable form, we assume that there exists a set of v polynomials, namely $\{g_1, \ldots, g_v\} \subset \mathbb{R}[\mathbf{x}]$, such that

$$V = \{x \in \mathbb{R}^n \mid g_1(x) \geqslant 0, \ldots, g_v(x) \geqslant 0\},$$

that is, V is a *closed basic semialgebraic set*. Then $V \subseteq \mathscr{B}[c_1, \ldots, c_m, r]$ if, and only if,

$$\mathscr{K}[c_1, \ldots, c_m, r] := \{x \in \mathbb{R}^n \mid p_i(x) > 0, g_j(x) \geqslant 0, \forall i \in \{1, \ldots, m\}, \forall j \in \{1, \ldots, v\}\} = \emptyset.$$

The set $\mathscr{K}[c_1, \ldots, c_m, r]$ represents the region of V not covered by $\mathscr{B}[c_1, \ldots, c_m, r]$. Note that the inequalities $p_i(x) > 0$ can be replaced by $p_i(x) \geqslant 0$ and $\overline{p}(x) := \prod_{i=1}^m p_i(x) \neq 0$. To simplify the notation, let us define $q_i := p_i$, for $i = 1, \ldots, m$, and $q_{m+i} = g_i$, for $i = 1, \ldots, v$, with $\eta := m + v$. Hence,

$$\mathscr{K}[c_1, \ldots, c_m, r] = \{x \in \mathbb{R}^n \mid q_i(x) \geqslant 0, \overline{p}(x) \neq 0, \forall i \in \{1, \ldots, \eta\}\}. \tag{23}$$

Then, we can rewrite (22) as

$$\begin{array}{ll} \underset{c_i \in \mathbb{R}^n, r \in \mathbb{R}}{\text{Minimize}} & r, \\ \text{subject to} & \mathscr{K}[c_1, \ldots, c_m, r] = \emptyset, \\ & r \geqslant 0. \end{array} \tag{24}$$

At first glance, formulation (24) may not seem so pleasant, but fortunately there are several results from the last few decades which manage to give infeasibility certificates for *basic semialgebraic sets*, in particular, for $\mathscr{K}[c_1, \ldots, c_m, r]$. Before introducing them, we recall two redundancy sets for the solutions of systems of polynomial inequalities. The set of all polynomials in n variables of degree at most $2d$ that can be written as a sum of squares (s.o.s.) of other polynomials will be denoted by $\Sigma_{n,2d}$, and $\Sigma_n := \bigcup_{d \in \mathbb{N}} \Sigma_{n,2d}$.

Definition 4.1. *The quadratic module generated by $q_1, \ldots, q_\eta$ is the set*

$$\mathsf{qmodule}(q_1, \ldots, q_\eta) := \{s_0 + q_1 s_1 + \cdots + q_\eta s_\eta \mid s_i \in \Sigma_n, \forall i \in \{1, \ldots, \eta\}\}.$$

Definition 4.2. *The preorder generated by $q_1, \ldots, q_\eta$ is the set*

$$\mathsf{preorder}(q_1, \ldots, q_\eta) := \{s_0 + \sum_i q_i s_i + \sum_{i \neq j} q_i q_j s_{ij} + \sum_{i \neq j \neq k} q_i q_j q_k s_{ijk} + \ldots\},$$

where each of the 2^η terms is free of squares and s_i, s_{ij} etc. are all s.o.s. of polynomials. To put it briefly,

$$\mathsf{preorder}(q_1, \ldots, q_\eta) = \left\{ \sum_{\Lambda \subseteq \{1, \ldots, \eta\}} (s_\Lambda \prod_{i \in \Lambda} q_i) \mid s_\Lambda \in \Sigma_n, \forall \Lambda \subseteq \{1, \ldots, \eta\} \right\}.$$

The next result, introduced in [54], provides infeasibility certificates for $\mathscr{K}[c_1,\ldots,c_m,r]$.

Theorem 4.1 (Stengle's Positivstellensatz)**.** *The set $\mathscr{K}[c_1,\ldots,c_m,r]$ is empty if, and only if, there exist $P \in \mathsf{preorder}(q_1,\ldots,q_\eta)$ and an integer $b \geq 0$ such that*

$$P + \overline{p}^{2b} = 0.$$

In [23], it was proved that deciding whether a polynomial is an s.o.s. or not is intrinsically related to the possibility of finding a positive semidefinite matrix that represents it, in the following sense:

Theorem 4.2 (Gram representation)**.** *Define $N := \binom{n+d}{d}$. A real polynomial s of degree $2d$ is an s.o.s. polynomial if, and only if, there exists a real symmetric $N \times N$ (Gram) matrix $S \succeq 0$ such that*

$$s = [\boldsymbol{x}]_d^T S [\boldsymbol{x}]_d,$$

where $[\boldsymbol{x}]_d := (1, x_1, x_2, \ldots, x_n, x_1x_2, \ldots, x_n^d)^T$ is a (lexicographically) ordered array containing every monomial of degree at most d.

