Gaussian Mixture Separation and Denoising on Parameterized Varieties

A thesis presented for the academic degree of

Doctor of Natural Sciences (Dr. rer. nat.)

by

Alexander Filipe Taveira Blomenhofer

to the

Universität Konstanz

Faculty of Sciences
Department of Mathematics and Statistics

Konstanz, 2022

Dedicated to the memory of my granny Idalina,
my parents Eduarda and Helmut
and my dear brother Tobias,
who kept supporting and encouraging me.

Acknowledgements

This work would not have been possible without the continued support of Pravesh Kothari, who introduced me to the majority of the problems considered in this thesis, and of Mateusz Michałek, whose impressive knowledge about the depths of algebraic geometry helped to a large part with solving them. These two have become pseudo-supervisors of mine and played a role equally important to my actual supervisor, Markus Schweighofer, whom I also wish to thank for the large amount of conceptual freedom he gave me throughout my entire PhD, in particular by letting me choose topics on my own right from the very beginning and also for the joint experience that teaching together was.

Markus Schweighofer had been teaching several of my undergraduate courses and as such had a major impact on my decision back then to switch from physics to mathematics. He introduced me to Sum-of-Squares programming as well as Real Algebraic Geometry and had an overall immense impact on shaping my mathematical background. My PhD fell into the time where he became a full-time dad of a very young daughter (and by now also a little son) and as a result, I experienced an environment of great creative freedom, where in spite of his other obligations, he still managed to have an ear for me every once in a while.

I met Pravesh Kothari when he visited Konstanz one day in the era before the pandemic. Pravesh had hour-long encouraging discussions with me, introducing me to Gaussian mixtures, denoising and more, thus providing me with a never-ending stream of problems to work on. He also aquainted me with the work of Ankit Garg on powers-of-forms decompositions, which inspired the largest part of Chapter 3.

Furthermore I enjoyed a lot working with Alex Casarotti and Alessandro Oneto, which both had a great impact on this work and, among many other things, noticed e.g. the connection of Fröberg's conjecture with powers-of-forms identifiability, which is central to the identifiability proofs in Chapter 3 and our joint work [12] with Mateusz. I hope to meet the two of them at some point in person. And of course I wish to thank my office colleague Julian Vill, with whom I had frequent, involved discussions on mathematics, some of which also turned into a joint paper, [92] of course again under the involvement and guidance of Mateusz. Mateusz Michałek arrived in Konstanz a few months before the last year of my PhD and, via a fortunate overlap of research interests, became my de facto supervisor for the remaining part.

I also want to thank Greg Blekhermann, Nick Vannieuwenhoven, João Gouveia and Oliver Schnürer for some helpful advice and guidance as well as Claus Scheiderer and Monique Laurent for kindly agreeing to chair the defense committee or being part of the examination board, respectively.

I owe special thanks and gratitude to Austin Conner, Julian Vill, Thorsten Mayer, Alison Surey, Jakob Everling and Sofie Vaas, who, on very short notice, agreed to proofread (fragments of) this thesis and gave me very valuable feedback! With my (partially former) office neighbours Thorsten Mayer and Christoph Schulze, I had lots of interesting discussions that I will hold in good memory, as well as with the other people on our floor who, occasionally, had time for a chat, e.g. Lothar Sebastian Krapp, Patrick Michalski, Laura Wirth

and Sarah Hess.

To honor the indirect contributions to this work, let me dedicate this paragraph to the friends I have made over the past years in Konstanz since undergraduate days. I especially wish to thank my friends from undergraduate days Florian Ginzel (with whom I really should finish our joint video game project), Jonas Blessing, Sebastian Kümpel, Lena Reichle and Tobias Schmude (whom I really should write again at some point) as well as those I met during later years.

I am also very grateful to have seen several students grow to full-fledged mathematicians during my time in Konstanz, e.g. my PhD "successor" David Sawall and all the students who have been repeated guests in my classes or assisted us with teaching, e.g. Mario Bauer and Jakob Everling, who always knew to ask the right questions. Particularly I enjoyed working with some of those students who wrote their Bachelor's or Master's theses under the supervision of Markus, e.g. Pirmin Klink, Svenja Eggi, Johannes Buchwald, Alison Surey and Raphael Nenno.

And lastly, I want to express my gratitude towards my family, particularly my brother Tobias and my parents Eduarda Martins Taveira and Helmut Blomenhofer, who accompanied me from the very beginning of my journey, and, with unmatched reliability, always had an ear for me for the past 27 years. Well, 23 in case of you, Tobi!

Overview of appendices, code, data and figures

This thesis consists of the present document, several figures, data files with results of numerical experiments and various Jupyter notebooks that contain the code for these experiments. Referring to the external files is done by citing [88]. This reference contains a GitHub url by which all files can be accessed. Should this link ever not work for you, please drop me an email.

The supplementary files are organized in a folder `appendices`. Subfolders distinguish the files thematically: `appendices/identifiability` contains code which calculates the dimension of certain secant varieties and code which generates parametrizations for Gaussian moment varieties. `appendices/gm-plots` creates some graphics visualizing Gaussian mixtures. Both are used in Chapter 3, most notably in Section 3.3. `appendices/pointedsos` contains some numerical experiments on pointedness of Sum-of-Squares cones in relation to the space recovery algorithm from Section 3.4. `appendices/ra` contains the numerical experiments of Chapter 4 on Riemannian Approximation (RA). The thematical subfolders are split into up to three more subfolders `code`, `data` and `figures`, where `code` contains the notebooks, `data` contains tables with results of numerical experiments and `figures` contains plots.

All programming was done with the `Julia` programming language [8]. Particularly used were the packages `MultivariatePolynomials` [56] and `JuMP` [31]. The interior point solver `MOSEK` [66] was used to solve all occurring semidefinite programs.

Zusammenfassung – German summary

Diese Arbeit untersucht die algebraische und geometrische Struktur der Momente Gaußscher Mischverteilungen: Letztere sind ein weitverbreitetes statistisches Modell mit Anwendungen im maschinellen Lernen, welches unter anderem für *Clusteranalysen* genutzt wird. Clusteranalyse ist die Aufgabe, einen Datensatz auf "sinnvolle" Weise in eine vorgegebene Anzahl m von *Punktwolken* einzuteilen und ist eines der klassischen unüberwachten Lernprobleme.

Eine bislang weitgehend offene Frage war, ob und inwiefern endlich viele Stichproben einer Gaußschen Mischverteilung deren Parameter (annähernd) eindeutig bestimmen. Diese Frage der sogenannten *Identifizierbarkeit* der Parameter lässt sich über die Momente der Mischverteilung in einem Rahmen behandeln, der die stochastischen Elemente eliminiert und das Identifizierbarkeitsproblem auf eine Fragestellung der algebraischen Geometrie zurückführt, nämlich: Welche Dimensionen haben die Sekantenvarietäten der Gaußschen Momentenvarietät?

Mit Casarotti, Michałek und Oneto beantworte ich die Frage der Identifizierbarkeit für *generische zentrierte* Gaußsche Mischverteilungen (Kapitel 3) für die überwiegende Anzahl der Sekantenvarietäten und leite dann den nichtzentrierten Fall für Momente von Grad 6 in Abschnitt 3.3 her.

Identifizierbarkeit hat nicht zwangsläufig algorithmische Konsequenzen: In der Tat ist es gerade bei dieser Art von Problemen häufig zu beobachten, dass es signifikante "Zonen des Unwissens" gibt, d.h. Klassen von Mischverteilungen, für die eine eindeutige Bestimmung der Parameter informationstheoretisch zwar möglich ist, aber kein Algorithmus bekannt ist, der jene eindeutige Lösung berechnet. Die wichtigste Komplexitätsgröße hierbei ist der *Rang* einer Darstellung als Momentenform einer Gaußschen Mischverteilung, welcher linear mit der Anzahl der Parameter korreliert. Algorithmen für sehr kleine Ränge (im Verhältnis zur Anzahl der Koordinaten der Datenpunkte) sind bekannt. Für sehr große Ränge *kann* in der Regel keine eindeutige Lösung existieren. Dazwischen klafft eine große Lücke, für die wir Identifizierbarkeit zeigen konnten, aber dennoch keine Algorithmen bekannt sind.

Eine Schwierigkeit hierbei ist, dass die Momente zentrierter Gaußscher Mischverteilungen ein ausdrucksstarkes Berechenbarkeitsmodell emulieren können: Sie bilden eine spezielle Klasse sogenannter *arithmetischer Schaltkreise.* Könnte man zeigen, dass bestimmte Polynome in diesem Berechenbarkeitsmodell nur durch eine sehr große Menge zentrierter Gaußscher Momente darstellbar sind, d.h. einen großen Rang haben, so hätte dies signifikante Konsequenzen für die Schaltkreiskomplexitätstheorie.

Bisherige algorithmische Herangehensweisen hatten sich daher häufig einer probabilistischen Analyse bedient, die naturgemäß keine Konsequenzen für *konkrete* Familien von Instanzen impliziert. Speziell für den Fall zentrierter Gaußscher Mischverteilungen hat dies aber bislang keine unteren Rangschranken geliefert, welche einen Algorithmus für konkrete Instanzen derselben Ränge unplausibel erscheinen lassen würden. Ich verbessere in Abschnitt 3.4 die Rangschranken gegenüber den besten bisher bekannten Algorithmen für n-variate zentrierte Normalverteilungen von $\mathcal{O}(n/\log(n)^2)$ auf $n-2$ (für Momente sechster Ordnung). Statt einer probabilistischen Analyse arbeite ich mit deutlich konkreteren Voraussetzungen, zB mit der Annahme, dass die Kovarianzformen eine reguläre Sequenz bilden und dass eine ihrer gemeinsamen Niveauflächen einen reellen Punkt enthält. Unterhalb der Schranke für den Rang m sind diese Voraussetzungen auf einer typischen Menge von m-Tupeln reeller positiv definiter Formen erfüllt und die Analyse ist deutlich weniger technisch als in den drei Vorgängerarbeiten [37], [36] und [5]. Interessanterweise stellt jedoch keine der vier Arbeiten in wirklich allen Aspekten eine Verbesserung gegenüber einer der anderen dar, was in Abschnitt 3.4 erläutert wird.

Kapitel 4 befasst sich dann mit einem anderen Seitenstrang meiner Forschung, nämlich dem Problem, einen Punkt auf die Bildmenge einer polynomialen Abbildung zu projizieren. Stammt der Ausgangspunkt aus einer *Tubenumgebung* der Bildmenge und ist generisch, so kann eine spezielle, dort betrachtete Hierarchie semidefiniter Programme die eindeutige orthogonale Projektion auf die Bildmenge stets in endlich vielen Hierarchieschritten exakt berechnen (bis auf numerische Fehler). Interessant ist hier aber vor allem, wie viele Hierarchieschritte für bestimmte polynomiale Abbildungen nötig sind. In Kapitel 4 gebe ich unter anderem Argumente dafür, dass die Anzahl der nötigen Hierarchieschritte einen wichtigen Komplexitätsparameter für bestimmte Inversenprobleme darstellt, die wiederum Relevanz in der algebraischen Statistik haben. Mithilfe numerischer Experimente berechne ich diesen *Riemannschen Approximationsgrad* für einige Varietäten, die im Zusammenhang mit Matrix- und Tensorrangapproximation stehen.

Das letzte Kapitel 5 behandelt schließlich ein anderes Problem im Zusammenhang mit arithmetischer Schaltkreiskomplexitätstheorie, nämlich, mittels eines deterministischen Algorithmus zu prüfen, ob die Determinante eines linearen Matrixpolynoms konstant Null ist. Ist dies der Fall, so nennen wir jenes lineare Matrixpolynom *singulär*. Ich führe eine Hierarchie spektraedrisch beschreibbarer Mengen ein, welche die Menge der nichtsingulären Matrixpolynome approximiert.

Contents

“I sometimes find, and I am sure you know the feeling, that I simply have too many thoughts and memories crammed into my mind.”

— Albus P. W. B. Dumbledore, [80].

1 Preface

This thesis largely resulted from my occupation with *Gaussian mixtures*, a highly expressive statistical model that has seen successful use in several Machine Learning applications, particularly clustering, but also bears some surprisingly deep connections to open questions in algebraic geometry [33],[34],[67] and circuit complexity [36] [35]. In this work, we will visit Gaussian mixtures and related problems from the perspective of (Real) Algebraic Geometry.

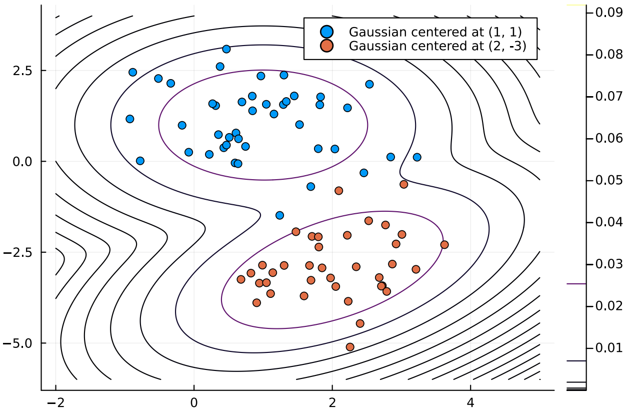

Figure 1.1: Samples from a mixture of two Gaussians on $\mathbb{R}^2$: Each of the 80 samples was chosen by selecting one of the two Gaussians with probability $\frac{1}{2}$ and then sampling the selected Gaussian. (cf. 2.5.9). The contour lines are the level sets of the probability density function of the mixture.

Gaussian mixtures provide a canonical distributional model for data sets that consist of several clusters.[50] Figure 1.1 shows samples drawn from a Gaussian mixture along with the probability density function (pdf) of the mixture distributions. Given a data set, applications typically aim to compute a Gaussian mixture that minimizes a certain “distance” or “loss” function to the empirical data. Such optimization problems are called *Gaussian mixture models.* The resulting solutions may then be used to assign probabilities to each point in the data set that it belongs to some cluster, with respect to the model. These probabilities effectively provide a soft-clustering of the data set. Clustering has significant applications in the era of big data and was one of the first large-scale uses of Machine Learning techniques, notably e.g. for Collaborative filtering in Amazon’s recommender systems. [58] [41]

However, a widespread problem in Machine Learning is that we only know partially what we are doing. Even in an idealized setting, a lot of things can go wrong when drawing conclusions from data, but aside from the philosophical difficulties, one often also has to deal with significant computational and algebraic hurdles. Many of the standard clustering procedures (such as the k-means algorithm) are computationally efficient, but based on heuristics. They typically can compute *some* solution very fast, but do not provide much of a justification that they found a good one. This is not only the case with heuristics, but also with many local optimization procedures for nonconvex problems. Gaussian mixtures, on the contrary, provide a solid probabilistic model that allows to perform statistical inference on data. The downside is that with the high expressive strength of Gaussian mixture models, there comes significant computational and algebraic complexity.

One question is e.g. under which conditions the optimal parameters of a Gaussian mixture model are unique. This is closely related to the question of *identifiability.* In the worst imaginable case, a data set could be equally well described by two mixtures with vastly different parameters, voiding any hope for inference. The loss function is typically computed with respect to the samples and there are various different options to choose. A nice algebraic framework emerges when considering moments: The moments of truncated degree are elements of some finite-dimensional subspace of the polynomial ring that may be endowed with an inner product. Degree-d moments of an n-variate Gaussian distribution form a real, Zariski-dense subset of some affine algebraic variety $\widehat{\mathrm{GM}_d(\mathbb{C}^n)}$, and a mixture of $m \in \mathbb{N}$ Gaussians is an element of the m-th (affine) secant variety $\sigma_m(\widehat{\mathrm{GM}_d(\mathbb{C}^n)})$. The loss function is then some Euclidean distance to the given empirical moments. If the data distribution follows a Gaussian mixture model, then, given sufficiently many samples, the empirical moment form will be a *general* point very close to some element of the set of Gaussian mixture moments with high probability (and will thus project to a general point of the set of moments). As samples are only used to compute the empirical moments, such a *moment based* framework allows to forget about the statistical aspects of the problem and study it only from the perspective of algebraic geometry and polynomial optimization.

A moment based Gaussian mixture model may be split into two subproblems: First, compute the projection $\pi(f)$ of some empirical moment form f to the set of Gaussian mixture moments. Then, find a representation $\pi(f) = \sum_{i=1}^{m} f_i$ as a sum of moment forms f_i of Gaussians $\mathcal{N}(\mu_i, \Sigma_i), i \in \{1, \dots, m\}$ with expectations μ_i and covariance matrices Σ_i. *Identifiability* is essentially the question whether such latter representations as sums of elements of a fixed variety are unique. *Generic identifiability* is the question whether *the general element*[1] of the m-th affine secant variety has a unique representation as a sum of m elements of $\widehat{\mathrm{GM}_d(\mathbb{C}^n)}$.

With Casarotti, Michałek and Oneto [12], we showed that mixtures of *centered* Gaussians are generically identifiable for most possible *ranks* m from their moments of some even degree at least 6. This results covers all degrees where identifiability beyond rank 1 is possible and also covers an asymptotically optimal range of ranks m in relation to the dimension n and the degree. In Section 3.2, I provide an exposition of these results. The arguments are based on the classical Terracini analysis for secant varieties, but new arguments [19] allow to pass from finite-to-one to one-to-one identifiability. Section 3.2 establishes a connection between identifiability for mixtures of Gaussians, powers-of-forms decompositions as well as Fröberg's conjecture on the Hilbert series of an ideal given by m general forms. Section 3.3 sketches some generalizations towards mixtures of general *noncentered* Gaussians.

Generic identifiability does not necessarily have algorithmic consequences for the decomposition problem. In fact, it is very typical for this type of problems that there exist significant "gaps" where one can show identifiability holds as a theoretical property, but there is no known efficient algorithm computing the unique decomposition. To some extent, this is expected, since algorithms could potentially show rank lower bounds for *specific* families of forms, which could have nontrivial consequences in circuit complexity, if these bounds are sufficiently large. [36] Nevertheless, many existing algorithms used identifiability as an implicit assumption, since they attempt to extract the parameters of the individual Gaussians one-by-one [37], [36], [5].[2] In Section 3.4, I give a Semidefinite Programming based algorithm for centered Gaussians, improving the maximum admissible number of Gaussians m in degree 6 from $m \in \mathcal{O}(n/\log(n)^2)$ (due to [5]) to $m = n - 2$ in a typical real case. The algorithm achieves the currently best-known rank threshold for mixtures of centered Gaussians from their moments of (minimal) degree 6 and works more generally also in the framework of powers-of-forms decomposition.

Local projections to parameterized varieties naturally occurred in the context of Gaussian mixture models, but are of independent interest. Chapter 4 develops a semidefinite approach to quantify the complexity of local projection towards the image set of a polynomial map, which leads to the notion of *Riemannian Approximation degrees.*

Finally, the last and very short Chapter 5 is dedicated to another class of objects in relation to circuit complexity, namely *singular matrix spaces.* Deciding membership in the class $\mathrm{SING}_{n,m}$ of singular matrix spaces is equivalent to the infamous problem of *determinant identity testing* (DIT), which has an almost trivial, efficient, randomized algorithm, but no known deterministic one with reasonable efficiency. Chapter 5 studies a semidefinite hierarchy for (DIT) and shows a finite convergence result.

[1] The terminology is justified as these varieties are irreducible.

[2] Of course, each such algorithm is an identifiability proof on its own, but limited to its own setting. Efficient algorithms with provable guarantees usually restrict the number m of Gaussians quite heavily.

"However it may (be) prove(n), one must tread the path that need chooses!"

— Gandalf, [89].

2 General Notions and Notation

2.1. Basics

$\mathbb{N} = \{1, 2, 3, 4, \ldots\}$ denotes the set of natural numbers. All serious work of this thesis is done over either the real field $\mathbb{R}$ or the complex field $\mathbb{C}$, but some elementary propositions are more generally formulated for arbitrary fields K (of characteristic 0).

To the standard $\mathcal{O}$-notation from complexity theory for functions $f, g \colon \mathbb{N} \to \mathbb{R}$,[1] the following nonstandard additions are used: $f \in \theta(g)$ is sometimes also written as $f \approx g$. Similarly, $f \in \mathcal{O}(g)$ and $f \in \Omega(g)$ may be written as $f \ll g$ and $f \gg g$, respectively. $f \in \theta^{\#}(g)$ signifies that $\lim_{n\to\infty} f(n)/g(n) = 1$. This is stronger than $f \in \theta(g)$, which just signifies that $\limsup_{n\to\infty} f(n)/g(n)$ and $\liminf_{n\to\infty} f(n)/g(n)$ both exist and are nonzero constants.

For multilinear maps $T \colon U_1 \times U_2 \times \ldots \to K, (x_1, x_2, \ldots) \mapsto T\langle x_1, x_2, \ldots\rangle$ of K-vector spaces $U_1, U_2, \ldots$, angle bracket notation is frequently used to emphasize linearity in certain arguments. This is predominantly useful to distinguish when certain functions have both linear and nonlinear arguments, e.g. for the Hessian form of some function f, we would write $\operatorname{Hess} f(z)\langle v, w\rangle$ for the Hessian of f *at* z in *directions* v, w. When there are no nonlinear arguments, the bracket notation has no purpose, but is kept for convenience and consistency.

For integration operators $\mathbb{E}_\mu$ of measures μ, square brackets are used instead of parantheses, e.g. I write $\mathbb{E}_\mu[p]$ instead of $\mathbb{E}_\mu(p)$ for the integral of some polynomial p with respect to μ. This bracket notation is also used in two more contexts: For certain functionals introduced in Section 2.3, called *pseudo expectation operators*, with the intent to convey the wishful thinking that they might be integration operators of measures and for entry- or coefficient-wise application of a linear functional.

For $n \in \mathbb{N}_0$ and $x \in \mathbb{R}^n$, $\|x\|$ denotes the Euclidean norm of x. Thus $\|x\|^2 = x_1^2 + \ldots + x_n^2$, which is given by the standard inner product $\langle x, y\rangle := x^T y$,

[1] And slightly more general settings.

where $x, y \in \mathbb{R}^n$. Typical norms used on matrices A are the spectral norm $\|A\|_{\mathrm{spec}}$ and the Frobenius norm $\|A\|_{\mathrm{F}}$. For some polynomial functions on $\mathbb{R}^n$, I extend their scope towards vectors of polynomials. E.g. if $X = (X_1, \dots, X_n)$ is a vector of unknowns, then let us understand $\|X\|^2 = X_1^2 + \ldots + X_n^2 = \langle X, X\rangle$.

In the context of algebraic geometry, topological notions such as *dense, irreducible, closed, open, continuous* etc. shall refer by default to the Zariski topology, introduced in Definition 2.2.1. However, in the context of real manifolds, sums of squares and optimization, the standard topology is the Euclidean one. Should it ever feel unclear, I will add a clarification.

Lastly, there are too many notions of *duality* present in this work. To distinguish, we denote dual cones with an asterisk superscript $*$, but dual spaces with a $\vee$: Thus if U is a vector space over some field K, then $U^\vee$ denotes the space of linear functionals $U \to K$, which is called the *dual space* of U. If W is a subspace of U, then $W^\perp \subseteq U^\vee$ denotes the *conormal space* of W with respect to U, i.e. the space of linear functionals $U \to K$ that vanish on W. If $K \in \{\mathbb{R}, \mathbb{C}\}$ and U is an inner product space, then a different convention is used: For a subspace W of an inner product space $(U, \langle\cdot,\cdot\rangle)$, $W^\perp = \{x \in U \mid \forall w \in W\colon \langle x, w\rangle = 0\}$ denotes the *normal space* of W, sometimes also called the *orthogonal complement.* The *dual cone* C^* of a convex cone C in a real vector space U is the set of all functionals $U \to \mathbb{R}$ which attain nonnegative values on C. This set is a convex cone in $U^\vee$. For finite-dimensional $\mathbb{R}$-vector spaces, the bidual cone C^{**} is not seen as a cone in $U^{\vee\vee}$, but rather in U, i.e. $C^{**} = \{x \in U \mid \forall L \in C^*\colon L(x) \geq 0\}$. Note that any subspace W of an $\mathbb{R}$-vector space U is also a convex cone and it holds that $W^* = W^\perp$.

2.2. Algebraic Geometry

We will introduce some fundamental notions and notation regarding systems of polynomial equations and their solution sets that should be familiar from standard courses on algebraic geometry. The conventions roughly follow [44, Chapter I], but we deviate in some minor aspects. E.g. we need some concepts also over the real field, whereas [44] focuses on algebraically closed fields.

2.2.1 Definition: Let K a field and U an affine or projective space over K. A subset $V \subseteq U$ is called a *closed subvariety* of U, if it is the zero set of some set of polynomial equations.[2] The set of closed subvarieties $\mathcal{Z}_U$ of U defines a topology by declaring $\mathcal{Z}_U$ to be the closed sets. We call it the Zariski-topology on U. If U is an affine space, then we call $V \subseteq U$ a *quasi-affine* variety, if V is an open subset of some closed subvariety of U. Similarly, if U is a projective space, then we call $V \subseteq U$ a *quasi-projective* variety, if V is an open subset of some closed subvariety of U. We call V a variety, if V is either a quasi-affine or a quasi-projective variety in some (affine or projective, respectively) space U and if the context clarifies whether the affine or projective setting is meant. If the field is unspecified, it is always the complex numbers.

2.2.2 Remark: In this work, we will deal with both affine cones and convex cones. An *affine cone* is to be understood as a quasi-affine *subvariety* W of some affine space, say $\mathbb{C}^N$ for some $N \in \mathbb{N}_0$, which is closed under multiplication with

[2] Homogeneous equations in the case that U is projective.

complex scalars, i.e. if for each $\lambda \in \mathbb{C}$ it holds $\lambda W \subseteq W$, then W is an affine cone. A convex cone W, on the other hand, is a *subset* of some real affine space, say $\mathbb{R}^N$ for some $N \in \mathbb{N}_0$, which is closed under multiplication with *real, nonnegative* scalars and under addition, that is, $W + W \subseteq W$. Lastly, a *real affine cone* is the set of real points of some complex affine cone.

2.2.3 Proposition and Definition: Let U an affine space and $V \subseteq U$ an affine cone. For $z \in U \setminus \{0\}$, we denote by $[z] \subseteq U$ the unique line through z and the origin. The set

$$\mathbb{P}(V) := \{[z] \mid z \in V\}$$

is called the *projectivization* of V. It is a quasi-projective variety and closed, if V is closed.

2.2.4 Proposition and Definition: Let U a projective space whose representatives stem from the affine space $\widehat{U}$ and $V \subseteq U$ a quasi-projective variety. The set

$$\widehat{V} := \{z \mid [z] \in V\} \cup \{0\} \subseteq \widehat{U}$$

is called the *affine cone* of V. It is a quasi-affine variety and closed, if V is closed.

2.2.5 Notation: Let U an affine or projective space over some infinite field K. By $K[U]$ we denote the ring of polynomial functions on U. With respect to a basis $\mathcal{B}$ of the dual space U^*, we may choose a vector of algebraic unknowns $Z = (Z_b)_{b\in\mathcal{B}}$ and identify $K[U]$ with the ring of polynomials $K[Z]$ in the unknowns Z. This identification is via the evaluation homomorphism

$$K[Z] \to K[U], Z_b \mapsto b \qquad (b \in \mathcal{B}) \tag{2.1}$$

which is a ring isomorphism. By convention, we usually use the letters Z or X as the vectors of unknowns and silently assume that a basis for U with the corresponding identification is chosen without mentioning it. Both $K[Z]$ and $K[U]$ are canonically graded rings: $K[Z]$ with respect to the polynomial degree, whereas for $k \in \mathbb{N}_0$, the k-th graded component of $K[U]$ consists of the *k-homogeneous* forms on U, i.e. those $f \in K[U]$ which satisfy

$$\forall \lambda \in K : \forall x \in U : f(\lambda x) = \lambda^k f(x)$$

It is clear that the ring isomorphism from Equation (2.1) respects these graduations. We denote by $S^k(U)$ the subspace of $K[U]$ of k-homogeneous forms on U and by $K[Z]_k$ the subspace of $K[Z]$ of *k-forms* in Z, i.e. $K[Z]_k$ is the span of all monomials of degree k in Z. Clearly $K[Z]_k \cong S^k(U)$ for all $k \in \mathbb{N}_0$. For $D \in \mathbb{N}_0$, we write $K[U]_{\leq D} := \sum_{k=0}^{D} S^k(U)$ and $K[Z]_{\leq D} := \sum_{k=0}^{D} K[Z]_k$. $\operatorname{mon}_k(Z)$ denotes the set of monomials of degree k in the variables Z.

2.2.6 Proposition: Let U an affine space over some field K and $k \in N_0$. Then $S^k(U^\vee)$ is spanned by k-th powers of linear forms on U.

2.2.7 Definition: Let U an affine space over some field K. For $h \in U$, the *directional derivative* $\partial_h \colon K[U] \to K[U]$ is defined on powers of linear forms given by $\ell \in S^1(U)$ and $k \in \mathbb{N}_0$ via

$$\partial_h \ell^k = k\ell(h)\ell^{k-1}$$

This is well-defined and by linear continuation, it determines ∂_h as a map $K[U] \to K[U]$.

Proof. Uniqueness of the linear continuation is due to Proposition 2.2.6. Existence is left as an exercise to the reader. In case of struggle, it is suggested to open a book on calculus. □

2.2.8 PROPOSITION:

$$\mathrm{Sym}_d(U^\vee) \to S^d(U), \ell^{\otimes d} \mapsto \ell^d = \begin{pmatrix} U \to K \\ x \mapsto \ell(x)^d \end{pmatrix} \tag{2.2}$$

is an isomorphism of K-vector spaces.

2.2.9 DEFINITION: Let U an affine space over some field K. Let $I(V) \subseteq K[U]$ denote the vanishing ideal of V. The *tangent space* T_zV of V in a point $z \in V$ is defined as

$$T_zV = \{h \in U \mid \forall f \in I(V) : \partial_h f(z) = 0\}$$

2.2.10 DEFINITION: Let U a projective space over some field K and V a quasi-projective variety. The *tangent space* T_zV of V in a point $[z] \in V$ is defined as the projectivization of the tangent space $T_z\widehat{V} \subseteq \widehat{U}$ of the corresponding affine cone. One quickly verifies that this definition does not depend on the choice of the representative of $[z]$, as $I(\widehat{V})$ is homogeneous.

2.2.11 REMARK: (a) If in the setting of Definition 2.2.9, $f_1, \ldots, f_r \in K[U]$ generate $I(V)$ for some $r \in \mathbb{N}_0$ then we see that

$$T_zV = \{h \in U \mid \forall i \in \{1, \ldots, r\} : \partial_h f_i(z) = 0\} = \ker Jf(z)$$

is precisely the kernel of the Jacobian map

$$Jf(z) \colon U \to K^r, h \mapsto \begin{pmatrix} \partial_h f_1(z) \\ \vdots \\ \partial_h f_r(z) \end{pmatrix} \tag{2.3}$$

of $f := (f_1, \ldots, f_r)$ at z. With respect to a basis of U, one usually describes the Jacobian as a matrix.

(b) Assume now V from Definition 2.2.9 is irreducible. The dimension of V is at most the dimension of $T_z(V)$ for any $z \in V$. Points z where

$$\dim V = \dim T_z(V)$$

are called *smooth points* of V. The other points are called *singular points.* The singular points form a proper Zariski-closed subset of V. V is called *smooth*, if all its points are smooth points. The set of all smooth points of V is called the *smooth locus* of V.

