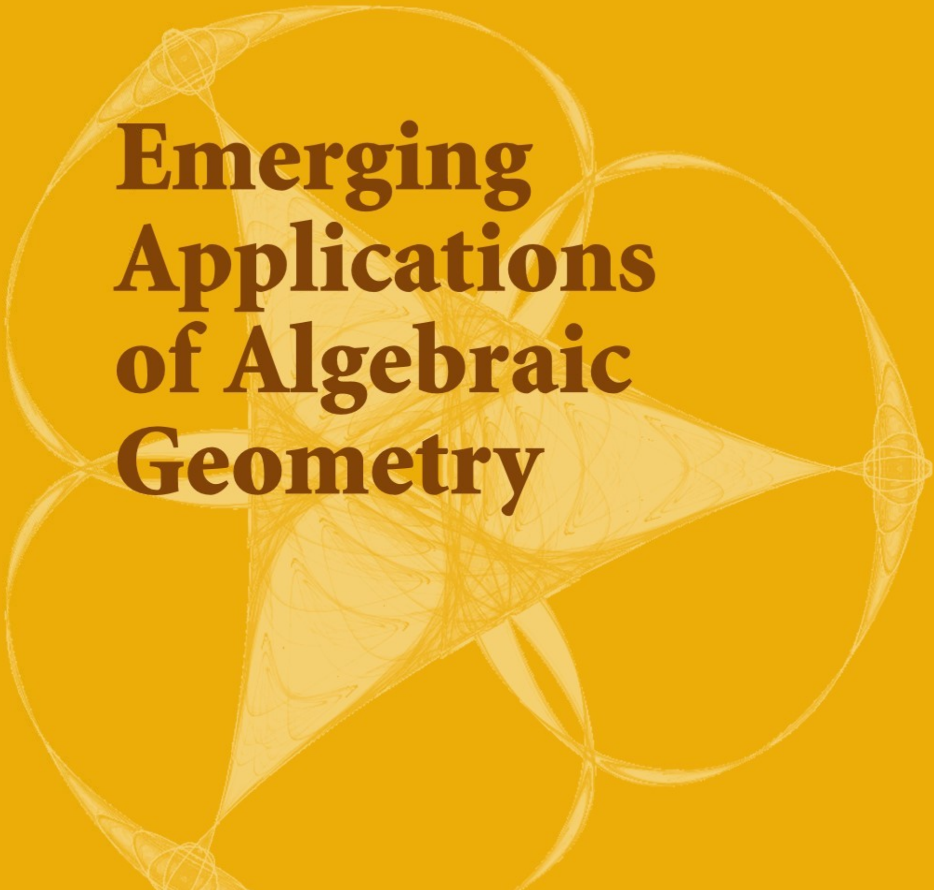
THE IMA VOLUMES IN MATHEMATICS AND ITS APPLICATIONS

EDITORS Mihai Putinar
Seth Sullivant

Emerging Applications of Algebraic Geometry



Springer

The IMA Volumes in Mathematics and its Applications

Volume 149

Series Editors
Douglas N. Arnold   Arnd Scheel

For other titles published in this series, go to
www.springer.com/series/811

Mihai Putinar Seth Sullivant
Editors

Emerging Applications of Algebraic Geometry

Springer

Editors
Mihai Putinar
Department of Mathematics
University of California
Santa Barbara, CA 93106
USA
http://www.math.ucsb.edu/~mputinar/

Seth Sullivant
Department of Mathematics
North Carolina State University
Raleigh, North Carolina 27695
USA
http://www4.ncsu.edu/~smsulli2/

Series Editors
Douglas N. Arnold
Arnd Scheel
Institute for Mathematics and its Applications
University of Minnesota
Minneapolis, MN 55455
USA

ISSN: 0940-6573
ISBN: 978-0-387-09685-8 e-ISBN: 978-0-387-09686-5
DOI: 10.1007/978-0-387-09686-5

Library of Congress Control Number: 2008932579

Mathematics Subject Classification (2000): 13P10, 14P10, 44A60, 52B12, 52C45, 62-09, 62H17, 90C22, 90C26, 92-08, 92D20, 93B25, 93B28

Camera-ready copy provided by the IMA.

9 8 7 6 5 4 3 2 1

springer.com

FOREWORD

This IMA Volume in Mathematics and its Applications

EMERGING APPLICATIONS OF ALGEBRAIC GEOMETRY

contains papers presented at two highly successful workshops, Optimization and Control, held January 16–20, 2007, and Applications in Biology, Dynamics, and Statistics, held March 5–9, 2007. Both events were integral parts of the 2006-2007 IMA Thematic Year on "Applications of Algebraic Geometry." We are grateful to all the participants for making these workshops a very productive and stimulating events. The organizing committee for the first workshop are Dimitris Bertsimas (Sloan School of Management, Massachusetts Institute of Technology), J. William Helton (Mathematics, University of California - San Diego), Jean Bernard Lasserre (LAAS-CNRS, France), and Mihai Putinar (Mathematics, University of California - Santa Barbara), and for the second workshop, Serkan Hosten (Mathematics, San Francisco State University), Lior Pachter (Mathematics, University of California - Berkeley), and Seth Sullivant (Mathematics, North Carolina State University). The IMA thanks them for their excellent work.

We owe special thanks to Mihai Putinar and Seth Sullivant for their superb role as editors of these proceedings. We take this opportunity to thank the National Science Foundation for its support of the IMA.

Series Editors

Douglas N. Arnold, Director of the IMA

Arnd Scheel, Deputy Director of the IMA

PREFACE

Algebraic geometry, that noble and refined branch of pure mathematics, resurfaces into novel applications. The present volume contains only a portion of these emerging new facets of modern algebraic geometry, reflected by the interdisciplinary activities hosted during the academic year 2006–2007 by the IMA- the Institute for Mathematics and its Applications, at the University of Minnesota, Minneapolis.

What has algebraic geometry to do with non-convex optimization, the control theory of complicated engineering systems, phylogenetic tree reconstruction, or the statistical analysis of contingency tables? The answer is: quite a lot. And all this is due to innovative ideas and connections discovered in the last decade between pure and applied mathematics. The reader of the present volume will find detailed and informative answers to all of the above questions, presented in a self-contained form by experts working on the boundary of algebraic geometry and one specific area of its applications.

Two major conferences organized at IMA: *Optimization and Control* (January 2007) and *Applications in Biology, Dynamics, and Statistics* (March 2007) gathered mathematicians of all denominations, engineers, biologists, computer scientists and statisticians, linked by the common use of a few basic methods of algebraic geometry. Among the most important of these methods: Positivstellensätze over the real field, elimination theory, and invariant theory, all with a computational/numerical component predominant in the background.

In order to explain why optimization and control theory appear naturally connected to real algebraic geometry one has to go back a century ago and recall that both the positivity of regular functions (polynomials for instance) and of linear transformations in a inner product space are the byproduct of a universal structure: they are representable by simple operations involving sums of squares in the respective algebras. Heuristically speaking, when proving inequalities one most often completes squares. It was Hilbert who unveiled the central role of sums of squares decompositions (see his 17-th problem) and their logical ramifications. On the algebraic side, his ideas were crystalized and developed by Emil Artin into the theory of real closed fields. Second, Hilbert's footprint was cast in the spectral theorem for self-adjoint linear transformations, which is essentially equivalent to an infinite sum of squares decomposition of a specific bilinear form. Third, Hilbert's program of revising the logical foundations of mathematics found a solid validation in the discovery due to Tarski, that any proposition of the first order involving only real variables is decidable (in particular implying that one can verify that every system of polynomial inequalities in $\mathbb{R}^d$ has a solution). Progress in real algebraic geometry rendered Tarski's theorem even more precise with the discovery of the so-called Positivstellensätze, that is theorems which quantify in algebraic certificates the existence of solutions of polynomial inequalities (e.g.: $b^2 - c \geq 0$ implies that there exists a real x satisfying $x^2 - 2bx + c = 0$).

It comes then as no surprise that the extremal problem

$$p^* = \min\{p(u);\ u \in S\},$$

where p is a polynomial and S is a subset of $\mathbb{R}^d$ defined by finitely many polynomial inequalities can be converted into algebraic identities which depend only on the coefficients of p and the minimal value p^*. At this point optimization theorists (and by extension control theory engineers) became aware that one can solve such extremal problems by global algebraic means, rather than by, for instance, the time consuming inner point technique.