Thus, based on [51], for every fixed degree $d > 0$ and $b \geq 0$, (24) can be written in the form

$$\begin{array}{ll} \underset{c_i \in \mathbb{R}^n, r \in \mathbb{R}}{\text{Minimize}} & r, \\ \text{subject to} & \mathsf{coeffs}\left(\overline{p}^{2b} + \displaystyle\sum_{\Lambda \subseteq \{1,\ldots,\eta\}} [\mathbf{x}]_d^T S_\Lambda [\mathbf{x}]_d \prod_{i \in \Lambda} q_i\right) = 0, \\ & S_\Lambda \in \mathbb{S}_+^N, \forall \Lambda \subseteq \{1,\ldots,\eta\}, \\ & r \geqslant 0, \end{array} \tag{25}$$

where $\mathsf{coeffs}(p)$ represents the vector of coefficients of the polynomial p. Formulation (25) is an NLSDP with 2^η matrix variables and, even if the fixed degrees do not lead to a *Positivstellensatz* certificate of emptyness for every feasible covering, they work for sufficiently large ones and thus bound the quality of the solution. In a similar fashion of Lasserre hierarchy for s.o.s. optimization, one can solve (25) starting from $d = 1$ and then increasing it until the desired accuracy is obtained. This procedure is expected to stop in a few levels, even though the known bounds for d are unpracticably large. Technical details can be found in [39, 40].

Although the formulation presented above is the most desired one in terms of corresponding exactly to the true covering problem, it involves a very large number of variables and constraints. The difficulty appears due to the presence of the unusual requirement $\overline{p}(x) \neq 0$ in the definition of the set $\mathscr{K}[c_1,\ldots,c_m,r]$ in (23). In our experiments, we will deal with a simplified version of the problem where this requirement is dropped. The corresponding model is then much simpler. This situation corresponds to the problem of covering a compact set with open balls. That is, let us consider the problem of finding a minimal covering for V made of open balls with common radius, namely, for

$$B_i(c_i,r) := \{x \in \mathbb{R}^n \mid p_i(x,c_i,r) < 0\}, i = 1,\ldots,m,$$

their union will be denoted by $\mathscr{B}(c_1,\ldots,c_m,r)$. Then $V \subseteq \mathscr{B}(c_1,\ldots,c_m,r)$ if, and only if, the set

$$\mathscr{K}(c_1,\ldots,c_m,r) := \{x \in \mathbb{R}^n \mid q_i(x) \geqslant 0, \forall i \in \{1,\ldots,\eta\}\} \tag{26}$$

is empty. Of course, the problem

$$\begin{array}{ll} \underset{c_i \in \mathbb{R}^n, r \in \mathbb{R}}{\text{Minimize}} & r, \\ \text{subject to} & \mathscr{K}(c_1,\ldots,c_m,r) = \emptyset, \\ & r \geqslant 0, \end{array} \tag{27}$$

has no solution, but every strictly feasible covering of (24) is also feasible in (27). There is a simplified *Positivstellensatz* for this case, assuming an additional hypothesis over $\mathscr{K}(c_1,\ldots,c_m,r)$. The set $\mathsf{qmodule}(q_1,\ldots,q_\eta)$ is said to be *archimedean* if there is a polynomial $p \in \mathsf{qmodule}(q_1,\ldots,q_\eta)$ such that $\{x \in \mathbb{R}^n \mid p(x) \geqslant 0\}$ is compact or, equivalently, if there exists $M > 0$ such that $b_M(x) := M - \|x\|_2^2 \in \mathsf{qmodule}(q_1,\ldots,q_\eta)$.

Theorem 4.3 (Putinar's Positivstellensatz)**.** *Given $p \in \mathbb{R}[\boldsymbol{x}]$, suppose that $\mathscr{K}(c_1,\ldots,c_m,r)$ is a compact set and that $\mathsf{qmodule}(q_1,\ldots,q_\eta)$ is archimedean. If $p(x)$ is positive on $\mathscr{K}(c_1,\ldots,c_m,r)$, then $p \in \mathsf{qmodule}(q_1,\ldots,q_\eta)$.*

See [52] for details.