2.2.12 PROPOSITION: Let $f \colon V_1 \to V_2$ a morphism of varieties. Fix $p \in V_1$ and $q := f(p) \in V_2$. Then f induces a natural linear map

$$T_f(p) \colon T_pV_1 \to T_qV_2$$

of the tangent spaces. In (local) affine coordinates, i.e. if $V_1 \subseteq \mathbb{C}^n$ and $V_2 \subseteq \mathbb{C}^N$, this map is represented by the *Jacobian matrix* $Jf(p) = (\partial_j f_i)_{i\in\{1,\dots,N\},j\in\{1,\dots,n\}}$ (just like in Remark 2.2.11(a), but note that this time, $f_1, \dots, f_N$ have a different role).

Proof. Write $V_2 = V(g)$ for some finite-length vector g of polynomials. Then $T_qV_2 = \ker Jg$. By the chain rule of differentiation, the matrix factorization $J(g \circ f)(p) = Jg(f(p)) \cdot Jf(p)$ holds true. Thus the map $h \mapsto Jf(p)h$ maps vectors in the kernel of $J(g \circ f)(p)$ to vectors in T_qV_2. It suffices thus to show that any $h \in T_pV_1$ is contained in the kernel of $J(g \circ f)(p)$. However, since any $x \in V_1$ satisfies the equations $g(f(x)) = 0$, these equations describe a supervariety V' of V_1. Naturally, we have $\ker J(g \circ f)(p) = T_pV' \supseteq T_pV_1$. □

2.2.13 Definition: A morphism $f\colon V_1 \to V_2$ of varieties is called *dominant*, if its image is dense.

2.2.14 Theorem: (Sard, complex algebraic version) Let $f\colon V_1 \to V_2$ a morphism of irreducible varieties. Then there is a dense open subset $\mathcal{U}$ of V_2 such that if $q \in \mathcal{U}$ and $p \in f^{-1}(\{q\})$ is a smooth point of V_1, then

$$T_f(p)\colon T_pV_1 \to T_qV_2$$

is surjective. If V_1 is smooth, then the fibers $f^{-1}(\{q\})$ are smooth for $q \in \mathcal{U}$.

Proof. This formulation is taken from some of McKernan's lecture notes, cf. [63, Lecture 9, Theorem 9.2]. Hartshorne proves it for the case of smooth varieties V_1 and V_2 in [44, III 10.4(i) $\Longrightarrow$ (iii)]. The nonsmooth version can e.g. be obtained by replacing V_2 with its smooth locus $\mathcal{S}(V_2)$ and V_1 with the intersection of its smooth locus $\mathcal{S}(V_1)$ with the preimage of $\mathcal{S}(V_2)$ under f. □

Note that the conclusion of Theorem 2.2.14 is nontrivial only if the morphism therein is dominant, cf. Definition 2.2.13.

2.2.15 Proposition and Definition: (Fiber Dimension) [44, Exercise 3.22, Chapter II] Let W_1, W_2 be irreducible varieties and $\gamma\colon W_1 \to W_2$ a dominant morphism. Then for a general point $y \in W_2$,

$$\dim \gamma^{-1}(y) = \dim W_1 - \dim W_2 \tag{2.4}$$

In particular, the value $\dim \gamma^{-1}(y)$ is generically constant on W_2 and called the *generic fiber dimension* of γ. Furthermore, for each $k \in \mathbb{N}_0$, the set of points $\{y \in \operatorname{im}\gamma \mid \dim \gamma^{-1}(y) \geq k\}$ is a (Zariski) closed subset of $\operatorname{im}\gamma$. In particular, if $y_1, y_2, \dots$ is a sequence of points in $\operatorname{im}\gamma$ converging to $y \in \operatorname{im}\gamma$ with respect to the Euclidean topology, then we have:

$$\dim \gamma^{-1}(y) \geq \limsup_{k\to\infty} \dim \gamma^{-1}(y_k). \tag{2.5}$$

Secant Varieties and Identifiability

The theory of secant varieties is the main algebraic tool to study certain decomposition problems that occur in relation to Gaussian mixtures. In Chapter 3, that theory will be used extensively. This section roughly follows Chiantini and Ottaviani [21], [22], [23], but deviates in some conventions. Also, some new and very important work due to Casarotti and Mella is used [19]. For full disclosure, note that in some previous articles I might have written an introductory section to secant varieties with some superficial similarities, cf. [12].

2.2.16 Definition: (Secant variety) Let W be a variety embedded in an affine or projective space and $m \in \mathbb{N}_0$. The closure of the union of subspaces spanned by m elements of W:

$$\bigcup_{x_1,\dots,x_m \in W} \langle x_1, \dots, x_m \rangle$$

is called the k-th *secant variety* of W and is denoted $\sigma_m(W)$.

2.2.17 Remark: In terms of affine cones, the secant variety has a convenient parameterization: Let W and m be as in Definition 2.2.16. Then $\widehat{\sigma_m(W)}$ is the closure of the image of the *summation map*

$$\psi_{m,W} \colon \widehat{W}^m \to \widehat{\sigma_m(W)}, (x_1, \dots, x_m) \mapsto \sum_{i=1}^{m} x_i \tag{2.6}$$

2.2.18 Notation: For a set A, recall that A^m is the m-fold cartesian product whereas by $A^{\underline{m}}$ we denote the set of all m-element multisets with elements from A, canonically identified with $A^m/\mathfrak{S}_m$, where $\mathfrak{S}_m$ is the group of permutations of $\{1, \dots, m\}$. Note that the map $\psi_{m,W}$ from Remark 2.2.17 is invariant under permutations of its arguments and can therefore be canonically seen as a function on $W^{\underline{m}}$. We will not formally do so, as this would just lead to more annoying notation, but later when we examine under which conditions $\psi_{m,W}$ is a one-to-one map, we do always mean one-to-one up to permutations, i.e. as a map on $W^{\underline{m}}$.

2.2.19 Lemma: (Terracini) Let W be a variety and consider for $m \in \mathbb{N}$ the secant $\sigma_m(W)$. For general points $x_1, \dots, x_m \in W$ and general $x \in \langle x_1, \dots, x_m \rangle \subseteq \sigma_m(W)$, we have that

$$T_x \sigma_m(W) = \sum_{i=1}^{m} T_{x_i} W$$

Proof. Consider the dominant map from Equation (2.6)

$$\psi_{m,W} \colon \widehat{W}^m \to \widehat{\sigma_m(W)}, (x_1, \dots, x_m) \mapsto \sum_{i=1}^{m} x_i$$

Let $\mathcal{S}_1$ be the smooth locus of $\widehat{\sigma_m(W)}$, $\mathcal{S}_2$ the smooth locus of $\widehat{W}$ and consider the dense open set $\mathcal{T} := \psi_{m,W}^{-1}(\mathcal{S}_1) \cap \mathcal{S}_2^m$ of $\widehat{W}^m$. By [44, 10.5.], we may choose $\mathcal{U} \subseteq \mathcal{T}$ dense open such that $\psi_{m,W}|_{\mathcal{U}} \colon \mathcal{U} \to \mathcal{S}_1$ is a smooth map. The domain and codomain of $\psi_{m,W}|_{\mathcal{U}}$ are smooth quasi-projective varieties by construction. Thus by [44, 10.4(i) $\Longrightarrow$ (iii)], for any $(x_1, \dots, x_m) \in \mathcal{U}$ the tangent space at $(x_1, \dots, x_m)$, which is $\bigoplus_{i=1}^m T_{x_i} \widehat{W}$, maps surjectively to the tangent at $\psi_{m,W}((x_1, \dots, x_m))$, which is $T_{x_1+\ldots+x_m} \widehat{\sigma_m(W)}$, under the induced map $T_{\psi_{m,W}}$ of tangent spaces. This induced map is given by the Jacobian of $\psi_{m,W}$ and therefore an easy calculation shows that the image of $T_{\psi_{m,W}}$ is also equal to $\sum_{i=1}^m T_{x_i} \widehat{W}$. For general $(x_1, \dots, x_m) \in \mathcal{U}$ and general $\lambda_1, \dots, \lambda_m \in \mathbb{C}$, it holds that $(\lambda_1 x_1, \ldots, \lambda_m x_m) \in \mathcal{U}$ and so the same follows for $x = \sum_{i=1}^m \lambda_i x_i \in \langle x_1, \ldots, x_m \rangle$. Thus $T_x \widehat{\sigma_m(W)} = \sum_{i=1}^m T_{\lambda_i x_i} \widehat{W} = \sum_{i=1}^m T_{x_i} \widehat{W}$. □

2.2.20 Proposition and Definition: (Expected Dimension) Let W be an irreducible variety and $m \in \mathbb{N}_0$ such that for general $x_1, \ldots, x_m \in W$, the tangent spaces at $x_1, \ldots, x_m$ are skew. Then the map

$$\psi_{m,W} \colon \widehat{W}^m \to \widehat{\sigma_m(W)}, (y_1, \ldots, y_m) \mapsto \sum_{i=1}^m y_i \tag{2.7}$$

is generically finite-to-one. The value $m(\dim(W)+1)-1$ is called the *expected dimension* of $\sigma_m(W)$. It gives an upper bound for the dimension of $\sigma_m(W)$ and if and only if it is attained, the map $\psi_{m,W}$ is generically finite-to-one. Otherwise, W is called *m-defective*.

Proof. We prove everything in affine notation. Let $x_1, \ldots, x_m \in \widehat{W}$ be general points. In the affine setting, skewness means that

$$\sum_{i=1}^m T_{x_i} \widehat{W} = \bigoplus_{i=1}^m T_{x_i} \widehat{W} \tag{2.8}$$

By Lemma 2.2.19 and generality of $x_1, \ldots, x_m$, the left hand side of Equation (2.8) has the dimension of $\widehat{\sigma_m(W)}$, while $\bigoplus_{i=1}^m T_{x_i} \widehat{W} \cong T_{(x_1, \ldots, x_m)} \widehat{W}^m$ has the dimension of $\widehat{W}^m$. Since $\sum_{i=1}^m x_i$ is general in $\sigma_m(W)$, the fiber dimension formula 2.2.15 yields that

$$0 = \dim \psi_{m,W}^{-1}(\sum_{i=1}^m x_i) = \dim \widehat{W}^m - \dim \widehat{\sigma_m(W)} = \dim \bigoplus_{i=1}^m T_{x_i} \widehat{W} - \dim \sum_{i=1}^m T_{x_i} \widehat{W}$$

if and only if the spaces $T_{x_1} \widehat{W}, \ldots, T_{x_m} \widehat{W}$ are skew. This characterizes when the preimage of a general point is finite. For the additional claim, we have by Terracini's Lemma 2.2.19 that

$$\dim \sigma_m(W) = \dim \sum_{i=1}^m T_{x_i} W \leq \left(\dim \bigoplus_{i=1}^m T_{x_i} \widehat{W} \right) - 1$$

where the rightmost term is the expected dimension. Whenever equality holds in the above equation, $\psi_{m,W}$ is thus generically finite-to-one. □

Let us introduce some useful terminology around secants, decompositions and identifiability.

2.2.21 Definition: Let W a variety, $m \in \mathbb{N}$ and consider the summation map $\psi_{m,W} \colon W^{\underline{m}} \to \sigma_m(W)$ from Remark 2.2.17. We call $p \in \sigma_m(W)$ an identifiable element of the m-th secant of W, if $\#\psi_{m,W}^{-1}(\{p\}) = 1$. We say that W is *generically m-identifiable*, if the generic element of $\sigma_m(W)$ is identifiable. In that case, we also call the map $\psi_{m,W}$ generically identifiable. In slight abuse of terminology, we also sometimes call $\sigma_m(W)$ generically identifiable, even though identifiability is a property of W and m.

2.2.22 Proposition and Definition: Let W a subvariety of some projective space U. For $[f]$ in the space $\langle W\rangle$ generated by W, the minimum number $m \in \mathbb{N}$ such that f lies in the image of $\psi_{m,W}$ is called the *rank* of $[f]$ with respect to W (or the W-rank of $[f]$). The smallest $m \in \mathbb{N}_0$ such that $[f]$ lies in $\sigma_m(W)$ is called the W*-border rank* of f. The smallest $m \in \mathbb{N}$ such that $\sigma_m(W) = \langle W\rangle$ is called the *generic rank* of decompositions with respect to W. If the space $\langle W\rangle$ is not equal to U, we call W a *degenerate* variety. We will be exclusively concerned with nondegenerate varieties, for which the generic rank is thus the smallest number $m \in \mathbb{N}$ such that $\sigma_m(W) = U$.

Proposition 2.2.20 gives a tool to examine whether a sum of m elements of W has only finitely many such representations. Identifiability appears more complicated, since the question is a priori not about the dimension of the generic fibers of $\psi_{m,W}$, but about their cardinality, which is not as well-behaved. However, over a series of recent work (e.g. [21], [22], [23], [19]), techniques have been developed that reduce the question of identifiability to another problem of determining dimension. One crucial tool for this is the tangential contact locus, which we introduce in the following.

2.2.23 Proposition and Definition: Let W a smooth closed variety, $m \in \mathbb{N}_0$ and $x = (x_1, \dots, x_m) \in W^m$ points of W with skew tangent spaces. The projective closed subvariety of

$$\Gamma_W(x) := \{y \in W \mid T_yW \subseteq \sum_{i=1}^{m} T_{x_i}W\}$$

which consists of the irreducible components of $\Gamma_W(x)$ passing through at least one of $x_1, \dots, x_m$ is called the (m-th) *tangential contact locus* to W at x and denoted $\mathcal{C}_W(x)$.

Proof. We only have to show that $\Gamma_W(x)$ is a closed variety. To this end, let $n := \dim \widehat{W} \subseteq \mathbb{C}^N$ for some $N \in \mathbb{N}$. For $[y] \in W$, the condition $T_y\widehat{W} \subseteq \sum_{i=1}^m T_{x_i}\widehat{W}$ is equivalent to $T_y\widehat{W}^\perp \supseteq (\sum_{i=1}^m T_{x_i}\widehat{W})^\perp$. Let $k := \operatorname{codim} \widehat{W}$ and choose $f_1, \dots, f_k \in I(\widehat{W})$ such that $\nabla f_1, \dots, \nabla f_k$ are linearly independent vectors of polynomials. Identify the dual space $(\mathbb{C}^N)^\vee$ with $\mathbb{C}^N$ via the nondegenerate symmetric bilinear form $\mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}, (x, y) \mapsto x^T y$. Then the vectors $\nabla f_1(y), \dots, \nabla f_k(y)$ form a basis of $T_y\widehat{W}^\perp$. Let $k' := \operatorname{codim} \sum_{i=1}^m T_{x_i}\widehat{W} \leq k$ and let $u_1, \dots, u_{k'}$ be a basis for $(\sum_{i=1}^m T_{x_i}\widehat{W})^\perp$. Then for each $y \in W$, $y \in \Gamma_W(x)$ is equivalent to vanishing of the $(k + 1) \times (k + 1)$ minors of

$$(\nabla f_1(y), \dots, \nabla f_k(y), u_1, \dots, u_{k'}) \tag{2.9}$$

which gives a system of polynomial equations for $\Gamma_W(x)$. □

The following lemma is called either the *generalized trisecant lemma*, as it is a generalization of a classical result known as the *trisecant lemma*, or, more straightforward, the *multisecant lemma*. The proof is very involved and beyond our scope, so we will refer the reader to [48], [49] for an overview. [20] also gives a sketch reducing the multisecant lemma to the classical trisecant lemma.

2.2.24 Lemma: [20, Proposition 2.6] (Multisecant Lemma) Let $n, N \in \mathbb{N}$, W an irreducible, nondegenerate subvariety of $\mathbb{P}^N$ of dimension n and $m \in \mathbb{N}_0$ such that $m \leq N - n$. Let $x_1, \ldots, x_m$ general points of W. Then

$$\langle x_1, \ldots, x_m \rangle \cap W = \{x_1, \ldots, x_m\} \tag{2.10}$$

Casarotti and Mella [19] recently obtained an extremely useful result that allows to prove generic identifiability as a consequence of expected dimension of the next-order secant.

2.2.25 Theorem: [19, Introduction] Let W be a smooth variety of dimension $n \in \mathbb{N}$ and let $m \in \mathbb{N}$. Assume that the m-th secant variety is of (expected) dimension $m(n+1) - 1$ and $m > 2n$. Then W is $(m-1)$-identifiable.

An older approach due to Chiantini and Ottaviani uses the tangential contact locus and a semicontinuity argument (as in Proposition and Definition 2.2.15) to reduce to the study of *one* specific decomposition $t = x_1 + \ldots + x_m$. I will present this approach by giving slightly elaborated versions of the proof sketches in [21] and [22] with the intent to give the reader a thorough intuition. However, note that expanding the arguments to every last detail would be beyond the scope, which is also why there is no proof of Theorem 2.2.25 included.

2.2.26 Proposition ([22, Proposition 2.3]): Let W be an irreducible, nondegenerate variety of dimension $n \geq 2$, which is not m-defective. If the generic element of $\sigma_m(W)$ is not identifiable, then for general $x \in W^m$ and some $i \in \{1, \ldots, m\}$, the tangential contact locus to W at x must contain a curve through x_i.

Proof. We use affine notation. We have to show that there exists a dense open subset $\mathcal{U}' \subseteq \widehat{W}^m$ such that the claim holds true for all $x \in \mathcal{U}'$. First, there exists of course a dense open subset $\mathcal{U} \subseteq \widehat{W}^m$ such that for all $x \in \mathcal{U}$:

(i) $x_1, \ldots, x_m$ are linearly independent.

(ii) The conclusion of Terracini's lemma holds true for x, i.e. for general $z \in \langle x_1, \ldots, x_m \rangle$,

$$T_z \widehat{\sigma_m(W)} = \bigoplus_{i=1}^{m} T_{x_i} \widehat{W} \tag{2.11}$$

(iii) The multisecant lemma 2.2.24 holds true for x, i.e.

$$\langle x_1, \ldots, x_m \rangle \cap \widehat{W} = \mathbb{C} \cdot \{x_1, \ldots, x_m\} \tag{2.12}$$

By Lemma 2.2.29, we may choose a dense open subset $\mathcal{U}' \subseteq \mathcal{U}$ such that decompositions of elements of the image of $\mathcal{U}'$ under $\psi_{m,W}$ have all their possible decompositions in $\mathcal{U}$.

Now, let $x \in \mathcal{U}'$ and let $y \in \mathcal{U}$ be in the preimage of $\psi_{m,W}(x)$, i.e. assume $t := \sum_{i=1}^{m} x_i = \sum_{i=1}^{m} y_i$ for some tuple $y = (y_1, \ldots, y_m) \in \mathcal{U}$ with $y \neq x$. By (i) the points $x_1, \ldots, x_m$ are linearly independent. Thus there must be at least one $i \in \{1, \ldots, m\}$ such that $y_i \notin \mathbb{C} \cdot \{x_1, \ldots, x_m\}$: Otherwise we could write $y_i = \rho_i x_i$ (after permutation) for some $\rho_1, \ldots, \rho_m \in \mathbb{C}$ and all $i \in \{1, \ldots, m\}$.

Then of course $\sum_{i=1}^{m}(1-\rho_i)x_i = 0$ ↯. Thus assume without loss of generalty y_1 is not a multiple of any of the $x_1, \ldots, x_m$. Consider the mixtures

$$t(\rho) := \sum_{i=1}^{m} \rho_i y_i, \qquad (\rho \in \mathbb{C}^n) \tag{2.13}$$

and let $L := T_{t(\rho)}\sigma_m(W)$. Since the conclusion of Terracini's Lemma (2.11) holds true on $\mathcal{U}$, the space L is constant with respect to general ρ and equal to $\sum_{i=1}^{m} T_{x_i}W$! For most choices of ρ, $t(\rho) \notin \langle x_1, \ldots, x_m\rangle$: Indeed, if there were m linearly independent choices of ρ such that $t(\rho) \in \langle x_1, \ldots, x_m\rangle$, then also $y_1, \ldots, y_m \in \langle x_1, \ldots, x_m\rangle$, contradicting the generalized trisecant lemma. We may thus choose (locally) a curve $\rho(u)$ with parameter u in some neighbourhood of 0 and $\rho(0) = (1, \ldots, 1)$ such that

$$t(\rho(u)) := \sum_{i=1}^{m} \rho(u)_i y_i \notin \langle x_1, \ldots, x_m\rangle$$

for $t \neq 0$. From the implicit function theorem, we get that there exists a curve $x(u)$ with parameter u in a neighbourhood of 0 and $x(0) = (x_1, \ldots, x_m)$ such that locally around $(x_1, \ldots, x_m)$,

$$t(\rho(u)) = \sum_{i=1}^{m} x_i(u)$$

is the only decomposition of $t(\rho(u))$ as a sum of m elements of W. Note that the condition of the implicit function theorem is fulfilled by (2.11). Not all curves $x_1(u), \ldots, x_m(u)$ can have their image contained in $\mathbb{C}\cdot\{x_1, \ldots, x_m\}$ since $\sum_{i=1}^{m} x_i(u) \notin \langle x_1, \ldots, x_m\rangle$. Thus, without loss of generality, $x_1(u)$ is not constant modulo scalars. Then for all u in a neighbourhood of 0,

$$T_{x_1(u)}W \subseteq T_{t(\rho(u))}\sigma_m(W) = L$$

Thus $\Gamma_W(x_1, \ldots, x_m) = \{z \in W \mid T_zW \subseteq L\}$ contains a curve through x_1 and so the tangential contact locus (cf. 2.2.23) does, too. □

2.2.27 Theorem ([21, Proposition 2.4]): Let W be a nondegenerate, irreducible smooth variety and $m \in \mathbb{N}_{>0}$. Consider the following statements:

(i) The m-th secant map $\psi_{m,W}$ is generically identifiable.

(ii) For every m *general* points $x_1, \ldots, x_m \in W$, $T_{x_1}W, \ldots, T_{x_m}W$ are skew spaces and the dimension of $\mathcal{C}_W(x_1, \ldots, x_m)$ at every x_i is zero.

(iii) There exist m *specific points* $x_1, \ldots, x_m \in W$ with skew tangent spaces

$$T_{x_1}W, \ldots, T_{x_m}W$$

such that the dimension of $\mathcal{C}_W(x_1, \ldots, x_m)$ at a specific x_i is zero.

Then we have (iii) $\Longrightarrow$ (ii) $\Longrightarrow$ (i).

Proof. **(iii) $\implies$ (ii):** Let $x = (x_1, \dots, x_m) \in W^m$ be such that the tangent spaces at $x_1, \dots, x_m$ are skew and $\mathcal{C}_W(x)$ has dimension 0 at x_1. There is an open neighbourhood $U \subset W^m$ of x, such that $T_u := \bigoplus T_{u_i}\widehat{W} = \sum T_{u_i}\widehat{W}$ is a vector bundle over U. We may consider a variety Y of pairs $(u, z) \in U \times W$ such that $T_z W \subset \mathbb{P}(T_u)$ with a natural projection map $\pi : Y \to U$. Each fiber of π is equal to $\mathcal{C}_W(u)$. Thus, the fiber of π over x has dimension zero at (x, x_1). By Proposition 2.2.15, this must be also locally true, and hence (ii) must hold for general x.
(ii) $\implies$ (i): The contraposition is proven in Proposition 2.2.26. □

2.2.28 Lemma: Let $f\colon W_1 \to W_2$ be a dominant morphism between two irreducible *affine* varieties W_1, W_2 of the same dimension. Let $\mathcal{U} \subseteq W_1$ a dense open subset of W_1. Then, there exists a dense open subset $\mathcal{V}$ of W_2 such that

$$f^{-1}(\mathcal{V}) \subseteq \mathcal{U}$$

Proof. Consider

$$\overline{f(W_1 \setminus \mathcal{U})} \subset W_2$$

which is a subvariety of W_2 of lower dimension, since the closed set $W_1 \setminus \mathcal{U}$ has dimension strictly smaller than W_1 and thus, by assumption, strictly smaller than W_2. Since f is dominant, the interior of $\operatorname{im} f$ is nonempty. Therefore,

$$\mathcal{V} := (\operatorname{im} f)^\circ \setminus \overline{f(W_1 \setminus \mathcal{U})} \subseteq f(\mathcal{U}) \subseteq W_2$$

is a dense open subset of W_2 and by construction,

$$f^{-1}(\mathcal{V}) \subseteq \mathcal{U},$$

for if $w \in f^{-1}(\mathcal{V})$ was not in $\mathcal{U}$, then $v := f(w) \in f(W_1 \setminus \mathcal{U}) \cap \mathcal{V} = \emptyset$. □

2.2.29 Lemma: Let W be an irreducible variety, $m \in \mathbb{N}_0$ and $\mathcal{U}$ be an open dense subset of $\widehat{W}^m$. Consider the map

$$\psi_{m,W}\colon \widehat{W}^m \to \widehat{\sigma_m(W)}, (x_1, \dots, x_m) \mapsto \sum_{i=1}^m x_i$$

from Remark 2.2.17. Assume $\sigma_m(W)$ has the expected dimension. Then there exists a dense open subset $\mathcal{U}' \subseteq \mathcal{U} \subseteq \widehat{W}^m$ such that for all $x \in \mathcal{U}'$:

$$\psi_{m,W}^{-1}(\sum_{i=1}^m x_i) \subseteq \mathcal{U}$$

Proof. First, we note that the map $\psi_{m,W}$ is a dominant morphism. Furthermore, $\widehat{W}^m$ and $\widehat{\sigma_m(W)}$ have the same dimension by assumption. Therefore we can apply Lemma 2.2.28 to obtain a dense open subset $\mathcal{V}$ of $\widehat{\sigma_m(W)}$ such that

$$\psi_{m,W}^{-1}(\mathcal{V}) \subseteq \mathcal{U}$$

Now, choose $\mathcal{U}' := \psi_{m,W}^{-1}(\mathcal{V})$. $\mathcal{U}'$ is nonempty and open, thus dense, satisfies $\mathcal{U}' \subseteq \mathcal{U}$ and it holds

$$\psi_{m,W}^{-1}(\psi_{m,W}(\mathcal{U}')) \subseteq \mathcal{U}$$

□

2.2.30 Remark: If in the setting of Proposition 2.2.26, the generic element of $\sigma_m(W)$ is not identifiable, then the tangential contact locus to W at x must in fact contain a curve through x_i for *every* $i \in \{1, \dots, m\}$. This is a monodromy argument and mentioned in the original proof (cf. [22, Proposition 2.3]): As the monodromy group acts transitively on a general fiber, if we have a curve through one point x_1, there must also exist a curve through any other x_i. At this point, the author admits that he does not understand monodromy theory and will therefore refrain from further explanations. The interested reader is referred e.g. to Cifani et al. [24, Section 2.2], which in turn refers to Harris [42]. It is possible to prove everything we need with the slightly weaker formulation from Proposition 2.2.26. In fact, in Theorem 3.2.9, the specific instance will be so symmetric that it suffices to show local zero-dimensionality at 3 points.

2.2.31 Lemma: Let W a variety and $m \in \mathbb{N}_{\geq 2}$. If W is generically m-identifiable, then W is also generically $(m-1)$-identifiable.

Proof. To the contrary, assume W was generically m- but not $(m-1)$-identifiable. Let $\mathcal{U}$ denote a dense open subset of $\widehat{\sigma_m(W)}$ such that every element of $\mathcal{U}$ has precisely one decomposition as a sum of m elements of $\widehat{W}$. The map

$$\rho\colon \widehat{W}^m \to \widehat{\sigma_{m-1}(W)}, (x_1, \dots, x_m) \mapsto \sum_{i=1}^{m-1} x_i \tag{2.14}$$

is dominant. Its image thus intersects any dense open subset $\mathcal{V}$ of $\widehat{\sigma_{m-1}(W)}$. In particular, we may choose such a subset $\mathcal{V}$ where all elements have more than one decomposition. Choose an element x in the preimage of $\mathcal{V}$ under ρ intersected with the preimage of $\mathcal{U}$ under $\psi_{m,W}$. By construction, there exist $y_1, \dots, y_{m-1} \in \widehat{W}$ such that $x_1 + \dots + x_{m-1} = y_1 + \dots + y_{m-1}$ and y is not just a permutation of $(x_1, \dots, x_{m-1})$. But then we have $\psi_{m,W}(x) = x_1 + \dots + x_m = y_1 + \dots + y_{m-1} + x_m$. This is a contradiction, since $\psi_{m,W}$ maps x to an identifiable element. ↯ □

Graded ideals and Hilbert series

Throughout this section, let $n \in \mathbb{N}$, $X = (X_1, \dots, X_n)$ variables and K a field. For a polynomial $f \in K[X]$ and $d \in \mathbb{N}_0$, we denote by $f_{=d}$ its d-homogeneous part, i.e. the sum of all terms of f of degree equal to d.

2.2.32 Proposition and Definition: An ideal $I \subseteq K[X]$ is called *homogeneous*, if one of the two equivalent conditions hold:

(a) I is generated by (finitely many) homogeneous polynomials

(b) For each $f \in K[X]$,

$$f \in I \iff \forall d \in \mathbb{N}_0 : f_{=d} \in I \tag{2.15}$$

For any homogeneous ideal I, the (total) degree is constant on equivalence classes modulo I and thus well-defined on elements of the quotient algebra $K[X]/I$. Via the degree, $K[X]/I$ becomes a graded algebra.

2.2.33 Proposition and Definition: For an ideal $I \subseteq K[X]$ and $d \in \mathbb{N}_0$, we call $I_d := I \cap K[X]_d$ the *degree-d part* of the ideal I. Whenever I is generated by homogeneous polynomials $f_1, \dots, f_m \in K[X]$, $m \in \mathbb{N}_0$, it holds

$$I_d = \{\sum_{i=1}^{m} f_i h_i \mid \forall i \in \{1, \dots, m\} : f_i h_i \in K[X]_d\} \tag{2.16}$$

2.2.34 Definition: Given a homogeneous ideal $I \subseteq K[X]$, the *Hilbert series* of the *ideal* I in the scalar unknown T is

$$\operatorname{HS}(I, T) := \sum_{d \in \mathbb{N}_0} \dim(I_d) T^d \in \mathbb{Z}[[T]] \tag{2.17}$$

On the other hand, the Hilbert series of the associated *quotient ring* is

$$\operatorname{HS}(K[X]/I, T) := \sum_{d \in \mathbb{N}_0} \dim(K[X]_d / I_d) T^d \in \mathbb{Z}[[T]] \tag{2.18}$$

Note that the coefficients of both Hilbert series are nonnegative integers. They are given by the *Hilbert function* of I and $K[X]/I$, respectively, which are

$$\operatorname{HF}_I \colon \mathbb{N}_0 \to \mathbb{N}_0, d \mapsto \dim(I_d) \tag{2.19}$$

$$\operatorname{HF}_{K[X]/I} \colon \mathbb{N}_0 \to \mathbb{N}_0, d \mapsto \dim(K[X]_d / I_d) \tag{2.20}$$

Clearly, it holds for any $d \in \mathbb{N}_0$ that $\operatorname{HF}_I(d) + \operatorname{HF}_{K[X]/I}(d) = \dim K[X]_d$.