An unexpected turn into this history was marked by an application of duality theory of locally convex spaces leading to the simplification of the known Positivstellensätze, by interpreting them as solvability criteria for some associated moment problems. In their own, moment problems have a glorious past and an unexpected vitality. Once the relation between real algebraic geometry and moment problems was clarified, this has greatly enhanced the applications of real algebraic geometry to optimization. These days we witness a proliferation of relaxation methods in polynomial minimization problems based on the introduction of new, virtual moment variables. The articles contained in this volume amply illustrate this new direction of research.

The emergence of algebraic geometry as a tool in statistics began with the work of Diaconis and Sturmfels whose research led to the construction of Markov chains for performing Fisher's exact test. They showed that the local moves in these Markov chains are the generators of an associated toric ideal. It was quickly realized that the zero set of this toric ideal in the probability simplex is precisely the set of distributions in the corresponding log-linear model.

The observation that the zero set of a toric ideal in the probability simplex is a statistical model led to the realization that many other statistical models have the structure of a semi algebraic set which can be exploited for various purposes. The foremost examples are hierarchical and graphical models, which include as special cases many of the statistical models used in computational biology.

Of course, this description of the emergence of algebraic geometry as a tool in statistics and its applications is brief and highly incomplete. In particular, it leaves out the algebraic developments in the design of experiments, the use of polynomial dynamical systems over finite fields to investigate gene regulatory networks, polyhedral tools for parametric inference, optimization problems in statistical disclosure limitation, and a slew of other connections between statistics and algebraic geometry, some of which are highlighted in this volume.

A short description of the articles that appear in the volume follows. For a general perspective, more details and a better description we recommend the reader to consult the introductions to any one of the contributions.

1. The article **Polynomial optimization on odd-dimensional spheres** by John d'Angelo and Mihai Putinar contains a basic Positivstellensatz with supports on an odd dimensional sphere, written in terms of hermitian sums of squares and obtained with methods of functional analysis, complex analysis and complex geometry.
2. The survey **Engineering systems and free semi-algebraic geometry** by Mauricio de Oliveira, John W. Helton, Scott McCullough and Mihai Putinar illustrates the ubiquity of positivity computations in a finitely generated free *-algebra in the study of dimensionless matrix inequalities arising in engineering systems.
3. The article **Algebraic statistics and contingency table problems: estimation and disclosure limitation** by Adrian Dobra, Stephen Fienberg, Alessandro Rinaldo, Aleksandra Slavkovic, and yi Zhou describes the use of algebraic techniques for contingency table analysis, and provides many open problems, especially in the area of statistical disclosure limitation.
4. Nicholas Eriksson's article **Using invariants for phylogenetic tree reconstruction** describes some of the mathematical challenges in using the polynomials that vanish on probability distributions of phylogenetic models as tools for inferring phylogenetic trees.
5. Abdul Jarrah and Reinhard Laubenbacher's article **On the algebraic geometry of polynomial dynamical systems** describes their work with collaborators on using dynamical systems over finite fields to model biological processes, as well as the mathematical advanced made on studying the dynamical behavior of special classes of these dynamical systems.
6. Jean Bernard Lasserre, Monique Laurent and Philip Rostalski in their article **A unified approach to computing real and complex zeros of zero dimensional ideals** introduce a powerful and totally new idea of finding the support of ideals of finite codimension in the polynomial ring. They propose a relaxation based on finitely many moments as auxiliary variables.
7. Monique Laurent's ample survey **Sums of squares, moment matrices and optimization of polynomials** is an invaluable source, as the only one of this kind, for the rapidly growing relaxation methods aimed at solving in optimal time minimization problems on polynomial functions restricted to semi-algebraic sets. All presented in a very clear, systematic way, with a spin towards the computational complexity, algorithmic aspects and numerical methods.

8. Claus Scheiderer's survey **Positivity and sums of squares: a guide to recent results** contains a lucid and comprehensive overview of areas of real algebraic geometry contingent to the concept of positivity. This unique synthesis brings the subject up to our days, is informative for all readers and it is accessible to non-experts.
9. Konrad Schmüdgen's **Non-commutative real algebraic geometry: some basic concepts and first ideas** is an essay about the foundations of real algebraic geometry over a *-algebra, based on the author's recent results in enveloping algebras of Lie algebras.
10. Bernd Sturmfels' article **Open problems in algebraic statistics** provides a summary of his lecture in the workshop *Applications in Biology, Dynamics, and Statistics* and covers a number of mathematical open problems whose solutions would greatly benefit this emerging area.

The completion of this volume could have not been achieved without the enthusiasm and professionalism of all contributors. The invaluable technical support of the experienced staff at IMA led to a high quality production of the book, in record time. We warmly thank them all.

Mihai Putinar
Department of Mathematics
University of California at Santa Barbara
http://math.ucsb.edu/~mputinar/

Seth Sullivant
Department of Mathematics
North Carolina State University
http://www4.ncsu.edu/~smsulli2/

CONTENTS

Foreword v

Preface vii

Polynomial optimization on odd-dimensional spheres 1
John P. D'Angelo and Mihai Putinar

Engineering systems and free semi-algebraic geometry 17
Mauricio C. de Oliveira, J. William Helton, Scott A. McCullough, and Mihai Putinar

Algebraic statistics and contingency table problems: Log-linear models, likelihood estimation, and disclosure limitation 63
Adrian Dobra, Stephen E. Fienberg, Alessandro Rinaldo, Aleksandra Slavkovic, and Yi Zhou

Using invariants for phylogenetic tree construction 89
Nicholas Eriksson

On the algebraic geometry of polynomial dynamical systems 109
Abdul S. Jarrah and Reinhard Laubenbacher

A unified approach to computing real and complex zeros of zero-dimensional ideals 125
Jean Bernard Lasserre, Monique Laurent, and Philipp Rostalski

Sums of squares, moment matrices and optimization over polynomials 157
Monique Laurent

Positivity and sums of squares: A guide to recent results 271
Claus Scheiderer

Noncommutative real algebraic geometry some basic concepts and first ideas 325
Konrad Schmüdgen

Open problems in algebraic statistics 351
Bernd Sturmfels

List of workshop participants 365

POLYNOMIAL OPTIMIZATION ON ODD-DIMENSIONAL SPHERES

JOHN P. D'ANGELO* AND MIHAI PUTINAR†

Abstract. The sphere S^{2d-1} naturally embeds into the complex affine space $\mathbb{C}^d$. We show how the complex variables in $\mathbb{C}^d$ simplify the known Striktpositivstellensätze, when the supports are resticted to semi-algebraic subsets of odd dimensional spheres. We also illustrate the subtleties involved in trying to control the number of squares in a Hermitian sum of squares.

Key words. Positive polynomial, Hermitian square, unit sphere, plurisubharmonic function, Cauchy-Riemann manifold.

AMS(MOS) subject classifications. Primary 14P10, Secondary 32A70.

1. Introduction. A deep observation going back to the work of Tarski in the 1930-ies implies the decidability (in a very strong, algorithmic sense) of the statement "there exists a real solution $x \in \mathbb{R}^n$ of a system of real polynomial inequalities". What is called today the algebraic certificate that such a system has a solution is condensed into the basic Positiv- and Null-stellensätze, discovered by Stengle in the 1970-ies (see for instance for details [23]).

All these topics are relevant to and depend on the structure of the convex cone $\mathbb{R}_+(K)$ of positive polynomials on a given basic semi-algebraic set $K \subset \mathbb{R}^n$ and on algebraic refinements of it, such as the pre-order associated to a system of defining equations of K and the corresponding quadratic module, both regarded as convex cones of the polynomial algebra. Thus, it comes as no surprise that duality techniques of functional analysis (i.e. the structure of non-negative functionals on these convex cones) reflect at the level of purely algebraic statements the mighty powers coming from a different territory. More precisely the spectral theorem for commuting tuples of self-adjoint operators simplifies some of the known Positivstellensätze, see for instance [30, 26].

The present note is an illustration of the latter transfer between Hilbert space techniques and real algebraic geometry. We specialize to basic semi-algebraic sets $K \subset S^{2d-1} \subset \mathbb{R}^{2d}$of an odd-dimensional sphere, and prove that the expected Positivstellensatz (i.e. the structure theorem of a positive polynomial on such a set) can further be simplified by means of the induced complex variables $\mathbb{R}^{2d} = \mathbb{C}^d$. To this aim we greatly benefit from the

*Department of Mathematics, University of Illinois at Urbana-Champaign, Urbana, IL 61801 (jpda@math.uiuc.edu). The first author was partially supported by the National Science Foundation grant DMS-0500765.