Corollary 4.1. *Suppose that $\mathscr{K}(c_1,\ldots,c_m,r)$ is a compact set. Then $\mathscr{K}(c_1,\ldots,c_m,r)$ is empty if, and only if, $-1 \in \mathsf{qmodule}(q_1,\ldots,q_\eta,b_M)$, for some $M>0$ such that $\mathscr{K}(c_1,\ldots,c_m,r) \subseteq \{x \in \mathbb{R} \mid b_M(x) \geqslant 0\}$.*

Proof. If there are $s_0,\ldots,s_\eta,s_M \in \Sigma_n$ such that $s_0+q_1s_1+\cdots+q_\eta s_\eta+b_Ms_M=-1$, then $\mathscr{K}(c_1,\ldots,c_m,r)$ must be empty. Otherwise, for any $x \in \mathscr{K}(c_1,\ldots,c_m,r)$,

$$0 \leqslant s_0(x)+q_1(x)s_1(x)+\cdots+q_\eta(x)s_\eta(x)+b_M(x)s_M(x)<0$$

would hold, as $\mathscr{K}(c_1,\ldots,c_m,r)=\{x \in \mathbb{R}^n \mid b_M(x) \geqslant 0, q_i(x) \geqslant 0, \forall i \in \{1,\ldots,\eta\}\}$. Conversely, if $\mathscr{K}(c_1,\ldots,c_m,r)$ is empty, the polynomial $p: x \mapsto p(x)=-1$ is strictly positive on $\mathscr{K}(c_1,\ldots,c_m,r)$; this is a vacuous truth. Hence $-1 \in \mathsf{qmodule}(q_1,\ldots,q_\eta,b_M)$, since $\mathsf{qmodule}(q_1,\ldots,q_\eta,b_M)$ is archimedean. □

When V is defined by at least one polynomial g_i such that $\{x \in \mathbb{R}^n \mid g_i(x) \geqslant 0\}$ is compact; for example, when V is a closed ball or a solid heart-shaped figure, then $\mathscr{K}(c_1,\ldots,c_m,r)$ is empty if, and only if, $-1 \in \mathsf{qmodule}(q_1,\ldots,q_\eta)$. Moreover, similarly to Stengle's theorem, Putinar's result leads to a reformulation of (27) as an NLSDP of the following form

$$\begin{array}{ll} \underset{c_i \in \mathbb{R}^n, r \in \mathbb{R}}{\text{Minimize}} & r, \\ \text{subject to} & \mathsf{coeffs}(s_0+s_1q_1+\cdots+s_\eta q_\eta+1)=0, \\ & S_i \in \mathbb{S}^N_+, \forall i \in \{0,\ldots,\eta\}, \\ & r \geqslant 0, \end{array} \tag{28}$$

where each polynomial s_i, with $i \in \{0,\ldots,\eta\}$, is to be understood as the polynomial $s_i(x) := [\mathbf{x}]_d^T S_i [\mathbf{x}]_d$. The problem has $\eta+1$ matrix variables, which is much better than 2^η from (25). We expected that an algorithm that aims at solving (28) would find a good approximation to a solution of (25), hence this is the formulation we use in our numerical experiments.

5 Numerical experiments

In this section, we give some implementation details for the application to the covering problem. We build the model based on Putinar's Positivstellensatz described in (28) using the MATLAB package YALMIP (more details in [41]). It has a structure made to represent semidefinite variables symbolically, called *sdpvar*, and a function to extract coefficients of polynomials even when they are symbolic. This is useful because the polynomial variables $x_1,\ldots,x_n$ are only used for ordering and grouping the problem variables. In general, finding a closed expression for each coefficient of the polynomials defined by the Positivstellensatz is not a simple task; also, we make use of a gradient-based (first order) method, which requires the partial derivatives of each coefficient with respect to the relevant variables, which we present below.

Define $X_{2d} := [\mathbf{x}]_d [\mathbf{x}]_d^t$, then the partial derivatives of the coefficients of Putinar's polynomial are given as follows:

- Derivative with respect to the radius:

$$\frac{\partial}{\partial r}\mathsf{coeffs}(s_0+s_1q_1+\cdots+s_\eta q_\eta+1)=\mathsf{coeffs}(-\sum_{k=1}^m \langle S_k, X_{2d}\rangle).$$

- Derivatives with respect to the centers:

$$\frac{\partial}{\partial c_{ij}}\mathsf{coeffs}(s_0+s_1q_1+\cdots+s_\eta q_\eta+1)=\mathsf{coeffs}(2(c_{ij}-x_j)\langle S_i, X_{2d}\rangle),$$

  for all $(i,j) \in \{1,\ldots,m\} \times \{1,\ldots,n\}$.

- Derivatives with respect to the Gram matrices:

$$\frac{\partial}{\partial S_k^{(i,j)}}\mathsf{coeffs}(s_0+s_1q_1+\cdots+s_\eta q_\eta+1)=\mathsf{coeffs}(q_k X_{2d}^{(i,j)}),$$

  for all $k \in \{0,\ldots,\eta\}$ and $(i,j) \in \{1,\ldots,N\} \times \{1,\ldots,N\}$, with $g_0 := 1$.

The extracted coefficients are then written in the form $H(y)$ and its jacobian matrix $J_H(y)$, where $y := (r,c,S)^T$ is a vector of variables that stores r, c_{ij} and $S_k^{(i,j)}$ in array form.

