Polynomial Matrix Inequalities within Tame Geometry

Christos Aravanis,1† Johannes Aspmann,2‡ Georgios Korpas,3§ and Jakub Marecek3§

1 The University of Sheffield International College, United Kingdom
2 School of Mathematics and Hamilton Mathematics Institute, Trinity College Dublin, Ireland
3 Department of Computer Science and Artificial Intelligence Center, Czech Technical University in Prague, Czech Republic

Abstract

Polynomial matrix inequalities can be solved using hierarchies of convex relaxations, pioneered by Henrion and Lasserre. In some cases, this might not be practical, and one may need to resort to methods with local convergence guarantees, whose development has been rather ad hoc, so far. In this paper, we explore several alternative approaches to the problem, with non-trivial guarantees available using results from tame geometry.

I. INTRODUCTION

Polynomial matrix inequalities (PMI) are a class of (generally) non-convex optimization problems [1,2], with extensive applications in control theory [2,3], structural engineering and design optimization [4], and quantum information theory [5,6]. Unfortunately, even the verification of local optimality of PMI is an NP-hard problem [7], and polynomial-time solvers with guarantees of global convergence to global optima are hence unlikely.

Nevertheless, there are several approaches to solving PMIs in the literature, which we survey below. These include polynomial-time methods [3,8–14], which have only guarantees of local convergence to local optima, developed in an ad-hoc fashion. There are also methods based on solving ever larger convexifications [2,15–18], which have guarantees of asymptotic convergence to global optima. A number of branch-and-bound [19,20] and branch-and-cut [21] approaches have also been considered. Overall, however, progress has been rather intermittent.

In this paper, we explore several approaches to the problem, which have not been explored previously, with guarantees on their performance (e.g., local convergence, rates thereof) available using results from tame geometry [22,23], a fast-developing field on the interface of topology, functional analysis, and optimization. We also try to reason about the suitability of the proposed approaches in various settings.

Technically, this relies on an extension of the reasoning about the tame nature of the feasible set of a constrained optimization problem and the graph of an unconstrained optimization problem to reasoning about tame representation of a (generalized) Lagrangian of a constrained optimization problem, which could be of some interest beyond PMI.

A. The Role of Tame Geometry

In [24], Grothendieck advocated for the construction of a theory of tame topological and geometrical structures motivated by several issues within analysis. He also initiated a program for its development. An important notion for the development of tame geometry is the stratification of smooth manifolds, that is, colloquially, the breaking up of manifolds into submanifolds of lower dimension. As a matter of fact, many “nice” topological spaces, including moduli spaces of curves and, more prominently, solution sets of systems of polynomial equations and semialgebraic sets, provide a landscape where stratifications are straightforward and understandable. Semialgebraic sets are very natural in convex optimization theory and are prototypical examples of sets with tame topology, also known as tame sets. Tame sets are now understood more broadly in the context of o-minimal structures, which we introduce in Sec. II. As we will explain, a set is called tame if it is definable on an o-minimal structure, and similarly, a function is tame if its graph corresponds to a tame set. Many different o-minimal structures have been constructed that generalize semialgebraic sets. Because of the tameness of these spaces, important results in analysis and optimization theory follow.

Tame geometry and the theory of o-minimal structures have been used extensively for obtaining new results in geometric topology [24], analysis and subgradient calculus [23,25–28] as well as complex algebraic geometry and Hodge theory [29,32] with implications reaching as far out as consistency arguments for effective field theories in the context of string theory [33,34].

Tame geometry has also found important applications in optimization theory. It is now understood that semialgebraic functions, semianalytic functions, and more generally functions definable in o-minimal structures form a class of functions in which first-order optimization methods are applicable [22,23] in the following sense: for any locally-Lipschitz Whitney stratifiable function, there exists a descent guarantee along any subgradient trajectory, which, in turn, results in convergence guarantees...
for stochastic subgradient methods [35]. Similarly, Newton’s method converges fast [36] on locally-Lipschitz tame functions, which turn out to be semismooth. To appreciate this, consider the three qualitatively different types of behavior, as far as convergence guarantees are concerned:

- Convex and tame: convergence to optimality along a finite trajectory [37].
- Convex and non-tame: many methods fail to converge [38], due to the lack of a finite trajectory.
- Tame and non-convex: methods adding noise to subgradients can produce global optima asymptotically. Without the addition of noise, we obtain convergence to a critical point in finite time.

See [22] and [23] for two state-of-the-art analyses. We expect many more algorithms and their analyses to appear, especially in relation to optimization in deep learning, where an important class of problems [23] has been shown to be tame.

In this paper, we introduce tame representations of non-convex polynomial matrix inequality problems by formulating the generalized Lagrangian \( \tilde{L} \) associated to the Lagrangian \( L \) of a PMI, as exposed in Eqs. (23) and (24). Our approach is inspired by the representability of the Lagrangian \( \tilde{L} \) formulating the generalized Lagrangian non-convex polynomial matrix inequality problems by shown to be tame.

The goal of this paper is to bring together tame structures and PMIs. In this section, we collect the necessary definitions about tame structures. Since semialgebraic structures are core examples of tame structures, we include related definitions too.

### Notation II.1

In the following, the \( n \)-dimensional Euclidean space is denoted by \( \mathbb{R}^n \). For a subset \( X \subseteq \mathbb{R}^n \) denote by \( \mathcal{P}(X) \) its powerset, that is, the set of all subsets. The graph of a map \( f : X \rightarrow Y \) is the set \( \Gamma(f) \) defined by

\[
\Gamma(f) = \{(x, y) \in X \times Y : y = f(x)\}.
\]

#### A. Semialgebraic structures

In this paper, we introduce the notions of semialgebraic sets and of semialgebraic functions necessary for what follows. For more details, see [42] Chapter 2.

A subset \( X \subseteq \mathbb{R}^n \) is said to be semialgebraic if it is a Boolean combination of sets of solutions of polynomial equations and polynomial inequalities with real coefficients. More precisely, if \( X \) can be represented as

\[
X = \bigcup_{i=1}^{s} \bigcap_{j=1}^{t} \{x \in \mathbb{R}^p : p_{ij}(x) \neq 0\},
\]

where \( p_{ij} \) are real polynomials, \( \neq \) is one of \( >, <, = \) and \( s \) and \( t \) are finite numbers.

Examples of semialgebraic sets are algebraic sets, meaning sets of all zeros of real polynomials on \( n \) variables. More examples of semialgebraic sets are obtained.
by taking finite unions, finite intersections, Cartesian products, and complements of semialgebraic sets.

A function \( f : X \to Y \) between semialgebraic sets is semialgebraic if its graph \( \Gamma(f) \) is a semialgebraic set. For example, real polynomials in \( n \) variables are semialgebraic functions. It is worth mentioning that, naturally, the exponential map, \( \exp : \mathbb{R} \to \mathbb{R} \), is not a semialgebraic function.