†Mathematics Department, University of California, Santa Barbara, CA 93106 (mputinar@math.ucsb.edu). The second author was supported in part by the Institute for Mathematics and its Applications (Minneapolis) with funds provided by the National Science Foundation and by the NSF grant DMS-0701094.

works of several functional analysts interested into some tuples of Hilbert space operators known as spherical isometries, and separately from the works of geometers and complex analysts dealing with Cauchy-Riemann manifolds and proper holomorphic maps between Euclidean balls. The following pages reveal some fascinating new frontiers for future research, such as the rigidity phenomenon discussed in the last section.

2. Preliminaries. Let $\mathbb{C}^d$ denote complex Euclidean space with Euclidean norm given by $|z|^2 = \sum_{j=1}^{d} |z_j|^2$. The unit, odd dimensional sphere

$$S^{2d-1} = \{z \in \mathbb{C}^d;\ |z| = 1\}$$

is a particularly important example of a Cauchy-Riemann (usually abbreviated CR) manifold. This note will show how one can study problems of polynomial optimization over semi-algebraic subsets of S^{2d-1} by using the induced Cauchy-Riemann structure. Our results can be regarded as multivariate analogues of classical phenomena about positive trigonometric polynomials, known for a long time in dimension one ($d = 1$). They are also related to results concerning proper holomorphic mappings between balls in different dimensional complex Euclidean spaces and the geometry of holomorphic vector bundles. See [9] for an exposition of these connections.

A polynomial map $p : \mathbb{C}^d \times \mathbb{C}^d \to \mathbb{C}$ is called *Hermitian symmetric* if

$$p(z, \overline{w}) = \overline{p(w, \overline{z})}$$

for all z and w. By polarization one can recover a Hermitian symmetric polynomial from its real values $p(z, \overline{z})$. We therefore work on the diagonal (where $w = z$) and let $\mathcal{H} \subset \mathbb{C}[z, \overline{z}]$ denote the space of Hermitian symmetric polynomials on $\mathbb{C}^d$. Note that $\mathcal{H}$ is a real algebra, naturally isomorphic to the polynomial algebra $\mathbb{R}[x, y]$, where $z = x + iy \in \mathbb{R}^d + i\mathbb{R}^d$. Henceforth we will freely identify a Hermitian symmetric polynomial $P(z, \overline{z})$ with its real form $p(x, y) = P(x + iy, x - iy)$.

We denote by $\Sigma^2\mathcal{H}$ the convex cone consisting of sums of squares of Hermitian polynomials. We denote by $\Sigma_h^2\mathcal{H}$ the convex cone consisting of polynomials which are squared norms of (holomorphic) polynomial mappings. Thus $R \in \Sigma_h^2\mathcal{H}$ if there exist polynomials $p_j \in \mathbb{C}[z]$ such that

$$R(z, \overline{z}) = \sum_{j=1}^{m} |p_j(z)|^2.$$

See [11] and [1] for various characterizations of $\Sigma_h^2\mathcal{H}$. We have the containment

$$\Sigma_h^2\mathcal{H} \subset \Sigma^2\mathcal{H},$$

simply because

$$|p|^2 = \left(\frac{p+\overline{p}}{2}\right)^2 + \left(\frac{p-\overline{p}}{2i}\right)^2 = u^2 + v^2,$$

where u and v are the real and imaginary parts of p. The containment is strict as illustrated by the following two examples.

EXAMPLE a). In one variable we define a polynomial R by

$$R(z,\overline{z}) = (z+\overline{z})^2 = 4x^2.$$

It is evidently a square but not in $\Sigma_h^2\mathcal{H}$. Note that the zero set of an element in $\Sigma_h^2\mathcal{H}$ must be a complex variety and thus cannot be the imaginary axis.

EXAMPLE b). In two variables we define $R(z,\overline{z}) = (|z_1|^2 - |z_2|^2)^2$. Again R lies in $\Sigma^2\mathcal{H}$ but not in $\Sigma_h^2\mathcal{H}$. Here one can observe that elements of $\Sigma_h^2\mathcal{H}$ must be plurisubharmonic but that R is not. In 4.3 we will show additionally that R cannot be written as a squared norm on the unit sphere.

In this paper we are primarily concerned with optimization on the sphere. We therefore first let $I = I(S^{2d-1})$ be the ideal of $\mathcal{H}$ consisting of all polynomials vanishing on S^{2d-1}. We then define

$$\mathcal{H}(S^{2d-1}) = \mathcal{H}/I,$$

and regard it as a space of polynomial functions defined on the sphere. As a matter of fact, each real-valued polynomial p has a representative in $\mathcal{H}(S^{2d-1})$, when p is regarded as a function on the sphere.

In analogy with the above notations we denote by $\Sigma^2\mathcal{H}(S^{2d-1})$ the convex cone consisting of sums of squares of Hermitian polynomials on the sphere. We denote by $\Sigma_h^2\mathcal{H}(S^{2d-1})$ the convex hull of Hermitian squares:

$$\Sigma_h^2\mathcal{H}(S^{2d-1}) = \mathrm{co}\{|p(z)|^2;\ p \in \mathbb{C}[z]\} \mod I.$$

A polynomial that is positive on the sphere must agree with the squared norm of a holomorphic polynomial mapping there. In the final section we naturally ask what is the minimum number of terms in the representation as a squared norm. This difficult question is particularly natural for at least three reasons: there is considerable literature on this problem in the real case, the Hermitian case is closely connected with difficult work on the classification of proper holomorphic mappings between balls, and finally because number-theoretic analogues such as Waring's problem have appealed to mathematicians for centuries. We consider only the second issue here. In Section 5 we provide a striking non-trivial example illustrating some of the subtleties.

Let us return to the general situation and recall a classical one-dimensional result which is guiding our investigation. We include its elementary proof for convenience.

LEMMA 2.1 (Riesz-Fejér). *A non-negative trigonometric polynomial is the squared modulus of a trigonometric polynomial.*

Proof. Let $p(e^{i\theta}) = \sum_{-d}^{d} c_j e^{ij\theta}$ and assume that $p(e^{i\theta}) \geq 0, \quad \theta \in [0, 2\pi]$. Since p is real-valued $c_{-j} = \overline{c_j}$ for all j. We set $z = |z|e^{i\theta}$ and extend p to the rational function defined by $p(z) = \sum_{-d}^{d} c_j z^j$. It follows that $p(z) = \overline{p(1/\overline{z})}$; furthermore its zeros and poles are symmetrical (in the sense of Schwarz) with respect to the unit circle.

Write $z^d p(z) = q(z)$. Then q is a polynomial of degree $2d$ whose modulus $|q|$ satisfies $|q| = |p| = p$ on the unit circle. In view of the mentioned symmetry one finds

$$q(z) = cz^{\nu} \prod_j (z - \lambda_j)^2 \prod_k (z - \mu_k)(z - 1/\overline{\mu_k}),$$

where c is a constant, $|\lambda_j| = 1$ and $0 < |\mu_k| < 1$.

Evaluating on the circle and using $|\zeta^2| = |\zeta|^2$ we obtain

$$p(e^{i\theta}) = |p(e^{i\theta})| = |q(e^{i\theta}| = |c| \prod_j |e^{i\theta} - \lambda_j|^2 \prod_k \frac{|e^{i\theta} - \mu_k|^2}{|\mu_k|^2},$$

and hence p is the squared modulus of a trigonometric polynomial. □

This fundamental lemma has deeply influenced twentieth century functional analysis. For instance the Riesz-Fejér Lemma is equivalent to the spectral theorem for unitary operators; see [29].

When invoking duality, the above is not less interesting. It was in this form that Riesz-Fejér Lemma was first generalized to an arbitrary dimension.

LEMMA 2.2. *Let $L \in \mathcal{H}(S^{2d-1})'$ be a linear functional which is non-negative on $\Sigma_h^2 \mathcal{H}(S^{2d-1})$. Then L is represented by a positive Borel measure supported on the sphere.*

The proof has implicitly appeared in the works of Ito [17], Yoshino [33], Lubin [22] and Athavale [3], all dealing with subnormality criteria for commuting tuples of bounded linear operators. Without aiming at completeness, here is the main idea.