Next, we give some general numbers. Recall that $N = \binom{n+d}{d}$. Then, for the Putinar model, we have $\frac{N(N+1)}{2}(\eta+1) + nm+1$ real variables and one integer parameter d, the degree of Putinar's polynomial is $d_P := 2d + \max_{i\in\{1,\dots,\eta\}}\{\deg(q_i)\}$, so the number of nonlinear equality constraints is $\binom{n+d_P}{d_P}$. Note that only $mn+1$ variables are geometrically relevant (c and r), while the remaining ones (S) are feasibility certificates.

The nonlinear constraints $H(y) = 0$ will be penalized, while the (convex) semidefinite constraints will be kept within the subproblems, that is, at each iteration, we solve the problem

$$\begin{array}{ll} \underset{y\in\mathbb{R}^N}{\text{Minimize}} & L_{\rho^k}(y,\lambda^k) := f(y) + \dfrac{1}{2\rho^k}\left\{\|\lambda^k + \rho^k H(y)\|^2\right\}, \\ \text{subject to} & S_i \succeq 0, \forall i \in \{0,\dots,\eta\}, \\ & r \geqslant 0, \end{array} \tag{29}$$

where $y = (r,c,S)^T$ and $f(y) = r$. We employ the Spectral Projected Gradient (SPG) method [19] for solving the Augmented Lagrangian subproblems (29) with its recommended parameters.

Let us now describe some details about the implementation.

1. **Modelling part:** The inputs are the dimension n, the number of balls m, half of the fixed degree for the s.o.s. polynomials d, and the chosen version of the Positivstellensatz; in case of Stengle's, choosing a degree b is also required. The code then writes the functions $H(y)$ and $J_H(y)$ in the format of a Fortran subroutine supported by the Fortran implementation [1] of the Augmented Lagrangian with subproblems solved by SPG. This modelling code was done in MATLAB 2015a using YALMIP - version R20180926. We used BLAS/LAPACK version 3.8.0. We make use of the Fortran compiler *gfortran 7.3.0* with the optimization flag *-O3*.

2. **Solving part:** The initial penalty parameter is $\rho^0 = 2$ and if the infeasibility in iteration k is larger than $\rho_{\mathsf{frac}} := 0.9$ times the infeasibility in iteration $k-1$, then the penalty parameter is multiplied by $\rho_{\mathsf{mult}} := 1.05$. Else, it remains unchanged. In our tests we considered the initial Lagrange multipliers approximations for the equality constraints as $\lambda^0 := 0$.

   The maximum number of outer iterations is fixed as $OMAX = 2000$ and the sum of all SPG iterations is limited to $TMAX = 10^6$. The required decay of the objective function is $\varepsilon_{\mathsf{obj}} = 10^{-8}$ and the desired feasibility is $\varepsilon_{\mathsf{feas}} = 10^{-8}$. M is set to 10, as usual.

   The program stops when either one of these occur:

   - The objective function have not decreased enough, that is $|f(y^k) - f(y^{k-1})| < \varepsilon_{\mathsf{obj}}$, for $N_{\mathsf{stat}} = 5$ consecutive iterations and feasibility is at least what was required, that is, $\|H(y^k)\|_\infty < \varepsilon_{\mathsf{feas}}$;
   - The maximum number of outer iterations $OMAX$ is reached;
   - The maximum number of total SPG iterations $TMAX$ is reached.

3. **Computing a feasible initial covering:** The initial point must also be provided by the user, in general, but in two dimensions, our choice consists on placing the initial centers $c_1^0,\dots,c_m^0$ over the vertices of an m-sided regular polygon inscribed in the circle centered at the origin, with radius 0.5. The initial common radius for those circles is defined as $r^0 := 1.0$. The initial point does not need to be feasible nor close to a minimizer, but feasibility certificates are computed (if they exist) by fixing $(r^0, c_1^0,\dots,c_m^0)$ in y and solving the following problem:

   $$\begin{array}{ll} \underset{S_i\in\mathbb{S}^N}{\text{Minimize}} & \dfrac{1}{2}\|H(y)\|_2^2, \\ \text{subject to} & S_i \succeq 0, \forall i \in \{0,\dots,\eta\}, \end{array} \tag{30}$$

   where $y = (r,c,S)^T$. This is done using SPG itself, and since $H(y) = \mathsf{coeffs}(s_0 + s_1 q_1 + \cdots + s_\eta q_\eta + 1)$ is now an affine function of the variables in the entries of S_i, the objective function of (30) is quadratic. Hence, this problem is expected to be easily solved. In all tests made it has converged in a single iteration.

We cover the closed unit circle centered at the origin using three other circles. Due to rotation symmetry of the solution, we arbitrarily fix one coordinate of a circle in order to guarantee unicity of solution. The optimal radius is given by $r^\star = 0.75$. Below, we present a comparison between different degrees for this test case. Note that the model is sharper with respect to the covering problem the larger the degree $2d$ is, hence we expect to approach r^* with the increase of the degree considered.