B. Tame structures and definable maps

In this section, we introduce tame structures also known as \( \omega \)-minimal structures which will be of particular interest in Section IV. We start with the definition of an \( \mathbb{R} \)-structure and then a tame structure is defined as an \( \mathbb{R} \)-structure which satisfies an extra condition. For more details on tame structures and its applications, see [43, 22, 44, Chapter 1].

Definition 1 An \( \mathbb{R} \)-structure \( S \) is a collection of sets \( \{S^p\}_{p \geq 1} \) where each \( S^p \) is a subset of the powerset of \( \mathbb{R}^p \) such that the following hold:

\[ A_1 : \text{Each } S^p \text{ contains all zero sets of finite collections of polynomials in real variables } p. \]

\[ A_2 : \text{For every linear map } L : \mathbb{R}^p \to \mathbb{R}, \text{ the half-plane } \{ x \in \mathbb{R}^p | L(x) \geq 0 \} \text{ belongs to the set } S^p. \]

\[ P_1 : \text{For every } p \geq 1, \text{ the } S^p \text{ is closed under set-theoretic unions, intersections, and complements.} \]

\[ P_2 : \text{If } X \in S^p \text{ and } Y \in S^q \text{ then } X \times Y \in S^{p+q}. \]

\[ P_3 : \text{If } T : \mathbb{R}^p \to \mathbb{R}^q \text{ is an affine map and } X \in S^p, \text{ then } T(X) \in S^q. \]

Example 2 One of the most fundamental examples of an \( \mathbb{R} \)-structure is the collection of real semialgebraic sets \( \mathbb{R}_{\text{semialg}} \), given as polynomial inequalities in real variables.

Definition 3 Let \( S \) be a fixed \( \mathbb{R} \)-structure.

1. \( A \) set \( X \) is said to be \( S \)-definable if \( X \) belongs to some \( S^p \).

2. A map \( f : X \to Y \) between \( S \)-definable sets is said to be an \( S \)-definable map if its graph \( \Gamma(f) \) is \( S \)-definable.

Example 4 In the following several examples of \( S \)-definable sets and \( S \)-definable functions are provided.

1. Any polynomial with real coefficients is a \( S \)-definable map.

2. Let \( X \xrightarrow{f} Y \xrightarrow{g} Z \) be two \( S \)-definable maps. Then its composition \( g \circ f \) is an \( S \)-definable map.

3. For a \( S \)-definable map \( f(x,y) \), the \( \inf_y f(x,y) \) is \( S \)-definable.

4. The image and pre-image of a definable set under a definable map are both definable sets.

5. The inverse of an injective and definable function is also a definable function, see [43, Lemma 2.3 (iii)].

New \( \mathbb{R} \)-structures are obtained from old ones as follows: for an \( \mathbb{R} \)-structure \( S \) and a collection \( A = \{A_p\}_{p \geq 1} \) where \( A_p \in \mathcal{P}(\mathbb{R}^p) \), denote by \( S(A) \) the smallest structure that contains \( S \) and the sets \( A_p \). Then, we say that \( S(A) \) is obtained from \( S \) by adjoining the collection \( A \).

Example 5 A restricted real analytic function is a function \( f : \mathbb{R}^p \to \mathbb{R} \) with the property that there exists a real analytic function \( f \) defined in an open neighborhood \( U \) of the hypercube \( C_n := [-1,1]^n \) such that

\[
    f = \begin{cases} 
        \bar{f}, & \text{if } x \in C_n \\
        0, & \text{if } x \in \mathbb{R}^p \setminus C_n.
    \end{cases}
\]

Adjoining the structure \( \mathbb{R}_{\text{semialg}} \) with the graphs of all restricted real analytic functions, one obtains the \( \mathbb{R} \)-structure \( \mathbb{R}_{\text{an}} \).

Example 6 Adjoining the structure \( \mathbb{R}_{\text{semialg}} \) with the graph of the exponential map \( \mathbb{R} \to \mathbb{R}, t \mapsto e^t \) one obtains the \( \mathbb{R} \)-structure \( \mathbb{R}_{\text{exp}} \).

We are really interested in structures that are tame, also known as \( \omega \)-minimal which we will define now.

Definition 7 Let \( S \) be an \( \mathbb{R} \)-structure, as given in Definition 1. Then, \( S \) is said to be a tame structure or an \( \omega \)-minimal structure if the following property is satisfied:

\[ T : \text{Any set } X \in S^1 \text{ is a finite union of singletons } \{\ast \} \text{ and open intervals } (a,b) \text{ where } -\infty \leq a < b \leq \infty. \]

In other words, tame structures are \( \mathbb{R} \)-structures which satisfy an extra property \( T \) as defined above.

Example 8 Examples of tame structures are the following.

1. The collection \( \mathbb{R}_{\text{semialg}} \) of real semialgebraic sets, defined in Example 4.

2. The \( \mathbb{R} \)-structure \( \mathbb{R}_{\text{an}} \), defined in Example 5.

3. The \( \mathbb{R} \)-structure \( \mathbb{R}_{\text{exp}} \), defined in Example 6.

Notation II.2 A function definable in some tame structure \( S \) will be referred to as a tame map.
III. BACKGROUND: POLYNOMIAL MATRIX INEQUALITIES

This section is devoted to introducing the types of problems we are interested in, namely polynomial matrix inequalities. After introducing the general definitions, we discuss some related previous results.

A. Generalities

Let \( S^m \subseteq \mathbb{R}^{m \times m} \) denote the space of real \( m \times m \) symmetric matrices. We denote the subset of positive definite (semidefinite) \( m \times m \) matrices by \( \mathcal{S}_+^m \subseteq \mathcal{S}_d^m \) \( (\mathcal{S}_+^m \subseteq \mathcal{S}^m) \). For a matrix \( A \in \mathcal{S}_+^m \) \((A \in \mathcal{S}_d^m) \) we write \( A \succ 0 \) \((A \succeq 0) \). Finally, a semialgebraic set \( K \) is defined as

\[
K = \{ x \in \mathbb{R}^k : g_1(x) \geq 0, \ldots, g_q(x) \geq 0 \},
\]

where \( g_1, \ldots, g_q \in \mathbb{R}[x] \).

**Definition 9 (PMI)** Let \( y \in \mathbb{R}^t, \ x \in \mathbb{R}^k \), a polynomial \( b \in \mathbb{R}[y] \) and a polynomial matrix \( P : \mathbb{R}^k \times \mathbb{R}^t \to \mathcal{S}^m \). Then, a polynomial matrix inequality optimization problem is the following optimization problem:

\[
B^* := \inf_{y \in \mathbb{R}^t} b(y) \text{ such that } P(x,y) \succeq 0.
\]

We will sometimes refer to the polynomial \( b \) in the above definition as a cost function and to a polynomial matrix inequality optimization problem as a PMI problem.

In general, \( P(x,y) \) is defined by \( P_0(x) - \sum_{i=1}^t P_i(x)y_i \) and \( \|y\|_\infty < 1 \). In general, verifying polynomial matrix inequalities is an NP-hard problem \[7\], and Prob. (4) is generally intractable.