Proof (Sketch). Let L be a non-negative functional on $\Sigma_h^2 \mathcal{H}(S^{2d-1})$. Fix a polynomial $p \in \mathbb{C}[z]$ and consider the functional

$$f(r_1^2, ..., r_d^2) \mapsto L(f|p(z)|^2), \quad f \in \mathbb{R}[r_1^2, ..., r_d^2],$$

where $r_j^2 = |z_j|^2$. Since

$$1 - |z_j|^2 = \sum_{k \neq j} |z_k|^2,$$

$$L\Big(\prod_j [(1-r_j^2)^{n_j} r_j^{2m_j}]|p|^2\Big) \geq 0, \qquad n_j, m_j \geq 0.$$

By a classical result of Haviland, see for instance [2], there exists a positive Borel measure $\mu_{|p|^2}$ on the simplex Δ defined by

$$\Delta = \{(r_1^2, ..., r_d^2);\ r_1^2 + ... + r_d^2 = 1\},$$

with the property

$$L(f|p(z)|^2) = \int_\Delta f d\mu_{|p|^2}.$$

The total mass of $\mu_{|p|^2}$ is $L(|p|^2)$.

By polarization, one can define complex valued measures by

$$L(fp\overline{q}) = \int_\Delta f d\mu_{p\overline{q}}, \quad f \in \mathbb{R}[r_1^2, ..., r_d^2],\ p, q \in \mathbb{C}[z],$$

so that the sesqui-linear kernel $(p, q) \mapsto \mu_{p\overline{q}}$ is positive semi-definite.

In short, the functional L can be extended to the linear space of functions (on the sphere) of the form

$$F(r, z) = \sum_{|\alpha| \leq n} c_\alpha(r) z^\alpha,$$

where $c_\alpha(r)$ are bounded, Borel measurable functions on the simplex Δ. The extended functional $\tilde{L}$ still satisfies

$$\tilde{L}(|F(r, z)|^2) \geq 0.$$

Next we pass to polar coordinates $z_j = r_j \omega_j$, $|\omega_j| = 1$ and remark that multiplication by ω_j satisfies the isometric condition

$$\tilde{L}(|\omega_j F(r, z)|^2) = \tilde{L}(|F(r, z)|^2).$$

Thus, we can further extend the functional $\tilde{L}$ to all polynomials in r and $\omega, \overline{\omega}$, so that

$$\tilde{L}(|\omega_j^{-1} F(r, z)|^2) = \tilde{L}(|F(r, z)|^2)$$

and

$$\tilde{L}(|p(r, \omega, \overline{\omega})|^2) \geq 0.$$

We refer to [33] or [31] for the details how this extension is constructed. By rewriting the latter positivity condition we have in particular

$$\tilde{L}(|h(z, \overline{z})|^2) \geq 0, \quad h \in \mathbb{C}[z, \overline{z}],$$

whence, by the Stone-Weierstrass Theorem and the Riesz Representation Theorem, the functional $\tilde{L}$ is represented by a positive Borel measure, supported on the sphere.

The representing measure is unique by the Stone-Weierstrass Theorem. □

3. A Striktpositivstellensatz. We now turn to the basic question considered in this paper. We are given a finite set of real polynomials in $2d$ variables $p, q_1, ..., q_r$, or equivalently, Hermitian symmetric polynomials in d complex variables. We suppose that $p(z, \overline{z})$ is strictly positive on the subset of S^{2d-1} where each q_j is nonnegative. Can we write p as a weighted sum of squared norms with q_i as weights, as the real affine Striktpositivstellensatz (see for instance [23]) suggests? The answer is yes, and we can offer at least two different reasons why it is so.

THEOREM 3.1. *Let* $p, q_1, ..., q_r \in \mathbb{R}[x, y]$, *where* $x + iy = z \in \mathbb{C}^d$. *If*

$$(|z| = 1,\ q_i(z, \overline{z}) \geq 0, 1 \leq i \leq r) \ \Rightarrow \ (p(z, \overline{z}) > 0),$$

then

$$p \in \Sigma_h^2 + q_1\Sigma_h^2 + ... + q_r\Sigma_h^2 + I(S^{2d-1}).$$

First we discuss the history of such Hermitian squares decompositions, in the case where there are no constraints. A Hermitian symmetric polynomial p is called bihomogeneous of degree (m, m) if

$$p(\lambda z, \overline{\lambda z}) = |\lambda|^{2m} p(z, \overline{z})$$

for all complex numbers λ and all $z \in \mathbb{C}^d$. The values of a bihomogeneous polynomial are determined by its values on the sphere. When p is bihomogeneous and strictly positive on the sphere, Quillen [28] proved that there is an integer k and a homogeneous polynomial vector-valued mapping $h(z)$ such that

$$|z|^{2k} p(z, \overline{z}) = |h(z)|^2.$$

This result was discovered independently by the first author and Catlin [6] in conjunction with the first author's work on proper mappings between balls in different dimensions. The proof in [6] uses the Bergman projection and some facts about compact operators, and it generalizes to provide an isometric imbedding theorem for certain holomorphic vector bundles [7].

It is worth noting that the integer k and the number of components of h can be arbitrarily large, even for polynomials p of total degree four in two variables. The result naturally fits into the phenomena encoded into the old or recent Positivestellensätze, see for instance [23]. For the specific case of Hermitian polynomials on spheres see [8] for considerable discussion and generalizations.

Using a process of bihomogenization, Catlin and the first author (see [6, 8] and [9]) proved that if p is arbitrary (not necessarily bihomogeneous) and strictly positive on the sphere, then p agrees with a squared norm on the sphere; in other words, $p \in \Sigma_h^2 + I(S^{2d-1})$. Thus Theorem 1 holds when there are no constraints. Our proof of Theorem 1 first considers the case of no constraints, but we approach this case in a completely different manner.

Strict positivity is required for these results. The polynomial $(|z_1|^2 - |z_2|^2)^2$ is bihomogeneous and nonnegative everywhere, but there is no element in Σ_h^2 agreeing with it on the sphere. See Example 4.3 below.

Proof of Theorem 1. Suppose first that no q_i's are present and assume by contradiction that $p \notin \Sigma_h^2$, all regarded as elements of $\mathcal{H}(S^{2d-1})$. Since the constant function 1 belongs to the algebraic interior of the convex cone $\mathcal{H}(S^{2d-1})$, the separation lemma due to Eidelheit-Kakutani [12, 18] provides a linear functional $L \in \mathcal{H}(S^{2d-1})'$, satisfying both $L(1) > 0$ and

$$L(p) \le 0 \le L|_{\Sigma_h^2}.$$

According to Lemma 2, there exists a positive Borel measure μ, supported on the unit sphere, which represents L. Therefore

$$0 \geq L(p) = \int p d\mu > 0,$$

a contradiction.

The proof of the general case is similar, with the difference that we have to prove that the support of the measure μ is contained in the nonnegativity set defined by the functions q_i. To this aim, fix an index i, and remark that

$$\int q_i |p|^2 d\mu \geq 0$$

for all $p \in \mathbb{C}[z]$. Now, by the first case, every positive polynomial $P(z, \overline{z})$ is in the convex hull of the Hermitian squares, whence

$$\int q_i P(z, \overline{z}) d\mu \geq 0$$

whenever $P(z, \overline{z}) > 0$ on the sphere, that is whenever $P(z, \overline{z}) \geq 0$ on the sphere. In view of Stone-Weierstrass Theorem, every continuous functions f on the sphere can be uniformly approximated by real polynomials. In particular, we infer

$$\int q_i f^2 d\mu \geq 0, \quad f \in C(S^{2d-1}).$$

But this inequality holds only if the support of μ is contained in the nonnegativity set $q_i(z, \overline{z}) \geq 0$. $\square$

4. Examples.

4.1. Optimization on the closed disk. The following simple example shows that Hermitian sums of squares do not suffice as positivity certificates on more general semi-algebraic sets. Specifically, let

$$P(z,\overline{z}) = 1 - \frac{4}{3}|z|^2 + a|z|^4,$$

with $\frac{1}{3} < a$. Note that

$$P(z,\overline{z}) = \left(1 - \frac{2}{3}|z|^2\right)^2 + \left(a - \frac{4}{9}\right)|z|^4,$$

and hence $P \in \Sigma^2\mathcal{H}$ when $a \geq \frac{4}{9}$. Hence we assume $\frac{1}{3} < a < \frac{4}{9}$. The polynomial $1 - \frac{4}{3}t + at^2$ is decreasing for $0 < t < 1$ when $a < \frac{2}{3}$; therefore $|z| \leq 1$ implies $P(z,\overline{z}) \geq 1 + a - \frac{4}{3} > 0$.