$2d$	Solution $f(\bar{y})$	# Outer it.	$\|H(\bar{y})\|_\infty$
4	0.78433	257	8.4564D-09
6	0.77585	286	8.4115D-10
8	0.77426	342	3.4125D-09
10	0.77119	295	3.0167D-09
12	0.77121	343	1.7274D-09
14	0.76878	321	4.3909D-09
16	0.76767	306	1.6420D-09
18	0.76712	285	3.7080D-09

Table 1: Numerical results for Putinar's Positivstellensatz.

Unfortunately, the experiment results were less satisfactory than what we expected, because the highest degree from our tests, $2d = 18$, demands some computational effort both for modelling and solving, but the computed approximation is still quite far from the optimum. By avoiding discretizations, we expected to obtain an arbitrarily close approximation for the actual solution in an acceptable amount of steps. This could not be verified in practice, even though there is no theoretical impeditive at all. The objective function decreases quickly and monotonically in the first few iterations, but slows down considerably after that (see Figure 1); feasibility follows a similar pattern. The issue found is that subproblems are generally not being solved up to the desired tolerance for optimality at later iterations. This may be due to the inherent difficulty of the subproblems, where a second-order method should probably perform better. Also, the Augmented Lagrangian method seems to be very sensitive to the initial Lagrange multipliers approximation, which is not clear how to compute them at this moment. It could also be the case that the true degree $2d$ at the solution r^* is much larger than what is computationally solvable.

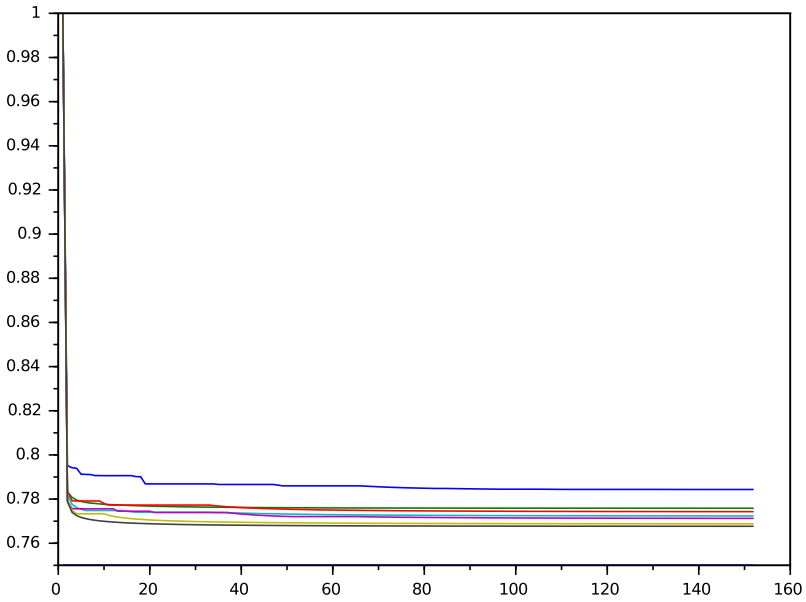

Figure 1: Radius per iteration, up to 157, degrees 4 to 16, up to down.

We have also done extensive tests relaxing several stopping criteria for the SPG solver, allowing oversolving the subproblems and iterating until the maximum number of outer iterations *OMAX* is reached, and also covering other objects. In general, the results were not much better than what was obtained after stopping under the usual criteria. We also performed some experiments with PENLAB (see [29]) via YALMIP, which is a general Augmented Lagrangian method for nonlinear semidefinite programming, but the solver could not find even a feasible point. This might indicate that our model provides problems that are inherently hard to solve. Let us recall that our model is very general, as it allows problems in any dimension, shape of the object to be covered and number of balls. We even do not need to use