**Definition 10** A feasible set for a PMI is the subset of the space of search variables that satisfies the PMI condition. In the notation of \[4\], this means the subset of \( \mathbb{R}^t \) such that the optimization problem has a solution that satisfies the PMI constraint \( P(x,y) \succeq 0 \).

A few remarks are due for some special cases of Prob. \[4\].
(1) If \( \deg(b) = \deg(P) = 1 \), then Prob. [4] corresponds to a convex linear matrix inequality (LMI). Precisely, given \( y \in \mathbb{R}^n \) and \( \{P_i\}_{i=0}^n \in S_+^m \), an LMI takes the form

\[
P_0 + y_1 P_1 + \ldots + y_n P_n \succeq 0,
\]

and the set of real vectors \( y \in \mathbb{R}^n \) that satisfy the matrix inequality above define a spectrahedron [47]. The set of real vectors \( x \) such that the pencil is positive semidefinite is a convex semialgebraic set called spectrahedron, described by a linear matrix inequality (LMI)

\[
P_0 + \sum_{i=1}^m x_i F_i + \sum_{j=1}^n y_j G_j + \sum_{i=1}^m \sum_{j=1}^n x_i y_j H_{ij} \succeq 0
\]

for \( F_i, G_j, H_{ij} \in S^m_+ \). In this case, Problem [4] is referred to as a bilinear matrix inequality (BMI).

(2) If \( \deg(P) = 2 \) and no square terms appear in it, then \( P \) takes the form

\[
P_0 + \sum_{i=1}^m x_i F_i + \sum_{j=1}^n y_j G_j + \sum_{i=1}^m \sum_{j=1}^n x_i y_j H_{ij} \succeq 0
\]

for \( F_i, G_j, H_{ij} \in S^m_+ \). In this case, Problem [4] is referred to as a bilinear matrix inequality (BMI).

(3) If the matrix \( P(x,y) := X \) is a symmetric positive definite matrix, i.e. \( X \in S^m_+ \), then Problem [4] is transformed to

\[
X^* := \inf_X Q(X) \quad \text{such that} \quad X \succ 0,
\]

where \( Q \) is a matrix-valued polynomial. Typically, this type of problem comes with some additional constraint, for example

\[
X^* := \inf_X Q(X) \\
\text{s.t.} \quad X \succeq 0, \\
\text{Tr}(X) = 1.
\]

(4) If \( P \) has a chordal sparsity graph with maximal cliques \( C_1, \ldots, C_t \), then \( P \) can be decomposed as [18]

\[
P(x,y) = S_0(x) + \sum_{j=1}^k g_j(x,y) S_j(x)
\]

for \( m \times m \) sum-of-squares (SOS) polynomial matrices \( \{S_j\}_{j=0}^k \) where \( S_i(x) = H(x)^t H(x) \) for some polynomial matrix \( H(x) \).

B. Convex Relaxations of PMIs

The objective function of a PMI is a polynomial, by definition, on an indeterminate \( X \in S_+^n \). One can leverage the power of polynomial optimization techniques, whose overview is in Table 1, to solve PMIs. Specifically, one can use convergent relaxations of PMIs [2] based on the scalar convergent relaxations of SDPs [48]. This method is global because it can solve PMI problems even when finite convergence occurs, while, in parallel, it provides a numerical certificate of global optimality. However, it is also local, given that the convex relaxations utilize SDPs.

In this section, we will review the moment relaxations of Lasserre [48] for PMI problems such as Problem [4], following Henrion and Lasserre [2]. Let \( b_d \) denote the vector of all possible monomials, up to degree \( d \), constructed out of the entries of \( x \in \mathbb{R}^n \). That is, the entries of \( b_d \)

\[
\begin{bmatrix}
1 \\
x_1 \\
x_1^2 \\
x_1 x_2 \\
x_2 \\
x_2 x_1 \\
x_2^2 \\
x_2 x_3 \\
x_2 x_3^2 \\
x_2 x_3 x_4 \\
x_2 x_3 x_4^2 \\
x_2 x_3 x_4 x_5 \\
x_2 x_3 x_4 x_5^2 \\
x_2 x_3 x_4 x_5 x_6 \\
x_2 x_3 x_4 x_5 x_6^2 \\
x_2 x_3 x_4 x_5 x_6 x_7 \\
x_2 x_3 x_4 x_5 x_6 x_7^2 \\
x_2 x_3 x_4 x_5 x_6 x_7 x_8 \\
x_2 x_3 x_4 x_5 x_6 x_7 x_8^2 \\
x_2 x_3 x_4 x_5 x_6 x_7 x_8 x_9 \\
x_2 x_3 x_4 x_5 x_6 x_7 x_8 x_9^2 \\
\vdots
\end{bmatrix}
\]

for what follows, consider the limit \( d \to \infty \) such that \( \lim_{d \to \infty} b_d := b \). Let \( y := \{y_a\}_{a \in \mathbb{N}^n} \) be a sequence of indeterminates labeled with the entries of \( b(x) \). The sequence \( \{y_a\}_{a \in \mathbb{N}^n} \) is referred to as the moment variable, by defining \( y := \int x^a d\mu \), for a Borel measure \( \mu \).

Similarly to the order-\( d \) generalized Hankel matrix, we can define the order \( d \)-moment matrix

\[
M_d(y) := \begin{pmatrix}
y(0,0) & y(1,0) & \cdots & y(0,1) \\
y(1,0) & y(2,0) & \cdots & y(1,1) \\
y(0,1) & y(1,1) & \cdots & y(0,2) \\
\vdots & \vdots & \ddots & \vdots \\
y(0,d) & y(1,d) & \cdots & y(0,2d)
\end{pmatrix}
\]

As before, for what follows, consider the limit \( d \to \infty \) and denote \( \lim_{d \to \infty} M_d(y) := M(y) \). Similarly, consider an infinite-dimensional basis represented by \( b(x) \) or, equivalently, \( y_a \). Any polynomial \( p \in \mathbb{R}[x_1, \ldots, x_n] \) can then be identified with the vector of its coefficients labelled by \( a \in \mathbb{N}^n \) as \( p(x) = \sum_{a \in \mathbb{N}^n} p_a x^a = (p,b) \). Similarly to this construction, we can define the linear mapping \( L_y(p) = \langle p, y \rangle \), while for polynomials \( p,q \in \mathbb{R}[x_1, \ldots, x_n] \) we can define the bilinear mapping \( L_y(p,q) = \langle p,M(y)q \rangle \).