On the other hand,

$$P \notin \Sigma_h^2 + (1 - |z|^2)\Sigma_h^2.$$

To see that P is not in this set, we apply the hereditary calculus. See [1] for details. We replace z with a contractive operator T and replace $\overline{z}$ with T^*. We follow the usual convention of putting all T^*'s to the left of the powers of T. If P were in this set, we would obtain

$$\|T\| \leq 1 \quad \Rightarrow \quad p(T,\overline{T}) \geq 0.$$

In particular let T be the 2×2 Jordan block with 1 above the diagonal. We obtain a contradiction by computing that $P(T,T^*)$ is the diagonal matrix with eigenvalues 1 and $-\frac{1}{3}$.

On the other hand, the larger convex cone $\Sigma^2 + (1 - |z|^2)\Sigma_h^2$ is appropriate in this case, see [25, 27].

4.2. Squared norms. Recall that $\Sigma_h^2\mathcal{H}$ denotes the convex cone consisting of polynomials which are squared norms of (holomorphic) polynomial mappings. In all dimensions the zero set of an element in $\Sigma_h^2\mathcal{H}$ must be a complex variety.

Suppose $R(z,\overline{z}) \geq 0$ for all z. Even in one dimension we cannot conclude that $R \in \Sigma_h^2\mathcal{H}$. We noted earlier, where $x = \mathrm{Re}(z)$, the example

$$R(z,\overline{z}) = (z + \overline{z})^2 = 4x^2.$$

The zero set of R is the imaginary axis, which has no complex structure. In one dimension of course, the zero set of an element in $\Sigma_h^2\mathcal{H}$ must be either all of $\mathbb{C}$ or a finite set.

Things are more complicated and interesting in higher dimensions. Consider the following example from [8]. Define a Hermitian bihomogeneous polynomial in three variables by

$$p(z,\overline{z}) = (|z_1z_2|^2 - |z_3|^4)^2 + |z_1|^8.$$

This polynomial p is nonnegative for all z, and its zero set is the complex plane given by $z_1 = z_3 = 0$ with z_2 arbitrary. Yet p is not a sum of squared moduli; even more striking is that p cannot be written as the quotient $\frac{|a(z)|^2}{|b(z)|^2}$ where a and b are sums of squared moduli. See [10] for additional information on this example and several tests for deciding whether a nonnegative polynomial R can be written as a quotient of squared norms. See [32] for a necessary and sufficient condition involving the zeroes of R.

We give an additional example in one dimension. Define p by

$$p(z, \overline{z}) = 1 + bz^2 + \overline{b}\overline{z}^2 + c|z|^2 + |z|^4.$$

The condition for being a quotient of squared norms is that one of the following three statements holds:

$$c > 2|b|^2 - 2,$$

$$b = 0, \quad c > -2,$$

$$|b| = 1, \quad c = 0.$$

The condition for being nonnegative is simpler: $c \geq 2|b| - 2$.

4.3. Proof of Example b). We claimed earlier that the polynomial $(|z_1|^2 - |z_2|^2)^2$ is bihomogeneous and nonnegative everywhere, but there is no element in Σ_h^2 agreeing with it on the sphere.

Proof. Put $R(z, \overline{z}) = (|z_1|^2 - |z_2|^2)^2$, and let $V(R)$ denote its zero set. We note that $V(R) \cap S^{2n-1}$ is the torus T defined by $|z_1|^2 = |z_2|^2 = \frac{1}{2}$. Suppose for some polynomial mapping $z \to P(z)$ we have $R = |P|^2$ on the unit sphere. Note first that the zero set of $|P|^2$ is a complex variety. We have $|P(z)|^2 = 0$ for $z \in T$. We claim that P is identically zero. For each fixed z_2 with $|z_2| = 1$, the vector-valued polynomial mapping $z_1 \to P(z_1, z_2)$ vanishes on the circle $|z_1|^2 = \frac{1}{2}$ and hence vanishes identically. Since z_2 was an arbitrary point with $|z_2|^2 = \frac{1}{2}$ we conclude that the mapping $(z_1, z_2) \to P(z_1, z_2)$ vanishes whenever $z_1 \in \mathbf{C}$ and z_2 lies on a circle. By symmetry it also vanishes with the roles of the variables switched. It follows that the zero set of P (which is a complex variety) is at least three real dimensions, and hence P vanishes identically. Since R does not vanish identically on the sphere we obtain a contradiction. □

4.4. Example. There exist non-negative polynomials R such that R is not in $\Sigma_h^2\mathcal{H}$, yet there is a positive integer N for which $R^N \in \Sigma_h^2\mathcal{H}$. The bihomogeneous polynomial R_λ given by

$$R_\lambda(z, \overline{z}) = (|z_1|^2 + |z_2|^2)^4 - \lambda|z_1 z_2|^4$$

satisfies this property whenever $\lambda < 8$. See [11] and [32]. For $\lambda < 16$, $R_\lambda > 0$ on the sphere. By Theorem 1 it agrees with a squared norm *on the sphere.*

5. Ranks of Hermitian forms on spheres and spectral values. Let R be a Hermitian symmetric polynomial. In this section we consider how many terms are needed to write R as an Hermitian sum of squares on the unit sphere. As we mentioned in the introduction, the real variables analogue of this problem is already quite appealing. It corresponds to the case when certain Hermitian forms are diagonal, and hence things are much easier. We therefore begin with a simple real variables example, observe an interesting phenomenon, and then turn to its Hermitian version.

Let p be a homogeneous polynomial of several real variables. Suppose $p(x) > 0$ for all x in the nonnegative orthant, the set where each $x_j \geq 0$. By a classical theorem of Polya, there is an integer d such that the polynomial f defined by $(\sum_i x_i)^d p(x) = f(x)$ has only positive coefficients. See [8] for considerable discussion of Polya's classical theorem and its Hermitian analogues. Here we will discuss a simple example where we are concerned with the number of terms involved.

Consider the one parameter family of real polynomials on $\mathbf{R}^2$ defined by $p_\lambda(x, y) = x^2 - \lambda xy + y^2$. Each p_λ is homogeneous of degree 2, and hence is determined by its values on the line given by $x + y = 1$. We ask when we can find a polynomial f with nonnegative coefficients and which agrees with p_λ on this line. We also want to know the minimum number N_λ of terms f must have. For $\lambda > 2$, the polynomial has negative values, and hence cannot be a sum of terms with positive coefficients. The same conclusion holds at the border case when $\lambda = 2$. When $\lambda < 2$, Polya's theorem guarantees that such an f exists. The number N_λ tends to infinity as λ tends to two. The value $\lambda = 1$ plays a surprising special role. On the line $x + y = 1$ we can write

$$x^2 - xy + y^2 = \frac{x^3 + y^3}{x + y} = x^3 + y^3.$$

Thus $N_\lambda = 2$ when $\lambda = 1$. On the other hand, $N_\lambda > 2$ for $0 < \lambda < 1$. Thus the minimum number of squares needed is not monotone in λ. This striking phenomenon also holds in the Hermitian case; we could create the Hermitian analogue simply by writing $x = |z_1|^2$ and $y = |z_2|^2$.

Let now R denote a Hermitian symmetric polynomial. Suppose $R \geq 0$ as a function. We write R in the form

$$R(z, \overline{z}) = \sum c_{\alpha\beta} z^\alpha \overline{z}^\beta$$

and we know that R is itself a squared norm if and only if the matrix of coefficients $(c_{\alpha\beta})$ is non-negative definite. In this case there is an integer N and holomorphic polynomials f_j such that

$$R(z, \overline{z}) = \sum_{j=1}^{N} |f_j(z)|^2 = |f(z)|^2.$$

By elementary linear algebra, the minimum possible N equals the rank of the matrix of coefficients. Thus the global problem is easy.