2.3. Semidefinite Forms and Sums of Squares

2.3.1 Definition: Let U a real, affine space. A polynomial $f \in \mathbb{R}[U]$ is a *sum of squares*, if there are $m \in \mathbb{N}_0$ and $q_1, \dots, q_m \in \mathbb{R}[U]$ such that

$$f = \sum_{i=1}^{m} q_i^2 \tag{2.21}$$

Sum-of-Squares give rise to a proof system for polynomial inequalities. The basic idea is to prove a polynomial inequality $f \geq g$ by writing $f - g$ as a sum of squares. It gets a slight bit more sophisticated when the inequality is not to be proven on all of U, but only on a semialgebraic subset. The following gives a thorough introduction of the proof system and the most relevant notations and geometric objects behind it. For more context on Sum-of-Squares as a proof system, the reader is referred to [40] and [70]. There are also really nice WIP online lecture notes by Barak and Steurer [86] and an extensive survey of Laurent [54].

2.3.2 Proposition and Definition: Let U a real, affine space.

(a) The set of all sums of squares of polynomials of degree (at most) $D \in \mathbb{N}_0 \cup \{\infty\}$ forms a cone in the real vector space $\mathbb{R}[U]_{\leq D}$, which we denote by $\mathrm{SOS}_D(U)$. The elements of its dual cone,

$$\mathrm{SOS}^*_D(U) = \{E \in \mathbb{R}[U]^\vee \mid \forall f \in \mathrm{SOS}_D(U) : E[f] \geq 0\}$$

are called *square-definite* functionals. A square-definite functional E satisfying $E[1] = 1$ is called a *pseudo-expectation operator*.

(b) More generally, the Sum-of-Squares cone of degree $D \in \mathbb{N}_0 \cup \{\infty\}$ over the *inequalities* $g_1 \geq 0, \dots, g_r \geq 0$ and the *equations* $h_1 = 0, \dots, h_k = 0$ (where $r, k \in \mathbb{N}_0$) is denoted by

$$\mathrm{SOS}_D(g_1 \geq 0, \dots, g_r \geq 0, h_1 = 0, \dots, h_k = 0)$$

and consists of all polynomials $f \in \mathbb{R}[U]$ which have a representation

$$f = \sum_{i=1}^{r} \eta_i g_i + \sum_{j=1}^{k} \nu_j h_j \tag{2.22}$$

where the $\eta_i \in \mathrm{SOS}(U)$ are *sums of squares* and $\nu_j \in \mathbb{R}[U]$ are polynomials such that

$$\forall i\colon \deg(\eta_i g_i) \leq D \quad \text{and} \quad \forall j\colon \deg(\nu_j h_j) \leq D \tag{2.23}$$

Writing $\mathcal{A} := \{g_1 \geq 0, \dots, g_r \geq 0, h_1 = 0, \dots, h_k = 0\}$ for the set consisting of these formal inequalities, we call the elements of its dual cone,

$$\mathrm{SOS}^*_D(\mathcal{A}) = \{E \in \mathbb{R}[U]^\vee \mid \forall f \in \mathrm{SOS}_D(\mathcal{A}) : E[f] \geq 0\}$$

the *square-definite* functionals respecting $\mathcal{A}$.

For brevity, we write "in:equality" whenever we do not want to specify whether something is an equation or an inequality. The notation $\mathrm{SOS}_D^*(\mathcal{A})$ intentionally suppressed the dependency on the space U as it is implicitly included in any nonempty set of in:equalities $\mathcal{A}$ as the domain of any of the polynomials contributing to any of the in:equalities. Let us also take the license to sometimes choose variables and adapt the notation accordingly. We will write $\mathcal{A}_{\geq}$ for the polynomials on the left hand sides of the formal inequalities of $\mathcal{A}$ and $\mathcal{A}_{=}$ for the polynomials defining the equations. Thus $\mathcal{A} = \{g \geq 0 \mid g \in \mathcal{A}_{\geq}\} \cup \{h = 0 \mid h \in \mathcal{A}_{=}\}$. In the interpretation of Sum-of-Squares as a *proof system*, the *axioms* are polynomial in:equalities forming a set $\mathcal{A} = \{h_1 \left\{\substack{\geq \\ \text{or} \\ =}\right\} 0, \dots, h_m \left\{\substack{\geq \\ \text{or} \\ =}\right\} 0\}$ and a proof of an inequality $f \geq g$ is to write $f - g$ as a sum of squares over the axioms, i.e. to show

$$f - g \in \mathrm{SOS}_D(\mathcal{A}) \tag{2.24}$$

D is called the *degree* of the proof. Typical notations to indicate that $f \geq g$ has a Sum-of-Squares proof are $f \succeq g$ or, if we do not want to suppress the axioms or the degree in the notation, $\mathcal{A} \vdash f \succeq_D g$, where $\mathcal{A}$ is a set of formal polynomial in:equalities. However, the reader should be alert that, similar to proofs in the regular mathematical proof system, it is neither common practice to indicate any usage of an axiom all the way through a proof nor is it necessary to write a Sum-of-Squares proof as a sequence of formal logical clauses. Instead, let us introduce the basic deduction rules in this chapter until they should be clear and natural to any reader and ultimately proceed to write down Sum-of-Squares proofs in natural language. The following are some elementary properties of Sum-of-Squares proofs.

2.3.3 Proposition: Let $\mathcal{A}$ a system of polynomial in:equalities in $\mathbb{R}[U]$. Then the following hold:

(a) **Transitivity**: If $\mathcal{A} \vdash f \succeq_{D_1} g$ and $\mathcal{A} \vdash g \succeq_{D_2} h$, then also $\mathcal{A} \vdash f \succeq_{\max\{D_1,D_2\}} h$ for all $f, g, h \in \mathbb{R}[U]$, $D_1, D_2 \in \mathbb{N}_0$.

(b) **Substitution Rule:** If $\mathcal{A} \vdash f \succeq_D g$ for some $f, g \in \mathbb{R}[U], D \in \mathbb{N}_0$, W is another affine real space and $s\colon W \to U$ is a polynomial map of degree at most d, then $s^*(\mathcal{A}) \vdash f \circ s \succeq_{Dd} g \circ s$. Here, $s^*(\mathcal{A})$ denotes the image of $\mathcal{A}$ under the pullback of the map s.[3]

(c) **Addition Rule and Multiplication Rule I:** If $\mathcal{A} \vdash f \succeq_{D_1} 0$ and $\mathcal{A} \vdash g \succeq_{D_2} 0$, then both $\mathcal{A} \vdash f + g \succeq_{\max\{D_1,D_2\}} 0$ and $\mathcal{A} \vdash f \cdot g \succeq_{D_1 D_2} 0$ for all $f, g \in \mathbb{R}[U]$, $D_1, D_2 \in \mathbb{N}_0$.

(d) **Multiplication Rule II:** If $\mathcal{A} \vdash f \succeq_{D_1} g$ and $\mathcal{A} \vdash h \succeq_{D_2} \ell \succeq 0$, then also $\mathcal{A} \vdash f \cdot h \succeq_{D_1 D_2} g \cdot \ell$ for all $f, g, h, \ell \in \mathbb{R}[U]$, $D_1, D_2 \in \mathbb{N}_0$.

(e) **Hierarchiality:** A Sum-of-Squares proof of some degree D is also a Sum-of-Squares proof of all higher degrees:

$$\mathrm{SOS}_0(\mathcal{A}) \subseteq \mathrm{SOS}_1(\mathcal{A}) \subseteq \mathrm{SOS}_2(\mathcal{A}) \subseteq \dots$$

[3] I.e. the set of in:equalities that we obtain by concatenating each polynomial contributing to these in:equalities with s.

> The inclusion chain for the dual cones is reversed:
>
> $$\mathrm{SOS}_0^*(\mathcal{A}) \supseteq \mathrm{SOS}_1^*(\mathcal{A}) \supseteq \mathrm{SOS}_2^*(\mathcal{A}) \supseteq \dots$$
>
> Note a technical subtlety: The square-definite functionals were defined on the entire polynomial algebra $\mathbb{R}[U]$ precisely to make this chain of inclusions true. However, we can still see them as finite objects, since the values on monomials of degree greater than D of some $E \in \mathrm{SOS}_D^*(\mathcal{A})$ can be those of a completely arbitrary functional and thus be disregarded in any analysis. Sometimes it is reasonable to do so and view the dual SOS cone $\mathrm{SOS}_D(U)^*$ instead as a subset of $\mathbb{R}[U]_{\leq D}^{\vee}$.

Note that in Proposition 2.3.3(d) can easily be obtained from (c). The following is a collection of some basic but important inequalities. As all of them make use of an inner product, let us switch to $U = \mathbb{R}^n$ from now on, with variables $X = (X_1, \dots, X_n)$.

2.3.4 Lemma: (SOS triangle inequality) Let $a, b \in \mathbb{R}^n$. Then

$$\|X - a\|^2 \preceq 2(\|X - b\|^2 + \|b - a\|^2)$$

Proof. Let Y denote another unknown vector of same dimension as X. We have $\|X - Y\|^2 = (X - Y)^T(X - Y) = X^TX + Y^TY - 2X^TY$. Furthermore, $-2X^TY \preceq X^TX + Y^TY$, since the right hand side minus the left hand side is $\|X + Y\|^2$. Thus

$$\|X - Y\|^2 \preceq 2(X^TX + Y^TY)$$

Now substituting $X \mapsto X - b$ and $Y \mapsto a - b$ yields the claim. □

2.3.5 Lemma: (SOS Cauchy-Schwarz-Inequality) Let $f, g \in \mathbb{R}[X]^n$. Then

$$(f^Tg)^2 \preceq \|f\|^2\|g\|^2$$

Proof. Let T an additional indeterminate. From $\|f - Tg\|^2 \succeq 0$ we obtain $f^Tf + T^2g^Tg \succeq 2Tf^Tg$. Substituting $T \mapsto \dfrac{f^Tg}{g^Tg}$ and multiplying by g^Tg we obtain

$$f^Tfg^Tg + (f^Tg)^2 \succeq 2(f^Tg)^2 \iff f^Tfg^Tg \succeq (f^Tg)^2$$

□

An important secondary type of Sum-of-Squares proofs are matrix Sum-of-Squares proofs. Their importance lies in the fact that they can give nonlinear relations in between the evaluations of a square-definite functional E, and they are usually obtained by finding Gram factorizations $G = B^TB$ for matrices G of polynomials. The most basic example is the dual Cauchy-Schwarz inequality, which in its simplest form says that

$$E[fg]^2 \leq E[f^2]E[g^2], \qquad (f, g \in \mathbb{R}[X])$$

for each $E \in \mathrm{SOS}_2(\mathbb{R}^n)$. In the following, we will give an instructive, albeit redundant, proof of this fact, where f and g might even be vectors of polynomials, and note that Lemma 2.3.6 is just a special case of the subsequent Lemma 2.3.7 where L can be taken as the identity matrix.

2.3.6 Lemma: (Dual Cauchy-Schwarz inequality) Let $f, g \in \mathbb{R}[X]^n$ vectors of polynomials of degree $D \in \mathbb{N}_0$ and $E \in \mathrm{SOS}^*_{2D,X}$. Then

$$E[f^Tg]^2 \leq E[f^Tf]E[g^Tg]$$

Proof. Let us claim that

$$A := \begin{pmatrix} f^Tf & f^Tg \\ f^Tg & g^Tg \end{pmatrix}$$

admits a factorization $A = B^TB$ where B is a matrix polynomial of degree at most D with two columns. Once this claim is shown, note that an immediate consequence is that for each $v \in \mathbb{R}^2$, $v^TAv = (Bv)^TBv$ is a sum of squares in $\mathbb{R}[X]_{\leq D}$. Therefore,

$$0 \leq E[v^TAv] = v^TE[A]v$$

and thus

$$E[A] = \begin{pmatrix} E[f^Tf] & E[f^Tg] \\ E[f^Tg] & E[g^Tg] \end{pmatrix} \succeq 0$$

is positive semidefinite. We deduce that the determinant $E[f^Tf]E[g^Tg] - E[f^Tg]^2$ of this matrix is nonnegative, which shows the claim. To construct the factorization, observe that

$$A = \underbrace{\begin{pmatrix} 1 & \dots & 1 & & & \\ & & & 1 & \dots & 1 \end{pmatrix}}_{=V^T} \underbrace{\begin{pmatrix} f_1^2 & & & f_1g_1 & & \\ & \ddots & & & \ddots & \\ & & f_n^2 & & & f_ng_n \\ f_1g_1 & & & g_1^2 & & \\ & \ddots & & & \ddots & \\ & & f_ng_n & & & g_n^2 \end{pmatrix}}_{:=M} \underbrace{\begin{pmatrix} 1 & \\ \vdots & \\ 1 & \\ & 1 \\ & \vdots \\ & 1 \end{pmatrix}}_{=:V}$$

so it suffices to factor M as

$$M = \begin{pmatrix} f_1 & & & \\ & f_2 & & \\ & & \ddots & \\ & & & f_n \\ g_1 & & & \\ & g_2 & & \\ & & \ddots & \\ & & & g_n \end{pmatrix} \begin{pmatrix} f_1 & & & & g_1 & & & \\ & f_2 & & & & g_2 & & \\ & & \ddots & & & & \ddots & \\ & & & f_n & & & & g_n \end{pmatrix}$$

□

The following is a much more general version of Lemma 2.3.6. The latter can be obtained by taking $m = \ell$ and $L = I_\ell$ as the identity matrix.

2.3.7 LEMMA: (Spectral norm bound) For $L \in \mathbb{R}^{m\times\ell}$, if $c \in \mathbb{R}_{\geq 0}$ is at least the spectral norm of L, then for any two vectors $f \in \mathbb{R}[Z]^m, g \in \mathbb{R}[Z]^\ell$ of polynomials of degree at most $d \in \mathbb{N}_0$ and any $E \in \mathrm{SOS}^*_{2d}(\mathbb{R}^n)$,

$$c^2 E[f^T f]E[g^T g] - E[f^T Lg]^2 \geq 0$$

Proof. It suffices to show that for any $E \in \mathrm{SOS}^*_{2d}(\mathbb{R}^n)$,

$$\begin{pmatrix} cE[f^T f] & E[f^T Lg] \\ E[f^T Lg] & cE[g^T g] \end{pmatrix} \succeq 0$$

since the claim is that the determinant of this matrix is nonnegative. To this end, consider the matrix

$$\mathcal{L}_c := \begin{pmatrix} cI_m & L \\ L^T & cI_\ell \end{pmatrix} \in \mathbb{R}^{\ell+m\times\ell+m}$$

which I claim is psd if and only if $c \in \mathbb{R}_{\geq 0}$ is at least the spectral norm of L. Indeed, it is easy to see that $\mathcal{L}_0$ is psd only if $L = 0$. For $c > 0$, by the Schur complement criterion, $\mathcal{L}_c$ is psd iff the Schur complement $cI_\ell - c^{-1}L^T I_m L$ is psd, which is the case iff $c^2 I_n \succeq L^T L$. The latter is satisfied iff c^2 is at least the largest eigenvalue of $L^T L$ and hence c is at least the spectral norm $\|L\|$ of L. Fix now some $c \geq \|L\|$ and a factorization $\mathcal{L}_c = M^T M$ for some $M \in \mathbb{R}^{\ell+m\times\ell+m}$. Clearly then

$$\begin{pmatrix} c\, f^T f & f^T Lg \\ f^T Lg & c\, g^T g \end{pmatrix} = (f^T\ g^T)\mathcal{L}_c \begin{pmatrix} f \\ g \end{pmatrix} = \left(M \begin{pmatrix} f \\ g \end{pmatrix} \right)^T \left(M \begin{pmatrix} f \\ g \end{pmatrix} \right)$$

has a factorization as a Gramian of a matrix of polynomials. Therefore, for each $v \in \mathbb{R}^2$,

$$v^T \begin{pmatrix} c\, f^T f & f^T Lg \\ f^T Lg & c\, g^T g \end{pmatrix} v \in \mathrm{SOS}_{2d}(\mathbb{R}^n)$$

and thus for each $E \in \mathrm{SOS}^*_{2d}(\mathbb{R}^n)$

$$\begin{pmatrix} c\, E[f^T f] & E[f^T Lg] \\ E[f^T Lg] & c\, E[g^T g] \end{pmatrix} \succeq 0$$

We deduce that the determinant of this matrix is nonnegative, i.e.

$$c^2 E[f^T f]E[g^T g] \geq E[f^T Lg]^2$$

□

The fact that we can certify spectral norm upper bounds in the Sum-of-Squares framework has significant implications for semidefinite relaxations of polynomial optimization problems, which are introduced in the next section. In particular, it also allows to upper-bound the value a pseudo-expectation takes on some polynomial via *flattenings* introduced below.

2.3.8 PROPOSITION AND DEFINITION: (Flattenings) For any $f \in S^d(\mathbb{R}^n)$, there is a unique *totally symmetric*[4] d-linear form $\mathcal{T}_f \colon \mathbb{R}^n \times \ldots \times \mathbb{R}^n \to \mathbb{R}$ such that for all $x \in \mathbb{R}^n$,

$$f(x) = \mathcal{T}_f\langle x, \ldots, x\rangle \tag{2.25}$$

[4] That is, $\mathcal{T}_f$ is symmetric under arbitrary permutations of indices.

Fix $k \in \{0, \dots, d\}$. We denote the k-mode-flattening of this d-form (with respect to an arbitrary index) to a bilinear map by $T_f \colon (\mathbb{R}^n)^{\otimes k} \times (\mathbb{R}^n)^{\otimes d-k} \to \mathbb{R}$. Then for all $x \in \mathbb{R}^n$,

$$f(x) = T_f\langle x^{\otimes k}, x^{\otimes d-k}\rangle \tag{2.26}$$

We can see T_f as a matrix by identifying $x^{\otimes k}$ and $x^{\otimes d-k}$ with the vectors of k-fold and $(d-k)$-fold products of entries of x, respectively. Define

$$\|f\|_{\text{flat}} := \|f\|_{k\text{-flat}} := \|T_f\|_{\text{spec}} \tag{2.27}$$

to be the *flattening norm* of f.[5] More generally, for an arbitrary polynomial f on $\mathbb{R}^n$ of degree at most d, set

$$\|f\|_{\text{flat}} = \max\{\|f_{=l}\|_{\text{flat}} \mid l \in \{0, \dots, d\}\} \tag{2.28}$$

where for $l \in \mathbb{N}_0$ $f_{=l}$ denotes the l-th homogeneous part of f. This defines a norm on the space of polynomials on $\mathbb{R}^n$ of degree at most d. Note that $\|f_{=0}\|_{\text{flat}} = |f(0)|$.

2.3.9 Lemma: Let $N, d, C \in \mathbb{N}$, $Z = (Z_1, \dots, Z_N)$ and $E \in \text{SOS}_{2d}(\|Z\|^2 \leq C)$ a pseudo-expectation operator such that there exists $z \in \mathbb{R}^n$ with $E[\|Z - z\|^2] = 0$. Then $E = \delta_z$.

Proof. Via substitution, wlog $z = 0$. Thus assume $E[\|Z\|^2] = 0$ and let $f \in \mathbb{R}[Z]_d$. By passing to homogeneous parts, wlog let f homogeneous of degree $k \in \{0, \dots, d\}$. For $k \leq 2$ the claim is easy and will be shown later. Thus let $k \geq 3$. Choose a matrix flattening T_f of f such that $f = T_f\langle Z^{\otimes k/2}, Z^{\otimes k/2}\rangle$, if k is even, or $f = T_f\langle Z^{\otimes\lceil k/2\rceil}, Z^{\otimes\lfloor k/2\rfloor}\rangle$ in general. By Lemma 2.3.7, we have

$$E[T_f\langle Z^{\otimes\lceil k/2\rceil}, Z^{\otimes\lfloor k/2\rfloor}\rangle]^2 \leq \|T_f\|^2 E[\|Z^{\otimes\lceil k/2\rceil}\|_F^2] E[\|Z^{\otimes\lfloor k/2\rfloor}\|_F^2] \tag{2.29}$$

Note that the Frobenius norm commutes with the (tensor) power, i.e. for each $l \in \mathbb{N}$ it holds $\|Z^{\otimes l}\|_F^2 = \|Z\|^{2l}$. Since E satisfies the Archimedeanity constraint $\|Z\|^2 \leq C$ and $\lceil k/2\rceil \geq 2$, it holds $E[\|Z^{\otimes\lceil k/2\rceil}\|_F^2] \leq C^{\lceil k/2\rceil - 2} E[\|Z\|^2] = 0$. Thus the right hand side in (2.29) is zero. But the left hand side is $E[f]^2$. Thus $E[f] = 0$, if f is homogeneous of degree at least 3. For quadratic homogeneous forms $f = T_f\langle Z, Z\rangle$, the proof is essentially the same, since $E[T_f\langle Z, Z\rangle]^2 \leq \|T_f\| E[\|Z\|^2]^2$. Finally for linear forms, note that

$$E[Z_i]^2 = E[e_i^T Z]^2 \leq E[1] E[Z^T Z] = 0 \tag{2.30}$$

by the dual Cauchy-Schwarz inequality Lemma 2.3.6. Thus for an arbitrary polynomial f of degree at most $2d$, $E[f] = E[f(0)] = f(0)E[1] = f(0)$, since $E[1] = 1$. But this means that $E = \delta_0$. □

2.3.10 Remark: (a) In Lemma 2.3.9, it is necessary to have an Archimedeanity constraint like "$\|Z\|^2 \leq C$". In the next section, we will get to know special Dirac measures supported "at infinity", which would otherwise satisfy the assumptions of Lemma 2.3.9 too, but not the conclusion.

[5] To be precise, there are several flattening norms, since for each l one may choose a $k_l \in \{1, \dots, l\}$ which is suppressed in the notation. It is common to choose the most "balanced" flattening, i.e. $k_l = \lfloor l/2\rfloor$.

(b) It is important that the maximum degree $2d$ is even, since for all odd degrees, (2.29) involves terms of degree larger than k.

As an application of Proposition and Definition 2.3.8, the following Lemma shows that Sum-of-Squares programming can be used to find a local minimizer of a polynomial which is strictly convex around some point z, provided it is told a starting value and a sufficiently tight ball constraint. It is not paradigmatic to use Sum-of-Squares for local optimization, since nonconvex methods have far superior performance in this setting. Thus for the time being, the following is a toy application merely illustrating the general spirit of Sum-of-Squares proof techniques. However, in Chapter 4, Lemma 2.3.11 will be applied in a "*semi-local*" framework, where local optimization methods are not directly applicable.

2.3.11 Lemma: Let $N, d \in \mathbb{N}$ with $d \geq 2$, $Z = (Z_1, \dots, Z_N)$, $g \in \mathbb{R}[Z]_{\leq d}$ and z a local minimizer of g such that $\operatorname{Hess} g(z) \succ 0$.

(a) There is some $\varepsilon > 0$ such that for any "starting value" $u \in B_\varepsilon(z)$, any pseudo-expectation $E \in \mathrm{SOS}^*_{2\lceil \frac{d}{2} \rceil}(\varepsilon^2 \geq \|Z - u\|^2)$ minimizing $E[g]$ satisfies

$$E = \delta_z \tag{2.31}$$

(b) More precisely, the above is true whenever $\|g_{\geq 3}\|_{\text{flat}} \dfrac{4\varepsilon^2}{1 - 4\varepsilon^2} < \lambda_{\min}$. Here, $\lambda_{\min}$ denotes the smallest Eigenvalue of $\operatorname{Hess} g(z)$ and $g_{\geq 3}$ is the sum of the homogeneous parts of g of degree at least 3.

Proof. We use a Taylor expansion around z to write g as

$$g - g(z) = (Z - z)^T \operatorname{Hess} g(z)(Z - z) + h(Z - z) \tag{2.32}$$

for some polynomial h in which all monomials have degree at least 3. Note that $\nabla g(z) = 0$, since z is a critical point of g. Let $\lambda_{\min}$ denote the smallest Eigenvalue of $\operatorname{Hess} g(z)$. Then $\operatorname{Hess} g(z) \succeq \lambda_{\min} I_S$ and thus

$$g - g(z) \succeq \lambda_{\min} \|Z - z\|^2 + h(Z - z) \tag{2.33}$$

Since E minimizes $E[g]$, we have $E[g] \leq g(z)$ and thus

$$0 \geq E[g] - g(z) \geq \lambda_{\min} E[\|Z - z\|^2] + E[h(Z - z)] \tag{2.34}$$

We will show that $E[\|Z - z\|^2] = 0$ by asserting that otherwise for sufficiently small $\varepsilon > 0$ the right hand side would attain positive values.

Denote for every $k \in \{3, \dots, d\}$ by h_k the k-th homogeneous part of h and choose a matrix flattening T_k of h_k such that $h_k = T_k \langle Z^{\otimes \lceil k/2 \rceil}, Z^{\otimes \lfloor k/2 \rfloor} \rangle$, similar to the previous proof.

Furthermore, let $\|T_k\|$ denote the spectral norm of each of these matrices $T_3, \dots, T_d$ and substitute $q := Z - z$. Using Lemma 2.3.7 with $L = T_k$ and the polynomial vectors $a = q^{\otimes \lfloor k/2 \rfloor}$ and $b = q^{\otimes \lceil k/2 \rceil}$ we obtain

$$E[h_k(q)]^2 = E[a^T T_k b]^2 \leq \|T_k\|^2 E[a^T a] E[b^T b] \tag{2.35}$$

Note that the Frobenius norm of the rank-1 tensor $q^{\otimes \ell}$ satisfies

$$\|q^{\otimes \ell}\|_F^2 = \|q\|^{2\ell} \tag{2.36}$$

for each $\ell \in \mathbb{N}_0$. By the SOS triangle inequality Lemma 2.3.4, it holds that $E \in \mathrm{SOS}_d^*(\|q\|^2 \leq 4\varepsilon^2)$, since

$$\|q\|^2 = \|Z - z\|^2 \preceq 2\|Z - u\|^2 + 2\|z - u\|^2 \tag{2.37}$$

and $E[\|Z - u\|^2 + \|z - u\|^2] \leq 2\varepsilon^2$. Applying (2.36) and (2.37) to (2.35) yields

$$E[h_k(q)]^2 \leq \|T_k\|^2 E[\|q\|^{2\lceil k/2\rceil}] E[\|q\|^{2\lfloor k/2\rfloor}] \leq (4\varepsilon^2)^{2k-4}\|T_k\|^2 E[\|q\|^2]^2 \tag{2.38}$$

Hence

$$\pm E[h_k(q)] \leq (4\varepsilon^2)^{k-2}\|T_k\| E[\|q\|^2] \tag{2.39}$$

Set $c := \max_{k\geq 3} \|T_k\|$. Using the geometric series, we can bound the sum over $k \geq 3$ of $(4\varepsilon^2)^{k-2}$ by $\dfrac{4\varepsilon^2}{1-4\varepsilon^2} \in \mathcal{O}(\varepsilon)$. We get

$$\pm E[h(q)] \leq c\frac{4\varepsilon}{1-4\varepsilon} E[\|q\|^2] \tag{2.40}$$

and thus in our initial estimate:

$$0 \geq E[g] - g(z) \geq (\lambda_{\min} - c\frac{4\varepsilon}{1-4\varepsilon}) E[\|Z - z\|^2] \tag{2.41}$$

If ε is small enough such that $\lambda_{\min} > \dfrac{4\varepsilon}{1-4\varepsilon}$, then this cannot be satisfied unless $E[\|Z-z\|^2] = 0$. By Lemma 2.3.9, this asserts both (a) and (b). The maximum degree of a Sum-of-Squares proof in the above argumentation is $2\lceil d/2\rceil$. □

Finally, let us introduce matrix pseudoinverses as a useful tool that will be needed in Section 2.4 and did not fit anywhere else.

2.3.12 Proposition and Definition: Let $m, n \in \mathbb{N}_0$ and $A \in \mathbb{R}^{n\times m}$. A^+ is called a *(Moore-Penrose) pseudoinverse* of A, if the following 4 conditions hold:

(i) $AA^+A = A$

(ii) $A^+AA^+ = A^+$

(iii) $(AA^+)^T = AA^+$

(iv) $(A^+A)^T = A^+A$

Every matrix $A \in \mathbb{R}^{n\times m}$ has a unique pseudoinverse. In addition, it holds that $(A^T)^+ = (A^+)^T$. Note that for invertible matrices A, the pseudoinverse coincides with the inverse. If A has full column rank, then it holds that $A^+ = (A^TA)^{-1}A^T$. Thus A^+ is a left-inverse for matrices of full column rank. In that case, we have $(AA^T)^+ = (A^+)^TA^+ = A(A^TA)^{-1}(A^TA)^{-1}A^T$ and therefore $A^T(AA^T)^+A = I_m$.

Spectrahedra, Interior point methods and Gram matrices

This section collects standard knowledge on the connection between polynomial optimization, moments, positive semidefinite matrices and Sum-of-Squares proofs. The underlying theory is huge, but for the limited purpose of this work, only the main notions and some instructive theorems are introduced, while scratching on the surface of their proofs. An in-depth treatment of the topic can be found in Laurent's survey [54].

For each $n \in \mathbb{N}_0$, let $S\mathbb{R}^{n \times n}$ denote the $\mathbb{R}$-vector space of symmetric $n \times n$ matrices.

2.3.13 Reminder: A matrix $G \in S\mathbb{R}^{n \times n}$ is called *positive semidefinite* (psd), denoted $G \succeq 0$, if one of the following equivalent conditions hold:

(a) $\forall v \in \mathbb{R}^n : v^T G v \geq 0$

(b) All eigenvalues of G are nonnegative.

(c) All coefficients of the polynomial $\det(G + TI_n)$ in T are nonnegative.

(d) There exists $m \in \mathbb{N}_0$ and $A \in \mathbb{R}^{m \times n}$ such that G has a *Gram factorization* $G = A^T A$.

(e) For all $m \geq \operatorname{rank} G$, there exists $A \in \mathbb{R}^{m \times n}$ such that $G = A^T A$.

(f) All principal minors of A are nonnegative.

(g) $X^T G X$ is a sum of rank(G)-many squares in $\mathbb{R}[X]_2 = \mathbb{R}[X_1, \dots, X_n]_2$.

(h) $\operatorname{sig} G = \operatorname{rank} G$, where $\operatorname{sig} G$ denotes the Sylvester signature of G.

For matrices $G, H \in S\mathbb{R}^{n \times n}$, we write $G \succeq H$ if $G - H \succeq 0$.

Proof. This is folk-lore, but can be found e.g. in Schweighofer's lecture notes [85, Proposition 2.3.2 and 2.3.3]. □

2.3.14 Remark: For any $n \in \mathbb{N}_0$ and any space $\mathcal{L} \subseteq S\mathbb{R}^{n \times n}$ of symmetric matrices given by a (suitably well-conditioned, reasonably sized) generating system or linear equations, interior point methods (IPMs) may compute a point in the relative interior of the convex set

$$\{G \in \mathcal{L} \mid G \succeq 0\} \tag{2.42}$$

Preimages of sets of the form (2.42) under linear maps are called *spectrahedra.* Images of spectrahedra under linear maps are called *spectrahedral shadows.* Optimizing a linear functional over a spectrahedral shadow is called a semidefinite program, or SDP for short. Renegar's book [76] gives an overview on IPMs and the surrounding duality theory. Under some assumptions, solving SDPs can be theoretically efficient, but state of the art solvers might struggle with large instances. At the same time, some applications stemming from convex hierarchies are notorious for producing SDPs with prohibitively large memory requirements. It is an oft-stated belief that SDPs can be solved in polynomial time in the matrix size. O'Donnell reminded the community that this is not obviously true [69], but of course this only means that we have to change the meaning of the word "efficient" to include algorithms that may involve solving SDPs whose matrix size is polynomial in the input.