It is straightforward to confirm that Eq. (12) is equivalent to

\[
M(y) = \int bb^T d\mu
\]

It immediately follows, for any two polynomials \( p \) and \( q \), \( L_y(pq) = \int pq d\mu \). For \( d \) finite, corresponding to monomial basis \( b_d \subset b \) of degree at most \( d \) and moment matrix \( M_d \subset M \) of order \( d \), it follows that \( M_d(y) \succeq 0 \) since for any \( p \in \mathbb{R}[x], \langle p,p \rangle_y = \langle p,M_d(y)p \rangle = \int p^2 d\mu \succeq 0 \). Similarly, for more general moments, if the measure \( \mu \) of \( y \) is in the semialgebraic set defined as \( \{x \in \mathbb{R}^n | g_i(x) \geq 0 \} \), \( \langle p,p \rangle_{g,y} \)
We formulate the notion of the non-necessarily bounded level of the relaxation hi-since they do not scale well given their super-exponential theoretical convergence guarantees. However, they do in-
the Newton method, gradient flows, and subgradients. For polynomial time methods such as
representations of PMIs which do involve (weaker) theo-
family of methods, for polynomial time methods such as
As in the scalar case, associated to a parent non-convex
problem \[ M(K) = \{ y \in R^n : G(y) \geq 0 \} \], the semialgebraic set \( \mathcal{P} \) of the matrix \( G \). Helton and Vinnikov considered this precise problem in \[33\].
\begin{equation}
\mathcal{K} := \{ x \in R^n : G(x) \geq 0 \}. 
\end{equation}

For a polynomial mapping \( P : R^n \rightarrow R^m \) of degree at most \( k \), we write \( x \mapsto P(x) = \mathbf{P}_k(x) \in R^m \), for some \( m \times s_k \) matrix \( \mathbf{P} \), where \( s_k \) is the dimension of the vector space \( R_k[x] \).

Similarly to the definition of the moment matrix for scalar variables, let \( M_k(y) := \{ y_\alpha \} |\alpha| \leq k \) be the or-
der \( k \) moment matrix associated with a sequence \( y \), and let \( M_k(G) \) denote the localizing matrix that, by abuse of notation, can be written as

\begin{equation}
M_k(G) = L_y (b_k b_k^\top \otimes G), 
\end{equation}

where \( M_k(y) \) and \( M_{k-d}(G) \) are the \( k \)-th order truncated moment and localizing matrices, respectively, associated with the sequence \( y \) and the matrix \( G \). We say that this program is a relaxation of the PMI Prob. (1), with \( f^{(k)} \leq f^* \) for all \( k \). Using results from \[49, 50\], Henrion and Lassere are able to prove that as \( k \rightarrow \infty \), \( f^{(k)} \) approaches \( f^* \) generalizing the Lasserre/SOS convergent relaxations to PMIs.

Since \[2\], several improvements, based on different Pos-
itivstellensätze, have been put forward, as summarized in Table 1. These state-of-the-art methods greatly improve the practicality of the methods of \[2\] and provide strong theoretical convergence guarantees. However, they do in-
crease the complexity of the implementation considerably since they do not scale well given their super-exponential or even unbounded runtime originating, for example, on the non-necessarily bounded level of the relaxation hi-
archy. Trying to find a computationally less resource
 intensive alternative, we formulate the notion of tame representations of PMIs which do involve (weaker) theoretical convergence guarantees, like the Henrion-Lassere
family of methods, for polynomial time methods such as
the Newton method, gradient flows, and subgradients.

\begin{enumerate}
\item PMIs should also be seen as non-linear SDPs \[31\]. There, first-order optimality conditions are well studied \[8, 13, 14\]. Although the direct application of the first-order optimality conditions \[52\] provides weaker local convergence guarantees than the global convergence guarantees of the previous section, they have a certain appeal to practitioners, who often prefer to rely on methods sacrificing guarantees of convergence to global optima for the sake of time efficiency. Indeed, the present-best numerical methods for PMIs \[3, 9, 40\] come only with local convergence guarantees.

\item Representations in Optimization Theory

It should be clear by now that PMIs form a generalization of LMIs, where the entries of the matrix \( P \), as in Eq. (5) are linear functions of the indeterminates \( x \in R^n \). In the literature of optimization theory, an interesting question is under which conditions a subset of \( R^n \) can be represented as an LMI. Helton and Vinnikov considered this pre-
cise problem in \[33\].

\textbf{Theorem 11 (LMI representations \[33\])} A subset \( \mathcal{C} \in R^m \), with \( 0 \) in the interior of \( \mathcal{C} \), admits an LMI representation, if it is a convex algebraic interior and the minimal defining polynomial \( p \) for \( \mathcal{C} \) satisfies the so-called RZ condition (with \( p(0) > 0 \)), wherein a real zero (RZ) polynomial \( p(x) \in R[x] \) for \( x \in R^m \) satisfies \( p(\mu x) = 0 \Rightarrow \mu \in R \).

See also \[57\]. Furthermore, in \[33\] it is also shown that, when \( m = 2 \), if \( p \) is a RZ polynomial of degree \( d \) and \( p(0) > 0 \), then \( \mathcal{C}_p \) has a monic LMI representation with \( d \times d \) matrices. In an important paper, Fawzi and El Din \[58\] bounded the size of the LMI representation of a convex set in dimension \( d \) from below by \( \Omega(\sqrt{\log d}) \). The representation of a certain subset of \( R^n \) as the feasible set of a convex optimization problem or the rep-
 resentation of the feasible set of an optimization problem as a subset of \( R^n \) with some extra properties is a powerful concept. Nie, in \[59\], considered the necessary conditions such that a PMI can be represented as an SDP, essen-
tially, finding a map that takes a non-convex problem to a convex one.

Related concepts of “MPC equivalence” \[60\], “reform-
ulation” \[61, 62\], and reduction \[63\] have been studied in control theory, discrete optimization, and global optimization, respectively. Similarly to Helton and Vinnikov, one can ask when a subset of \( R^n \) can be represented as a PMI, a problem that, to the best of our knowledge, is open. However, we find strikingly more powerful applications in asking when a PMI can be represented as a subset \( \mathcal{C} \) of \( R^n \) with some special “tame” property (in the sense of definable functions in o-minimal structures, cf. Definition 7).
TABLE I. Summary of Positivstellensätze for polynomials and polynomial optimization problems (POP), polynomial matrix inequalities (PMIs), and PMIs with structural sparsity, expanded from Zheng and Fantuzzi [18].

| Positivstellensatz | POP | PMIs | Sparse PMIs |
|--------------------|-----|------|-------------|
| Hilbert-Artin      | Artin [53] | Du [19] | Zheng & Fantuzzi [18, Thm. 2.2] |
| Reznick            | Reznick [51] | Dinh et. al. [17] | Zheng & Fantuzzi [18, Thm. 2.3] |
| Putinar            | Putinar [55] | Scherer & Hol [15] | Zheng & Fantuzzi [18, Thm. 2.4] |
| Dinh               | Lasserre [48] | Henrion & Lasserre [2] |
| Zheng & Fantuzzi   | Putinar & Vasilescu [56] | Dinh et. al. [17] | Zheng & Fantuzzi [18, Thm. 2.5] |

\[\inf_{y,\lambda} \sup_{y,\lambda} \mathcal{L}(x,y) = \inf_{y,\lambda} \mathcal{L}(x,y), \quad (21)\]

Instead, one could consider the “saddle-point problem” [66, 67]

or, to simplify theoretical analyses, consider \(\lambda = \infty\). By utilizing \(\mathcal{L}_\infty\) we achieve a reformulation of Prob. [4] as an unconstrained optimization problem that is equivalent to the constrained PMI [4].