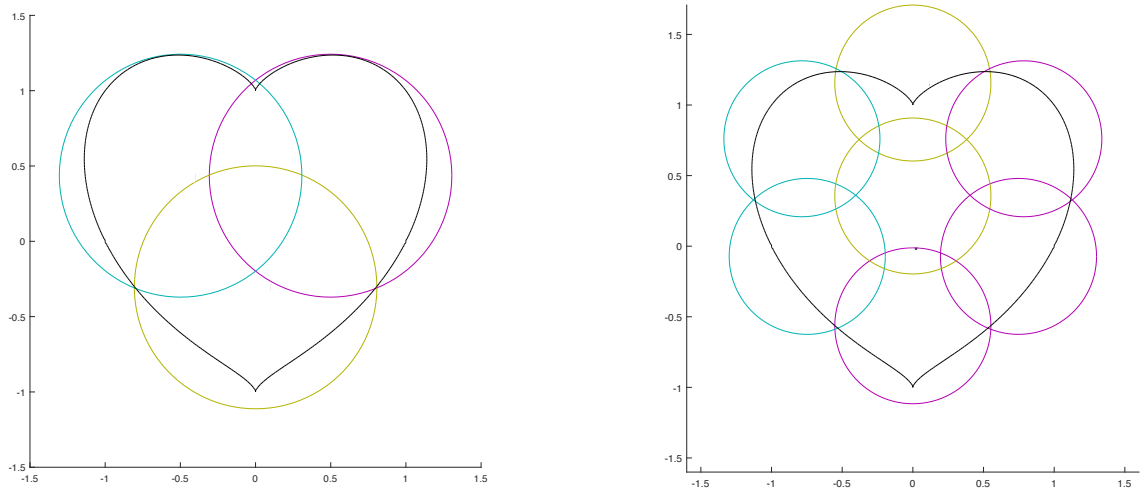

Figure 2: Solution found by an Inexact Restoration method for covering a heart-shaped object.

balls to cover the given object. Any basic semialgebraic set can be used in place of a ball. We achieve this generality while avoiding discretizations.

To exemplify the generality of the model, we present the following pictures illustrating the coverings obtained with three and seven circles (with squared radii equal to 0.6505 and 0.3051, respectively) of a heart-shaped object described by $\{(x,y) \in \mathbb{R}^2 \mid (x^2+y^2-1)^3 - x^2y^3 \leqslant 0\}$.

Note that our model provides certificates of feasibility of the given solution and certificates of approximate optimality via the approximate Lagrange multipliers, while we do not know any other technique for dealing with such problem besides straightforward discretization. These experiments can be found in details in [48], where an Inexact Restoration solver was implemented for the same model, but no proof of convergence was given.

6 Conclusions

We presented an Augmented Lagrangian method for solving nonlinear semidefinite programming problems which allows constrained subproblems. Our global convergence proof is based on the AKKT optimality condition, which is strictly stronger than the usually employed Fritz John optimality condition (or equivalently, where a KKT point is reached under MFCQ). This algorithm was designed in order to approach the covering problem by means of convex algebraic geometry tools, which provides a very general and versatile nonlinear semidefinite programming problem. The dimension and the shape of the regions that build the covering are left as input for the user, even though those factors have a strong influence on the problem structure. Some of the previous algorithms that aimed to cover arbitrary regions were developed for two and three dimensions using sphere coverings; we offer an alternative for higher dimension and our model gives freedom in the shape of the covering as well, not limited to balls. However, due to the size of the models and lack of information about good Lagrange multipliers estimates, or geometric insight about the feasibility certificates S, we are still not able to solve our model to a high accuracy. Further work should be done in order to compute these estimates and to presolve the constraints, hopefully avoiding its inherent difficulty. We advocate that the model is interesting in its own due to the use of somewhat heavy algebraic machinery. We believe that algebraic techniques can provide further insights for solving difficult optimization problems without resorting to discretizations, and we hope our work may inspire further investigations in this direction.

Acknowledgements. This work was Funded by FAPESP grants 2018/24293-0, 2017/18308-2, 2017/17840-2, 2016/16999-5, 2013/05475-7, CNPq and CAPES.

References

[1] R. Andreani, E.G. Birgin, J.M. Martínez, and M.L. Schuverdt. On augmented Lagrangian methods with general lower-level constraint. *SIAM Journal on Optimization*, 18(4):1286–1309, 2007.

[2] R. Andreani, E.G. Birgin, J.M. Martínez, and M.L. Schuverdt. Augmented Lagrangian methods under the constant positive linear dependence constraint qualification. *Mathematical Programming*, 111:5–32, 2008.

[3] R. Andreani, N.S. Fazzio, M.L. Schuverdt, and L.D. Secchin. A sequential optimality condition related to the quasinormality constraint qualification and its algorithmic consequences. *Optimization-online*, 2017.

[4] R. Andreani, G.Haeser, and D.S. Viana. Optimality conditions and global convergence for nonlinear semidefinite programming. *Mathematical Programming*, 2018.

[5] R. Andreani, G. Haeser, and J.M. Martínez. On sequential optimality conditions for smooth constrained optimization. *Optimization*, 60(5):627–641, 2011.

[6] R. Andreani, G. Haeser, A. Ramos, and P.J.S. Silva. A second-order sequential optimality condition associated to the convergence of algorithms. *IMA Journal of Numerical Analysis*, 37(4):1902–1929, 2017.

[7] R. Andreani, G. Haeser, M. L. Schuverdt, and P.J.S. Silva. Two new weak constraint qualifications and applications. *SIAM Journal on Optimization*, 22(3):1109–1135, 2012.

[8] R. Andreani, G. Haeser, M.L. Schuverdt, and P.J.S. Silva. A relaxed constant positive linear dependence constraint qualification and applications. *Mathematical Programming*, 135(1-2):255–273, 2012.