Things are considerably different on the sphere. For example, with the right choice of nonnegative constants c_j and integer K, the expression

$$1 - \sum_{j=1}^{K} c_j |z|^{2j}$$

will be strictly positive on the sphere while the underlying matrix of coefficients will have arbitrarily many negative eigenvalues. Suppose however that we write R as a squared norm $|f|^2$ on the sphere. What can we say about N, the rank of the coefficient matrix of $|f|^2$? This problem is difficult. The following example and the fairly detailed sketches of the proofs provide an accurate illustration of the subtleties involved.

5.1. Example. Let $n = 2$. Given N, is there a polynomial or rational function g from $\mathbf{C}^2$ to $\mathbf{C}^N$ such that $|g(z)|^2 = 1 - |\zeta z_1 z_2|^2$ on the sphere?

0) If $|\zeta|^2 \geq 4$, then for all N, the answer is no.

1) If $N = 1$, then the answer is yes only when $\zeta = 0$.

2) If $N = 2$, the answer is yes precisely when one of the following holds: $\zeta = 0$, $|\zeta|^2 = 1$, $|\zeta|^2 = 2$, $|\zeta|^2 = 3$.

3) For each ζ with $|\zeta|^2 < 4$, there is a smallest N_ζ for which the answer is yes. The limit as $|\zeta|$ tends to 2 of N_ζ is infinity.

Proof. We merely indicate the main ideas in the proofs and refer for complete details to some published articles. In general we are seeking a holomorphic polynomial mapping g such that

$$|g_1(z)|^2 + \ldots + |g_N(z)|^2 + |\zeta|^2 |z_1 z_2|^2 = 1$$

on the unit sphere. The components of g and the additional term $\zeta z_1 z_2$ define a holomorphic mapping from the n ball to the $N+1$ ball which maps the sphere to the sphere. Either such a map is constant or proper. We now consider the results for small N.

0) The maximum of $|\zeta z_1 z_2|^2$ on the sphere is 1 when $|z_1|^2 = |z_2|^2 = \frac{1}{2}$. Hence $|\zeta|^2 \leq 4$ must hold if the question has a positive answer. We claim that $|\zeta|^2 = 4$ cannot hold either. Suppose $|\zeta|^2 = 4$ and g exists. Then we would have

$$|g(z)|^2 + 4|z_1|^2 |z_2|^2 = 1 = (|z_1|^2 + |z_2|^2)^2$$

on the sphere, and hence

$$|g(z)|^2 = (|z_1|^2 - |z_2|^2)^2$$

on the sphere. By Example 3.b), no such g exists.

1) The only proper mappings from the 2-ball to itself are automorphisms, hence linear fractional transformations. Hence the term $\zeta z_1 z_2$ can arise only if $\zeta = 0$. When $\zeta = 0$ we may of course choose $g(z)$ to be (z_1, z_2).

2) This result follows from Faran's classification [14] of the proper holomorphic rational mappings from B_2 to B_3. First we mention that maps g and h are spherically equivalent if there are automorphisms u, v of the domain and target balls such that $h = vgu$. If g existed, then there would be a proper polynomial mapping h from B_2 to B_3 with the monomial $\zeta z_1 z_2$ as a component. It follows from Faran's classification that h would have to be spherically equivalent to one of the four mappings:

$$\begin{aligned} h(z_1, z_2) &= (z_1, z_2, 0) \\ h(z_1, z_2) &= (z_1, z_1 z_2, z_2^2) \\ h(z_1, z_2) &= (z_1^2, \sqrt{2} z_1 z_2, z_2^2) \\ h(z_1, z_2) &= (z_1^3, \sqrt{3} z_1 z_2, z_2^3). \end{aligned}$$

These four mappings provide the four possible values for $|\zeta|$. It turns out, however, that one can say more. In this case one can prove, via an analysis of the possible denominators, that h must be *unitarily* equivalent to one of these four maps. Since h must have the component $\zeta z_1 z_2$, one can compute that the only possible values of $|\zeta|$ are the four that occur in these formulas.

3) This conclusion appears in [8].

We discuss the situation further. There are certain *spectral* values ζ^* for which the value N_{ζ^*} is smaller than that of N_ζ for some ζ with $|\zeta| < |\zeta^*|$. If $\zeta = 1$ we can solve the problem with $N = 2$; take $g(z) = (z_1^2, z_2)$ for example. Yet, if $|\zeta| < 1$, then we cannot solve the problem when $N = 2$ unless $\zeta = 0$. The proof relied on Faran's determination of all proper (rational) mappings from B_2 to B_3. There are 4 spherical equivalence classes; each class contains a monomial map, but there are no families of maps. If we allow one larger target dimension, then we can get a one-parameter family of maps:

$$f(z) = (z_1, z_2^2, \cos(t) z_1 z_2, \sin(t) z_1 z_2^2, \sin(t) z_1^2 z_2).$$

From this formula we see that we can recover all values of $|\zeta|$ up to unity, but not beyond. We omit the details. The phenomenon that certain discrete values become possible before smaller values do continues as we increase N. If $N = 4$, for example, the answer is yes for $0 \leq |\zeta|^2 \leq 2$ and the following additional values for $|\zeta|^2$:

$$\frac{7}{2}, \frac{10}{3}, \frac{8}{3}, \frac{5}{2}.$$

We satisfy ourselves with explicit maps where the constants $\sqrt{\frac{7}{2}}$ and $\sqrt{\frac{10}{3}}$ arise as coefficients of $z_1 z_2$:

$$f(z) = \left(z_1^7, z_2^7, \sqrt{\frac{7}{2}}z_1z_2, \sqrt{\frac{7}{2}}z_1^5z_2, \sqrt{\frac{7}{2}}z_1z_2^5\right)$$

$$f(z) = (z_1^5, z_2^5, \sqrt{\frac{10}{3}}z_1z_2, \sqrt{\frac{5}{3}}z_1^4z_2\sqrt{\frac{5}{3}}z_1z_2^4).$$

Both these maps are proper from the two ball to the five ball.

The results in Example 4.1 illustrate clearly the difference between finding a representation as a squared norm of rank N and finding any representation as a squared norm. We close with an explanation of why we called this section *spectral values.* Given the polynomial $R(z, \overline{z})$, we solve the problem $R = |f|^2$ on the sphere as in [6] or [8]. We add a variable t to homogenize R; call the result R_h. We may choose C large enough such that the function

$$R_h(z, t, \overline{z}, \overline{t}) + C(|z|^2 - |t|^2)^m$$

is strictly positive on the sphere in $\mathbf{C}^{n+1}$. The underlying matrix of coefficients need not be non-negative definite. We then invoke [6] or [28] to find an integer d such that, after multiplication by $(|z|^2 + |t|^2)^d$, the underlying form is positive definite. We then dehomogenize and evaluate on the sphere. Thus the isolated values of $|\zeta|$ for which we can solve the problem in Example 4.1 are in fact vanishing eigenvalues of a Hermitian form; for nearby values of ζ the eigenvalues may become negative. If we multiply by higher powers of $|z|^2 + |t|^2$, then we can make these eigenvalues positive (an open condition), but other eigenvalues will generally vanish. Given ζ with $|\zeta| < 4$, there always exists an N_ζ, but N_ζ depends on ζ, rather than only on the dimension and degree of R. See [8] for lengthy discussion.

We close by mentioning that the proof of the Positivstellensatz cannot provide effective information on N_λ based upon only dimension and degree. One must take into account precise information about the size of R, and even then things are delicate. Example 4.1 relies on Faran's deep work. No general classification of proper polynomial mappings between balls exists that gives precise information on the relationship between the degree and the target dimension. See [9] for more information.

REFERENCES

[1] J. AGLER AND J.E. MCCARTHY, Pick interpolation and Hilbert function spaces. Graduate Studies in Mathematics. **44**. Providence, RI: American Mathematical Society (2002).

[2] N.I. AKHIEZER, The Classical Moment Problem. Oliver and Boyd, Edinburgh and London (1965).

[3] A. ATHAVALE, Holomorphic kernels and commuting operators. Trans. Amer. Math. Soc. **304**:101–110 (1987).

[4] J.A. BALL AND T.T. TRENT, Unitary colligations, reproducing kernel Hilbert spaces, and Nevanlinna-Pick interpolation in several variables. J. Funct. Anal. **157**(1):1–61 (1998).