To talk about a tame representation PMI [4], with a generalized Lagrangian as defined in [20] we first need to define what we mean by a representation. Informally, this is a reformulation of the problem. Formally, we define

**Definition 14 (Representation)** A map \(f : \mathbb{R}^n \to \mathbb{R}\) is said to be represented by the map \(g : \mathbb{R}^m \to \mathbb{R}\) if there exists a map \(h : \mathbb{R}^n \to \mathbb{R}^m\) such that \(f = g \circ h\) for all values \(\mathbb{R}^n\).

**Example 15 (Offset)** Let us continue with our planar example [13] with \(y + \lambda \mathbb{1}_{P(x,y) \geq 0}\), which we can consider as being the map \(f\) of Definition 14. One may choose the map \(g : \mathbb{R}^2 \to \mathbb{R}\) to be \(1 + y + \lambda \mathbb{1}_{P(x,y) \geq 0}\) and \(h : \mathbb{R}^2 \to \mathbb{R}^2\) to be the map \(h(x,y) := (x,y + 1)\), because \(y\) enters linearly in the objective of the PMI.

**Example 16 (Embedding)** For the example [13] we can also consider a function \(g : \mathbb{R}^3 \to \mathbb{R}\), whose arguments \((x,y,z) \in \mathbb{R}^3\) would still yield \(y + \lambda \mathbb{1}_{P(x,y) \geq 0}\). The function \(h : \mathbb{R}^2 \to \mathbb{R}^3\) would introduce a \(z = 0\).

Clearly, the examples above are unusual, in the sense that we do not benefit from any improved continuity properties of the functions \(g\). Still, the functions \(g\) represent function \(f\), considering that there exists a function \(h\) that makes it possible to obtain, for all \(y \in \mathbb{R}^2\), \(f(x) = g(h(y)) = g(y)\).

**Definition 17 (Minimizer Representation)** A function \(f : \mathbb{R}^n \to \mathbb{R}\) is said to be minimizer-represented by \(g : \mathbb{R}^m \to \mathbb{R}\) if there exists a function \(h : \mathbb{R}^n \to \mathbb{R}^m\) such that \(\inf f\) attains its value at \(x \in \mathbb{R}^n\) (“global minimizer”) that is the same as \(h\) applied to the global minimizer of \(\inf g\), whenever the global minimizers of \(f, g\) are unique. When the global minimizers of \(f, g\) are not unique, we require that this holds true for one pair of the global minimizers.
Example 18 (Irrelevant Offset) Let us revisit our Example 13, where \( q = 1 + y + \lambda \mathbb{1}_{P(x,y) \leq 0} \), i.e., adding the offset of 1. For minimizer representability, we could consider \( h \) to be the identity, because the offset of 1 would not affect the global minimizer.

In this paper, we will provide various representations of the Generalized Lagrangian of a PMI, wherein a representation of a PMI is called \textit{tame} if the graph of the representation of a Generalized Lagrangian \( \tilde{L}_\lambda \) of the PMI is definable in an o-minimal structure. Precisely, we have the following definition:

**Definition 19 (Tame representation)** Consider a PMI \( X \) with generalized Lagrangian \( L_\lambda \). Assume that \( L_\lambda \) is represented by the function \( \tilde{L}_\lambda \). Then, \( X \) is said to be tame-representable if the graph \( \Gamma_{\tilde{L}_\lambda} \) is definable in an o-minimal structure in the large limit.

In what follows, we present five distinct examples of tame representations for PMIs.

A. Tame representation using the characteristic polynomial

Recall that in Prob. (4), \( P(x,y) \in \mathbb{S}^m \), i.e., it is a symmetric matrix whose entries are polynomials in the indeterminates \( x \in \mathbb{R}^k \) and \( y \in \mathbb{R}^l \). The condition \( P(x,y) \geq 0 \), defines a semialgebraic set. To see this, we take the characteristic polynomial of the \( m \times m \) symmetric matrix \( P(x,y) \),

\[
p(t;x,y) := \det(t \mathbb{1}_m - P(x,y))
\]

\[
= t^m + \sum_{j=1}^{m} (-1)^j q_j(x,y) t^{m-j}.
\]  

(22)

for \( m \) polynomials \( q_j(x,y) \). The roots of the characteristic polynomial are the eigenvalues. Since the eigenvalues of a real symmetric positive semidefinite matrix are all real and non-negative, we have \( q_j(x,y) \geq 0 \) for all \( j \) due to the Descartes’ rule of signs, see [2]. We thus have:

**Proposition 20** One representation of the Generalized Lagrangian of the PMI (4) is

\[
\tilde{L}_\lambda(x,y) := b(y) - \lambda \prod_{j=1}^{m} q_j(x,y).
\]

(23)

in the large limit of \( \lambda \).

**Proof.** Note that even for a sufficiently large and finite \( \lambda \), the minimum of \( b(y) - \lambda \prod_{j=1}^{m} q_j(x,y) \) will have positive \( q_j(x,y) \) for all \( j \). By extending Example 15 to have the offset dependent on \( x, y \), we can construct \( h \) in Definition 14.

The obvious difficulty here is that in order to construct \( h \) in Definition 14, we would have to know the global minimizer. To address this difficulty, we can consider:

**Proposition 21** One minimizer-representation of the Generalized Lagrangian of the PMI (4) is

\[
\tilde{L}_\lambda(x,y) := b(y) - \lambda \prod_{j=1}^{m} q_j(x,y).
\]

(24)

in the large limit of \( \lambda \).

Example 22 Let us study the preceding representation for the Example 13. The characteristic polynomial, (22), is

\[
p(t;x,y) = t^2 - t(2 - x^2 - 16xy - y^2) + (16x^3y - 2x^2 - 16xy + 16xy^3 - y^2 + 1)
\]

We thus have the polynomials

\[
q_1 = 2 - x^2 - 16xy - y^2,
\]

\[
q_2 = 16x^3y - 2x^2 - 16xy + 16xy^3 - y^2 + 1,
\]

and the minimizer-representation of the generalized Lagrangian is simply

\[
\tilde{L}_\lambda = b(y) - \lambda q_1 q_2.
\]

The advantage of considering the minimizer-representation becomes apparent here: we only wish for the global minimizer to coincide, rather than for the value attained at the global minimizer to coincide.

**Proposition 23** The representation (23) of the generalized Lagrangian is definable in a tame structure.

**Proof.** The representation (23) is polynomial, and hence its graph is tame by definition.

Furthermore, semialgebraic sets are the smallest sets defined over an o-minimal structure, specifically (\( \mathbb{R}, +, \cdot \)). Therefore, the characteristic polynomial defines a tautological tame representation for the PMI of Prob. (4).

The example given here is perhaps one of the simplest to construct, but it suffers from some severe drawbacks when it comes to practical use. Namely, for large \( m \) the characteristic polynomial is a high-degree polynomial, which is inherently difficult (albeit not impossible, cf. [68]) to work with numerically.

B. Tame representation using factorisation

Factorization is a very useful concept throughout mathematical optimization and machine learning. For example, solutions to problems that concern optimization over a large square matrix, such as the matrices involved in the attention mechanism in transformers [69], often utilize factorization into a pair of matrices of much lower ranks.