2.3.15 Proposition and Definition: Let $N \in \mathbb{N}_0$ and $u = (u_1, \dots, u_N)$ some vector of distinct monomials in $X = (X_1, \dots, X_n)$. A matrix G is called a *Gram matrix representation* of the polynomial f with respect to u, if G is psd and

$$f = u^T G u \tag{2.43}$$

For any polynomial f whose monomials with nonzero coefficients are contained in $\{u_i u_j \mid i, j \in \{1, \dots, N\}\}$, it holds that f is a sum of squares if and only if it has a Gram matrix representation with respect to u.

Proof. This is classical, see e.g. [85, 2.6]. □

2.3.16 Proposition: Let $\mathcal{A}$ a system of polynomial in:equalities on the real affine space U. Let $d \in \mathbb{N}$. Then $\mathrm{SOS}_d(\mathcal{A})^*$ is a spectrahedron and $\mathrm{SOS}_d(\mathcal{A})$ is a spectrahedral shadow.

Proof. Let us show the claims without any in:equalities, as the general case is very similar. A functional E on $\mathbb{R}[U]_{\leq d}$ is square definite if and only if $E[p^2] \geq 0$ for any $p \in \mathbb{R}[U]$ with $\deg(p^2) \leq d$. Let $n := \dim(U)$ and choose variables $Z = (Z_1, \dots, Z_n)$ for U. For each $k \in \mathbb{N}$, let mon_k denote the vector of monomials in Z. In particular, for $d' := \lfloor d/2 \rfloor$, we may write $p = c_p^T \mathrm{mon}_{d'}$, where c_p denotes the coefficient vector of p w.r.t. the monomials. By linearity, it holds that $E[p^2] = c_p^T E[\mathrm{mon}_{d'} \mathrm{mon}_{d'}^T] c_p$ for each polynomial p. We therefore see that $E[p^2] \geq 0$ holds for all polynomials with $\deg(p) \leq d'$ if and only if $M_E := E[\mathrm{mon}_{d'} \mathrm{mon}_{d'}^T]$ is psd. M_E is called the *moment matrix* of E. The map $E \mapsto M_E$ is linear and surjects to the cone of psd matrices of appropriate size. Thus $\mathrm{SOS}_d(U)^*$ is a spectrahedron.

By Proposition and Definition 2.3.15, a polynomial f of degree at most d is a sum of squares if and only if it has a Gram matrix representation with respect to $\mathrm{mon}_{d'}$. The operation that maps a psd matrix G to the Sum-of-Squares polynomial $\mathrm{mon}_{d'}^T G \,\mathrm{mon}_{d'}$ represented by it is linear. Thus $\mathrm{SOS}_d(U)$ is a spectrahedral shadow. □

The hierarchy of Lasserre

Consider a polynomial optimization problem (POP) of the kind

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{s.t.} \quad & x \in U \\ \forall i \in \{1, \dots, m\} : \quad & g_i(x) \geq 0 \\ \forall i \in \{1, \dots, l\} : \quad & h_i(x) = 0 \end{aligned} \tag{2.44}$$

where U is a real affine space, $l, m \in \mathbb{N}_0$ and $f, g_1, \dots, g_m, h_1, \dots, h_m$ are polynomials on U. POPs can be highly difficult optimization problems both in theory and practice. They are often nonconvex and their amount of local minimizers can grow exponentially in the dimension. A general-purpose approximation scheme for POPs is *Lasserre's hierarchy.* It starts by an artificial convexification of the problem: In the first step, let us replace points $x \in U$ by evaluation maps

$\delta_x \colon U^\vee \to \mathbb{R}, \ell \mapsto \ell(x)$ to obtain the equivalent formulation

$$\begin{aligned} \text{minimize} \quad & \delta_x(f) \\ \text{s. t.} \quad & x \in U \\ \forall i \in \{1, \dots, m\} : \quad & \delta_x(g_i) \geq 0 \\ \forall i \in \{1, \dots, l\} : \quad & \delta_x(h_i) = 0 \end{aligned} \tag{2.45}$$

Notice that while the objective function is not linear in x, it is certainly linear on $\delta(U) = \{\delta_x \mid x \in U\}$. Thus we made the objective function linear, but the price is that the feasible region

$$\{L \in \delta(U) \mid \forall i : L(g_i) \geq 0, L(h_i) = 0\} \tag{2.46}$$

is now a complicated subset of $\mathbb{R}[U]^\vee$. Since a linear objective function attains the same optimal value on (2.46) that it does on its convex hull, we can now convexify the feasible region by replacing $\delta(U)$ with $\operatorname{conv} \delta(U)$. Note that

$$\{L \in \operatorname{conv} \delta(U) \mid \forall i : L(g_i) \geq 0, L(h_i) = 0\} \tag{2.47}$$

is indeed convex, since it is the intersection of $\operatorname{conv} \delta(U)$ with a polyhedron, and thus the convex hull of (2.46). Fortunately, $\operatorname{conv} \delta(U)$ has a convenient description by *Richter's* theorem:

2.3.17 Theorem: (Richter, cf. [77, Satz 4][6]) Let μ a (finite) Borel measure on the real affine space U and let S a finite-dimensional subspace of $\mathbb{R}[U]$ such that $\mathbb{E}_\mu[p] := \int_U p(x) d\mu$ exists for each $p \in S$. Then there exist $m \in \mathbb{N}_0$, $x_1, \dots, x_m \in U$ and $\lambda_1, \dots, \lambda_m \in \mathbb{R}_{\geq 0}$ such that

$$\mathbb{E}_\mu = \sum_{i=1}^{m} \lambda_i \delta_{x_i} \tag{2.48}$$

Furthermore, if a representation (2.48) exists, then there is one with $m \leq \dim S$ and such that all *nodes* $x_1, \dots, x_m$ are contained in the support of μ.

Let us write $\mathrm{MEAS}_d(S) \subseteq S^\vee$ for the cone of those linear functionals on the subspace $\mathbb{R}[U]_{\leq d}$ which are integration operators of (finite) Borel measures and for which all polynomials of degree at most d are integrable.

2.3.18 Proposition and Definition: Let S a subset of the real affine space U and $d \in \mathbb{N}_0$. Then

$$\mathrm{POS}_d(S) = \{f \in \mathbb{R}[U]_{\leq d} \mid \forall x \in S \colon f(x) \geq 0\} \tag{2.49}$$

denotes the cone of *nonnegative polynomials* on S of degree at most d. It holds

$$\mathrm{MEAS}_d(S)^* = \mathrm{POS}_d(S) \tag{2.50}$$

where we interpret the dual cone of $\mathrm{MEAS}_d(S)$ as a subset of $\mathbb{R}[U]_{\leq d}$ (rather than $\mathbb{R}[U]_{\leq d}^{\vee\vee}$). On the converse side, it holds that for any $S \subseteq U$,

$$\mathrm{POS}_d(S)^* = \operatorname{conv}(\delta_x \mid x \in \overline{\mathbb{P}_\mathbb{R}(S)}) \tag{2.51}$$

[6] Or [26, Theorem 19] for an English article stating the theorem. Note that Richter's theorem is formulated for an affine space of functions on a measurable space. To obtain this "polynomial" version, one applies it to a space of polynomial functions modulo the vanishing ideal of $\operatorname{supp} \mu$.

where $\overline{\mathbb{P}_{\mathbb{R}}(S)}$ denotes the (real) projective closure of the subset S in $\mathbb{P}_{\mathbb{R}}(U)$ with respect to the quotient topology of the sphere (note that topologically $\mathbb{P}_{\mathbb{R}}(\mathbb{R}^n) \cong \mathbb{S}^n/\{\pm 1\}$, where the latter has the metric $d(x,y)^2 := 1 - \langle x, y\rangle^2 = \|xx^T - yy^T\|^2_{\text{spec}}$).

Proof. Confer [78, Proposition 4.5] □

The "evaluations at infinity" are a just a minor nuisance since they usually do not obstruct any optimization, at least not on compact sets. If we ignore the minor issue that the cones of integration operators need not be closed, then this dual viewpoint tells us that we can interpret functionals that respect nonnegativity of polynomials as integration operators of measures.

In the previous section, we learnt to see Sum-of-Squares as a proof system to verify nonnegativity of polynomials. But if we see Sum-of-Squares polynomials as "provably nonnegative" polynomials, then what is a reasonable interpretation of the elements of the dual Sum-of-Squares cone, the square definite functionals? By Proposition 2.3.16 and Remark 2.3.14, square-definite functionals form a tractable envelope of the cone of expectation operators of measure, provided that square-definiteness is only required up to some fixed, not too large degree.

Since being a sum of squares is "slightly more" than being nonnegative, being square-definite is "slightly less" than being monotonic, i.e. mapping nonnegative polynomials to nonnegative values. It became a philosophical paradigma to view square-definite functional as *pseudo integration operators*, cf. e.g. the work of Barak and Steurer [86].

The Hierachy of Lasserre is a systematic way to exploit Sum-of-Squares proofs to solve a polynomial optimization problem. The basic idea is to look at the polynomial optimization problem in its dual convexified form

$$\begin{aligned} (P) \qquad \text{minimize} \quad & E[f] \\ \text{s.t.} \quad & E \in \mathrm{POS}_d(S)^* \\ & E[1] = 1 \end{aligned} \tag{2.52}$$

where $S = \{x \in U \mid \forall i : g_i(x) \geq 0, h_i(x) = 0\}$ is the feasible set of the constraints and $d = \deg(f)$. Now, replace the cone $\mathrm{POS}_d(S)^*$ by a dual Sum-of-Squares cone $\mathrm{SOS}_k(g_1 \geq 0, \dots, g_m \geq 0, h_1 = \dots = h_l = 0)^*$. Up to extension of functionals to the correct domain (cf. Proposition 2.3.3(e)), it holds that $\mathrm{POS}_d(S)^* \subseteq \mathrm{SOS}_k(g_1 \geq 0, \dots, g_m \geq 0, h_1 = \dots = h_l = 0)^*$. The optimization problem

$$\begin{aligned} (LP)_k \qquad \text{minimize} \quad & E[f] \\ \text{s.t.} \quad & E \in \mathrm{SOS}_k(g_1 \geq 0, \dots, g_m \geq 0, h_1 = \dots = h_l = 0)^* \\ & E[1] = 1 \end{aligned} \tag{2.53}$$

is called the degree-k Lasserre relaxation of (P), or the k-th level of the Lasserre hierarchy. The parameter $k \in \mathbb{N}$ with $k \geq \max \deg(f, g_1, \dots, g_m, h_1, \dots, h_l)$ is called the *relaxation degree.* Note that Lasserre's hierarchy does not just depend on the set S, but on its description via in:equalities.

Fundamental results in Real Algebraic Geometry suggest that this approach is much more than a heuristic. Artin's famous solution to Hilbert's 17th problem shows that every polynomial which is nonnegative on $\mathbb{R}^n$ is a sum of squares of rational functions. A series of classical results called *Positivstellensätze*, e.g. due to Krivine-Prestel [74], Schmüdgen [83] and Putinar [75], relate nonnegativity of polynomials on a semialgebraic set S described by a system of in:equalities $\mathcal{A}$ to the cone $\mathrm{SOS}(\mathcal{A})$. In fact, under mild conditions, one can show that the optimal values of $(LP)_k$ converge to the optimal value of (P). In Chapter 4 and Chapter 5, we will see two specific examples of polynomial optimization problems where (a variant of) Lasserre's hierarchy achieves even finite convergence.

2.4. Waring decompositions and powers of forms

Throughout this section, let K an arbitrary field of characteristic 0.

2.4.1 Notation: For $k, d \in \mathbb{N}$, let us write

$$V_{d,k} := V_{d,k}(U) := \{q^d \mid q \in \mathbb{P}(S^k(U))\} \tag{2.54}$$

for the projective set of d-th powers of k-forms on the K-vector space U. The space U is often suppressed in the notation.

2.4.2 Proposition: For $k, d, n \in \mathbb{N}$, the map

$$\iota\colon \mathbb{P}(S^k(\mathbb{C}^n)) \to V_{k,d} \subseteq \mathbb{P}(S^{kd}(\mathbb{C}^n)), p \mapsto p^d$$

is an embedding. In particular, $V_{k,d}$ is isomorphic to $\mathbb{P}(S^k(\mathbb{C}^n))$ and thus a smooth variety.

Proof. Clearly ι is bijective and to prove that it is an isomorphism onto the image it is enough to show that locally it is invertible. Let us choose variables $X = (X_1, \dots, X_n)$ and fix $q = p^d = \iota(p)$. Write $p = \left[\sum_{t \in \mathrm{mon}_k(X)} a_t t\right]$ and $q = \left[\sum_{t \in \mathrm{mon}_{kd}(X)} b_t t\right]$ with coefficients $a_t, b_t \in \mathbb{C}$. By making a general change of coordinates and rescaling, we may assume $a_{X_1^k} = 1$. Now ι restricts to a map of affine charts of projective spaces respectively given by $a_{X_1^k} = 1$ and $b_{X_1^{kd}} = 1$. It is an easy exercise to write down a recursive formula for the inverse map by first recovering the coefficients of monomials $m \in \mathrm{mon}_k(X)$ with low degree in $X_2, \dots, X_n$: First, the coefficients $a_{X_1^{k-1}X_i}$ may be reconstructed from $b_{X_1^{kd-1}X_i}$ for $i \in \{1, \dots, n\}$. Next, the coefficients $a_{X_1^{k-2}X_iX_j}$ may be computed from $b_{X_1^{kd-2}X_iX_j}$ for $i, j \in \{1, \dots, n\}$, as the only coefficients of p that can contribute to that are: $a_{X_1^{k-2}X_iX_j}$, which we want, and $a_{X_1^{k-1}X_i}$, $a_{X_1^{k-1}X_j}$ which are already computed, etc. For example, for $d = 3$ and $k = 2$ the inverse map is explicitly given by:

$$\begin{aligned} a_{X_1X_i} &= \frac{1}{3} b_{X_1^5X_i}, \\ a_{X_i^2} &= \frac{1}{3} b_{X_1^4X_i^2} - \frac{1}{9} b^2_{X_1^5X_i}, \\ a_{X_iX_j} &= \frac{1}{3} b_{X_1^4X_iX_j} - \frac{2}{9} b_{X_1^5X_i} b_{X_1^5X_j}, \end{aligned} \tag{2.55}$$

where $1, i, j$ are pairwise disjoint. □

2.4.3 Lemma: For $p \in S^k(\mathbb{C}^n)$, the tangent space of $V_{k,d}$ at p^d is

$$\{hp^{d-1} \mid h \in \mathbb{P}(S^k(\mathbb{C}^n))\} \tag{2.56}$$

Proof. The containment $T_pV_{k,d} \supseteq \{hp^{d-1} \mid h \in \mathbb{P}(S^k(\mathbb{C}^n))\}$ is seen by deriving the parametrization $p \mapsto p^d$. The other inclusion follows by dimension count, as by Proposition 2.4.2, $V_{k,d}$ is a smooth variety. □

2.4.4 Definition: Let U a K-vector space, $p \in S^d(U)$ and $m \in \mathbb{N}_0$. A multiset $\{\ell_1^d, \ldots, \ell_m^d\}$ of d-th powers of linear forms $\ell_1, \ldots, \ell_m$ on U is called a (rank-m) *Waring decomposition* of p, with weights $\lambda_1, \ldots, \lambda_m \in K$ if

$$p = \sum_{i=1}^{m} \lambda_i \ell_i^d \tag{2.57}$$

In that situation, m is called the *rank* of the decomposition. The *Waring rank* of p is the minimal number $n \in \mathbb{N}_0$ (cf. 2.2.6) such that p has a Waring decomposition of rank m.

2.4.5 Remark: Over an algebraically closed field K, the weights in Definition 2.4.4 are redundant parameters and thus omitted by convention. Over a real closed field, the weights can be assumed to be either 1 or -1 in the case of even d and completely omitted in the case of odd d. A consequence of Proposition 2.2.6 is that a Waring decomposition always exists and the maximal Waring rank of some $f \in S^d(U)$ is in fact $\dim S^d(U)$.

The bound from Proposition 2.2.6 for the rank of the decomposition turns out to be too much by roughly a factor of n for *general* forms. The following celebrated theorem states that in all but a few exceptional cases, a general element of $S^d(U)$ will have exactly the rank that one would expect from counting and comparing parameters on both sides of Equation (2.57).

2.4.6 Theorem (Alexander-Hirschowitz, informal, cf. e.g. [53]): For all but finitely many values of $(n, d) \in \mathbb{N} \times \mathbb{N}_{\geq 3}$, the general n-variate d-form $f \in S^d(\mathbb{C}^n)$ has rank

$$\lceil \dim S^d(\mathbb{C}^n)/n \rceil \tag{2.58}$$

which is thus the generic rank of $V_{1,d}(\mathbb{C}^n)$ in all but finitely many exceptional cases.

Subsequently, people first examined skewness of tangent spaces for subgeneric ranks, i.e. when the general element of $\sigma_m(\widehat{V_{1,d}(\mathbb{C}^n)})$ has only finitely many representations as a sum of m d-th powers of linear forms (cf. Proposition and Definition 2.2.20) and then went on to examine under which circumstances such a representation is unique. A line of work [21], [22], [23] completed the classification of cases where identifiability does not hold for all subgeneric ranks. The interested reader is referred to the original publications. The main purpose of this exposition is to provide some context for what we will be doing in Chapter 3 on powers-of-forms decompositions.

2.4.7 Theorem (Chiantini-Ottaviani-Vannieuwenhoven, informal, [23]): For all but finitely many values of $(n, d) \in \mathbb{N} \times \mathbb{N}_{\geq 3}$ and any $m \in \mathbb{N}$ which is strictly smaller than the generic rank of $V_{1,d}(\mathbb{C}^n)$, the latter variety is m-identifiable, i.e. a sum

$$f = \ell_1^d + \ldots + \ell_m^d \in S^d(\mathbb{C}^n) \tag{2.59}$$

of d-th powers of *general* linear forms has no decomposition as a sum of m d-th powers of linear forms other than (2.59).

For both Theorem 2.4.6 and Theorem 2.4.7, the exceptional cases are known explicitly. We do not list them here for brevity and refer the interested reader to [23, Introduction and Theorem 1.1].

Algorithms for Waring Decomposition

Let now $K \in \{\mathbb{R}, \mathbb{C}\}$. We want to give a brief overview on classical algorithms for the following problem: Given a d-form p which has a representation

$$p = \sum_{i=1}^{m} \lambda_i \ell_i^d \tag{2.60}$$

for some $m \in \mathbb{N}_0$, with weights $\lambda_1, \dots, \lambda_m$, output $\ell_1^d, \dots, \ell_m^d$ (in any order). This is the *Waring decomposition problem* for p. Identifiability (cf. 2.2.21) is a crucial first necessary condition for this task to even make sense. Please also remember our convention for the weights in the real and complex case 2.4.5 to see that the problem is not overparameterized.

Real decomposition in minimal degree $d = 3$ is the minimal interesting degree where identifiability can hold. We focus on algorithms for the real field $\mathbb{R}$. Since d is odd, we will not need any weights. The following is a classical result. It is commonly, but not necessarily correctly, attributed to R. Jennrich (via Harshman, [43])[7]. It is to be understood both as an algorithmic decomposition theorem and as a proof of uniqueness of the decomposition. For the decomposition, classical algorithms employ simultaneous matrix diagonalization, cf. e.g. the work of Leurgans, Ross and Abel [57]. I will give a few different proofs, starting with an extremely simple algorithm based on Sums-of-Squares programming. The seconds one uses simultaneous diagonalization.

2.4.8 Theorem: Let $m \leq n$ and $\ell_1, \dots, \ell_m \in S^1(\mathbb{R}^n)$ linearly independent. Then there exists an efficient algorithm that solves the Waring decomposition problem for $p = \sum_{i=1}^m \ell_i^3$. In particular, p has Waring rank m and $(\ell_1^3, \dots, \ell_m^3)$ is the unique Waring decomposition of p.

Algorithmic Proof. (SoS proof of Theorem 2.4.8). Choose general $v \in \mathbb{S}^{n-1}$, e.g. by sampling v at random from a continuous distribution. Compute

$$\partial_v p = 3 \sum_{i=1}^{m} \ell_i(v) \ell_i^2$$

and likewise, $\partial_w p$. Via solving an SDP, compute some relative interior point of the intersection of the space $\{\partial_w p \mid w \in \mathbb{R}^n\}$ with the Sum-of-Squares cone of degree 2. We thus computed a quadratic form of the kind $\sum_{i=1}^m \ell_i(w)\ell_i^2$ where $\ell_1(w) > 0, \dots, \ell_m(w) > 0$. Consider the optimization program

$$\begin{aligned} \text{maximize} \quad & E[\sum_{i=1}^{m} \ell_i(v)\ell_i^2] \\ \text{s.t.} \quad & E[\sum_{i=1}^{m} \ell_i(w)\ell_i^2] = 1 \\ & E \in \mathrm{SOS}_2(\mathbb{R}^n) \end{aligned} \tag{2.61}$$

[7] Jennrich's work is credited in a publication due to R. Harshman, which separately describes a uniqueness result and an algorithm for the polyadic decomposition of (not necessarily symmetric) 3-tensors. The uniqueness result resembles Theorem 2.4.8, but the algorithm of Jennrich is more akin to what nowadays would be called Alternating Least Squares and does not have much to do with the uniqueness result.

which is a Sum-of-Squares optimization program. Note that any feasible solution satisfies

$$\begin{aligned} E[\sum_{i=1}^m \ell_i(v)\ell_i^2] &= E[\sum_{i=1}^m \frac{\ell_i(v)}{\ell_i(w)}\ell_i(w)\ell_i^2] \\ &\leq \left(\max_{i\in\{1,\dots,m\}} \frac{\ell_i(v)}{\ell_i(w)}\right) E[\sum_{i=1}^m \ell_i(w)\ell_i^2] \\ &= \left(\max_{i\in\{1,\dots,m\}} \frac{\ell_i(v)}{\ell_i(w)}\right) \end{aligned}$$

Note that the first estimate uses that $E[\ell_i^2] \geq 0$ for each $i \in \{1,\dots,m\}$. Thus we see that any feasible solution attaining the value $c := \max_{i\in\{1,\dots,m\}} \frac{\ell_i(v)}{\ell_i(w)}$ for the objective must be optimal. Note that the values $\frac{\ell_1(v)}{\ell_1(w)}, \dots, \frac{\ell_m(v)}{\ell_m(w)}$ are pairwise distinct and there is no division by zero due to the general choice of v and w. Wlog reorder $\ell_1,\dots,\ell_m$ such that the ratio is maximal for $i = 1$, i.e. $c = \frac{\ell_1(v)}{\ell_1(w)}$. It is an easy exercise to see that a feasible solution attains this value (and is thus optimal) if and only if $E[\ell_1^2] = 1$ and $E[\ell_i^2] = 0$ for all $i \geq 2$. Compute an optimal solution E^* via semidefinite programming. Now, we are almost done: For $j \in \{1,\dots,n\}$, let e_j denote the j-th standard basis vector and compute $E[\partial_{e_j} p] = 3\frac{\ell_1(e_j)}{\ell_1(w)}$. Up to scaling, we thus computed $(\ell_1(e_1),\dots,\ell_1(e_n))$. Hence we obtain a linear form $l_1 \in \mathbb{R}^\times \ell_1$ which is a scalar multiple of ℓ_1. Solving the program again with the added constraint $E[l_1] = 0$ gives us another linear form l_2, wlog with $l_2 \in \mathbb{R}^\times \ell_2$ and so forth...

After computing $l_1 \in \mathbb{R}^\times \ell_1, \dots, l_m \in \mathbb{R}^\times \ell_m$, we can obtain the missing scalars by solving a linear system. Note that as $\ell_1,\dots,\ell_m$ are linearly independent, so are $l_1^3,\dots,l_m^3$ and therefore there exist unique $\mu_1,\dots,\mu_m \in \mathbb{R}$ such that $p = \sum_{i=1}^m \mu_i l_i^3$. Computing these and setting $\ell_i := \sqrt[3]{\mu_i} l_i$ for each $i \in \{1,\dots,m\}$ gives the desired output. □

The next proof (thoroughly) follows Barak and Steurer [86, Lecture 7.4] and is based on simultaneous matrix diagonalization. Note that Barak and Steurer assume either orthogonal decomposability or that an additional matrix of "degree-2 moments" is given. The latter is an equivalent condition, since it allows to perform a procedure called *whitening* which orthonormalizes the rank-1 components. This assumption is commonly satisfied in statistical applications, but one can get rid of it by solving an SDP.

Algorithmic Proof. (Proof of Theorem 2.4.8 via simultaneous diagonalization). We start again by deriving the input $p = \sum_{i=1}^m \ell_i^3$ in some random directions $v \in \mathbb{S}^{n-1}$ and then dividing by 3 to obtain expressions $p_v = \sum_{i=1}^m \ell_i(v)\ell_i^2$. By Proposition 2.2.8, we may compute $\sum_{i=1}^m \ell_i(v)\ell_i^{\otimes 2}$ from $\sum_{i=1}^m \ell_i(v)\ell_i^2$. By duality, we may replace the linear forms $\ell_1,\dots,\ell_m$ by vectors $a_1,\dots,a_m \in \mathbb{R}^n$. Extending this map $(\mathbb{R}^n)^\vee \to \mathbb{R}^n$ to the space of 2-tensors, $\ell_i^{\otimes 2}$ is mapped to $a_i^{\otimes 2}$ which can be interpreted as the matrix $a_i a_i^T$ for each $i \in \{1,\dots,m\}$. Thus, we have access to the space formed by the matrices $M(v) = \sum_{i=1}^m (a_i^T v) a_i a_i^T$. Note that this space is equal to $\mathcal{L} := \{\sum_{i=1}^m \lambda_i a_i a_i^T \mid \lambda_1,\dots,\lambda_m \in \mathbb{R}\}$. This argument crucially uses that $a_1,\dots,a_m$ are linearly independent. The matrix space $\mathcal{L}$ contains a psd matrix of rank m and we may compute such a matrix

G via semidefinite programming, just as we did in the previous proof. The psd matrix G defines an inner product on the subspace orthogonal to $\ker G$ by letting $\langle x, y\rangle_G := x^T G^+ y$, where G^+ denotes the Moore-Penrose pseudoinverse of G, defined in 2.3.12. With respect to this inner product, $a_1, \dots, a_m$ are orthogonal: Indeed, write $A = (a_1, \dots, a_m)$. Choose positive $\lambda_1, \dots, \lambda_m$ such that $G = \sum_{i=1}^m \lambda_i a_i a_i^T$ and note that $G = AD^2A^T$, where $D = \operatorname{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_m})$. Since AD has full column rank, by 2.3.12 we have $G^+ = (AD(AD)^T)^+ = ((AD)^+)^T(AD)^+ = (D^{-1}A^+)^T D^{-1}A = (A^+)^T D^{-2} A^+$.

Therefore, $\langle a_i, a_j\rangle_G = e_i^T A^T G^+ A e_j = e_i^T (A^+A)^T D^{-2} A^+ A e_j = e_i^T D^{-2} e_j = \frac{1}{\lambda_i}\delta_{ij}$ for $i, j \in \{1, \dots, m\}$ by the properties of the pseudoinverse. Thus, $a_1, \dots, a_m$ form a generalized Eigenbasis of every element of $\mathcal{L}$. Now, choose some $v \in \mathbb{S}^{n-1}$ at random. Then $M(v) = A \operatorname{diag}(a_1^T v, \dots, a_m^T v)A^T$ has one-dimensional Eigenspaces, as its (generalized) Eigenvalues $a_1^T v, \dots, a_m^T v$ with respect to G are pairwise distinct. Thus, after computing an Eigenbasis $b_1, \dots, b_m$ of $M(v)$, we have $b_1 \in \mathbb{R}^\times a_1, \dots, b_m \in \mathbb{R}^\times a_m$. The missing scalars may be computed by Linear Algebra as in the first proof. □

One last proof uses the fact that a quadratic Sum-of-Squares form s does not have representations $s = \sum_{i=1}^m \sigma_i \ell_i^2$ as signed sums of squares, with signs $\sigma_1, \dots, \sigma_m \in \{\pm 1\}$, unless all the signs are nonnegative or there is a linear dependency between the linear forms $\ell_1, \dots, \ell_m$. Note that a curious problem of independent interest is to ask the same question in degree 4: For which $n \in \mathbb{N}$ and $m \in \{2, \dots, \binom{n+1}{2}\}$ can a signed sum of squares $\sum_{i=1}^m \sigma_i q_i^2$ of *general* quadratic forms $q_1, \dots, q_m \in S^2(\mathbb{R}^n)$ not be SOS, unless all signs are one? I will leave that as a brain-candy for the reader to think about. In case you know the answer, please email me.

Algorithmic Proof. Similar to the two previous proofs, note that we can obtain the space $\mathcal{L} = \operatorname{span}_{\mathbb{R}}(\ell_1^2, \dots, \ell_m^2)$ by applying derivations in n linearly independent directions. By the Multisecant Lemma 2.2.24, this space intersects the closed affine variety of squares of linear forms $\mathrm{V}_{1,2} = \{\ell^2 \mid \ell \in S^1(\mathbb{C}^n)\}$ only in the (real) multiples of $\ell_1^2, \dots, \ell_m^2$, as even the complex space generated by $\ell_1^2, \dots, \ell_m^2$ does so. Finding the squares in a subspace of the polynomial ring is a rank-1 SDP: Compute a system F of linear equations describing $\mathcal{L}$. Then we search for the solutions of

$$\begin{array}{rr} \text{find} & \ell^2 \in \mathrm{V}_{1,2} \\ \text{s.t.} & \forall f \in F:\ f(\ell^2) = 0 \end{array} \tag{2.62}$$

To disable rescaling, we may choose a linear functional $g \in \mathrm{SOS}_2^*(\mathbb{R}^n)^\circ$ attaining positive values on any nonzero sums of squares and add the constraint $g(\ell^2) = 1$. Furthermore, we can add a random objective function $h \in S^2(\mathbb{R}^n)^\vee$ to Equation (2.62). After doing so, we relax to a Sum-of-Squares program and obtain

$$\begin{array}{rr} \text{maximize} & h(s) \\ \text{s.t.} & s \in \mathrm{SOS}_2(\mathbb{R}^n) \\ & \forall f \in F:\ f(s) = 0 \\ & g(s) = 1 \end{array} \tag{2.63}$$

We claim that the optimal solutions for (2.63) are squares. By the constraints given by F, any feasible $s \in \mathrm{SOS}_2(\mathbb{R}^n)$ must lie in $\mathcal{L}$, thus there exist $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ such that $s = \lambda_1 \ell_1^2 + \ldots + \lambda_m \ell_m^2$. If any of $\lambda_1, \ldots, \lambda_m$ was negative, then the Sylvester signature of s would be smaller than its rank, which is impossible for a nonnegative quadratic form. Thus we are in fact optimizing over the polyhedral cone $\mathrm{cone}(\ell_1^2, \ldots, \ell_m^2)$ intersected with the affine hyperplane where $g(s) = 1$. The optimum must thus be attained on an extremal ray of $\mathrm{cone}(\ell_1^2, \ldots, \ell_m^2)$, i.e. on a multiple of one of $\ell_1^2, \ldots, \ell_m^2$. As h was random, the optimal value is exclusively attained on one specific extremal ray, wlog on ℓ_1^2. Thus we can iteratively obtain multiples of $\ell_1^2, \ldots, \ell_m^2$ and then obtain the missing scalars as in the first proof. □

2.4.9 Remark: All three algorithms overlap in some conceptual similarities, but there are noteworthy differences. The first and second algorithm use an SDP to get a maximum rank psd matrix in some space. In some statistical applications, one gets such a matrix for free. E.g. for Gaussian mixtures with identical covariances, cf. Section 3.1, one can take the matrix of degree-2 moments. On the other hand, searching for PSD quadratic forms in a subspace is the integral part of the third algorithm, regardless of whether second order moments are given or not. The rank-1 optimization problem that occurs in (2.62) is of independent interest: In the third proof, we used the assumption of linear independence at two crucial points: First, to show that we obtain $\mathcal{L}$, second, to show $\mathcal{L} \cap \mathrm{SOS}_2(\mathbb{R}^n) = \mathrm{cone}(\ell_1^2, \ldots, \ell_m^2)$ and thus the relaxation of the rank-1 problem is exact. Therefore, the third proof suggests two natural barriers for Waring decomposition in the case $m > n$: First, can we still compute the space $\mathcal{L} = \mathrm{span}(\ell_1^2, \ldots, \ell_m^2)$? Second, given a basis for $\mathcal{L}$, can we still solve the rank-1 SDP from Equation (2.62) efficiently? Note that determining the rank of a 3-form is NP-hard in the worst case (cf. the article of Hillar and Lim [45]), so we would expect at least one of the two problems to be hard for larger values of m, if not both!