[5] A. Barvinok, Integration and optimization of multivariate polynomials by restriction onto a random subspace. Foundations of Computational Mathematics **7**:229–244 (2007).

[6] D.W. Catlin and J.P. D'Angelo, A stabilization theorem for Hermitian forms and applications to holomorphic mappings, Math. Res. Lett. **3**(2):149–166 (1995).

[7] ———, An isometric imbedding theorem for holomorphic bundles. Math. Res. Lett. **6**(1):43–60 (1999).

[8] J.P. D'Angelo, Inequalities from Complex Analysis, Carus Mathematical Monographs, No. **28**, Mathematical Assn. of America (2002).

[9] ———, Proper holomorphic mappings, positivity conditions, and isometric imbedding J. Korean Math Soc. **40**(3):341–371 (2003).

[10] ———, Complex variables analogues of Hilbert's seventeenth problem, International Journal of Mathematics **16**:609–627 (2005).

[11] J.P. D'Angelo and D. Varolin, Positivity conditions for Hermitian symmetric functions, Asian J. Math. **8**:215–232 (2004).

[12] M. Eidelheit, Zur Theorie der konvexen Mengen in linearen normierten Räumen. Studia Math. **6**:104–111 (1936).

[13] J. Eschmeier and M. Putinar, Spherical contractions and interpolation problems on the unit ball. J. Reine Angew. Math. **542**:219–236 (2002).

[14] J. Faran, Maps from the two-ball to the three-ball, Inventiones Math. **68**:441–475 (1982).

[15] C. Foias and A.E. Frazho, The commutant lifting approach to interpolation problems. Operator Theory: Advances and Applications, **44**. Basel. Birkhäuser (1990).

[16] J.W. Helton, S. McCullough, and M. Putinar, Non-negative hereditary polynomials in a free *-algebra. Math. Zeitschrift **250**:515–522 (2005).

[17] T. Ito, On the commutative family of subnormal operators, J. Fac. Sci. Hokkaido Univ. **14**:1–15 (1958).

[18] S. Kakutani, Ein Beweis des Satzes von M. Eidelheit über konvexe Mengen. Proc. Imp. Acad. Tokyo **13**, 93–94 (1937).

[19] M.G. Krein, M.A. Naimark, The method of symmetric and Hermitian forms in the theory of separation of the roots of algebraic equations. (Translated from the Russian by O. Boshko and J. L. Howland). Linear Multilinear Algebra **10**:265–308 (1981).

[20] J.E. McCarthy and M. Putinar, Positivity aspects of the Fantappiè transform. J. d'Analyse Math. **97**:57–83 (2005).

[21] J.B. Lasserre, Global optimization with polynomials and the problem of moments. SIAM J. Optim. **11**(3):796–817 (2001).

[22] A. Lubin, Weighted shifts and commuting normal extension, J. Austral. Math. Soc. **27**:17–26 (1979).

[23] A. Prestel and C.N. Delzell, Positive polynomials. From Hilbert's 17th problem to real algebra. Springer Monographs in Mathematics. Berlin: Springer (2001).

[24] A. Prestel, Representation of real commutative rings. Expo. Math. **23**:89–98 (2005).

[25] M. Putinar, Sur la complexification du problème des moments. C. R. Acad. Sci., Paris, Série I **314**(10):743–745 (1992).

[26] ———, Positive polynomials on compact semi-algebraic sets. Indiana Univ. Math. J. **42**(3):969–984 (1993).

[27] ———, On Hermitian polynomial optimization. Arch. Math. **87**:41–5 (2006).

[28] D.G. Quillen, On the representation of Hermitian forms as sums of squares. Invent. Math. **5**:237–242 (1968).

[29] F. Riesz and B.Sz.-Nagy, Functional analysis. Transl. from the 2nd French ed. by Leo F. Boron. Reprint of the 1955 orig. publ. by Ungar Publ. Co., Dover Books on Advanced Mathematics. New York: Dover Publications, Inc. (1990).

[30] K. SCHMÜDGEN, The K-moment problem for compact semi-algebraic sets, Math. Ann. **289**:203–206 (1991).

[31] B.SZ.-NAGY AND C. FOIAS, Analyse harmonique des opérateurs de l'espace de Hilbert. Budapest: Akademiai Kiado; Paris: Masson et Cie, 1967.

[32] D. VAROLIN, Geometry of Hermitian algebraic functions: quotients of squared norms. American J. Math., to appear.

[33] T. YOSHINO, On the commuting extensions of nearly normal operators, Tohoku Math. J. **25**:163–172 (1973).

ENGINEERING SYSTEMS AND FREE SEMI-ALGEBRAIC GEOMETRY*

MAURICIO C. DE OLIVEIRA†, J. WILLIAM HELTON‡, SCOTT A. MCCULLOUGH§, AND MIHAI PUTINAR¶

(To Scott Joplin and his eternal RAGs)

Abstract. This article sketches a few of the developments in the recently emerging area of real algebraic geometry (in short RAG) in a free* algebra, in particular on "noncommutative inequalities". Also we sketch the engineering problems which both motivated them and are expected to provide directions for future developments. The free* algebra is forced on us when we want to manipulate expressions where the unknowns enter naturally as matrices. Conditions requiring positive definite matrices force one to noncommutative inequalities. The theory developed to treat such situations has two main parts, one parallels classical semialgebraic geometry with sums of squares representations (Positivstellensätze) and the other has a new flavor focusing on how noncommutative convexity (similarly, a variety with positive curvature) is very constrained, so few actually exist.

1. Introduction. This article sketches a few of the developments in the recently emerging area of real algebraic geometry in a free* algebra, and the engineering problems which both motivated them and are expected to provide directions for future developments. Most linear control problems with mean square or worst case performance requirements lead directly to matrix inequalities (MIs). Unfortunately, many of these MIs are badly behaved and unsuited to numerics. Thus engineers have spent considerable energy and cleverness doing non-commutative algebra to convert, on an ad hoc basis, various given MIs into equivalent better behaved MIs.

A classical core of engineering problems are expressible as linear matrix inequalities (LMIs). Indeed, LMIs are the gold standard of MIs, since they are evidently convex and they are the subject of many excellent numerical packages. However, for a satisfying theory and successful numerics a convex MI suffices and so it is natural to ask:

How much more restrictive are LMIs than convex MIs?

It turns out that the answer depends upon whether the MI is, as is the case for systems engineering problems, fully characterized by perfor-

*The authors received partial support respectively from the Ford Motor Co., NSF award DMS-0700758 and the Ford Motor Co., NSF award DMS-0457504 and NSF award DMS-0701094.

†Mechanics and Aerospace Engineering Department, UC at San Diego, La Jolla, CA 92093 (mauricio@ucsd.edu).

‡Department of Mathematics, University of California at San Diego, La Jolla, CA 92093 (helton@math.ucsd.edu).

§Department of Mathematics, University of Florida, Gainesville, FL 32611-8105 (sam@math.ufl.edu).

¶Department of Mathematics, University of California, Santa Barbara, CA 93106 (mputinar@math.ucsb.edu).

mance criteria based on L^2 and signal flow diagrams (as are most textbook classics of control). Such problems have the property we refer to as "*dimension-free*".

Indeed, there are two fundamentally different classes of linear systems problems: dimension free and dimension dependent. A dimension free MI is a MI where the unknowns are g-tuples of matrices which appear in the formulas in a manner which respects matrix multiplication. Dimension dependent MIs have unknowns which are tuples of numbers.

The results presented here suggest the surprising conclusion that for dimension free MIs convexity offers no greater generality than LMIs. Indeed, we conjecture:

Dimension free convex problems are equivalent to an LMI.

The key ingredient in passing from convex MIs to LMIs and proving their equivalence lies in the recently blossoming and vigorously developing direction of semi-algebraic in a free $*$ algebra; i.e., semi-algebraic geometry with variables which, like matrices, do not commute. Indeed at this stage there are two main branches of this subject. One includes non-commutative Positivstellensätze which characterize things like one polynomial p being positive where another polynomial q is positive. The other classifies situations with prescribed curvature.