In addition, in the context of this article, the exploitation of the sparsity in Prob. (9) can be seen as a factorization problem. As another example, one can replace
the positive semidefinite constraint \( X \in S^n_+ \) with the low-rank factorization \( X \rightarrow vv^\top \), in an approach now known as the Burer-Monteiro technique \cite{70, 71}.

Given that a factorization is found, the analysis of optimization over such a factorization might be non-trivial, given that it converts a convex programming problem to a non-convex one. Consider that one of the globally-optimal solutions \( X^* \) has some rank \( r \geq 0 \). Instead of an \( m \times m \) matrix \( X \), we can consider its factorization

\[
X^* := \inf_v Q(vv^\top), \quad vv^\top = X.
\]

s.t. \( \operatorname{Tr}(v^\top v) = 1 \),

where \( v \in \mathbb{R}^{\times m} \).

**Proposition 24** A minimizer-representation of the generalized Lagrangian for the above problem is given by \cite{72}

\[
\widetilde{L}_\lambda(v) = Q(vv^\top) - \lambda \left( \operatorname{Tr}(v^\top v) - 1 \right),
\]

if and only if

\[
\nabla_x Q(x) - \lambda = S_v \succeq 0,
\]

for Lagrangian multipliers \( \lambda \) satisfying

\[
S_v v = 0.
\]

This is proven in \cite{72}. Some of these analyses could be simplified considerably by considering:

**Proposition 25** The graph of \( Q(vv^\top) \) is semialgebraic and thus tame.

Of course, this is trivial due to the polynomial nature of the objective function. With Proposition 25, we thus know that there exists a descent guarantee along any subgradient trajectory \cite{35}, which could strengthen Theorem 4.1 in \cite{74}. We also know that Newton’s method converges fast on the factorization \cite{36}, which has been observed empirically \cite{73}, but which has not been proven, yet. Plausibly, more recent analyses \cite{74, 75} e.g. could also benefit.

In practice, however, such a PMI representation is not particularly convenient since the optimal rank for the decomposition is not known a priori. Therefore, factorization provides a family of possible representations, parametrized by the rank \( r \) of \( v \). The actual representation \( \tilde{h} \) required might be hard to estimate.

**Remark 26** Notice that this approach has rich links to numerical algebraic geometry and its uses in linear semidefinite programming \cite{76, 77}. Notably, for \( 0 \leq r \leq m - 1 \), \cite{76, 77} consider the real algebraic set

\[
D_r := \{ x \in \mathbb{R}^k : \operatorname{rank}(X) \leq r \}.
\]

This is an algebraic variety, known as the determinantal variety. We expect that the representation through factorisation presented in this Section would mainly be appropriate for problems with rank deficient optimizers \cite{72}, i.e., where the optimum lies in \( D_r \).

**C. Tame representation using log-det barrier**

Perhaps the most natural tame representation utilizes the log-det barrier function \( \phi \) \cite{39, Proposition 5.4.5} \cite{80, Chapter 11}, which for a positive definite matrix \( X \) is defined by

\[
\phi(X) := \begin{cases} -\log \det X, & X > 0, \\ +\infty, & \text{otherwise.} \end{cases}
\]

This is known \cite{39, Remark 5.4.1} to be the best possible self-concordant barrier function for the cone of positive semidefinite matrices.

As above, the idea in using the log-det barrier is to reformulate the PMI Prob. \cite{7} as an unconstrained problem. We reformulate the Generalized Lagrangian \( \tilde{L}_\lambda \) as

\[
\tilde{L}_\lambda = b(X) + \lambda \phi(X).
\]

Essentially, the log-det barrier approximates the non-smooth indicator function of the Generalized Lagrangian \cite{20} with a smooth function. The problem to solve becomes

\[
\inf_x \tilde{L}_\infty.
\]

At the same time:

**Theorem 27** The map \( b(X) + \lambda \phi(X) \) is definable in a tame structure.

**Proof.** The log function is the inverse of the exp and a definable map due to Example \cite{9} and Example \cite{7}. Hence, the graph of \( \phi \) is tame as it is the composition of two definable maps. To conclude the proof, recall that \( b(X) \) is a polynomial hence a definable map.

With this result, we conclude that the log-det barrier function provides a tame representation of the PMI (Prob. \cite{7}).

Notice that leading first-order solvers for PMI, such as PENBMI \cite{3, 9, 40}, come with ad hoc proofs of local convergence, without using the results from tame geometry \cite{87}. At the same time, results from tame geometry \cite{80} could be used to show that Newton’s method converges fast, for instance.

**D. Tame representations using alternative barrier functions**

While the log-det barrier function is the natural choice for numerical optimization methods \cite{3, 9}, a wealth of other options exist \cite{39, Chapter 5} \cite{52, Section 3.1}. Let us consider rational powers of the determinant, instead of the logarithm of the determinant above:

\[
\phi(X) := \begin{cases} -\det^r X, & X > 0, \\ +\infty, & \text{otherwise.} \end{cases}
\]
for some rational $0 < r < 1$ following Example 3.1 of [59]. This is [59] p. 239] positive-definite representable, which should allow for analyses similar to the sections above, e.g., Section IV A.

E. Tame representation using a bound

In PMI (Prob. [4]), let us denote the global optimum of the objective function $b^\ast$. If we knew the optimum $b^\ast$, we could constrain the feasible set of the optimization problem such that all feasible points were the global optima: $\{P(x, y) \geq 0\} \cap \{b(y) \leq b^\ast\}$. We do not know $b^\ast$ a priori, but we can work with an estimate $\hat{b}$. Then, repeatedly deciding the emptiness of the intersection:

$$\{P(x, y) \geq 0\} \cap \{b(y) \leq \hat{b}\}$$

is clearly equivalent to solving the original problem [4]. Trivially, one may consider, e.g., bisection on $\hat{b}$.

As an example, whenever the intersection (32) is non-empty, multiply $\hat{b}$ by a number between 1 and 2. Whenever the intersection (32) is non-empty, halve $\hat{b}$.

Thus:

**Proposition 28** There exists $\hat{b}$ such that one minimizer-representation of the Generalized Lagrangian of the PMI [4] is

$$\tilde{\mathcal{L}}_\hat{b}(x, y) := \min_{x, y} 1$$

s.t. $P(x, y) \geq 0$

$$b(y) \leq \hat{b}.$$  

Such a representation suffers from a number of issues:

- the fact that $\hat{b}$ is hard to find, in general;
- testing feasibility of semidefinite programming is hard, in general;
- while the intersection of two semialgebraic sets is semialgebraic, and $P(x, y) \geq 0$ is semialgebraic, we may need one of the representations of Sections IV A IV C in a numerical routine.

Still, the convexifications of Section III B can be seen as using this minimizer representation by asking for the non-negative polynomial $b(y) - b^\ast$ to belong to the truncated quadratic module. Thus, this representation may be one of the most often used, in practice.