2.5. Moments and Mixtures of random vectors

Throughout, we will need some concepts from stochastics, but not enough to justify an introduction of the entire measure theoretic formalism of measured/probability spaces etc. For an introduction to the latter formalism, the curious reader is referred to [51] and [13] and to [46] for a logic-centered approach. Thus let us consider random variables Y in a highly simplified formalism, where we assume that we have certain magical, unexplained operators $\mathbb{P} = \mathbb{P}_Y$ and $\mathbb{E} = \mathbb{E}_Y$, where $\mathbb{P}$ assigns a *probability* $\mathbb{P}(A)$ to certain subsets A of a vector space U (or logical clauses describing them, e.g. for $U = \mathbb{R}^2$ and a random variable Y on U, $\mathbb{P}(Y_1 \geq 0) = \mathbb{P}_Y(\mathbb{R}_{\geq 0} \times \mathbb{R}))$ and $\mathbb{E}$ is an *expectation operator* that maps polynomial expressions in random vectors on U to elements of $\mathbb{R}[U]$.

$\mathbb{E}$ and $\mathbb{P}$ are typically denoted without subscripts, since notation such as $\mathbb{E}[Y_1^2 Y_2]$ and $\mathbb{P}(Y_1 \geq 1)$ makes it clear and very intuitive with respect to which random variables the probability is taken. If not mentioned otherwise, let us always assume that all our random variables have the following nice properties:

(a) All random variables attain values on some real, affine space U.

(b) For any random variable Y, we assume that there is an operator $\mathbb{E}$ on the algebra $\mathbb{R}[Y]$. This is saying that we assume that $\mathbb{E}[f(Y)]$ exists for *any* polynomial $f \in \mathbb{R}[U]$. We will denote expectation operators always by the same symbol $\mathbb{E}$, even though they may have different scopes.

We will write down all the following notation for the real affine space $U = \mathbb{R}^n$ with variables $X = (X_1, \ldots, X_n)$.

2.5.1 Notation: For a random variable Y on $\mathbb{R}^n$ we will write

$$\mathbb{E}[Y] = (\mathbb{E}[Y_1], \ldots, \mathbb{E}[Y_n]) \in \mathbb{R}^n$$

Similarly, for some $f \in \mathbb{R}[X, Y]$, where $X = (X_1, \ldots, X_n)$ are unknowns, if $f = \sum_{\alpha,\beta \in \mathbb{N}_0^n} f_{\alpha,\beta} X^\alpha Y^\beta$ for some $f_{\alpha,\beta} \in \mathbb{R}$, we write

$$\mathbb{E}[f] = \sum_{\alpha \in \mathbb{N}_0^n} f_{\alpha,\beta} \mathbb{E}[Y^\beta] X^\alpha$$

Thus, we see $\mathbb{E}$ as a “general-purpose” operator that can be applied to scalars, vectors and polynomials alike. In the latter two cases, it acts entry- or coefficient-wise, respectively. This might be considered abuse of notation, but it also simplifies notation by a lot.

2.5.2 Proposition and Definition: Let $n \in \mathbb{N}_0$, Y a random variable on $\mathbb{R}^n$ and let $\alpha \in \mathbb{N}_0^n$. $\mathbb{E}[Y^\alpha]$ is called the *α-moment* of Y. For $d \in \mathbb{N}_0$, we call

$$\mathcal{M}_d(Y) := \mathbb{E}[\langle X, Y \rangle^d] \in \mathbb{R}[X]_d \tag{2.64}$$

the *degree-d moment form* of Y. The coefficients of the degree-d moment form are, up to multinomial coefficients, the α-moments of Y where α ranges over all multi-indices with $|\alpha| = d$.

Proof. By the multinomial theorem,

$$\langle X, Y\rangle^d = \sum_{|\alpha|=d} \binom{d}{\alpha} X^\alpha Y^\alpha \tag{2.65}$$

Thus the α-coefficient of $\mathcal{M}_d(Y)$ is $\binom{d}{\alpha}\mathbb{E}[Y^\alpha]$. □

2.5.3 Definition: Let $n \in \mathbb{N}_0$ and Y a random variable on $\mathbb{R}^n$. Let $X = (X_1, \ldots, X_n)$ a vector of unknowns. The *moment generating series* of Y is defined as the formal power series

$$\mathbb{E}[\exp(\langle Y, X\rangle)] = \sum_{d=0}^{\infty} \frac{1}{d!}\mathbb{E}[\langle Y, X\rangle^d] = \sum_{d=0}^{\infty} \frac{1}{d!}\mathcal{M}_d(Y) \in \mathbb{R}[[X]] \tag{2.66}$$

where the expected values are to be understood coefficient-wise in X.

The moment generating series is a formal power series constructed from all the moment forms of a random variable. It sometimes admits useful representations that make it easy to memorize all the moments.

Dirac and Gaussian random variables

We introduce important families of random variables. The first one is the class of *Dirac random variables* or constant random variables. These are essentially not random at all, as they always attain the same value.

2.5.4 Definition: Let $v \in U$. A random variable Y on U is *Dirac distributed* with parameter v, if for all $f \in \mathbb{R}[U]$ it holds that

$$\mathbb{E}[f(Y)] = f(v) = \delta_v(f) \tag{2.67}$$

Observe that the moment generating series of a *Dirac random variable* Y with respect to $v \in \mathbb{R}^n$ is simply $\exp(\langle v, X\rangle)$.

2.5.5 Reminder: An n-variate *Gaussian normal distribution* $\mathcal{N}(\ell, q)$ on $\mathbb{R}^n$ is given by a pair (ℓ, q) where ℓ is a linear form on $\mathbb{R}^n$ and $q \in \mathrm{SOS}_2(\mathbb{R}^n)$ is a psd quadratic form. In the literature, q is usually required to be positive definite, but whether or not q has a nontrivial kernel, the pair (ℓ, q) always defines a normal distribution on the affine subspace given by the *mean vector* $(\ell(e_1), \ldots, \ell(e_n))$ (e_i is the i-th coordinate vector) plus the orthogonal complement of the kernel of q (in the maximal degenerate case, i.e. when $q = 0$, this definition gives the Dirac distribution at the mean vector). The moment generating series of $\mathcal{N}(\ell, q)$ is equal to $\exp(\ell + \frac{1}{2}q)$. Gaussian normal distributions are very often also described in terms of their *mean vector* $\mu \in \mathbb{R}^n$ and psd *covariance matrix* $\Sigma \in \mathbb{R}^{n\times n}$. The relationship to our parametrization is that $\ell = \mu^T X$ and $q = X^T \Sigma X$.

Empirical Moments

In many applications, the moments of a probability distribution are approximately known from empirical observations. Assume Y is a random variable on $\mathbb{R}^n$ and we are given a multiset $\mathcal{Y}$ of iid samples drawn from Y. Then e.g.

$\mathbb{E}[Y]$ can be approximated by $\frac{1}{\#\mathcal{Y}}\sum_{y\in\mathcal{Y}} y$ with high probability over the choice of samples. The following elementary proposition can be obtained by employing the Markov-Chebyshev-Chernov inequality. It gives a quantitative relation between the number of samples needed to guarantee a given approximation accuracy $\varepsilon \in \mathbb{R}_{>0}$. Summarized, it says that if we have at least $N \in \Omega(n^3\sigma^2\varepsilon^{-2})$ iid samples, we can estimate the mean of an n-dimensional random variable with marginal variances bounded by σ^2 up to error ε with high probability:

2.5.6 Proposition: Let Y a random variable taking values in $\mathbb{R}^n$ and satisfying the variance bound $\mathbb{E}[(Y_i - \mathbb{E}[Y_i])^2] \leq \sigma^2$ for all $i \in \{1,\dots,n\}$. Let $\overline{Y}$ denote the average over $N \in \mathbb{N}$ iid copies of Y. Then, we have for any $\varepsilon \in \mathbb{R}_{>0}$ that

$$\mathbb{P}[\|\overline{Y} - \mathbb{E}[Y]\|^2 \geq \varepsilon] \leq \frac{n^3\sigma^2}{\varepsilon^2 N} \tag{2.68}$$

Proof. It holds that $\mathbb{P}[\|\overline{Y} - \mathbb{E}[Y]\|^2 \geq \varepsilon] \geq \mathbb{P}[\exists i\colon (\overline{Y}_i - \mathbb{E}[Y_i])^2 \geq \varepsilon/n]$ by an averaging argument. With the union bound, we get

$$\mathbb{P}[\exists i\colon (\overline{Y}_i - \mathbb{E}[Y_i])^2 \geq \varepsilon/n] \leq \sum_{i=1}^{n} \mathbb{P}[(\overline{Y}_i - \mathbb{E}[Y_i])^2 \geq \varepsilon/n] \tag{2.69}$$

The claim now follows by applying Chebyshev's inequality to each addend on the right side. □

2.5.7 Remark: Proposition 2.5.6 shows that we may estimate the moments of a random vector up to a given degree from sufficiently many samples. If nothing about the variances of the random variables Y^α is known, then Proposition 2.5.6 is only a qualitative statement. In many cases however, e.g. when Y is some fixed value in $\mathbb{R}$ plus a centered Gaussian noise term with variance σ^2, we know that $\mathbb{E}[Y^{2k}] \in \mathcal{O}_\sigma(\sigma^{2k})$ for all $k \in \mathbb{N}$. This qualitative behaviour for the growth of higher order moments is very natural. It is captured in the notion of *subgaussian distributions*, which we will introduce next. Note that Gaussian distributions are subgaussian distributions as trivially seen from the moment series (cf. 2.5.5) and Definition 2.5.8.

2.5.8 Definition: A random variable Y on $\mathbb{R}^n$ is called (σ-)*subgaussian*, if there exists $\sigma \in \mathbb{R}$ such that for each $i \in \{1,\dots,n\}$ and all $k \in \mathbb{N}_0$ it holds that

$$\mathbb{E}[Y_i^{2k}] \leq \sigma^{2k} k^k \tag{2.70}$$

Note that for $k = 1$ this implies $\mathbb{E}[Y_i^2] \leq \sigma^2$, so σ^2 needs to be a bound on the variances. Note further that as $\mathbb{E}[|Y_i|^k]^2 \leq \mathbb{E}[Y_i^{2k}]$ by Cauchy-Schwarz, condition (2.70) also imposes bounds on the odd-order moments.

For subgaussian random variables, we can quantify the "price" (i.e. the amount of samples) that it costs to compute higher order moments by just one parameter σ. Note that in computational algebraic statistics, samples are a resource just as time, space and randomness are in the classical theory of computing. In analogy to the notions of time-complexity and space-complexity of a computational problem, the *sample-complexity* of an algorithm for a statistical problem is, informally, the number of samples it needs to compute its output (with high success probability over the given samples). For σ-subgaussian distributions, there exists a straightforward connection between the asymptotical order of the sample complexity (in σ) and moments: If an algorithm should work with sample-complexity $\mathcal{O}(\sigma^k)$, then it may compute and use the moments of degree at most k by Proposition 2.5.6. For some problems it is possible to show "converse" results, in the sense that there exists an algorithm of optimal sample complexity order $k \in \mathbb{N}_0$ that does nothing with the samples but to use them to compute the moments up to a certain degree. E.g. this was done in [6]. We call such an algorithm *moment-based*, as all the interesting algorithmic computations could also be carried out if instead of samples one was just given access to the (slightly noisy) moments.

Mixtures of random variables

2.5.9 Definition: Let $\mathcal{Y}$ a family of random variables. A *mixture random variable* over $\mathcal{Y}$ is given by a tuple $(Y_1, \ldots, Y_m) \in \mathcal{Y}^m$ of $m \in \mathbb{N}_0$ random variables on $\mathbb{R}^n$, where $n \in \mathbb{N}_0$, together with associated weights $\lambda_1, \ldots, \lambda_m \geq 0$ summing up to 1. m is called the *rank* of the mixture representation. With such a model, we associate a random variable Y that is sampled by choosing $i \in \{1, \ldots, m\}$ with probability λ_i and then sampling Y_i. Let us denote this random variable by $Y = \lambda_1 Y_1 \oplus \ldots \oplus \lambda_m Y_m$.

2.5.10 Remark: If the random variables of a mixture model $(Y_1, \ldots, Y_m)$ as in Definition 2.5.9 have probability density functions $\rho_1, \ldots, \rho_m$, respectively, then the probability density function of Y is $\lambda_1 \rho_1 + \ldots + \lambda_m \rho_m$. The random variable $\lambda_1 Y_1 \oplus \ldots \oplus \lambda_m Y_m$ associated with a mixture model should not be confused with the sum $\lambda_1 Y_1 + \ldots + \lambda_m Y_m$. A more straightforward way to denote the random variable associated to a mixture model would be $Y_{i(\lambda)}$, where $i(\lambda)$ is a random variable on $\{1, \ldots, m\}$ attaining the value $j \in \{1, \ldots, m\}$ with probability λ_j. However, we will not use this formal notation because it looks weird and unintuitive.

2.5.11 Remark: Instead of specifying the family $\mathcal{Y}$ of random variables for a mixture, I will usually only specify the family of their distributions. This has of course no impact on the moments.

With any class of random variables $\mathcal{Y}$ and any rank $m \in \mathbb{N}$, it is possible to formulate an associated (rank-m) *mixture model*, which is a statistical model that aims to describe a sample set in the "closest" possible way by a mixture of at most m random variables over $\mathcal{Y}$. Semi-formally, mixture models are optimization problems which aim to find random variables $Y_1, \ldots, Y_m \in \mathcal{Y}$ and associated weights $\lambda_1, \ldots, \lambda_m$ such that some kind of "distance" or loss function is minimized between $\lambda_1 Y_1 \oplus \ldots \oplus \lambda_m Y_m$ and some fixed, known empirical data u (e.g. u can be some vector of empirical moments). Some mixture models admit a formulation as a polynomial optimization problem via moments, as long as the distance function can be expressed as a polynomial in the moments of $Y_1, \ldots, Y_m$. Indeed, note that the moments of a mixture random variable are a convex combination of the moments of the random variables that contribute to the mixture.

2.5.12 Remark: For a mixture model $(Y_1, \ldots, Y_m)$ with weights $\lambda_1, \ldots, \lambda_m$ as in Definition 2.5.9 and any polynomial f for which all $\mathbb{E}[f(Y_i)]$ exist for

$i \in \{1, \ldots, m\}$, it holds

$$\mathbb{E}[f(Y)] = \sum_{i=1}^{m} \lambda_i \mathbb{E}[f(Y_i)] \tag{2.71}$$

In particular, if all moments of all $Y_1, \ldots, Y_m$ exist, then the moment generating series of $Y = \lambda_1 Y_1 \oplus \ldots \oplus \lambda_m Y_m$ is

$$\sum_{i=1}^{m} \lambda_i \mathbb{E}[\exp(\langle Y_i, X\rangle)] \tag{2.72}$$

Another very related problem that is naturally associated with a class $\mathcal{Y}$ of random variables is the corresponding *mixture moment problem*:

2.5.13 Problem: Given $d, n, m \in \mathbb{N}$ and $L \in \mathbb{R}[X]^{\vee}_{\leq d}$, when does there exist a mixture random variable $Y = \lambda_1 Y_1 \oplus \ldots \oplus \lambda_m Y_m$ of rank at most m over $\mathcal{Y}$ such that for all $f \in \mathbb{R}[X]_{\leq d}$:

$$L(f) = \mathbb{E}[f(Y)] = \sum_{i=1}^{m} \lambda_i \mathbb{E}[f(Y_i)]$$

and if such a representation exists, when is it unique?

2.5.14 Remark: It does not hurt to also allow formal mixtures $\lambda_1 Y_1 \oplus \ldots \oplus \lambda_m Y_m$ of random variables where the weights do not sum up to one. Generally, such a positive multiple of a random variable will not have any probabilistic meaning, but on the level of measures and moments, it makes sense to multiply a probability measure with a positive scalar. These formal objects are mainly introduced to be able to write $\frac{1}{m}(Y_1 \oplus \ldots \oplus Y_m)$ rather than $\frac{1}{m}Y_1 \oplus \ldots \oplus \frac{1}{m}Y_m$.

One very important special case of Definition 2.5.9 are *Gaussian mixtures* defined below.

2.5.15 Definition: (cf. Definition 2.5.9) A *Gaussian mixture* (in dimension $n \in \mathbb{N}_0$ of rank $m \in \mathbb{N}_0$) is a mixture over the family

$$\{\mathcal{N}(\ell, q) \mid \ell \in S^1(\mathbb{R}^n), q \in S^2(\mathbb{R}^n)\} \tag{2.73}$$

There exist various special types of Gaussian mixtures. The associated mixture models can vary quite strongly both in terms of computational tractability and in expressiveness. The following definition highlights two that will be relevant for us.

2.5.16 Definition: A *mixture of centered Gaussians* is a mixture over

$$\{\mathcal{N}(0, q) \mid q \in S^2(\mathbb{R}^n)\} \tag{2.74}$$

A *mixture of Gaussians with identical covariance* q is a mixture over the family

$$\{\mathcal{N}(\ell, q) \mid \ell \in S^1(\mathbb{R}^n)\} \tag{2.75}$$

Gaussian mixtures with identical covariance matrices

The first special case of Gaussian mixtures from Definition 2.5.16 will occupy us for the major part of Chapter 3. The second special case from 2.5.16 is comparatively well-understood, as the associated mixture model admits a reduction to best fixed-rank tensor approximation: By Reminder 2.5.5 and Remark 2.5.12, the degree-3 moment form of a mixture $\mathcal{N}(\ell_1, q) \oplus \ldots \oplus \mathcal{N}(\ell_m, q)$ is

$$\sum_{i=1}^{m} \ell_i^3 + 3\sum_{i=1}^{m} q\ell_i \tag{2.76}$$

Via a simple substitution, the decomposition problem for such mixtures of Gaussians translates to Waring decomposition: Shifting the origin such that $\sum_{i=1}^{m} \ell_i = 0$ turns (2.76) into a sum of third powers of linear forms and Theorem 2.4.8 may be applied. Performing this shift is approximately possible as an algorithmic operation, as $\sum_{i=1}^{m} \ell_i$ is simply the moment form of degree 1 of the mixture, which may be estimated from samples. Identifiability of $V_{1,3}$ is thus the relevant question for this special type of Gaussian mixtures, cf. Theorem 2.4.7, [23]. Recovering the parameters algorithmically is possible with any sufficiently noise stable algorithm for Waring decomposition. The algorithms presented in the proofs of Theorem 2.4.8 were all analyzed without noise. However, there exist algorithms for Waring decomposition in the undercomplete setting $m \leq n$ which tolerate input noise of order inverse polynomial in the dimension [3], [60].

2.5.17 Remark: If one does not care about quantitative statements on the amount of tolerated noise, it is surprisingly simple to get around a noise-stability analysis. This is achieved by splitting the problem into three subproblems: Estimation, projection and decomposition. We already know how to decompose forms of small Waring rank. Projection will have its own Chapter 4 dedicated to it, but for the time being, let me provide a short proof sketch: After estimating a linear form $\bar{l}$ from samples which is almost $\bar{l} \approx \sum_{i=1}^{m} \ell_i$, one can shift the space such that $\bar{l} = 0$. Here and in the following, the $\approx$-symbol denotes that the difference of both sides lies in $\mathcal{O}(\varepsilon)$, where ε is the noise from estimation. Then, after the shift, the true degree-3 moments $\mathcal{M}_3(Y)$ of the mixture random variable $Y = \frac{1}{m}(\mathcal{N}(\ell_1, q) \oplus \ldots \oplus \mathcal{N}(\ell_m, q))$ will satisfy $\mathcal{M}_3(Y) \approx \sum_{i=1}^{m} \ell_i^3$. Similarly, $\mathcal{M}_2(Y) \approx \sum_{i=1}^{m} \ell_i^2$. By estimation (cf. 2.5.6), with high probability over sufficiently many iid samples from Y, we may thus compute a psd matrix $G \approx \sum_{i=1}^{m} \ell_i^{\otimes 2}$ and a cubic form $t \approx \sum_{i=1}^{m} \ell_i^3$. By computing a Gram factorization $G^+ = B^T B$ of the pseudoinverse of G and applying the linear transformation B, we can transform t into an *orthogonally decomposable* tensor, similar to what we did in the second proof of Theorem 2.4.8. It is known that the variety of orthogonally decomposable tensors is an irreducible component of the variety cut out by Robeva's quadratic equations in $S^3(\mathbb{R}^n)$, cf. [79, Lemma 3.7] and [9, Proposition 4.2.1]. These $N \in \mathbb{N}$ quadratic equations $q_1, \ldots, q_N$ may be used to define the Lagrangian $\mathcal{L}(f, \lambda) = \|f - t\|^2 + \sum_{i=1}^{N} \lambda_i q_i$, where $f \in S^3(\mathbb{R}^n)$, $\lambda \in \mathbb{R}^N$ and $\|\cdot\|^2$ denotes some inner product norm on $S^3(\mathbb{R}^n)$. Local optimization will compute the nearest critical point of $\mathcal{L}(f, \lambda)$. In Chapter 4, more precisely Remark 4.3.7, we will see that this critical point is the unique orthogonal projection of t to the set of orthogonally decomposable tensors, with high probability over the (sufficiently many) samples.

"It would be so nice if something made sense for a change"

— Alice in Wonderland, [39].

3 Sums of Powers of Forms

Throughout this chapter, we will study problems of the following type: Given $n, m, k, d \in \mathbb{N}$, a field K of characteristic 0 and a sum

$$\sum_{i=1}^{m} q_i^d \tag{3.1}$$

of d-th powers of n-variate k-forms $q_1, \ldots, q_m \in S^k(K^n)$.

3.0.1 Question: (a) When is the decomposition from (3.1) unique?

(b) When can we compute $\{q_1^d, \ldots, q_m^d\}$ from power sums like the one in (3.1)?

We will focus on the case of general parameters $q_1, \ldots, q_m$ and we give sufficient criteria for positive answers to both problems. Note that uniqueness of the decomposition in (3.1) is the natural first problem to examine, as Question 3.0.1(b) is only well-posed if Question 3.0.1(a) has a positive answer. But before that, let us briefly discuss reasons why you should care about powers-of-forms decompositions. It turns out that there is a connection between decompositions of the kind (3.1) and *mixtures of centered Gaussians* (cf. 2.5.15 and 2.5.16).

3.1. A primer on Gaussian Mixtures

Imagine the following situation: You are the teaching assistant for an undergraduate lecture and the overwhelming majority of your students performed miserably in the exam. Therefore, instead of letting only the few people with no less than 50 out of 100 points pass, as originally intended, you try searching for a reasonable passing threshold $c \in \{0, \dots, 100\}$ instead, as the original one was likely too hard.

The objective is to find the passing threshold via optimizing a statistical model. A reasonable threshold should fulfill some criteria: We certainly only want to let those pass that achieved a sufficient understanding of the lecture and distinguish them from those who should better repeat everything again.

It is likely impossible to give an objective answer on this, as in fact even the whole system of grading and the meaning of grades are subjects of ongoing debate [4], [52], [72]. The central question in this debate is about the relation between *grades* on one side and *achievement* or *understanding* on the other side. In Section 3.1 I will address some caveats and attempt to do justice to the various ideological positions on grading.

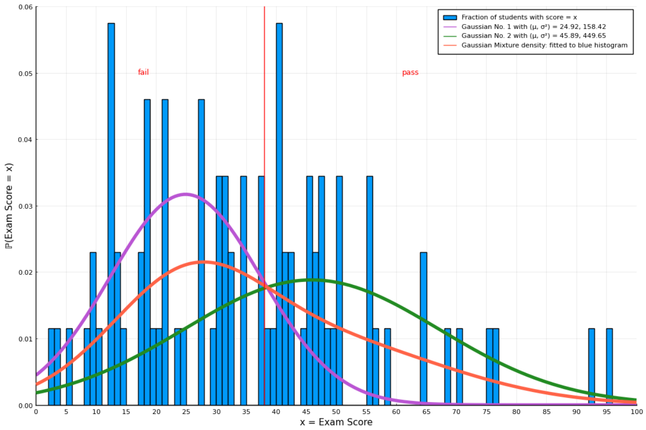

Figure 3.1: Gaussian Mixture Model computing passing threshold for an exam. Blue is the empirical probability density function. The green curve is the pdf of the Gaussian associated with A, the purple curve is the Gaussian of B and the orange curve is the pdf of the Gaussian Mixture. The passing threshold is the intersection point of the Gaussians, rounded to the next integer below. The thresholds for individual grades were done by hand.

For now, let us completely give up on any better way to understand a student's performance other than the exam grade. In particular, let us also assume that we are not allowed to look into the exams, by which we could make use of any personal knowledge. Then, we can only grade the students relative to each other. Let us split the lecture's attendees into two disjoint groups A and B, where A are the students of sufficient understanding and B is the complement of A, all the while pretending there was some objective way of splitting. We will give an justification for this a posteriori. Every student in group A performs in the exam acccording to some distribution Y_A and every student in group B performs according to some distribution Y_B. Thus a random student performs according to the distribution $\lambda_A Y_A \oplus \lambda_B Y_B$, where $\lambda_A = \frac{\#A}{\#A+\#B}$ and $\lambda_B = \frac{\#B}{\#A+\#B}$. Let us simplify by assuming Y_A and Y_B to be Gaussian distributions: Students from A are expected to perform better than students from B, but even a very good student might completely fail an exam. Thus it is natural to model an individual students exam performance by an expected value that has to do with their understanding plus some random, e.g. Gaussian, noise. The gross simplification of our model is to assume that all students from A have the same expected value, e.g. we think of $Y_A = \mathcal{N}(\mu_A, \sigma_A^2)$ as describing (the results of) a random student of sufficient understanding and $Y_B = \mathcal{N}(\mu_B, \sigma_B^2)$ describing a random student of nonsufficient understanding. It is clear that such a model will in general not give a faithful description of our data set, but note that this is not our goal: Instead, we want to use Maximum Likelihood Estimation to determine the passing grade. A student of grade x shall pass if and only if

$$\mathbb{P}(Y_A = x) \geq \mathbb{P}(Y_B = x)$$

i.e. if the model student of sufficient understanding is at least as likely to score x as the model student of nonsufficient understanding. What we introduced here is a special case of a *Gaussian mixture model*, see Definition 2.5.15. It remains to find reasonable values of $\lambda_A, \lambda_B, \mu_A, \mu_B, \sigma_A^2, \sigma_B^2 \in \mathbb{R}$. This is done by optimization, as $\lambda_A Y_A \oplus \lambda_B Y_B$ should be the best fit to the empirical distribution of exam scores. Turning on scikit-learn in Julia yields the result in Figure 3.1. Thus, in such a situation we would set the passing grade to 37, which looks reasonable.

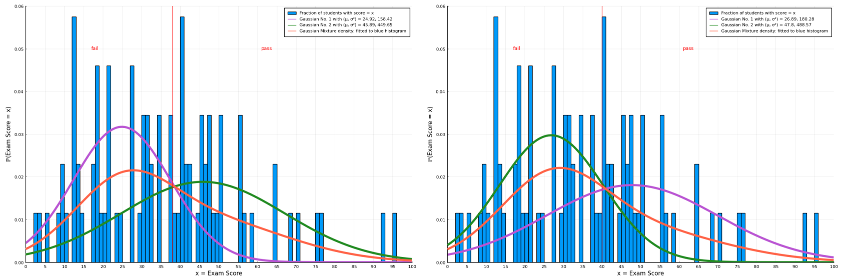

Figure 3.2: Running scikit-learn's algorithm for Gaussian mixture models on two different Julia kernel sessions gives two different answers for the passing grade.

Alas it is not that easy: The next day we rerun the code producing Figure 3.1 and we obtain a different result, as seen in Figure 3.2. What happened? First, note that we can certainly only trust the output of our model to be a valid passing threshold up to some error margin, as it depends on empirical data and e.g. running the same exam twice might give different empirical distributions and thus thresholds. But this is not the reason for what happened here: We ran the exact same algorithm on the exact same data and obtained different results. Two questions are apparent: First, does there even exist a unique optimal solution? Second, if so, can we compute it? Note the conceptual similarity with Question 3.0.1. One problem is that the local optimization procedures used by scikit-learn are not guaranteed to find optimal solutions.[1] Another problem is that the empirical distribution we try to approximate is very far away from being a mixture of two Gaussian distributions, but even if it was not, it would not be clear if a finite number of samples uniquely determined the parameters (up to small noise).

Gaussian Mixture Models

Gaussian Mixture Models date back to the 19th century, notably the work of K. Pearson ([71], see [1] for a modern exposition), who used them as a statistical tool to separate biological species. His aim was to find evidence for the then-new theory of evolution by examining crab populations. Without modern tools such as DNA analysis, he had to rely on statistical methods: Assuming a population of crabs consists of two different species, their measurable traits would not be explained by one, *simple* distribution, but from a convex combination $\lambda_1 Y_1 \oplus \lambda_2 Y_2$ of two *simple* distributions, each corresponding to one of the species. The key here is the word "simple", as what we consider to be a simple distribution has a great impact on the analysis. Pearson's quantification of the word "simple" was to assume the individual distributions for the crab traits he measured were Gaussian.

With the advent of Machine Learning, Gaussian Mixture models have risen in popularity, e.g. since they allow unsupervised clustering of a dataset and the resulting model gives probabilities for membership of a point in a cluster, as opposed to just hard-assigning points to clusters. Via Remark 2.5.7, such Gaussian mixture models admit a formulation as a polynomial optimization problem via moments. The moments of a Gaussian mixture are expressed in the following.