As of today there are numerous versions of the Positivstellensätze for a free $*$- algebra, with typically cleaner statements than in the commutative case. For instance, in the non-commutative setting, positive polynomials are sums of squares. Through the connection between convexity and positivity of a Hessian, non-commutative semi-algebraic dictates a rigid structure for polynomials, and even rational functions, in non-commuting variables. For instance, a noncommutative polynomial p has second derivative p'' which is again a polynomial. Further, if p is matrix convex (as defined below), then p'' is matrix positive (also defined below) and is thus a sum of squares. It is a bizarre twist that p'' can be sum of squares only if p has degree at most two (see §3. The authors suspect that this is a harbinger of a very rigid structure in a free $*$-algebra for "irreducible varieties" whose curvature is either nearly positive or nearly negative; but this is a tale for another day.

A substantial opportunity for noncommutative algebra and symbolic computation lies in numerical computation for problems whose variables are naturally matrices. The goal is to exploit this special structure to accelerate and to increase the allowable size of computation. This is the subject of Section 9.

This survey is not intended to be comprehensive. Rather its purpose is to provide some snippets of results in non-commutative semi-algebraic geometry and their related computer algebra and numerical algorithms, and of motivating engineering problems with the idea of entertaining and even piquing the readers interest in the subject. In particular, this article draws heavily from [HP07] and [HPMV]. Sometimes we shall abbreviate the word noncommutative to NC.

As examples of other important directions and themes, some of which are addressed in other articles in this volume, there is a non-commutative algebraic geometry based on the Weyl algebra and corresponding computer algebra implementations, for example, Gröbner basis generators for the Weyl algebra are in the standard computer algebra packages such as Plural/Singular. A very different and elegant area is that of rings with a polynomial identity, in short PI rings, e.g. $N \times N$ matrices for fixed N. While most PI research concerns identities, there is one line of work on polynomial inequalities, indeed sums of squares, by Procesi-Schacher [PS76]. A Nullstellensätz for PI rings is discussed in [Ami57].

As indicated LMIs play a large role in this paper, so now we describe them precisely.

1.1. LMIs and noncommutative LMIS. Since they play a central role in engineering and the study of convexity in the free $*$ setting, we digress, in the next subjection to define the notion of an LMI.

Given $d \times d$ symmetric (real entry) matrices $\Lambda_0, \Lambda_1, \ldots, \Lambda_g$, the function $L : \mathbb{R}^g \to S_d(\mathbb{R})$ given by

$$L(x) = \sum_{j=0} \Lambda_j x_j$$

is a classical linear pencil; and the inequality $L(x) \succeq 0$ is the classical (commutative) linear matrix inequality. Here $(x_1, \ldots, x_g) \in \mathbb{R}^g$.

In the non-commutative (dimension free) setting it is natural to substitute $X \in \mathbb{S}_n(\mathbb{R}^g)$ for the x above, obtaining the non-commutative version of a linear pencil. Namely, for each n a function $L : \mathbb{S}_n(\mathbb{R}^g) \to \mathbb{SR}^{n \times n}$

$$L_n(X) = L(X) = \sum \Lambda_j \otimes X_j.$$

The inequality $L(X) \succeq 0$ is what we will generally mean by LMI. And, as with polynomials, when we discuss LMIs and linear pencils it will be understood in the non-commutative sense.

EXAMPLE 1.1. For $x := (x_1, x_2)$ being either commuting or noncommuting variables L written as

$$L(x) := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 2 & 3 \\ 3 & 0 \end{pmatrix} x_1 + \begin{pmatrix} 3 & 5 \\ 2 & 0 \end{pmatrix} x_2$$

denotes a linear pencil or NC linear pencil. For $X := (X_1, X_2)$ with $X_j \in \mathbb{R}^{n \times n}$

$$\begin{aligned} L(X) &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 2 & 3 \\ 3 & 0 \end{pmatrix} \otimes X_1 + \begin{pmatrix} 3 & 5 \\ 2 & 0 \end{pmatrix} \otimes X_2 \\ &= \begin{pmatrix} I_n + 2X_1 + 3X_2 & 3X_1 + 5X_2 \\ 3X_1 + 2X_2 & I_n \end{pmatrix}. \end{aligned}$$

For $X := (X_1, X_2)$ with $X_j \in \mathbb{R}$, the set of solutions to $L(X) \succeq 0$ is

$$\mathcal{C} := \{(X_1, X_2): \quad 1 + 2X_1 + 3X_2 - (3X_1 + 5X_2)(3X_1 + 2X_2) \succeq 0.\}, \tag{1.1}$$

This last equivalence follows from taking an appropriate Schur complement which we now recall. The *Schur complement* of a matrix (with pivot γ^{-1}) is defined by

$$SchurComp \begin{pmatrix} \alpha & \beta \\ \beta^* & \gamma \end{pmatrix} := \alpha - \beta\gamma^{-1}\beta^*.$$

A key fact is: if γ is invertible, then the matrix is positive semi-definite if and only if $\gamma > 0$ and its Schur complement is positive semi-definite.

EXAMPLE 1.2. Apply this to the LMI in our example to obtain (1.1) for $X := (X_1, X_2)$ with $X_j \in \mathbb{R}^{n \times n}$ and X_j symmetric.

The Schur complement of $L(x)$ using the other pivot is the "rational expression"

$$I_n - (3X_1 + 2X_2)(I_n + 2X_1 + 3X_2)^{-1}(3X_1 + 5X_2).$$

1.2. Outline. The remainder of the survey is organized as follows. We expand upon the connection between systems engineering problems and dimension free MIs in Section 2. Convexity in the non-commutative (namely equal free $*$) setting is formalized in Section 3. This section also contains a brief glimpse into the *NCAlgebra* package. *NCAlgebra*, and the related *NCGB* (stands for non-commutative Gröbner basis) [HdOSM05] do symbolic computation in a free $*$- algebra and greatly aided the discovery of the results discussed in this survey. *NCAlgebra* and *NCGB* are free, but run under Mathematica which is not. Section 4 describes the engineering necessity for having a theory of matrix-valued non-commutative polynomials whose coefficients are themselves polynomials in non-commuting variables; much of the analysis of Section 3 carries over naturally in this setting. The shockingly rigid structure of convex rational functions is described in Section 5, with a sketch of proofs behind this "curvature oriented" non-commutative semi-algebraic geometry in §6. Section 7 discusses numerics designed to take advantage of matrix variables. Sections 8 gives the solution to the H^∞ control problem stated in §4.

Sections 9 describes noncommutative semi-algebraic geometry aimed at positivity and Positivstellensäten, this is an analogue of classical semi-algebraic geometry which is elegant though it does not have direct engineering applications.

1.3. Acknowledgments. The authors are grateful to Igor Klep for his comments generally and specifically for considerable help with Section 9 and allowing us to include his forthcoming Theorem 9.7.

2. Dimension free engineering: the map between systems and algebra. This section illustrates how linear systems problems lead to semi-algebraic geometry over a free or nearly free $*$- algebra and the role of convexity in this setting. The discussion will also inform the necessary further directions in the developing theory of non-commutative semi-algebraic needed to fully treat engineering problems.

In the engineering literature, the action takes place over the real field. Thus in much of this article, and in particular in this section, we restrict to real scalars. However,we do break from the engineering convention in that we will use A^* to denote the transpose of a (real entries) matrix and at the same time the usual involution on matrices with complex entries. Context will evidently determine the meaning.

The inner product of vectors in a real Hilbert space will be denoted $u \cdot v$.

2.1. Linear systems. A *linear system* $\mathfrak{F}$ is given by the linear differential equations

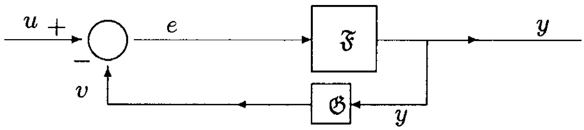
$$\begin{aligned} \frac{dx}{dt} &= Ax + Bu, \\ y &= Cx, \end{aligned}$$

with the vector

- $x(t)$ at each time t being in the vector space $\mathcal{X}$ called the *state space*,
- $u(t)$ at each time t being in the vector space $\mathcal{U}$ called the *input space*,
- $y(t)$ at each time t being in the vector space $\mathcal{Y}$ called the *output space*,

and A, B, C being linear maps on the corresponding vector spaces.

2.2. Connecting linear systems. Systems can be connected in incredibly complicated configurations. We describe a simple connection and this goes along way toward illustrating the general idea. Given two linear systems $\mathfrak{F}$, $\mathfrak{G}$, we describe the formulas for connecting them as follows.



Systems $\mathfrak{F}$ and $\mathfrak{G}$ are respectively given by the linear differential equations