V. CONCLUSIONS

We have shown how to reformulate PMIs [4] within tame geometry, to make it possible to utilize first-order algorithms with recently developed convergence guarantees [22, 23]. The interest for such reformulations lies precisely in the fact that if a representation of Prob. [4] is Whitney stratifiable, then first-order methods might be preferable compared to the convexifications of Sec. III.

One obvious question that we can ask is what representation is best suited to a particular application. It seems plausible that the representations of Secs. IV A and IV C are suitable only for full-rank minimizers, due to the nature of the Generalized Lagrangian. Similarly, the representation of Sec. IV B would be most useful for low-rank minimizers. As we can tell from Corollary 3.1 in [81], the low-rank minimizers may actually be quite rare, in general, although they are very desirable (and more common) in particular applications. See also [58]. If we do not have reasons to believe that the low-rank minimizers are available in a particular instance, perhaps Secs. IV A and IV C are a great starting point.

An independent consideration may arise when one considers an intersection with further linear matrix inequalities. The work of Waldspurger and Waters [82] showed that if the size of the factors is less than the square root of the number of linear constraints, poor behavior may arise in the representation of Sec. IV B even if we deal with linear matrix inequalities. This may suggest that the representation of Sec. IV B is particularly suitable for problems with few linear constraints in such extensions of polynomial matrix inequalities.

A related question asks how much we can relax the large $\lambda$ limit constraints in the representations. From some numerical experiments, it seems hard to attain the true global minimizers using any of the Generalized Lagrangians we have presented here for a finite value of $\lambda$, but for any given $\epsilon$ error, there is a finite value of $\lambda$ that suffices. Bounding the value of $\lambda$ as a function of $\epsilon$ such that we can guarantee that it suffices seems rather hard analytically. A more in-depth experimental study, or perhaps a “low-to” procedure, would be a good venture for future research.

Many natural questions arise in connection with the notion of a tame representation of sets, more generally. A question, probably of mathematical flavor, is whether the map $h$ is unique or not. Another natural question is whether every map $f$ is thus representable by some $h$, or what conditions ensure that $f$ is representable.

ACKNOWLEDGMENTS

The authors thank Didier Henrion for useful comments, insights, and suggestions.

J.A. is supported by the Government of Ireland Postgraduate Scholarship Programme GOIPG/2020/910 of the Irish Research Council. G.K. and J.M. acknowledge support of the OP RDE funded project CZ.02.1.01/0.0/0.0/16_019/0000765 “Research Center for Informatics”. J.M. acknowledges support of the Czech Science Foundation (22-15524S).
[1] J. G. VanAntwerp and R. D. Braatz, A tutorial on linear and bilinear matrix inequalities, Journal of Process Control 10, 363 (2000).

[2] D. Henrion and J.-B. Lasserre, Convergent relaxations of polynomial matrix inequalities and static output feedback, IEEE Transactions on Automatic Control 51, 192 (2006).

[3] D. Henrion, J. Löfberg, M. Kocvara, and M. Stingl, Solving polynomial static output feedback problems with penbmi, in Proceedings of the 44th IEEE Conference on Decision and Control (IEEE, 2005) pp. 7581–7586.

[4] M. Diehl, F. Jarre, and C. H. Vogelbusch, Loss of superlinear convergence for an sqp-type method with conic constraints, SIAM Journal on Optimization 16, 1201 (2006).

[5] R. W. Freund, F. Jarre, and C. H. Vogelbusch, Nonlinear semidefinite programming: sensitivity, convergence, and an application in passive reduced-order modeling, Mathematical Programming 109, 581 (2007).

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

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

[8] R. Andreani, W. Gómez, G. Haeser, L. M. Mito, and A. Ramos, On optimality conditions for nonlinear conic programming, Mathematics of Operations Research (2021).

[9] C. W. Scherer and C. W. J. Hol, Matrix sum-of-squares relaxations for robust semi-definite programs, Mathematical Programming 107, 189 (2005).

[10] T. H.-B. Dù, A note on positivstellensätze for matrix polynomials (2017).

[11] T. H. Diniz, M. T. Ho, and C. T. Le, Positivstellensätze for polynomial matrices, Positivity 25, 1295 (2021).

[12] Y. Zheng and G. Fantuzzi, Sum-of-squares chordal decomposition of polynomial matrix inequalities, Mathematical Programming 10.1007/s10107-021-01728-w (2021).

[13] K.-C. Goh, M. G. Safonov, and G. P. Papavassiliopoulos, Global optimization for the biaffine matrix inequality problem, Journal of global optimization 7, 365 (1995).

[14] P. Apkarian and H. D. Tuan, Robust control via concave minimization local and global algorithms, IEEE Transactions on Automatic Control 45, 299 (2000).

[15] M. Fukuda and M. Kojima, Branch-and-cut algorithms for the bilinear matrix inequality eigenvalue problem, Computational Optimization and Applications 19, 79 (2001).

[16] D. Davis, D. Drusvyatko, S. Kakade, and J. D. Lee, Stochastic subgradient method converges on tame functions, Foundations of computational mathematics 20, 119 (2020).

[17] J. Bolte and E. Pauwels, Conservative set valued fields, automatic differentiation, stochastic gradient methods and deep learning, Mathematical Programming 188, 19 (2021).

[18] A. Grothendieck, Around Grothendieck’s Esquisse D’un Programme, Vol. 1 (Cambridge University Press, 1997).

[19] J. Bolte, A. Danilidis, A. Lewis, and M. Shiota, Clarke subgradients of stratifiable functions, SIAM Journal on Optimization 18, 556 (2007).

[20] H. Attouch, J. Bolte, and B. F. Svaiter, Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized gaussian–seidel methods, Mathematical Programming 137, 91 (2011).

[21] J. Bolte, E. Pauwels, and R. Rios-Zertuche, Long term dynamics of the subgradient method for Lipschitz path differentiable functions, arXiv preprint arXiv:2006.00098 (2020).

[22] B. Bakker, Y. Brunebarbe, and J. Tsimerman, o-minimal gaga and a conjecture of griffiths (2018).

[23] B. Bakker, B. Klingler, and J. Tsimerman, Tame topology of arithmetic quotients and algebraicity of hodge loci (2018).

[24] B. Bakker, Y. Brunebarbe, B. Klingler, and J. Tsimerman, Definability of mixed period maps (2020).

[25] B. Bakker, T. W. Grimm, C. Schnell, and J. Tsimerman, Finiteness for self-dual classes in integral variations of hodge structure (2021).

[26] T. W. Grimm, Taming the landscape of effective theories (2021).

[27] T. W. Grimm, S. Lanza, and C. Li, Tameness, strings, and the distance conjecture (2022).

[28] J. W. Helton and V. Vinnikov, Linear matrix inequality representation of sets (2003).

[29] J. Bolte, A. Danilidis, and A. Lewis, Tame functions, Foundations of computational mathematics 11, 763 (2001).

[30] J. Bolte and E. Pauwels, Curiosities and counterexamples in smooth convex optimization (2020).
[39] Y. Nesterov and A. Nemirovskii, *Interior-point polynomial algorithms in convex programming* (SIAM, 1994).