3.1.1 Proposition: Consider a Gaussian Mixture model $Y = \lambda_1 Y_1 \oplus \ldots \oplus \lambda_m Y_m$, where $m, n \in \mathbb{N}_0$ and $Y_1 \sim \mathcal{N}(\ell_1, q_1), \ldots, Y_m \sim \mathcal{N}(\ell_m, q_m)$ are n-variate Gaussian distributions given by linear forms $\ell_1, \ldots, \ell_m \in S^1(\mathbb{R}^n)$ and psd quadratic forms $q_1, \ldots, q_m \in S^2(\mathbb{R}^n)$ and $\lambda_1, \ldots, \lambda_m \in \mathbb{R}_{>0}$ are weights summing up to 1. Then the moment generating series of Y is

$$\sum_{i=1}^{m} \lambda_i \exp(\ell_i + \frac{1}{2} q_i) = \sum_{d=0}^{\infty} \frac{1}{d!} \sum_{i=1}^{m} \lambda_i (\ell_i + \frac{1}{2} q_i)^d \tag{3.2}$$

3.1.2 Remark: Note that the degree-d part of the series in (3.2) is of the form

$$\frac{1}{d!} \sum_{i=1}^{m} \lambda_i \sum_{k=1}^{\lfloor d/2 \rfloor} 2^{-k} \widehat{\widehat{\binom{d}{k}}} q_i^k \ell_i^{d-2k} \tag{3.3}$$

Here, $\widehat{\widehat{\binom{d}{k}}} := \dfrac{d!}{k!(d-2k)!} = \frac{d!\binom{d-k}{k}}{(d-k)!}$ denotes expressions resembling binomial coefficients, which we nickname the *wambonomial* coefficients. As a concrete example, for $d = 6$, we have

$$\frac{1}{6!} \sum_{i=1}^{m} \lambda_i (\ell_i^6 + 15 q_i \ell_i^4 + 45 q_i^2 \ell_i^2 + 15 q_i^3) \tag{3.4}$$

[1] Both examples used different random generator seeds.

Other examples are collected in Table 3.1. In the special case of centered Gaussians (cf. Definition 2.5.15(b)), the odd degree parts of the moment generating series vanish and the degree-$2d$ part is simply

$$\frac{1}{d!2^d}\sum_{i=1}^{m}\lambda_i q_i^d \tag{3.5}$$

In the centered case, we may ignore the scalar $\frac{1}{d!2^d}$ by homogeneity.

d	Gaussian moment form $\mathcal{M}_d(\mathcal{N}(\ell, q))$ of degree d
1	ℓ
2	$\ell^2 + q$
3	$\ell^3 + 3q\ell$
4	$\ell^4 + 6q\ell^2 + 3q^2$
5	$\ell^5 + 10q\ell^3 + 15q^2\ell$
6	$\ell^6 + 15q\ell^4 + 45q^2\ell^2 + 15q^3$
7	$\ell^7 + 21q\ell^5 + 105q^2\ell^3 + 105q^3\ell$
8	$\ell^8 + 28q\ell^6 + 210q^2\ell^4 + 420q^3\ell^2 + 105q^4$

Table 3.1: Moment forms (2.5.2) of *one* Gaussian distribution $\mathcal{N}(\ell, q)$ in degree $d \in \{1, \ldots, 8\}$, (cf. 3.1.2 and 2.5.3). Recall that the moment form is the degree-d part of the moment generating series rescaled by $d!$. The data of this table was generated by the notebook [88, `appendices/identifiability/code/gaussian-moment-derivatives.ipynb`].

Gaussian Mixtures with identical covariance

When all covariance forms (and for simplicity also the mixing weights) are equal, mixtures of Gaussians are algorithmically quite well-understood. In fact, a simple substitution translates the associated moment problem to tensor decomposition. This is clear from a statistical point of view, but can also be seen algebraically: if $q := q_1 = \ldots = q_m$, then the third-order moments attain the form

$$\sum_{i=1}^{m}\ell_i^3 + 3q\sum_{i=1}^{m}\ell_i$$

where the point $\sum_{i=1}^{m}\ell_i = \mathbb{E}[Y_1 + \ldots + Y_m]$ is known, since it is the vector of first-order moments of the mixture. Thus one can shift the space such that $\sum_{i=1}^{m}\ell_i = 0$ and perform classical tensor decomposition on this kind of Gaussian Mixture problem. The relevant identifiability question for this special type of Gaussian mixtures is therefore identifiability of the Veronese variety $V_{1,3}$, which was answered by Chiantini, Ottaviani and Vannieuwenhoven [23].

Mixtures of Centered Gaussians Recently, mixtures of *centered* Gaussians have gained attention [37], [36]. A Gaussian distribution $\mathcal{N}(\ell, q)$ is called *centered* if $\ell = 0$ (cf. 2.5.15). Mixtures of centered Gaussians are interesting both as a stepping stone to understand mixtures of arbitrary Gaussians and as a statistical model on its own: Assume we are observing a known signal x that is affected by various sources of Gaussian noise $Y_1 \sim \mathcal{N}(0, q_1), \ldots, Y_m \sim \mathcal{N}(0, q_m)$, where

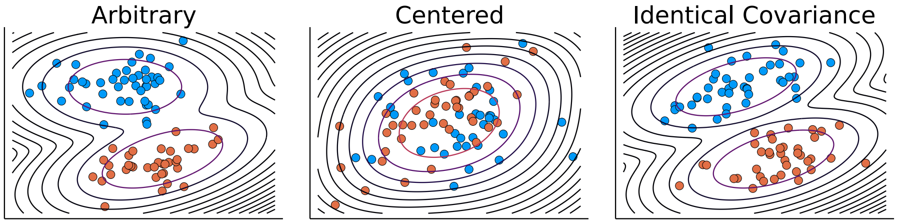

Figure 3.3: Comparison of different types of rank-2 Gaussian mixtures on $\mathbb{R}^2$: The left picture shows a Gaussian mixture in full generality with distinct centers $\mu_1, \mu_2 \in \mathbb{R}^2$ and distinct pd covariances $\Sigma_1, \Sigma_2 \in \mathrm{S}\mathbb{R}^{2\times 2}$. The middle one shows *centered* Gaussians with distinct covariance matrices and the right one shows Gaussians with distinct means but identical covariances. In all pictures, the dots correspond to 80 samples that were drawn from a bivariate mixture of two Gaussians, where the color indicates which of the two distributions was chosen in the sampling process (cf. 2.5.9). The contour lines are the level sets of the probability density function of the mixture.

the noise contributions are weighted by $\lambda_1, \ldots, \lambda_m \in \mathbb{R}_{\geq 0}$ with $\sum_{i=1}^m \lambda_i = 1$. Wlog normalize x to 0. Then we observe the random variable

$$\lambda_1 Y_1 \oplus \ldots \oplus \lambda_m Y_m \tag{3.6}$$

which by Remark 3.1.2 has the moment generating series

$$\sum_{d=0}^{\infty} \frac{1}{2^d d!} \sum_{i=1}^{m} \lambda_i q_i^d \tag{3.7}$$

The goal is to find out the individual covariance forms $q_1, \ldots, q_m$, given sufficiently many samples from (3.6). This gives sort of a “separation of noises”, i.e. mixture-of-centered-Gaussians models can explain a potentially complicated noise signal as a composition of “simple” (i.e. Gaussian) noises. Note that the latter application has nothing to do with clustering, even though it employs a Gaussian mixture model. As an aside, this demonstrates that it would be narrow-minded to reduce Gaussian mixture models solely to the purpose of clustering. Specifically for mixtures of centered Gaussians (and the more general powers of forms), there exist connections to circuit complexity, as described in [36, Introduction]. After estimating the moments up to a fixed degree $2d$, where $d \in \mathbb{N}_0$ (cf. Remark 2.5.7), the problem is purely algebraic and, after restricting to one, fixed degree $2d$, is a special case of Question 3.0.1, where $k = 2$, $K = \mathbb{R}$ and $q_1, \ldots, q_m$ are in addition assumed to be psd. The following theorem formalizes how generic identifiability for powers of quadratics will automatically yield identifiability results for mixtures of centered Gaussians.

3.1.3 Theorem: ([12, Theorem 2.2]) Let $n, m, d \in \mathbb{N}$ such that generic identifiability holds for d-th powers of quadratic forms in n variables of rank m (cf. e.g. 3.2.1 and 3.2.15).

Let $Y_1 \sim \mathcal{N}(0, q_1), \ldots, Y_m \sim \mathcal{N}(0, q_m)$ centered normal distributions given by general psd covariance forms $q_1, \ldots, q_m \in S^2(\mathbb{R}^n)$. Let $Z_1, \ldots, Z_m$ be any other centered Gaussian random vectors on $\mathbb{R}^n$ such that one of the following holds:

(a) the uniformly weighted Gaussian mixtures $Y = \frac{1}{m}(Y_1 \oplus \ldots \oplus Y_m)$ and $Z = \frac{1}{m}(Z_1 \oplus \ldots \oplus Z_m)$ agree on the moments of degree $2d$; or

(b) for general $\lambda_1, \ldots, \lambda_m \in \mathbb{R}_{>0}$ and for $\mu_1, \ldots, \mu_m \in \mathbb{R}_{\geq 0}$ both summing up to 1, the Gaussian mixtures $Y = \lambda_1 Y_1 \oplus \ldots \oplus \lambda_m Y_m$ and $Z = \mu_1 Z_1 \oplus \ldots \oplus \mu_m Z_m$ agree on the moments of degree $2d$ and $2d-2$.

Then $\{Y_1, \ldots, Y_m\} = \{Z_1, \ldots, Z_m\}$ and $Y = Z$. In case (b), the corresponding mixing weights are equal, too.

Proof. Let $p_1, \ldots, p_m \in S^2(\mathbb{R}^n)$ be quadratic psd forms such that

$$Z_1 \sim \mathcal{N}(0, p_1), \ldots, Z_m \sim \mathcal{N}(0, p_m)$$

In case (a), we denote $\lambda_i := \mu_i := \frac{1}{m}$ for each $i \in \{1, \ldots, m\}$, while in case (b) we fix $\lambda_1, \ldots, \lambda_m$ and $\mu_1, \ldots, \mu_m$ accordingly. Knowing that the degree $2d$ moments of $Y = \lambda_1 Y_1 + \ldots + \lambda_m Y_m$ and $Z = \mu_1 Z_1 + \ldots + \mu_m Z_m$ are equal, by Remark 3.1.2 we have

$$\sum_{i=1}^{m} (\sqrt[d]{\lambda_i} q_i)^d = \sum_{i=1}^{m} (\sqrt[d]{\mu_i} p_i)^d \tag{3.8}$$

where the quadratic forms $\sqrt[d]{\lambda_i} q_i$ are general for each $i \in \{1, \ldots, m\}$. By generic identifiability of powers of quadratics and Remark 3.1.4, we get

$$\{\sqrt[d]{\lambda_1} q_1, \ldots, \sqrt[d]{\lambda_m} q_m\} = \{\sqrt[d]{\mu_1} p_1, \ldots, \sqrt[d]{\mu_m} p_m\}$$

Note that for case (a) this is enough to conclude. In case (b), without loss of generality let us assume that

$$\sqrt[d]{\lambda_1} q_1 = \sqrt[d]{\mu_1} p_1, \ldots, \sqrt[d]{\lambda_m} q_m = \sqrt[d]{\mu_m} p_m$$

and write $\alpha_1 := \sqrt[d]{\frac{\lambda_1}{\mu_1}}, \ldots, \alpha_m := \sqrt[d]{\frac{\lambda_m}{\mu_m}}$. Since the degree $2d-2$ moments of Y and Z agree, we have

$$\sum_{i=1}^{m} \lambda_i q_i^{d-1} = \sum_{i=1}^{m} \mu_i p_i^{d-1}$$

where the $(d-1)$-st powers of the quadratic forms $q_1^{d-1}, \ldots, q_m^{d-1}$ are linearly independent. Substituting $p_i = \alpha_i q_i$, we obtain

$$\sum_{i=1}^{m} \lambda_i q_i^{d-1} = \sum_{i=1}^{m} \mu_i \alpha_i^{d-1} q_i^{d-1}$$

yielding $\mu_i (\frac{\lambda_i}{\mu_i})^{\frac{d-1}{d}} = \lambda_i$, i.e. $\mu_i = \lambda_i$ for each $i \in \{1, \ldots, m\}$. □

3.1.4 Remark: Note that the psd quadratic forms are a (Zariski) dense subset of $S^2(\mathbb{R}^n)$. Since the map $(q_1, \ldots, q_m) \mapsto \sum_{i=1}^{m} q_i^d$ is given by polynomials with rational coefficients, its image when restricted to real points (or even rational points) is (Zariski) dense in its complex image. Therefore, generic identifiability of complex powers of quadratic forms implies generic identifiability over the real field.

Caveats

In order to avoid stepping onto a minefield, let me briefly collect some positions on the matter of grading. According to [4, Introduction], for a long time, it was a common belief that teachers grade their students according to a normal distribution, where a large fraction of students have results clustering around the mean and a small fraction excels or underperforms. This "bell-curve paradigm" in grading has two facets: It can be purely descriptive and it can be a *normative practice*. By the latter one means that a distributional assumption (such as Gaussianity) is used to judge whether a certain way of grading a class is "correct" or socially acceptable. This practice can range from just sanity-checking or adjusting an exam that might have been too hard as done in Figure 3.1 to the other extreme of *curve grading*,[2] where the teacher fits a (normal) distribution to the data set and reads the grades from it. Curve grading can protect students from unintentionally hard exams, as the grading adjusts to the results, but can e.g. also punish good students for being in a class full of geniuses.

One might argue that distributional assumptions as a normative practice protect the value of grades: If only a small fraction of students in a class get a top grade, then this certifies that they performed better than most students in the class. A top grade is then a somewhat exclusive achievement and thus has value by rarity. This position typically worries about *grade inflation*: The disputed term (cf. [72] vs. [52]) describes the phenomenon when average grades in an educational system become better over time. Calling it an inflation connotes that grades lose value when higher grades are more common – which is far from clear and depends on what grades are supposed to express.

The counter position (e.g. [52]) argues that grade inflation can also occur as a natural consequence of an improving education system: If the system gets better, the average student will achieve better results. Thus, provided that the education goals stay constant, one expects to see a rise in grades. A grading system that does not reflect such collective improvement can be seen as demotivating, as it turns grades into a competitive zero-sum game.

A more fine-grained counterargument challenges the descriptive aspect of the bell-curve paradigm: Arthurs, Stenhaug, Karayed and Piech [4] argue that grades are better described by logit-normal distributions. Finally, let me disclose that I did not use the actual empirical distribution of exam grades, but a "hand-adjusted" similar-looking one in order to not accidentally disclose student information that could be present in an anonymized dataset.

[2] For which there are tutorials on the internet, cf. [64], [90]

3.2. Identifiability for Sums of Powers

This section is dedicated to answering Question 3.0.1(a), i.e. to examine for which numbers $n, m, k, d \in \mathbb{N}$ there exists generically a unique solution to the power-sum decomposition problem. Whenever that is the case, we say that *generic identifiability* holds for sums of dth powers of k-forms of rank m in n variables. The smallest nontrivial case (and arguably also the most interesting one) is $d = 3$. Indeed, note that for $d = 2$ we cannot hope for identifiability since $q_1^2 + q_2^2 = \frac{1}{2}(q_1 + q_2)^2 + \frac{1}{2}(q_1 - q_2)^2$. This section is based on joint work with Casarotti, Michałek and Oneto and overlaps to a very large part with our publication [12]. We used two different approaches to obtain identifiability that I will both explain in the following. The first approach is limited to the "minimal" case $(d, k) = (3, 2)$ and shows identifiability up to rank $m \leq \binom{n}{2} + 1$, for any $n \in \mathbb{N}$. Note this gives the wrong asymptotics, as the generic rank for cubes of quadratics has to be of the order $\mathcal{O}(n^4)$. The second result gives the correct asymptotics for *any* (d, k) with $d \geq 3$ and $k \geq 2$, but is vacuous for the first small values of $n \in \mathbb{N}$ (e.g. for $(d, k) = (3, 2)$ it does not say anything for $n \leq 16$). Further note that for fixed d, k, the finitely many values where our result is not applicable may be checked on a computer.

Overview of contributions

In the case of sextics as sums of cubes of quadratic forms, we obtain the following result about general identifiability. Note that "(b)" in the theorem below is only interesting for $n \leq 16$, as otherwise the range from "(a)" is strictly better.

3.2.1 Theorem: ([12, Theorem 1.1]) Let $n, m \in \mathbb{N}$ such that one of the following holds:

(a) $n > 16$ and $m \leq \binom{n+5}{6} / \binom{n+1}{2} - \binom{n+1}{2} - 1$; or

(b) $m \leq \binom{n}{2} + 1$.

Then, for general $q_1, \ldots, q_m \in S^2(\mathbb{C}^n)$, the sextic $t = \sum_{i=1}^m q_i^3$ has a unique representation as a sum of m cubes of quadratic forms, up to permutation and third roots of unity.

With the connection exposed in Section 3.1 and formalized in Theorem 3.1.3, a direct consequence is the following degree-6 identifiability result for mixtures of general centered Gaussians.

3.2.2 Corollary (cf. 3.1.3): For (m, n) in the same range as in Theorem 3.2.1, the parameters of a general centered Gaussian mixture of rank m are uniquely identifiable (up to permutation) from the mixture moments of degree 6 and 4.

Condition (a) of Theorem 3.2.1 may be generalized to guarantee generic identifiability for sums of arbitrary powers of higher degree forms (cf. Corollary 3.2.15). This is of independent interest, but specifically for higher order powers of quadratic forms, it has similar consequences regarding identifiability of Gaussian mixtures. The subsequent theorem summarizes the asymptotics of all the results we are going to show with respect to identifiability.

3.2.3 Theorem: (compare with [12, Corollary 4.4]) Fix $d, k \in \mathbb{N}$ with $d \geq 3$ and $k \geq 2$. There exists a rank threshold $m_{\max}(n) := m_{\max,k,d}(n) \in \Omega(n^{kd-k})$ such that for all $n \in \mathbb{N}$ and for all $m \leq m_{\max}(n)$, the following holds true:

For general $q_1, \ldots, q_m \in S^k(\mathbb{C}^n)$, the degree-$kd$ form $t = \sum_{i=1}^m q_i^d$ has a unique representation as a sum of m d-th powers of k-forms, up to permutation and dth roots of unity. For $k = 2$ and m up to $m_{\max}(n) \in \Omega(n^{2d-2})$, the parameters of a general centered Gaussian mixture of rank m are uniquely identifiable (up to permutation) from the mixture moments of degree $2d$ and $2d - 2$ (cf. 3.1.3).

Note that the threshold value $m_{\max}(n)$ up to which we show identifiability will be zero for the first few values of $n \in \mathbb{N}$. This is due to the dimensionality conditions that stem from the work of Casarotti and Mella (2.2.25) and of Nenashev (3.2.12). To make it explicit, the following theorem summarizes the identifiability results we obtain for mixtures of Gaussians from their moments.

3.2.4 Theorem: ([12, Theorem 1.3]) Let $n, d \in \mathbb{N}$ with $d \geq 3$ such that

$$3\binom{n+1}{2}^2 - 2\binom{n+1}{2} < \binom{n-1+2d}{2d} \tag{3.9}$$

and $m \in \mathbb{N}$ such that

$$m \leq \binom{n-1+2d}{2d} / \binom{n+1}{2} - \binom{n+1}{2} - 1$$

Then, for general $q_1, \ldots, q_m \in S^2(\mathbb{C}^n)$, the degree-$2d$ form $t = \sum_{i=1}^m q_i^d$ has a unique representation as a sum of m d-th powers of quadratic forms, up to permutation and dth roots of unity. In the same range of (m, n), the parameters of a general centered Gaussian mixture of rank m are uniquely identifiable (up to permutation) from the mixture moments of degree $2d$ and $2d - 2$ (cf. 3.1.3).

3.2.5 Remark: Most of our results are formulated for either the complex or the real field. However, complex generic identifiability implies generic identifiability over all fields of characteristic 0. For *algebraically closed* fields C of characteristic 0, this is clear by the Lefschetz principle / quantifier elimination (cf. [73, Theorem 3.3.4]), since for fixed m, n, k, d, D, the quantifiers in the first line of the formula

$$\begin{aligned} &\exists f \in C[S^k(C^n)^m]_{\leq D} : \forall q_1, \ldots, q_m, p_1, \ldots, p_m \in S^k(C^n) : \\ &\qquad \Bigg(\Bigg(f(q_1, \ldots, q_m) \neq 0 \quad \& \quad \sum_{i=1}^m q_i^d = \sum_{i=1}^m p_i^d\Bigg) \\ &\qquad \Longrightarrow \exists \sigma \in \mathfrak{S}_n : \forall i \in \{1, \ldots, m\} : \quad q_i = p_{\sigma(i)}\Bigg) \end{aligned} \tag{3.10}$$

can be reduced to finitely many quantifications over C (by quantifying over the coefficients of the q_i, p_i and f) and the quantifiers in the third line are finite. Note that f encodes the generality of $q = (q_1, \ldots, q_m)$, i.e. some dense open subset $\mathcal{U} = \{q \in S^k(C^n)^m \mid f(q) \neq 0\}$ of $S^k(C^n)^m$.

Now, for an arbitrary field L of characteristic 0 and $C := \overline{L}$, the argument is essentially the same as in Remark 3.1.4: We know that the image of the polynomial map $S^k(L^n)^m \to \sigma_m(V_{k,d}(C^n)), (q_1, \ldots, q_m) \mapsto \sum_{i=1}^m q_i^m$ is (Zariski) dense. Since $\sigma_m(V_{k,d}(C^n))$ contains a dense open subset $\mathcal{V}$ of identifiable forms, the intersection $\sigma_m(V_{k,d}(L^n))\cap\mathcal{V}$ is a dense open subset of $\sigma_m(V_{k,d}(L^n))$. If thus the general element of the latter is identifiable, then so is the general element of $\sigma_m(V_{k,d}(L^n))$.

First approach via specific decomposition

Recall that $V_{k,d} = \{q^d \mid q \in \mathbb{P}(S^k(\mathbb{C}^n))\}$ denotes the smooth projective variety of d-th powers of degree-k forms on $\mathbb{C}^n$ (cf. Proposition 2.4.2). We will eventually prove that general identifiability holds in a range of ranks which is asymptotically optimal, i.e. of the same asymptotic order as the generic rank. This is done in the *next* section. For the time being, let us start with a modest but instructive approach that assumes $(k, d) = (2, 3)$ and manages to prove identifiability for n-variate cubes of quadratics up to rank $m = \binom{n}{2} + 1$, via studying an explicit decomposition of a specific sextic form and employing the semicontinuity argument from Theorem 2.2.27: Indeed it suffices to find specific quadratics $q_1, \dots, q_m$ with skew tangent spaces and $m = \binom{n}{2} + 1$, such that the tangential contact locus at $q_1, \dots, q_m$ consists only of the points $q_1, \dots, q_m$.

We will use $X = (X_1, \dots, X_n)$ as variables. Given $q_1, \dots, q_m \in \mathbb{C}[X]_2$ we denote by $\widehat{\mathcal{C}}(q_1, \dots, q_m)$ the affine cone of the preimage of the tangential contact locus at the points $[q_1^3], \dots, [q_m^3]$ via the map ι from Proposition 2.4.2. This is slight abuse of notation, since it conflicts with our usual usage of the hat symbol. Similarly, we denote by $\widehat{\Gamma}(q_1, \dots, q_m) := \iota^{-1}(\widehat{\Gamma_{V_{2,3}}([q_1^3], \dots, [q_m^3])})$ the preimage via ι of $\Gamma_{V_{2,3}}$ at $q_1, \dots, q_m$, cf. 2.2.23. This notation suppresses the dependency on n. Written out explicitly, this yields:

$$\widehat{\Gamma}(q_1, \dots, q_m) = \left\{ p \in \mathbb{C}[X]_2 \mid \forall h \in \mathbb{C}[X]_2 : \exists h_1, \dots, h_m \in \mathbb{C}[X]_2 : p^2 h = \sum_{i=1}^{m} q_i^2 h_i \right\}.$$

3.2.6 Definition: For $i, j \in \{1, \dots, n\}$ with $i < j$, define

$$q_{ij} := (X_i + X_j)^2 \tag{3.11}$$

and let

$$\mathcal{B}_n := \{(X_i + X_j)^2 \mid i, j \in \{1, \dots, n\}, i < j\} \cup \{X_1^2\}$$

We call $\mathcal{B}_n$ the *binomial set* of quadratics in dimension n. Up to relabeling we can write $\mathcal{B}_n = \{q_1, \dots, q_{\binom{n}{2}+1}\}$, where the order of the elements is arbitrary.

3.2.7 Remark: For $n \geq 2$, the following is an easy observation:

$$\mathcal{B}_{n-1} \cup \{4X_1^2\} = \{p(X_1, \dots, X_{n-1}, X_1) \mid p \in \mathcal{B}_n\} \tag{3.12}$$

3.2.8 Theorem: (cf. [12, Theorem 4.10]) The tangent spaces at elements of the binomial set $\mathcal{B}_n$ are skew, i.e.

$$T_{q_1}\widehat{V} + \ldots + T_{q_{\binom{n}{2}+1}}\widehat{V} = T_{q_1}\widehat{V} \oplus \ldots \oplus T_{q_{\binom{n}{2}+1}}\widehat{V}$$

Proof. We proceed by induction on n. For $n \leq 5$, we verify the statement on a computer. The code may be found on GitHub [87]. Therefore we can assume that $n \geq 6$ and that the claim on $\mathcal{B}_k$ is true for all $k < n$. Let $h_1, h_{ij} \in S^2(\mathbb{C}^n)$, where $i, j \in \{1, \dots, n\}$ and $i < j$. Suppose that

$$0 = h_1 X_1^4 + \sum_{1 \leq i < j \leq n} h_{ij}(X_i + X_j)^4 \tag{3.13}$$

We have to show that $h_1 = h_{ij} = 0$ for all $1 \leq i < j \leq n$. Denote $h_{ji} := h_{ij}$. Since $\mathcal{B}_n$ is symmetric under permutations of $\{X_2, \dots, X_n\}$, without loss of generality it suffices to show that $h_{12} = 0$ and $h_{23} = 0$. Let us first consider the case $(i, j) = (2, 3)$. Since $n \geq 6$, we may apply the substitution

$$\varphi_4 \colon \mathbb{C}[X] \to \mathbb{C}[X_1, \dots, X_3, X_5 \dots, X_n], X_4 \mapsto X_1$$

to reduce to a case with one variable less. We obtain

$$0 = \varphi_4(h_1)X_1^4 + \sum_{\substack{1 \leq k < l \leq n \\ 4 \notin \{k,l\}}} \varphi_4(h_{kl})(X_k + X_l)^4 + \sum_{k=1}^{n} \varphi_4(h_{4k})(X_k + X_1)^4 \tag{3.14}$$

Now note that the form $(X_2 + X_3)^4$ can only occur on the first summation, yielding $\varphi_4(h_{23}) = 0$. Therefore by construction $(X_1 - X_4)$ divides the quadratic form h_{23}. Repeating this same argument with the substitutions

$$\varphi_5 \colon X_5 \mapsto X_1$$
$$\varphi_6 \colon X_6 \mapsto X_1$$

yields that $(X_1 - X_5)$ and $(X_1 - X_6)$ divide h_{23}, too. Since these linear forms are coprime, $(X_1 - X_4)(X_1 - X_5)(X_1 - X_6)$ must divide h_{23}, which for degree reasons is only possible if $h_{23} = 0$. By symmetry of $\mathcal{B}_n$, we get that $h_{ij} = 0$ for all pairs $\{i, j\}$ not containing 1. Thus Equation (3.13) simplifies to

$$0 = h_1 X_1^4 + \sum_{j=1}^{n} h_{1j}(X_1 + X_j)^4 \tag{3.15}$$

As for the $(i, j) = (1, 2)$ case: If h_{12} were not the zero form, then $h_{12}(X_1 + X_2)^4$ would contain a monomial of degree at least 4 in X_2. Since all other addends in Equation (3.15) can only contain monomials of degree at most 2 in X_2, the terms of degree at least 4 in X_2 from $h_{12}(X_1 + X_2)^4$ could not cancel with any other addend from (3.15). After a short argument left to the reader, this forces $h_{12} = 0$ and by symmetry thus $h_{13} = \ldots = h_{1n} = 0$. Finally, we also must have $h_1 = 0$, as it is the only remaining term in (3.15). □

Now we show that the tangential contact locus for the binomial set is zero dimensional at each point of the binomial set.

3.2.9 Theorem: ([12, Theorem 4.11]) For $n \in \mathbb{N}$, and each $q \in \{q_1, \dots, q_{\binom{n}{2}+1}\}$, locally around q, $\widehat{\mathcal{C}}(q_1, \dots, q_{\binom{n}{2}+1})$ only contains points from the line $\mathbb{C}q$.

Proof. We use affine notation and proceed by induction on the number n of variables. The base cases $n \leq 5$ were verified on a computer, see [87].

Thus let us assume $n \geq 6$. As $\mathcal{B}_n$ is invariant under permutations of $X_2, \ldots, X_n$, it suffices to show the claim at $q \in \{X_1^2, (X_1 + X_2)^2, (X_2 + X_3)^2\}$. In particular, we may assume that q is a polynomial in X_1, X_2, X_3. As we work locally around q, it does not matter whether we show the statement for $\widehat{\Gamma}$ or $\widehat{\mathcal{C}}$, cf. 2.2.23. We thus have to show that there exists a neighbourhood $\mathcal{U} \subseteq \mathbb{C}[X]_2$ of q such that $\mathcal{U} \cap \widehat{\Gamma}(q_1, \ldots, q_m) \subseteq \mathbb{C}q$. Consider the substitution

$$\varphi \colon \mathbb{C}[X] \to \mathbb{C}[X_1, \ldots, X_{n-1}]$$

that maps X_n to X_1. Note $\varphi(q) = q$, as $n \geq 6$. By induction hypothesis, we know there exists a neighbourhood $\mathcal{V} \subseteq \mathbb{C}[X_1, \ldots, X_{n-1}]_2$ of q such that $\mathcal{V} \cap \widehat{\Gamma}(\mathcal{B}_{n-1}) \subseteq \mathbb{C}q$. This means that for all

$$p \in \varphi^{-1}(\mathcal{V}) \cap \widehat{\Gamma}(q_1, \ldots, q_m)$$

there is $\lambda \in \mathbb{C}$ such that $\varphi(p) = \lambda q$. In other words,

$$(X_1 - X_n)|(p - \lambda q)$$

Repeating the same argument with the substitution φ' that maps X_{n-1} to X_1, we obtain another neighbourhood $\mathcal{V}'$ with the property that for each

$$p \in \varphi'^{-1}(\mathcal{V}') \cap \widehat{\Gamma}(q_1, \ldots, q_m)$$

it holds that $\varphi'(p) = \lambda' q$.