[9] R. Andreani, J.M. Martínez, A. Ramos, and P.J.S. Silva. A cone-continuity constraint qualification and algorithmic consequences. *SIAM Journal on Optimization*, 26(1):96–110, 2016.

[10] R. Andreani, J.M. Martínez, A. Ramos, and P.J.S. Silva. Strict constraint qualifications and sequential optimality conditions for constrained optimization. *Mathematics of Operations Research*, 43(3):693–717, 2018.

[11] R. Andreani, J.M. Martínez, and L.T. Santos. Newton's method may fail to recognize proximity to optimal points in constrained optimization. *Mathematical Programming*, 160:547–555, 2016.

[12] R. Andreani, L.D. Secchin, and P.J.S. Silva. Convergence properties of a second order augmented Lagrangian method for mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, 3(28):2574–2600, 2018.

[13] K. Bezdek. Optimal covering of circles. *Ph.D. Thesis, University of Budapest*, 1979.

[14] K. Bezdek. Über einige kreisüberdeckungen. *Beiträge Algebra Geom.*, 14:7–13, 1983.

[15] E.G. Birgin, J.L. Gardenghi, J.M. Martínez, S.A. Santos, and Ph.L. Toint. Evaluation complexity for nonlinear constrained optimization using unscaled KKT conditions and high-order models. *SIAM Journal on Optimization*, 26:951–967, 2016.

[16] E.G. Birgin, G. Haeser, and A. Ramos. Augmented Lagrangians with constrained subproblems and convergence to second-order stationary points. *Computational Optimization and Applications*, 69(1):51–75, 2018.

[17] E.G. Birgin, N. Krejic, and J.M. Martínez. On the minimization of possibly discontinuous functions by means of pointwise approximations. *Optimization Letters*, 11(8):1623–1637, 2017.

[18] E.G. Birgin and J.M. Martínez. *Practical Augmented Lagrangian Methods for Constrained Optimization*. SIAM Publications, 2014.

[19] E.G. Birgin, J.M. Martínez, and M. Raydan. Nonmonotone spectral projected gradient methods on convex sets. *SIAM Journal on Optimization*, 10(4):1196–1211, 2000.

[20] E.G. Birgin, J.M. Martínez, and M. Raydan. Spectral Projected Gradient Methods: Review and Perspectives. *Journal of Statistical Software, Articles*, 60(3):1–21, 2014.

[21] J.F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. New York, 2000.

[22] L.F. Bueno, G. Haeser, and F.N. Rojas. Optimality conditions and constraint qualifications for Generalized Nash Equilibrium Problems and their practical implications. *SIAM Journal on Optimization*, 29(1):31–54, 2019.

[23] M.D. Choi, T.Y. Lam, and B. Reznick. Sums of squares of real polynomials. *Proc. Symp. Pure Math.*, 58(2):103–126, 1995.

[24] M. Kočvara and M. PENNON - a code for convex nonlinear and semidefinite programming. *Optimization Methods and Software*, 18(3):317–333, 2010.

[25] M. Kočvara and M. Stingl. Pennon - a generalized augmented lagrangian method for semidefinite programming. *High performance algorithms and software for nonlinear optimization (eds. G. Di Pillo and A. Murli),Kluwer Academic Publishers*, pages 297–315, 2003.

[26] J. H. Demyanov. On the maximization of a certain nondifferentiable function. *Journal of Optimization Theory and Applications*, 7:75–89, 1971.

[27] M.A. Diniz-Ehrhardt, M.A. Gomes-Ruggiero, J.M. Martínez, and S.A. Santos. Augmented Lagrangian algorithms based on the Spectral Projected Gradient method for solving nonlinear programming problems. *Journal of Optimization Theory and Applications*, 123(3):497–517, 2004.

[28] J. Dutta, K. Deb, R. Tulshyan, and R. Arora. Approximate KKT points and a proximity measure for termination. *Journal Global Optimization*, 56(4):1463–1499, 2013.

[29] J. Fiala, M. Kocvara, and M. Stingl. Penlab: A matlab solver for nonlinear semidefinite optimization. *arXiv:1311.5240*, 2013.

[30] Z. Gáspár and T. Tarnai. Covering a square by equal circles. *El. Math.*, 50:167–170, 1995.

[31] Z. Gáspár, T. Tarnai, and K. Hincz. Partial covering of a circle by equal circles. part i: The mechanical models. *Journal of Computational Geometry*, 5:104–125, 2014.

[32] Z. Gáspár, T. Tarnai, and K. Hincz. Partial covering of a circle by equal circles. part ii: The case of 5 circles. *Journal of Computational Geometry*, 5:126–149, 2014.

[33] G. Haeser. A second-order optimality condition with first- and second-order complementarity associated with global convergence of algorithms. *Computational Optimization and Applications*, 2018.

[34] G. Haeser, O. Hinder, and Y. Ye. On the behavior of Lagrange multipliers in convex and non-convex infeasible interior point methods. *Optimization online*, 2017.