[40] M. Kočvara and M. Stingl, *Pennon: A code for convex nonlinear and semidefinite programming*, Optimization methods and software **18**, 317 (2003).

[41] J. Bolte, A. Daniilidis, O. Ley, and L. Mazet, Characterizations of Lojasiewicz inequalities: subgradient flows, talweg, convexity, Transactions of the American Mathematical Society **362**, 3319 (2010).

[42] J. Bochnak, M. Coste, and M.-F. Roy, *Real algebraic geometry*, Vol. 36 (Springer Science & Business Media, 2013).

[43] L. van den Dries, *Tame topology and o-minimal structures*, London Mathematical Society Lecture Note Series, Vol. 248 (Cambridge University Press, Cambridge, 1998) pp. x+180.

[44] L. I. Nicolaescu, *Tame flows*, Mem. Amer. Math. Soc. **208**, vi+130 (2010).

[45] B. Klinger, *Tame geometry and hodge theory* (2020).

[46] J. C. Ponce-Campuzano and M. ´A. Maldonado-Aguilar, *C. W. J. Hol and C. W. Scherer, Sum of squares relaxations for polynomial semidefinite programming*, in *Mathematical Programming* **362**, 3319 (2010).

[47] D. Henrion, S. Naldi, and M. S. E. Din, *Exact algorithms for linear matrix inequalities* (2015).

[48] J. B. Lasserre, Global optimization with polynomials and the problem of moments, SIAM Journal on Optimization **11**, 796 (2001).

https://doi.org/10.1137/S1052623400366802.

[49] C. W. J. Hol and C. W. Scherer, A sum-of-squares approach to polynomial semidefinite programming, in *Mathematical Theory of Networks and Systems (MTNS)* (2004).

[50] C. Hol and C. Scherer, A sum-of-squares approach to fixed-order h∞ synthesis, in *Lecture notes in control and information sciences* (2005) pp. 45–71.

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

[52] J. Fiala, M. Kočvara, and M. Stingl, *Pennon* a matlab solver for nonlinear semidefinite optimization, arXiv preprint arXiv:1311.5240 (2013).

[53] E. Artin, Über die zerlegung definierter funktionen in quadrat, Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg **5**, 100 (1927).

[54] B. Reznick, Uniform denominators in hilbert quadratics, Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg **5**, 100 (1927).

[55] M. Putinar, *Positive polynomials on compact semi-algebraic sets*, Indiana University Mathematics Journal **42**, 960 (1993).

[56] M. Putinar and F.-H. Vasilescu, Positive polynomials on semi-algebraic sets, *Comptes Rendus de l’Académie des Sciences - Series I - Mathematics* **328**, 585 (1999).

[57] P. Brändén, Obstructions to determinantal representability, Advances in Mathematics **226**, 1202 (2011).

[58] H. Fawzi and M. S. El Din, A lower bound on the positive semidefinite rank of convex bodies, SIAM Journal on Applied Algebra and Geometry **2**, 126 (2018).

[59] J. Nie, *Polynomial matrix inequality and semidefinite representation* (2009).

[60] M. Zanon and S. Gros, A new dissipativity condition for asymptotic stability of discounted economic mpc, *Automatica* **141**, 110287 (2022).

[61] W. P. Adams, R. J. Forrester, and F. W. Glover, Comparisons and enhancement strategies for linearizing mixed 0-1 quadratic programs, *Discrete Optimization* **1**, 99 (2004).

[62] L. Liberti, Reformulations in mathematical programming: Definitions and systematics, RAIRO-Operations Research **43**, 55 (2009).

[63] H. S. Ryoo and N. V. Sahinidis, A branch-and-reduce approach to global optimization, Journal of global optimization **8**, 107 (1996).

[64] J. B. Lasserre, A lagrangian relaxation view of linear and semidefinite hierarchies, 52nd IEEE Conference on Decision and Control , 1966 (2013).

[65] D. Henrion and J.-B. Lasserre, Inner approximations for polynomial matrix inequalities and robust stability regions, IEEE Transactions on Automatic Control **57**, 1456 (2011).

[66] M. Benzi, G. H. Golub, and J. Liesen, *Numerical solution of saddle point problems*, Acta numerica **14**, 1 (2005).

[67] J. Nie, Z. Yang, and G. Zhou, The saddle point problem of polynomials, Foundations of Computational Mathematics , 1 (2021).

[68] D. Armentano, C. Beltrán, P. Bürgisser, F. Cucker, and M. Shub, A stable, polynomial-time algorithm for the eigenpair problem, *Journal of the European Mathematical Society* **20**, 1375 (2018).

[69] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, Attention is all you need, in *Advances in Neural Information Processing Systems*, Vol. 30, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Inc., 2017).

[70] S. Burer and R. D. Monteiro, *A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization*, Mathematical Programming **95**, 329 (2003).

[71] S. Burer and R. D. Monteiro, Local minima and convergence in low-rank semidefinite programming, *Mathematical programming* **103**, 427 (2005).

[72] M. Journé, F. Bach, P.-A. Absil, and R. Sepulchre, *Low-rank optimization for semidefinite convex problems*, arXiv preprint arXiv:0807.4423 (2008).

[73] J. Liu, A. C. Liddell, J. Mareček, and M. Takić, Hybrid methods in solving alternating-current optimal power flows, IEEE Transactions on Smart Grid **8**, 2988 (2017).

[74] N. Boumal, V. Voroninski, and A. Bandeira, The non-convex Burer-Monteiro approach works on smooth semidefinite programs, Advances in Neural Information Processing Systems **29** (2016).

[75] N. Boumal, V. Voroninski, and A. S. Bandeira, Deterministic guarantees for burer-monteiro factorizations of smooth semidefinite programs, Communications on Pure and Applied Mathematics **73**, 581 (2020).

[76] S. Naldi, *Exact algorithms for determinantal varieties and semidefinite programming*, Ph.D. thesis, Toulouse, INSA (2015).

[77] D. Henrion, S. Naldi, and M. S. E. Din, Spectra – a maple library for solving linear matrix inequalities in exact arithmetic, Optimization Methods and Software **34**, 62 (2019). https://doi.org/10.1080/10556788.2017.1341505.

[78] S. Naldi, *Solving rank-constrained semidefinite programs*
in exact arithmetic, Journal of Symbolic Computation \textbf{85}, 206 (2018).

[79] D. Henrion, S. Naldi, and M. S. El Din, Exact algorithms for semidefinite programs with degenerate feasible set, Journal of Symbolic Computation \textbf{104}, 942 (2021).

[80] S. Boyd and L. Vandenberghe, \textit{Convex Optimization} (Cambridge University Press, 2004).

[81] D. Amelunxen and P. Bürgisser, Intrinsic volumes of symmetric cones and applications in convex programming, Mathematical Programming \textbf{149}, 105 (2015).

[82] I. Waldspurger and A. Waters, Rank optimality for the Burer–Monteiro factorization, SIAM Journal on Optimization \textbf{30}, 2577 (2020) https://doi.org/10.1137/19M1255318