Let $\mathcal{U} = \varphi^{-1}(\mathcal{V}) \cap \varphi'^{-1}(\mathcal{V}')$, then for each $p \in \mathcal{U} \cap \widehat{\Gamma}(q_1, \ldots, q_m)$, we can find $\lambda, \lambda' \in \mathbb{C}$ and linear forms $\ell, \ell' \in \mathbb{C}[X]_1$ such that

$$\lambda q + \ell(X_1 - X_n) = p = \lambda' q + \ell'(X_1 - X_{n-1}) \tag{3.16}$$

Finally, we have that ℓ has to be a polynomial in the variables $\{X_1, X_{n-1}\}$: Indeed, if a variable X_j for some $j \notin \{1, n-1\}$ occurred in ℓ, then the monomial $X_j X_n$ on the left hand side of (3.16) could not cancel with any other terms on the left-hand side, but does also not occur on the right hand side. It follows that p is a polynomial in $\{X_1, X_2, X_3, X_{n-1}, X_n\}$. Thus we reduced to the case of 5 variables and proved the claim. □

3.2.10 Remark: (a) Our results do *not* imply that the mixture of cubes of quadratics $\sum_{q \in \mathcal{B}_n} q^3$ has a unique decomposition as a sum of $\binom{n}{2} + 1$ cubes of quadratics! In [23], the authors consider some sufficient criteria for the identifiability of specific tensors that maybe, albeit with unnegligible effort, could be transferred to the setting of cubes of quadratics. We did not do any work regarding specific identifiability for cubes of quadratics.

(b) We verify Theorem 3.2.8 and Theorem 3.2.9 for $n = 5$ on a computer. This readily implies the statements for all dimensions at most 5, since $\mathcal{B}_1 \subseteq \mathcal{B}_2 \subseteq \ldots$. The code is publicly available on GitHub, see [87]. This base case can be verified using only methods of Numerical Linear Algebra (such as determining dimensions of certain vector spaces of polynomials) and should therefore be easy to reproduce independently.

(c) We could also deduce identifiability from Theorem 3.2.8 together with Theorem 2.2.25, albeit only up to rank $\binom{n}{2}$. This almost voids the need to prove Theorem 3.2.9 directly, but would have been less instructive.

Higher order identifiability and Fröberg's conjecture

Recall that the (affine) tangent space to $V_{k,d}$ at $q = p^d$ is given by

$$T_p V_{k,d} = \{p^{d-1}h \mid h \in S^k(\mathbb{C}^n)\} = (p^{d-1})_{dk} \tag{3.17}$$

(cf. Lemma 2.4.3 and 2.2.33). Therefore, by Terracini's Lemma (2.2.19, 2.2.20), in order to prove that secant varieties of $V_{k,d}$ have the expected dimension, we only have to show that the tangent spaces $T_{p_1^d}V_{k,d}, \ldots, T_{p_m^d}V_{k,d}$ are skew for a general choice of $p_1, \ldots, p_m$. That is equivalent to say that the degree-kd part of the ideal $(p_1^{d-1}, \ldots, p_m^{d-1})$ has maximal dimension, cf. 2.2.33.

Such ideals generated by, in a certain sense, "general" polynomials have been studied since a few decades in the context of *Fröberg's conjecture* stated below.

3.2.11 Conjecture: (Fröberg, [33]) Let $m \in \mathbb{N}$ and $g_1, \ldots, g_m$ general forms of degrees $a_1, \ldots, a_m$. Let $I = (g_1, \ldots, g_m) \subseteq \mathbb{C}[X]$. Then the Hilbert series of I is given by the formula

$$\mathrm{HS}(\mathbb{C}[X]/I, T) = \left[\frac{\prod_{i=1}^m (1 - T^{a_i})}{(1-T)^n}\right] \tag{3.18}$$

The "$[\cdot]$" in Conjecture 3.2.11 means that each negative coefficient in the power series expansion of the fraction is replaced by zero. There are various variants in the literature, some of which can be obtained by specializing the class of forms. E.g. we do not need the case of (absolutely) general forms, but rather forms that are general *within the class of powers of forms*, and we can assume all forms to be of the same degree. The analogous of Fröberg's conjecture for this case would state that whenever $p_1, \ldots, p_m$ are forms of *same* degree $k \geq 2$ and $d \in \mathbb{N}$, then

$$\mathrm{HS}(\mathbb{C}[X]/(p_1^d, \ldots, p_m^d), T) = \left[\frac{(1 - T^{dk})^m}{(1-T)^n}\right] \tag{3.19}$$

This is in fact precisely [68, Conjecture 2]. Note that the variant of Fröberg's conjecture for powers of forms is formally not a special case of the standard Fröberg's conjecture, due to the different notions (relative vs. absolute) of genericity. Fortunately, for the powers-of-forms setting and *most* values of m, (3.19) holds true, which follows from a much more general result proven by Nenashev [67]:

3.2.12 Theorem: [67, Theorem 1] Let $a \in \mathbb{N}$ and let $\mathcal{D}$ be a class[3] of degree-a forms which is closed under the canonical action of $\mathrm{GL}_n(\mathbb{C})$ on forms. For $m, h \in \mathbb{N}$, general $g_1, \ldots, g_m \in \mathcal{D}$, and $I = (g_1, \ldots, g_m)$, as long as

$$m \leq \frac{\dim S^{a+h}(\mathbb{C}^n)}{\dim S^h(\mathbb{C}^n)} - \dim S^h(\mathbb{C}^n) \qquad \text{or} \tag{3.20}$$

$$m \geq \frac{\dim S^{a+h}(\mathbb{C}^n)}{\dim S^h(\mathbb{C}^n)} + \dim S^h(\mathbb{C}^n) \tag{3.21}$$

[3] The paper [67] does not specify what a "class" is supposed to be in this context (note that "general $g \in \mathcal{D}$" needs to make sense), but we are safe if we assume $\mathcal{D}$ e.g. to be an algebraic variety.

it holds that the T^{a+h}-coefficients of

$$\mathrm{HS}(\mathbb{C}[X]/I,T) \quad \text{and} \quad \left[\frac{(1-T^a)^m}{(1-T)^n}\right] \tag{3.22}$$

are equal. Thus, formulated in terms of the ideal rather than the quotient (cf. 2.2.34),

$$\mathrm{HF}_I(a+h) = \dim(g_1,\dots,g_m)_{a+h} = m\cdot\dim S^h(\mathbb{C}^n), \text{ for } m \text{ as in 3.20} \tag{3.23}$$

$$\mathrm{HF}_I(a+h) = \dim(g_1,\dots,g_m)_{a+h} = \dim S^{a+h}(\mathbb{C}^n), \text{ for } m \text{ as in 3.21} \tag{3.24}$$

3.2.13 Remark: A consequence of Theorem 3.2.12 is that there are *at most* $2\cdot\dim S^h(\mathbb{C}^n)$ possible values for m for which (3.19) might fail. To see the connection between (3.22) and (3.23)/(3.24), note that the coefficient of T^{a+h} in $\frac{(1-T^a)^m}{(1-T)^n}$ is simply $\dim S^{a+h}(\mathbb{C}^n) - m\cdot\dim S^h(\mathbb{C}^n)$, which follows from playing around with binomial coefficients.

Theorem 3.2.12 may be applied with $a=(d-1)k, h=k$ and $\mathcal{D} := V_{k,(d-1)}$:

3.2.14 Theorem: Let $m,n,d\in\mathbb{N}$ with $d\geq 3$ and $m\leq \frac{\dim S^{kd}(\mathbb{C}^n)}{\dim S^k(\mathbb{C}^n)} - \dim S^k(\mathbb{C}^n)$. Then, the m-th secant variety of $V_{k,d}$ has the expected dimension. That is,

$$\dim\sigma_m(V_{k,d}) = m\cdot\dim S^k(\mathbb{C}^n) - 1 \tag{3.25}$$

Proof. Choose general forms $p_1,\dots,p_m\in S^k(\mathbb{C}^n)$. By Terracini's Lemma 2.2.19 and Lemma 2.4.3, the tangent space at $t := p_1^d + \ldots + p_m^d$ to $\sigma_m(V_{k,d})$ is the degree-kd part of the ideal $I = (p_1^{d-1},\dots,p_m^{d-1})$ (cf. 2.2.33). The latter has dimension $m\cdot\dim S^k(\mathbb{C}^n)$ by Theorem 3.2.12, which is the (affine) expected dimension. □

Now we combine Theorem 3.2.14 and Theorem 2.2.25 as in [12] to obtain

3.2.15 Corollary: The m-th secant variety of $V_{k,d}$ is generically m-identifiable for all $m\leq \frac{\dim S^{kd}(\mathbb{C}^n)}{\dim S^k(\mathbb{C}^n)} - \dim S^k(\mathbb{C}^n) - 1$ provided that $2\cdot(\dim S^k(\mathbb{C}^n)-1) < \frac{\dim S^{kd}(\mathbb{C}^n)}{\dim S^k(\mathbb{C}^n)} - \dim S^k(\mathbb{C}^n)$.

Here, it was used that generic m-identifiability implies generic $(m-1)$-th identifiability, cf. Lemma 2.2.31. In particular, in the $(k,d)=(2,3)$ case, we obtain the condition (a) of Theorem 3.2.1. Indeed, note that the condition required by Corollary 3.2.15, i.e.,

$$2\left(\binom{n+1}{2}-1\right) < \frac{\binom{n+5}{6}}{\binom{n+1}{2}} - \binom{n+1}{2},$$

holds if and only if $n>16$. As a corollary, we prove the claim from Theorem 3.2.1 under Condition (a), which completes the missing part of the proof of Theorem 3.2.1. It also proves Theorem 3.2.3. Note that in the special case $k=2$, the condition of Corollary 3.2.15 simplifies to

$$3\binom{n+1}{2}^2 - 2\binom{n+1}{2} < \binom{n-1+2d}{2d}$$

and thus we also completed the proof of Theorem 3.2.4. In conclusion, I think it is apt to say that we gave a thorough answer to Question 3.0.1(a) and the corresponding identifiability question for mixtures of centered Gaussians.

3.3. Mixtures of noncentered Gaussians

As a last application, we show how the identifiability results for powers of quadratics together with classical results on the nondefectivity of the secants of the Veronese imply generic identifiability for mixtures of *noncentered* Gaussians from their moments of degree 6 in a large range of ranks. Note that numerical experiments suggest that degree 6 is not the minimal degree where we can hope for nontrivial identifiability results, but degree 5 is.

Degree 6 identifiability

3.3.1 Definition: The degree-6 *Gaussian moment variety* $\mathrm{GM}_6(U)$ of a (complex) affine space U is the closure of

$$\{\ell^6 + 15q\ell^4 + 45q^2\ell^2 + 15q^3 \mid \ell \in U^\vee, q \in S^2(U)\} \subseteq S^6(U) \tag{3.26}$$

3.3.2 Proposition: $\mathrm{GM}_6(U)$ is the closure of the image of

$$s\colon U^\vee \times S^2(U) \to S^6(U), (\ell, q) \mapsto \ell^6 + 15q\ell^4 + 45q^2\ell^2 + 15q^3 \tag{3.27}$$

For general (ℓ, q), the tangent space at $s(\ell, q)$ can thus be computed by deriving curves $t \mapsto s(\ell + th, q + tp)$, cf. Theorem 2.2.14. Note that $\operatorname{im} T_s(\ell, q)$ equals

$$\{(\ell^5 + 10q\ell^3 + 15q^2\ell)h + (\ell^4 + 6q\ell^2 + 3q^2)p \mid h \in S^1(U), p \in S^2(U)\}$$

where T_s denotes the linear map of (affine) tangent spaces induced by s.

3.3.3 Remark: From deriving by the q-parameter in direction p, one would actually obtain the term $(15\ell^4 + 90q\ell^2 + 45q^2)p$. The common factor of $15 = \binom{6}{2}$ was substituted away into p to make all subsequent calculations easier to read. Similarly, a common factor of 6 was substituted away into h from the ℓ-derivative in Proposition 3.3.2. How this works for general degrees is explained in Table 3.2.

3.3.4 Lemma: Let U a complex vector space of dimension n, $\ell_1, \ldots, \ell_m \in U^\vee$ general and $q_1, \ldots, q_m \in S^2(U)$ general. Then the tangent spaces $T_{\ell_i,q_i} \mathrm{GM}_6(U)$, where $i \in \{1, \ldots, m\}$, are skew spaces for some $m \in \theta(n^4)$.

Proof. It suffices to show that $\operatorname{im} T_s(\ell_1, q_1) + \ldots + \operatorname{im} T_s(\ell_m, q_m)$ has the maximum possible dimension $m(\dim(U) + \binom{\dim(U)+1}{2})$ by Proposition 3.3.2, where s denotes the parametrization from 3.3.2. This dimension can only be smaller on specific instances of (ℓ_i, q_i). We are to show that a general *variable-split* instance will achieve maximum dimension.

Choose variables $(X, Y) = (X_1, \ldots, X_n, Y_1, \ldots, Y_n)$ if $\dim(U) = 2n$ is even or $Y = (Y_1, \ldots, Y_{n+1})$ if $\dim(U) = 2n + 1$ is odd. We are to show that the sum $\operatorname{im} T_s(\ell_1, q_1) + \ldots + \operatorname{im} T_s(\ell_m, q_m)$ has the maximum possible dimension for general $\ell_i \in \mathbb{C}[Y]_1$ and general $q_i \in \mathbb{C}[X]_2$. Thus, let us assume in the following that $\ell_1, \ldots, \ell_m$ only depend on the variables in Y and $q_1, \ldots, q_m$ only depend on the variables in X. One obtains the equation:

$$0 = \sum_{i=1}^{m} (\ell_i^5 + 10q_i\ell_i^3 + 15q_i^2\ell_i)h_i + \sum_{i=1}^{m} (\ell_i^4 + 6q_i\ell_i^2 + 3q_i^2)p_i \tag{3.28}$$

where $h_i \in \mathbb{C}[X,Y]_1$ and $p_i \in \mathbb{C}[X,Y]_2$. To show: $h_1 = \ldots = h_m = 0$ and $p_1 = \ldots = p_m = 0$. Let us split each $h_i = h_i(X,0) + h_i(0,Y) =: h_{i,X} + h_{i,Y}$ into a part that only depends on X and a part that only depends on Y. Similarly, let us split $p_i = p_{i,X} + p_i\langle X,Y\rangle + p_{i,Y}$ into a part of degree 2 in X, a part of degree 2 in Y and a part $p_i\langle X,Y\rangle$ that is bilinear in X and Y. Now, if an expression of the kind (3.28) vanishes, in particular, the part of degree 6 in X has to vanish. Due to the split variables, only one term can contribute to this part and we obtain f

$$0 = 3\sum_{i=1}^{m} q_i^2 p_{i,X} \tag{3.29}$$

(3.29) is a a sum of tangents to the cubes-of-quadratics variety in variables X at general points. By Theorem 3.2.14, we deduce $p_{i,X} = 0$ for all i. Next, we look at the terms whose (X,Y)-degrees are $(4,2)$. We obtain the equation:

$$0 = 15\sum_{i=1}^{m} q_i^2 \ell_i h_{i,Y} + 3\sum_{i=1}^{m} q_i^2 p_{i,Y} \tag{3.30}$$

Note that the term $\sum_{i=1}^{m} q_i \ell_i^2 p_{i,X}$ cannot make a contribution since we just showed that all $p_{i,X}$ vanish. Thus, again by Theorem 3.2.14, we obtain that $-5\ell_i h_{i,Y} = p_{i,Y}$ for all i. Let us plug in this newly-obtained identity into the part of Y-degree 6 from (3.28), which is

$$0 = \sum_{i=1}^{m} \ell_i^5 h_{i,Y} + \sum_{i=1}^{m} \ell_i^4 p_{i,Y} \tag{3.31}$$

to obtain

$$0 = \sum_{i=1}^{m} \ell_i^5 h_{i,Y} - 5\sum_{i=1}^{m} \ell_i^5 h_{i,Y} = -4\sum_{i=1}^{m} \ell_i^5 h_{i,Y} \tag{3.32}$$

It is classical that general tangent spaces to the Veronese variety $V_{1,6}$ are skew, as long as m is subgeneric, which asymptotically is satisfied for some $m \in \theta(n^5)$. Thus we conclude that $h_{i,Y} = 0$ for all i (and thus $p_{i,Y} = -4\ell_i h_{i,Y} = 0$ by what was previously shown). Let us repeat the same procedure with the part of degree $(5,1)$ in (X,Y), which is

$$0 = 15\sum_{i=1}^{m} q_i^2 \ell_i h_{i,X} + 3\sum_{i=1}^{m} q_i^2 p_i\langle X,Y\rangle \tag{3.33}$$

This readily gives that $p_i\langle X,Y\rangle = -5\ell_i h_{i,X}$ for all i. We can plug that into

$$0 = \sum_{i=1}^{m} \ell_i^5 h_{i,X} + \sum_{i=1}^{m} \ell_i^4 p_i\langle X,Y\rangle \tag{3.34}$$

which is the degree-$(1,5)$ part of (3.28) to obtain

$$0 = -4\sum_{i=1}^{m} \ell_i^5 h_{i,X} \tag{3.35}$$

Hence also $h_{i,X} = 0$ and thus $h_i = 0$ for all i. As we have $p_i\langle X,Y\rangle = -4\ell_i h_{i,X} = 0$ for all i, we also get $p_1 = \ldots = p_m = 0$, since it was shown that all X-homogeneous components of the p_i vanish. □

3.3.5 Theorem: $\mathrm{GM}_6(\mathbb{C}^n)$ is m-identifiable for some semi-increasing function $m \in \theta(n^4)$.

Proof. Combine Lemma 3.3.4 with Theorem 2.2.25. □

3.3.6 Remark: Note that for the range of ranks m in the above identifiability proof, the cubes-of-quadratics variety is the bottleneck: We showed identifiability for cubes of n-variate quadratics up to rank $m \in \mathcal{O}(n^4)$, whereas $V_{1,6}(\mathbb{C}^n)$ is identifiable up to rank $m \in \mathcal{O}(n^5)$. We can split the variables unevenly, i.e. choose $c = c_n \in [0, 1]$ such that approximately a c-fraction of the variables goes to Y and a $(1-c)$ fraction goes to X. Particularly, choosing $c \approx \sqrt[5]{1/n}$, we obtain for large n with this trick that identifiability holds up to the correct constant, i.e. for ranks up to a function $m_{\max}(n) \in \theta^{\#}(\dim S^6(\mathbb{C}^n)/\dim \mathrm{GM}_6(\mathbb{C}^n))$.

3.3.7 Remark: Similar arguments could in principle be made for Gaussian moment varieties of any even degree $2d \geq 6$. The procedure is to start with the term of highest degree in X, which is always $\sum_{i=1}^m q_i^{d-1} p_i(X, 0)$, and then work along a similar pattern, using skewness of general tangents of for $V_{2,d}$ and $V_{1,2d}$ in the process. The only thing one needs to watch out for when repeating the scheme of the proof of Lemma 3.3.4 for general degrees $2d$ is that none of the relevant coefficients in the analogous identities of (3.32) and (3.35) accidentally sum up to zero. A day might come where I will write down this proof in full generality. But it is not this day!

For odd degrees, note that a reduction to powers of forms identifiability cannot achieve optimal asymptotics. Indeed, for degree $2d + 1$, the highest powers of the q_i in a tangent space expression occur in the ℓ-derivative and thus look like $q_i^d h_i$, where h_i are *linear* forms rather than quadratics! The spaces $\langle q_1^d h \mid h \in U^\vee\rangle, \ldots, \langle q_m^d h \mid h \in U^\vee\rangle$ might not intersect until $m \in \theta(n^{2d})$ from counting parameters, but tangents of the q_i can only be skew up to $m \in \theta(n^{2d-1})$. While Nenashev's Theorem 3.2.12 applies in much greater generality and can show *something* also for odd degrees, it is not apparent to me that the variable splitting argument can achieve optimal asymptotics for odd degrees. The subsequent section will collect numerical results suggesting that degree 5 is minimal for any meaningful identifiability results. Section 3.3 then examines the case of degree 7 and shows identifiability up to ranks $\approx n^4$, which is roughly a factor of n less than the conjectured optimal range.

Numerical Experiments

As an asymptotically formulated statement, Theorem 3.3.5 might be trivial for the first finitely many values of $n = \dim(U)$. Note that Theorem 3.2.14 was used, which makes nontrivial claims only if there are at least 17 X-variables after splitting as in the proof of Lemma 3.3.4. The first value of n where nontrivial results from this type of proof are obtained thus depends on the splitting strategy, as discussed in Remark 3.3.6. E.g. for even (X, Y)-splitting, one would need at least 33 variables, for uneven splitting likely less. Analyzing such fine-grained aspects of the proof of Theorem 3.3.5 would be cumbersome. Instead, for $d \in \{5, 6\}$, I let a computer check skewness of tangent spaces for parameters (ℓ_i, q_i) sampled with Gaussian random coefficients, $i \in \{1, \ldots, m\}$, where $m = \lfloor \dim S^d(U)/\dim \mathrm{GM}_d(U)\rfloor$ is an upper bound for the maximum value where tangent spaces are skew obtained from counting parameters. The results

d	s_d	$\partial_\ell s_d$ (normalized)	$\partial_q s_d$ (normalized)
1	ℓ	1	0
2	$\ell^2 + q$	ℓ	1
3	$\ell^3 + 3q\ell$	$\ell^2 + q$	ℓ
4	$\ell^4 + 6q\ell^2 + 3q^2$	$\ell^3 + 3q\ell$	$\ell^2 + q$
5	$\ell^5 + 10q\ell^3 + 15q^2\ell$	$\ell^4 + 6q\ell^2 + 3q^2$	$\ell^3 + 3q\ell$
6	$\ell^6 + 15q\ell^4 + 45q^2\ell^2 + 15q^3$	$\ell^5 + 10q\ell^3 + 15q^2\ell$	$\ell^4 + 6q\ell^2 + 3q^2$
7	$\ell^7 + 21q\ell^5 + 105q^2\ell^3 + 105q^3\ell$	$\ell^6 + 15q\ell^4 + 45q^2\ell^2 + 15q^3$	$\ell^5 + 10q\ell^3 + 15q^2\ell$

Table 3.2: The Gaussian moment variety in arbitrary degree d is given by the polynomials $\mathcal{M}_d(\mathcal{N}(\ell, q))$ from Table 3.1 and Remark 3.1.2, which define a map $s_d \colon U^\vee \times S^2(U) \to S^d(U), (\ell, q) \mapsto s_d(\ell, q)$. We can interpret them as bivariate polynomials in variables (ℓ, q). Their partial derivatives with respect to ℓ and q are in the third and fourth column, respectively, after a normalization which cancels out a common integer divisor of all coefficients: d for the ℓ-derivative and $d(d-1)/2$ for the q-derivative. The alert reader might have noticed some diagonal patterns in the table, i.e. $\partial_q s_d = (d/2)\partial_\ell s_{d-1} = \binom{d}{2} s_{d-2}$ for $d \in \{3, \dots, 7\}$.

may be observed in Table 3.3 and indicate that degree 5 is minimal for identifiability beyond rank 1. Table 3.2 contains polynomials giving parametrizations for the Gaussian moment varieties and their tangent spaces. For degrees $d \leq 3$, even the parameters of a *single* Gaussian are not identifiable. For $d \leq 2$, this is trivial for degree reasons. For $d = 3$, the moment form is $s_3(\ell, q) = \ell^3 + 3q\ell = \ell(\ell^2 + 3q)$ and it turns out that the map s_3 is degenerate, i.e. its image has dimension smaller than its domain: Indeed, note that for general (ℓ, q), both $s_3(\ell, q)$ and $s_3(c\ell, \frac{1}{c}q + \frac{1/c - c^2}{3}\ell^2)$ are equal for each $c \in \mathbb{C}^\times$, thus the general preimage contains a curve. (Of course, the parameters of a rank-1 mixture are always identifiable from the moments of degree 1 *and* 2, but our analysis restricts to moments of one, fixed degree).

$\mathrm{GM}_4(U)$ is generically 1-identifiable if $\dim U \geq 3$: Assume $\ell_1^4 + 6q_1\ell_1^2 + q_1^2 = \ell_2^4 + 6q_2\ell_2^2 + q_2^2$. Write $a_i := \ell_i^2 + 11q_i$ and $b_i := \ell_i^2 - 5q_i$ for $i \in \{1, 2\}$ and note that $\ell_i^4 + 6q_i\ell_i^2 + q_i^2 = (a_i + b_i)(a_i - b_i)$. Further assume that ℓ_1, q_1 are general. It suffices to show that $a_1 + b_1$ and $a_1 - b_1$ are prime, as then the statement follows from uniqueness of the prime factorization. Since a reducible quadratic form is a product of two linear forms, only quadratic forms of rank at most 2 can be reducible. Due to generality, $a_1 \pm b_1$ have full rank equal to $\dim U \geq 3$.

Numerical evidence that $\mathrm{GM}_4(U)$ is not 2-identifiable is collected in Table 3.4. In fact, for $\dim(U) \geq 4$ it suggests that the defect of the rank-2 secant is precisely one. There is likely a proof behind this, albeit it might be more difficult than in the centered case. In this work though, I do not want to focus too much on specific nonidentifiable cases. At least qualitatively, nonidentifiability in low degrees should have been apparent already in the age of Pearson [71], even though his original work considered univariate Gaussians and thus inhomogeneous moments. While I did not find a reference for the homogeneous degree-4 case, I expect it to be known. E.g. in Améndola, Ranestad and Sturmfels [2, Theorem 13], nonidentifiability from inhomogeneous degree-3 moments from rank 2 onwards is shown as a theorem.

$\mathrm{GM}_5(\mathbb{C}^n)$**-secants**				$\mathrm{GM}_6(\mathbb{C}^n)$**-secants**			
n	rank	secant dim.	exp. dim.	n	rank	secant dim.	exp. dim.
2	1	5	5	2	1	5	5
3	2	18	18	3	3	27	27
4	4	56	56	4	6	84	84
5	6	120	120	5	10	200	200
6	9	243	243	6	17	459	459
7	13	455	455	7	26	910	910
8	18	792	792	8	39	1716	1716
9	23	1242	1242	9	55	2970	2970
10	30	1950	1950	10	77	5005	5005
11	39	3003	3003	11	104	8008	8008
12	48	4320	4320	12	137	12330	12330
13	59	6136	6136	13	178	18512	18512
14	72	8568	8568	14	228	27132	27132
15	86	11610	11610	15	287	38745	38745
16	102	15504	15504	16	357	54264	54264
17	119	20230	20230	17	438	74460	74460
18	139	26271	26271	18	534	100926	100926
19	161	33649	33649	19	644	134596	134596
20	184	42320	42320				

Table 3.3: Numerical experiments show that for $n = \dim(U) \in \{2, \ldots, 19\}$, the Gaussian moment variety is nondefective up to the maximum possible rank $m = \lfloor \dim S^d(U) / \dim \mathrm{GM}_d(U) \rfloor$ for both $d = 5$ (left) and $d = 6$ (right). The value for $(d, n) = (6, 20)$ is missing since I ran out of memory. The notebook generating the data may be found in the appendix in [88, `identifiability/code/secant-nondefective-pcountrank.ipynb`]

Degree 7 identifiability

In odd degrees, the variable splitting argument does apparently not give the correct asymptotics, but seems to lose a factor of n. The following proof demonstrates this for degree 7. The broad scheme of the proof should be also applicable for higher odd degrees, but not for the minimal degree 5, which is a special case.

3.3.8 Theorem: $\mathrm{GM}_7(\mathbb{C}^n)$ is m-identifiable for some semi-increasing function $m \in \theta(n^4)$.

Proof. Choose split variables $(X, Y) = (X_1, \ldots, X_{n'}, Y_1, \ldots, Y_{n-n'})$ for some $n' \in \{1, \ldots, n\}$. Write $N := n - n'$. Consider $q_1, \ldots, q_m$ general in $\mathbb{C}[X]_2$ and $\ell_1, \ldots, \ell_m$ general in $\mathbb{C}[Y]_1$. Assume that $h_i \in \mathbb{C}[X, Y]_1$ and $p_i \in \mathbb{C}[X, Y]_2$ are such that

$$0 = \sum_{i=1}^{m} (\ell_i^6 + 15 q_i \ell_i^4 + 45 q_i^2 \ell_i^2 + 15 q_i^3) h_i + \sum_{i=1}^{m} (\ell_i^5 + 10 q_i \ell_i^3 + 15 q_i^2 \ell_i) p_i \quad (3.36)$$

Let us split each $h_i = h_i(X, 0) + h_i(0, Y) =: h_{i,X} + h_{i,Y}$ into a part that only depends on X and a part that only depends on Y. Similarly, split $p_i = p_{i,X} + p_i\langle X, Y\rangle + p_{i,Y}$, where $p_i\langle X, Y\rangle$ denotes the part of p_i which is bilinear

n	$\dim \mathrm{GM}_4(U)$	$\dim \sigma_2(\mathrm{GM}_4(U))$	$2(\binom{n+1}{2}+n)$
2	5	5	10
3	9	15	18
4	14	27	28
5	20	39	40
6	27	53	54
7	35	69	70
8	44	87	88
9	54	107	108
10	65	129	130

Table 3.4: Table showing the dimensions of $\mathrm{GM}_4(U)$ and its second secant variety, where $n = \dim U$. The third column contains the expected dimension of $\sigma_2(\mathrm{GM}_4(U))$. The table collects numerical evidence that $\mathrm{GM}_4(U)$ is m-defective for $m \geq 2$. For $n \geq 4$ and $m = 2$, it appears that the defect is precisely one. Code generating the data may be found in the notebook [88, `appendices/identifiability/code/secant-defective-deg4.ipynb`].

in (X, Y). Let us split (3.36) into the parts of fixed (X, Y)-degree and look first at the part of degree 7 in X, which is

$$0 = 15 \sum_{i=1}^{m} q_i^3 h_{i,X} \tag{3.37}$$

By Nenashev's Theorem 3.2.12 applied on the ideal $(q_1, \ldots, q_m) \subseteq \mathbb{C}[X]$, such expressions are skew even up to $m \approx n'^6$. Thus we conclude that $h_{i,X} = 0$ for all i. Now look at the terms of degree 7 in Y and obtain

$$0 = \sum_{i=1}^{m} \ell_i^5 (\ell_i h_{i,Y} + p_{i,Y}) \tag{3.38}$$

Nenashev's Theorem 3.2.12 yields that $p_{i,Y} = -\ell_i h_{i,Y}$ for all i as long as $m \in \mathcal{O}(n^5)$, since $\mathcal{D} := \{\ell^5 \mid \ell \in \mathbb{C}[Y]_1\}$ is a class closed under linear transformations of the Y-variables. Since all $h_{i,X} = 0$, the term of degree 6 in Y has only one contribution

$$0 = \sum_{i=1}^{m} \ell_i^5 p_i \langle X, Y \rangle \tag{3.39}$$

Plugging in any $x \in \mathbb{C}^n$ yields

$$0 = \sum_{i=1}^{m} \ell_i^5 p_i \langle x, Y \rangle \tag{3.40}$$

Apply now Nenashev's Theorem 3.2.12 to the ideal $(\ell_1^5, \ldots, \ell_m^5) \subseteq \mathbb{C}[Y]$ in degree 6, (which we can for some $m \in \theta(N^5)$) to conclude that $p_i\langle x, Y \rangle = 0$ for all i. Since x was arbitrary, this implies that all $p_i\langle X, Y \rangle = 0$. Continue with the term of degree 6 in X:

$$0 = \sum_{i=1}^{m} q_i^2 (q_i h_{i,Y} + \ell_i p_{i,X}) \tag{3.41}$$

As long as $m \in \mathcal{O}(n'^4)$, evaluating in some point y yields that $q_i h_{i,Y}(y) + \ell_i(y) p_{i,X} = 0$, which is again by 3.2.12 applied to the degree-6 component of $(q_1^2, \dots, q_m^2)$. Note that we loose an asymptotically nonnegligible factor in this step. Since y was arbitrary, we conclude $q_i h_{i,Y} = -\ell_i p_{i,X}$ for all i. The term of degree 5 in X is trivial, but the term of degree 5 in Y yields

$$0 = \sum_{i=1}^{m} (15 q_i \ell_i^4 h_{i,Y} + \ell_i^5 p_{i,X} + 10 q_i \ell_i^3 p_{i,Y}) \tag{3.42}$$

which by the previously shown can be simplified to

$$0 = \sum_{i=1}^{m} (5 q_i \ell_i^4 h_{i,Y} + \ell_i^5 p_{i,X}) \tag{3.43}$$

and then again to

$$0 = -4 \sum_{i=1}^{m} \ell_i^5 p_{i,X} \tag{3.44}$$

Evaluating in some $x \in \mathbb{C}^n$ yields a vanishing linear combination $\sum_{i=1}^m p_{i,X}(x) \ell_i^5 = 0$, and thus $p_{i,X}(x) = 0$ for all i, since $\ell_1^5, \dots, \ell_m^5$ are linearly independent for $m \in \mathcal{O}(N^5)$. Since x was arbitrary, this implies that all $p_{i,X}$ are zero. Due to $q_i h_{i,Y} = -\ell_i p_{i,X} = 0$ thus also $h_{i,Y} = 0$ and $p_{i,Y} = -\ell_i h_{i,Y} = 0$. This concludes the proof. As for the optimal variable splitting, note that the tightest bound on m stems from (3.41) and is $m \in \mathcal{O}(n'^4)$. Thus it is reasonable to put most variables into X, e.g. choose $N \approx c_n n$ and $n' \approx (1 - c_n) n$ for $c_n = \sqrt[5]{1/n}$. However, note that in this case, getting a good constant is actually not too important since even the asymptotics itself are not optimal. □

3.3.9 Remark: Theorem 3.3.5 and Theorem 3.3.8 directly imply that the parameters of a uniformly weighted mixture of general Gaussians are uniquely determined both by their moments of degree 6 or 7, in their respective ranges of ranks. For non-uniformly weighted mixtures of Gaussians, moment forms of two different degrees are necessary, as otherwise the problem is overparameterized. This is just a slight bit more complicated than Theorem 3.1.3, but essentially the same argument: One starts with two potentially different sets of parameters and weights which have the same moments. The moments of highest degree are used to conclude that the parameters must be equal up to scalar multiples. Thus the identity of moments of degree one less (or two less) gives a linear combination of general moments of Gaussians, which are linearly independent, and then a comparison of coefficients shows that the parameters must be equal and the weights must be the same on both decompositions. Working out the details is left to the reader.