[35] G. Haeser and V.V. Melo. Convergence detection for optimization algorithms: approximate-KKT stopping criterion when Lagrange multipliers are not available. *Operations Research Letters*, 43(5):484–488, 2015.

[36] G. Haeser and M.L. Schuverdt. On approximate KKT condition and its extension to continuous variational inequalities. *Journal of Optimization Theory and Applications*, 149(3):528–539, 2011.

[37] A. Heppes and H. Melissen. Covering a rectangle with equal circles. *Periodica Mathematica Hungarica*, 1997.

[38] X.X. Huang, K.L. Teo, and X.Q. Yang. Approximate augmented lagrangian functions and nonlinear semidefinite programs. *Acta Mathematica Sinica*, 22(5):1283–1296, 2006.

[39] J.B. Lasserre. *Moments, Positive Polynomials and Their Applications*. Imperial College Press, Covent Garden, London, 2010.

[40] M. Laurent. *Sums of Squares, Moment Matrices and Optimization Over Polynomials*, pages 11–49. Springer New York, New York, NY, 2009.

[41] J. Löfberg. Yalmip : A toolbox for modeling and optimization in matlab. In *Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004.

[42] H.Z. Luo, H.X. Wu, and G.T. Chen. On the convergence of augmented lagrangian methods for nonlinear semidefinite programming. *Journal of Global Optimization*, 54(3):599–618, 2012.

[43] J.M. Martínez and B.F. Svaiter. A practical optimality condition without constraint qualifications for nonlinear programming. *Journal of Optimization Theory and Applications*, 118(1):117–133, 2003.

[44] H. Melissen. Packing and covering with circles. *Ph.D. Thesis, University of Utrecht*, 1997.

[45] J. B. M. Melissen and P. C. Schuur. Improved coverings of a square with six and eight equal circles. *Electr. J. Comb.*, 3, 1996.

[46] J. B. M. Melissen and P. C. Schuur. Covering a rectangle with six and seven circles. *Discrete Applied Mathematics*, 2000.

[47] L. Minchenko and S. Stakhovski. On relaxed constant rank regularity condition in mathematical programming. *Optimization*, 60(4):429–440, 2011.

[48] L. M. Mito. O problema de cobertura via geometria algébrica convexa. *Master's Thesis, University of São Paulo*, 2018.

[49] K. J. Nurmela. Covering a circle by congruent circular disks. *Ph.D. Thesis, Helsinki University of Technology*, 1998.

[50] K. J. Nurmela and P. R. J. Östergård. Covering a square with up to 30 equal circles. Technical report, Helsinki University of Technology, 2000.

[51] P.A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Math. Program., Ser. B*, 96:293–320, 2003.

[52] M. Putinar. Positive polynomials on compact semi-algebraic sets. *Indiana University Mathematics Journal*, 42(3):969–984, 1993.

[53] L. Qi and Z. Wei. On the constant positive linear dependence conditions and its application to SQP methods. *SIAM Journal on Optimization*, 10(4):963–981, 2000.

[54] G. Stengle. A nullstellensatz and a positivstellensatz in semialgebraic geometry. *Mathematische Annalen*, 207(2):87–97, 1974.

[55] D. Sun, J. Sun, and L. Zhang. The rate of convergence of the augmented lagrangian method for nonlinear semidefinite programming. *Mathematical Programming*, 114(2):349–391, 2008.

[56] J. Sun, L.W. Zhang, and Y. Wu. Properties of the augmented Lagrangian in nonlinear semidefinite optimization. *Journal of Optimization Theory and Applications*, 12(3):437–456, 2006.

[57] G. Fejes Tóth. Thinnest covering of a circle by eight, nine or ten congruent circles. *Mathematical Sciences Research Institute Publications, Cambridge University Press*, 52:361–376, 2005.

[58] N.V. Tuyen, J. Yao, and C. Wen. A note on approximate Karush-Kuhn-Tucker conditions in locally Lipschitz multiobjective optimization. *ArXiv:1711.08551*, 2017.

[59] H. Venceslau, D. Lubke, and A. Xavier. Optimal covering of solid bodies by spheres via the hyperbolic smoothing technique. *Optimization Methods and Software*, 30(2):391–403, 2014.

[60] H. Wu, H. Luo, X. Ding, and G. Chen. Global convergence of modified augmented lagrangian methods for nonlinear semidefinite programmings. *Computational Optimization and Applications*, 56(3):531–558, 2013.

[61] A. E. Xavier and A. A. F. Oliveira. Optimum covering of plane domains by circles via hyperbolic smoothing method. *Journal of Global Optimization*, 31:493–504, 2005.

[62] H. Yamashita and H. Yabe. A survey of numerical methods for nonlinear semidefinite programming. *Journal of the Operations Research Society of Japan*, 58(1):24–60, 2015.