3.4. An Algorithm for Sums of Powers

After giving a thorough answer to Question 3.0.1(a) in Section 3.2, let me attempt to also give a partial answer to Question 3.0.1(b). This section will develop semidefinite algorithms to decompose powers of forms and thus also mixtures of *centered* Gaussians. At the heart of both is the following:

3.4.1 Theorem: For any $n \in \mathbb{N}$, $m \in \{1, \ldots, n-2\}$, there is an efficient algorithm for the following problem: If $q_1, \ldots, q_m \in S^k(\mathbb{R}^n)$ are a regular sequence of forms such that $I := (q_1, \ldots, q_m)$ is radical and $V(I)$ has dense real points, compute the set $\{q_1, \ldots, q_m\}$ from inputs $q_1^2 + \ldots + q_m^2$ and $q_1^3 + \ldots + q_m^3$.

The conditions of Theorem 3.4.1 are satisfied on a typical subset of $S^k(\mathbb{R}^n)^m$, which appears to be quite large, cf. Remark 3.4.15. Except for the density-of-real-points assumption, all conditions of Theorem 3.4.1 hold for general choices of $q_1, \ldots, q_m$. While the algorithm is mainly intended for general power sum decomposition instances, it does not hurt to protocol the nondegeneracy conditions on $q_1, \ldots, q_m$ explicitly, since they are relatively easy to describe. If the regular sequence assumption is replaced by the weaker condition that none of the $q_1, \ldots, q_m$ are redundant to describe $V(I)$, then it is still possible to prove a recovery theorem which also makes use of the fourth power sum. This is done in the last subsection. As Gaussian mixtures are quite the active area of research, let me briefly review part of what is known on the side of provably efficient algorithms, before developing the original contributions.

Related work In 2015, Ge, Huang and Kakade [37] gave an algorithm that can decompose a mixture of n-variate Gaussians of rank $\Omega(1) \leq m \leq \mathcal{O}(\sqrt{n})$, not necessarily centered, in polynomial time (in the dimension n) in a smoothed analysis model, where the input are polynomially many samples from the mixture. Up to noise from estimation, their input is equivalent to the degree-6 moments of the mixture. Several years after, Garg, Kayal and Saha [36] studied powers-of-forms decompositions as in 3.0.1 for general k, d from the perspective of circuit complexity via the so-called affine projections of partials method and simultaneous decompositions of certain vector spaces. They derived algorithms from rank lower bound techniques where the maximum rank scaled with the order of the power.

In terms of techniques, a crucial step of Ge, Huang and Kakade's algorithm was to estimate the spaces spanned by the mean vectors and the covariances via repeated projections to lower-dimensional spaces. Subsequently, the authors tried to algorithmically invert the symmetrization operation $S^3(S^2(\mathbb{R}^n)) \to S^6(\mathbb{R}^n), q^{\otimes 3} \mapsto q^3$ on low-rank mixtures. Estimating the span of the covariances is also one of the key steps in the algorithm presented here as well as in a yet unpublished algorithm mentioned below due to Bafna, Hsieh, Kothari and Xu [5]. By "$\Omega(1) \leq m$" the reader is supposed to understand that [37] requires the rank of the mixture to be at least some constant $c \in \mathbb{N}$. The algorithm of Garg, Kayal and Saha [36] is motivated from lower-bound techniques in circuit complexity and can decompose sums of d-th powers of k-forms of rank $m \in n^{\frac{d}{1100k}}$ and sufficiently large n. Note that their method yields interesting results only if d is *very large* in relation to k.

From private discussion with Kothari, I know that there is an upcoming result [5] that can decompose sums of 3rd powers of quadratics[4] in the range $m \in \mathcal{O}(\frac{n}{\log(n)^2})$. While the authors initially use some distributional assumptions (i.e. that the covariance forms are chosen with normally distributed coefficients), the results apparently can be leveraged to a framework that resembles smoothed analysis.[5]

To guarantee relevancy for Gaussian mixtures, the authors use an argument that allows to pass from an average-case analysis to a smoothed setting. In contrast, I will face a similar problem in a non-probabilistic setting and guarantee relevancy for Gaussian mixtures via a substitution trick. However, in my case, this does not yield an algorithm for *general* choices of $q_1, \ldots, q_m \in S^2(\mathbb{R}^n)$ but rather only for *typical* choices. Thus, all four results are incomparable in the sense that none is a strict improvement of another: For degree-6 moments, whenever $q_1, \ldots, q_m$ have a common level set, the present algorithm is strictly better than any of the other three. It also has comparatively modest and explicit "nondegeneracy conditions". Ge, Huang and Kakade's [37] algorithm is still the only one that can deal with general *non-centered* Gaussians, albeit only in low (but not too low) rank. Bafna, Hsieh, Kothari and Xu [5] obtain better thresholds on the rank when decomposing sums of higher powers of forms, as long as the power is divisible by 3. Garg, Kayal and Saha's algorithm is the only of the four that works for *general* fields rather than the real field and does not have a requirement on the powers to be divisible by three. Also, Garg, Kayal and Saha use a different notion of nondegeneracy compared to [37] and [5] who encode nondegeneracy probabilistically by assuming random or smoothed inputs.

Another conceptual difference is that [37] and [5] are written from the statistical perspective, whereas [36] and the present work are written from an algebraic perspective. However, as all four algorithms are moment-based, the main difference is the lack of a noise analysis in [36] and here. For [36] this limits the applicability for statistical problems, as from private discussion I heard that the algorithm therein can only tolerate noise which is exponentially small in the dimension. For the present work, we remark that the only critical part of the procedure that could potentially be vulnerable to noise is the space recovery from Proposition 3.4.13.

One last difference I want to point out is that the algorithms derived in this section could potentially also work for certain quadratics of fixed rank at least 3, whereas the probabilistic framework of [37] and [5] implicitly rules out any quadratics that are not of full rank. If $q_1, \ldots, q_m$ are chosen as some minimal set of (rank-4) quadratic equations of some Veronese variety, then clearly all assumption but the regular sequence are fulfilled, so we might apply the fourth-powers algorithm from the last subsection. For general quadratics of fixed rank, the key thing to understand seems to be whether for an intersection of irreducible rank-k quadrics (where $k \geq 3$), the sum of their vanishing ideals is radical.

Relevance of contributions In a typical real case, I improve the rank threshold from $m \leq \mathcal{O}(\frac{n}{\log(n)^2})$ ([5]) or $\Omega(1) \leq m \leq \mathcal{O}(\sqrt{n})$ ([37]) to $1 \leq m \leq n-2$. Compared to the previous results, this gives an improvement of the asymptotic order *and* the constant factors, with a much simpler proof. On the contrary,

[4] And some other cases of powers of forms, which may have different ranges.

[5] Note that for a choice of Gaussian-random quadratic forms $q_1, \ldots, q_m$, the probability that all of them are positive definite will quickly decrease to zero as m increases.

the present result is unlikely to give better bounds with higher powers as inputs (which [5] does) and it is not clear whether it can be adapted to noncentered Gaussians just as [37].

The algebra generated by general quadratics

I will show that there is an algorithm that, whenever there are $m \in \{1, \dots, n-2\}$ forms $q_1, \dots, q_m \in S^k(\mathbb{R}^n)$ general with the property that $V_{\mathbb{R}}(q_1, \dots, q_m)$ is nonempty and $q_1^d + \dots + q_m^d$ is given as input for both values $d \in \{2, 3\}$, computes $\{q_1, \dots, q_m\}$. The basic idea is to first use sum-of-squares optimization to recover the space $\langle q_1, \dots, q_m \rangle$, then use a basis of this space to construct an algebra isomorphism that allows a reduction to undercomplete symmetric tensor decomposition, for which efficient and noise-stable algorithms are known.

Let us first discuss the latter step: It turns out that for $m \leq n$ and general choices of $q_1, \dots, q_m \in S^k(\mathbb{C}^n)$, if we are given a basis $u_1, \dots, u_m \in S^k(\mathbb{C}^n)$ for $\langle q_1, \dots, q_m \rangle$ then we can reduce to Waring decomposition: Any such basis gives algorithmic access to the evaluation map

$$\varphi \colon \mathbb{R}[X_1, \dots, X_m] \to \mathbb{R}[q_1, \dots, q_m], X_1 \mapsto q_1, \dots, X_m \mapsto q_m \tag{3.45}$$

which is clearly surjective. Now, since we are in a low-rank setting, this will turn out to be an *isomorphism* of graded $\mathbb{R}$-algebras, where $\langle q_1, \dots, q_m \rangle$ is declared to be the grade-1 component of $\mathbb{R}[q_1, \dots, q_m]$. Thus there exists the inverse isomorphism φ^{-1}. As $\varphi|_{\mathbb{R}[X]_{\leq 3}}$ is a linear map between $\mathbb{R}$-vector spaces, the inverse map, which is $\varphi^{-1}|_{\mathbb{R}[X]_{\leq 3k}}$, can be computed by inverting a linear system. Writing $\ell_i := \varphi^{-1}(q_i)$ for $i \in \{1, \dots, m\}$, note that φ^{-1} maps the input $\sum_{i=1}^m q_i^3$ to $\sum_{i=1}^m \ell_i^3$ from which we proceed as in one of the proofs of Theorem 2.4.8. The same argument works under the more explicit condition that $q_1, \dots, q_m$ form a *regular sequence*. In particular, this allows to also choose $q_1, \dots, q_m$ as general elements from a special class $\mathcal{D}$ of forms, e.g. one may take $\mathcal{D}$ as the class of k-forms of fixed Waring rank. The details of the reduction are elaborated in the final part of this section. For now, let us collect basic properties about the algebra $\mathbb{R}[q_1, \dots, q_m]$.

3.4.2 Definition: A closed subvariety V of $\mathbb{P}(\mathbb{C}^n)$ is called a *complete intersection*, if the homogeneous vanishing ideal $I(V)$ of V in $\mathbb{C}[X_1, \dots, X_n]$ can be generated by $\operatorname{codim}(V)$ elements.

3.4.3 Proposition: (Corollary of Bertini's theorem, cf. [44, II 8.4(d)]) Let the closed subvariety $V \subseteq \mathbb{P}(\mathbb{C}^n)$ be a complete intersection and let $\operatorname{codim}(V) \leq n - 2$. Then V is irreducible and nonsingular.

3.4.4 Proposition: Let $K \in \{\mathbb{R}, \mathbb{C}\}$ and $m, n, k \in \mathbb{N}_0$ such that $m \leq n$. Let $\mathcal{D} \subseteq S^k(K^n)$ an irreducible variety containing a regular sequence of length m and let $q_1, \dots, q_m$ general in $\mathcal{D}$, then the algebra $K[q_1, \dots, q_m]$ is isomorphic to the polynomial algebra $K[X_1, \dots, X_m]$ in m variables.

Proof. We have to show that there are no algebraic relation between $q_1, \dots, q_m$. To this end, let $d \in \mathbb{N}$ and $N := \{\alpha \in \mathbb{N}_0^m \mid |\alpha| = d\}$. We show that there are algebraic relations in degree kd over $K = \mathbb{C}$, which clearly also shows the claim over $K = \mathbb{R}$. Consider the variety

$$W_{m,d} = \{(\lambda, q) \in \mathbb{C}^N \times S^k(\mathbb{C}^n)^m \mid \sum_{\alpha \in \mathbb{N}_0^m, |\alpha|=d} \lambda_\alpha q^\alpha = 0\} \tag{3.46}$$

consisting of pairs of $(q_1, \ldots, q_m)$ and the coefficients λ of relations in between them. Let π denote the projection to the q-coordinates and consider any tuple of forms $q = (q_1, \ldots, q_m)$. Then the fiber $\pi^{-1}(\{q\})$ is a linear subspace of $\mathbb{C}^N$ corresponding to the algebraic relations of $q_1, \ldots, q_m$ in degree d. By semicontinuity of the fiber dimension (2.2.15), the dimension of the space $\pi^{-1}(\{q\})$ can only go up in the limit, so it suffices to find a specific instance of $q_1, \ldots, q_m$ where it is zero. But this is the case for any regular sequence and $\mathcal{D}$ contains a regular sequence by assumption. Since $\mathcal{D}$ is irreducible, it follows that *the* general element of $\mathcal{D}^m$ is a regular sequence. □

3.4.5 Remark: Proposition 3.4.4 applies e.g. to the irreducible classes $\mathcal{D}_{k,r}$ of k-forms of Waring rank at most r: Indeed, for any $k, r \in \mathbb{N}$, $X_1^k, \ldots, X_m^k \in \mathcal{D}_{k,r}$ is a regular sequence, as $K[X_1^k, \ldots, X_m^k]$ is trivially a polynomial ring.

3.4.6 Remark: For $m = n + 1$, $K \in \{\mathbb{R}, \mathbb{C}\}$ and general $q_1, \ldots, q_m \in S^k(K^n)$, $K[q_1, \ldots, q_m]$ is not a polynomial ring, as $I_{\text{rel}}(q_1, \ldots, q_m) = (f)$, i.e. the ideal of relations is a principal ideal generated by only one equation $f \in K[Z_1, \ldots, Z_m]$ in unknowns $Z = (Z_1, \ldots, Z_m)$. Indeed, one can easily see that the polynomial map

$$\underline{q}\colon K^n \to K^m, x \mapsto \begin{pmatrix} q_1(x) \\ \vdots \\ q_m(x) \end{pmatrix} \tag{3.47}$$

has n-dimensional image (e.g. by observing the Jacobian $J\underline{q}$ at a general point). The degree of f is the degree of the variety $V(f) \subseteq \mathbb{C}^m$, which is determined by intersecting $V(f)$ with a general subspace H of dimension 1. Denote by $\mathcal{L}$ the n-dimensional space of linear equations defining H. Choose a basis $\mathcal{L} = \langle \ell_1, \ldots, \ell_n \rangle$ where $\ell_1, \ldots, \ell_n \in \mathbb{C}[Z]_1$. Pulling back via $\underline{q}$ yields a system of n quadratic forms $\ell_i(q_1, \ldots, q_m) \in \mathbb{C}[X]_2$, $i \in \{1, \ldots, n\}$. By Bézout's theorem and genericity of H and $q_1, \ldots, q_m$, we obtain that this system has k^n (complex) solutions, so $\deg(f) = k^n$. For all $n \geq 2, k \geq 2$, this means that there exist no relations in between $q_1, \ldots, q_m$ of degree at most 3. Hence, the evaluation map $X_i \mapsto q_i, i \in \{1, \ldots, m\}$ still gives an isomorphism of the graded vector spaces $K[Z]_{\leq 3}$ and $K[q_1, \ldots, q_m]_{\leq 3}$.

This argument would allow to reduce to a Waring decomposition problem, given a basis of $\langle q_1, \ldots, q_m \rangle$. However, note that then we would not necessarily know how to solve for the Waring decomposition, as the assumption of Theorem 2.4.8 would be violated. Let us not give this case too much attention, since for the present algorithm to recover the space $\langle q_1, \ldots, q_m \rangle$, it does not appear that there are large typical neighbourhoods where it succeeds for $m > n$, cf. Remark 3.4.15.

Space recovery

The next step is to find a basis $u_1, \ldots, u_m$ of the space $\langle q_1, \ldots, q_m \rangle$. We will develop an algorithm, based on Semidefinite Programming, that, given $\sum_{i=1}^m q_i^2$ as input, succeeds in recovering the space for a typical subset of $S^k(\mathbb{R}^n)^m$, as long as $m \leq n-2$. The basic idea is that any functional E on $S^k(\mathbb{R}^n)$ that maps squares to something nonnegative satisfies

$$E(\sum_{i=1}^m q_i^2) = 0 \implies \forall i \in \{1, \ldots, m\} : \forall h \in S^k(\mathbb{R}^n) : E(q_i h) = 0 \qquad (3.48)$$

Therefore, such functionals are forced to vanish on the space $(q_1, \ldots, q_m)_{2k}$ (cf. 3.4.7) by just writing down a single equation in terms of the input. We will of course give formal proofs, but it is not too hard to intuitively accept that knowing the space of *all* functionals satisfying the right hand side of (3.48) is, up to Linear Algebra, equivalent to knowing the space $\langle q_1, \ldots, q_m \rangle$. However, of course there is a catch: It is not at all clear that we find enough linearly independent square-definite functionals E to span the entire conormal space $(q_1, \ldots, q_m)_{2k}^{\perp}$. In fact, the set of square-definite functionals vanishing on $\sum_{i=1}^m q_i^2$ might be trivial. E.g. for $k = 2$ this is the case if any of the q_i is positive definite. Say e.g. $q_1 \succeq c(X_1^2 + \ldots + X_n^2)$ for some $c \in \mathbb{R}_{>0}$. Then any such square-definite E satisfies $0 = E(\sum_{i=1}^m q_i^2) \geq c^2 E((X_1^2 + \ldots + X_n^2)^2) \geq 0$. A short exercise shows that $E[X_i^2 X_j^2] = 0$ for all $i, j \in \{1, \ldots, n\}$ implies $E = 0$.

We will therefore often use assumptions implying that $V(q_1, \ldots, q_m) \subseteq \mathbb{P}(\mathbb{C}^n)$ contains a real point. At first glance, this assumption appears to be completely unsuited for mixtures of centered Gaussians, as there all the quadratic forms are psd. However, a separate observation will allow us to perform the substitution $q_i' = q_i - \lambda Y$ for a new variable Y and a scalar λ of our choice. With that substitution in mind, we might alternatively read the common real zero condition as "$q_1, \ldots, q_m$ have a nonempty common level set".

3.4.7 Notation: Let R a commutative, graded $\mathbb{R}$-algebra with graded components $R_0, R_1, R_2, \ldots$. In this section, for $k \in \mathbb{N}_0$, we denote by

$$\Sigma_{R,2k} = \left\{ f \in R_{2k} \mid \exists N \in \mathbb{N}_0, p_1, \ldots, p_N \in R_k : f = \sum_{i=1}^N p_i^2 \right\} \qquad (3.49)$$

the *homogeneous Sums-of-Squares cone of* R in degree $2k$. If $R = \mathbb{R}[X]$ is the polynomial ring, we simply write Σ_{2k}, completely suppressing the dependency on the variables. For an ideal $I \subseteq R$, we denote by I_k the degree-k component of I, i.e. $I_k = I \cap R_k$.

3.4.8 Proposition: A subvariety $V \subseteq \mathbb{P}(\mathbb{C}^n)$ has dense real points $V_{\mathbb{R}} \subseteq \mathbb{P}(\mathbb{R}^n)$ if and only if every irreducible component of V contains a real point that is smooth in V. In particular, an irreducible nonsingular subvariety $V \subseteq \mathbb{P}(\mathbb{C}^n)$ has dense real points $V_{\mathbb{R}}$ if and only if it contains a real point.

3.4.9 Proposition: (Prop. 2.5. in [11]). Let $k, n \in \mathbb{N}$ and $V \subseteq \mathbb{P}(\mathbb{R}^n)$ a real projective subscheme with $\mathbb{Z}$-graded coordinate ring R such that for each $g \in R_k \setminus \{0\}$ the multiplication map

$$\eta_g \colon R_k \to R_{2k}, f \mapsto fg \qquad (3.50)$$

is injective. Then the following are equivalent:

(a) The cone $\Sigma_{R,2k}$ is *pointed*, i.e. it is closed and contains no lines.

(b) No nontrivial sum of squares of forms of degree k equals zero in R_{2k}.

3.4.10 Lemma: Let $k, m, n \in \mathbb{N}$ and $q_1, \ldots, q_m \in \mathbb{R}[X]_k$. Assume $(q_1, \ldots, q_m)$ is a prime ideal and $V(q_1, \ldots, q_m) \subseteq \mathbb{P}(\mathbb{C}^n)$ has dense real points. Let $R := \mathbb{R}[X]/(q_1, \ldots, q_m)$ with the grading inherited from $\mathbb{R}[X]$. Then the cone $\Sigma_{R,2k}$ is pointed.

Proof. First, observe that the map η from Proposition 3.4.9 is injective. Write $A := \mathbb{R}[X]$ and $I := (q_1, \ldots, q_m)$. Thus, we have to show there are no $g \in A_k \setminus I_k$ such that there exists $f \in A_k \setminus I_k$ with $gf \in I_{2k}$. That is clearly the case since I is even a prime ideal by assumption.

We now employ Proposition 3.4.9(b) $\Longrightarrow$ (a): Let $N \in \mathbb{N}_0$ and $s_1, \ldots, s_N \in \mathbb{R}[X]_k$ such that $\sum_{i=1}^N s_i^2 \in I_{2k}$. We have to show that $s_1, \ldots, s_N \in I_k$. For all $x \in V_{\mathbb{R}}(I)$, it holds that $0 = \sum_{i=1}^N s_i(x)^2$ and therefore $s_1, \ldots, s_N$ vanish on $V_{\mathbb{R}}(I)$. As $V_{\mathbb{R}}(I)$ is dense in $V(I)$, we conclude $s_1, \ldots, s_N \in I$, since I is a prime ideal and therefore radical. Hence $\Sigma_{R,2k}$ is pointed. □

The previous lemma simplifies by a lot when considering general forms:

3.4.11 Lemma: Let $k, m, n \in \mathbb{N}$ with $m \leq n-2$ and $q_1, \ldots, q_m \in \mathbb{R}[X]_k$ general. Let $R := \mathbb{R}[X]/(q_1, \ldots, q_m)$ with the grading inherited from $\mathbb{R}[X]$. If $\emptyset \neq V_{\mathbb{R}}(q_1, \ldots, q_m) \subseteq \mathbb{P}(\mathbb{R}^n)$, then the cone $\Sigma_{R,2k}$ is pointed.

Proof. For $n \leq 2$, it must be $m = 0$ and thus the statement is that $\Sigma_{\mathbb{R}[X],2k}$ is pointed, which is well-known. Thus wlog $n \geq 3$. Then the forms $q_1, \ldots, q_m$ are irreducible by generality. By Bertini's Theorem (cf. 3.4.3), the (scheme theoretic) intersection of their varieties is smooth and irreducible. Thus $I := (q_1, \ldots, q_m)$ is even a prime ideal. By Proposition 3.4.8, $V(I)$ has dense real points since $V(I)$ is smooth and contains a real point. Therefore, the assumptions of Lemma 3.4.10 are fulfilled. We conclude that the cone $\Sigma_{R,2k}$ is pointed. □

Similar to what we did in Lemma 3.4.11, the assumptions of the following Lemma 3.4.12 may also be replaced by "m at most $n-2$, $q_1, \ldots, q_m$ general and $V_{\mathbb{R}}(q_1, \ldots, q_m)$ nonempty as a projective set."

3.4.12 Lemma: Let $k, m, n \in \mathbb{N}$ and $q_1, \ldots, q_m \in \mathbb{R}[X]_k$. Assume $(q_1, \ldots, q_m)$ is a prime ideal and $V(q_1, \ldots, q_m) \subseteq \mathbb{P}(\mathbb{C}^n)$ has dense real points. Let $\mathcal{L} = (q_1, \ldots, q_m)_{2k}^{\perp} \subseteq \mathbb{R}[X]_{2k}^{\vee}$. Then $\mathcal{L}$ is spanned by $\mathcal{L} \cap \Sigma_{2k}^*$.

Proof. We have

$$\mathcal{L} = \{L \in \mathbb{R}[X]_{2k}^{\vee} \mid \forall h \in \mathbb{R}[X]_k : L(hq_1) = \ldots = L(hq_m) = 0\} \tag{3.51}$$

The canonical epimorphism $\pi \colon \mathbb{R}[X] \to R$ yields a linear map

$$\pi^* \colon R_{2k}^{\vee} \to \mathcal{L} \subseteq \mathbb{R}[X]_{2k}^{\vee}, L \mapsto L \circ \pi|_{\mathbb{R}[X]_{2k}} \tag{3.52}$$

It is a very easy exercise to see that this map is injective and surjects to $\mathcal{L}$. Thus, via the isomorphy $R_{2k}^{\vee} \cong \mathcal{L}$, it suffices to show that $R_{2k}^{\vee}$ is spanned by

$(\pi^*)^{-1}(\Sigma_{2k}^*) = \Sigma_{R,2k}^*$. Thus we have to show that $\Sigma_{R,2k}^*$ is not contained in a proper subspace of $R_{2k}^\vee$. To the contrary, assume $\Sigma_{2k}^* \subseteq H \subset R_{2k}^\vee$ for some proper subspace H. Recall that Σ_{2k} is pointed by Lemma 3.4.11. By biduality then $H^\perp \subseteq \Sigma_{2k}^{**} = \Sigma_{2k}$ (where we see $H^\perp = \{p \in R_{2k} \mid \forall L \in H : L(p) = 0\}$ as a subspace of R_{2k}). We used that Σ_{2k} is closed. But then $H^\perp = \{0\}$ must be trivial, as Σ_{2k} and thus H must not contain a line. □

3.4.13 Proposition: (Space recovery) Let $k, m, n \in \mathbb{N}$. Assume $q_1, \ldots, q_m \in \mathbb{R}[X]_k$ are such that $(q_1, \ldots, q_m)$ is a prime ideal and $V(q_1, \ldots, q_m) \subseteq \mathbb{P}(\mathbb{C}^n)$ has dense real points. Then, from input $\sum_{i=1}^m q_i^2$, we may compute a basis of the space $\langle q_1, \ldots, q_m \rangle$ via Semidefinite Programming and Linear Algebra.

Algorithmic Proof. We may instead compute a generating system for the conormal space

$$\langle q_1, \ldots, q_m \rangle^\perp = \{L \in S^k(\mathbb{R}^n)^\vee \mid L(q_1) = \ldots = L(q_m) = 0\}$$

as this will give a linear system to compute $\langle q_1, \ldots, q_m \rangle$. Now, let us start with the algorithm: Let $\sum_{i=1}^m q_i^2$ be given as input. By Semidefinite Programming, we may compute a maximal linear independent set $\mathcal{E} \subseteq \Sigma_{2k}^* \subseteq \mathbb{R}[X]_{2k}^\vee$ of functionals E satisfying $E(\sum_{i=1}^m q_i^2) = 0$. Denote $\mathcal{L} := (q_1, \ldots, q_m)_{2k}^\perp \subseteq \mathbb{R}[X]_{2k}^\vee$. By Cauchy-Schwarz, for any $E \in \Sigma_{2k}^*$ it holds that

$$E(\sum_{i=1}^m q_i^2) = 0 \iff \forall i \in \{1, \ldots, m\} : \forall h \in S^k(\mathbb{R}^n) : E(q_i h) = 0 \iff E \in \mathcal{L}$$

Hence we computed a maximal linear independent set in $\mathcal{L} \cap \Sigma_{2k}^*$. $\mathcal{E}$ is thus a basis of $\mathcal{L}$ by Lemma 3.4.12. Now, fix some nonzero $h \in S^k(\mathbb{R}^n)$ (e.g. $h = X_1^k + \ldots + X_n^k$) with respect to which we define the linear map

$$S^{2k}(\mathbb{R}^n)^\vee \to S^k(\mathbb{R}^n)^\vee, L \mapsto \widehat{L} := \begin{pmatrix} S^k(\mathbb{R}^n) \to \mathbb{R} \\ p \mapsto L(ph) \end{pmatrix} \tag{3.53}$$

The reader may observe that the map (3.53) is linear and surjective. Now, apply the map (3.53) to $\mathcal{E}$ to obtain a generating system of $\langle q_1, \ldots, q_m \rangle^\perp$, which gives a linear system to compute a basis for $\langle q_1, \ldots, q_m \rangle$. □

3.4.14 Remark: The asymptotically dominating part of the complexity of the procedure in Proposition 3.4.13 is to solve at most $\dim S^k(\mathbb{R}^n)$-many Sums-of-Squares programs in n variables and degree $2k$. However, if an interior point solver is used, which is strongly suggested, then it actually suffices to solve *one* Sums-of-Squares program in n variables and of degree $2k$. Without an objective function, interior point methods for conic programs will converge to a point in the relative interior of the feasible set. Some converge e.g. to the so-called *analytic center* (cf. [76, 2.3.2.]) of the cone. For each $E \in \Sigma_{2k}^* \cap (q_1, \ldots, q_m)_{2k}^\perp$, the (right-)kernel of the symmetric bilinear moment map $M_E \colon (p, q) \mapsto E(pq)$ is a superspace of $\langle q_1, \ldots, q_m \rangle$. The kernel of M_E is the same among all relative interior points E, as if some cone element E vanishes on the square of some form q while other cone elements do not, then the equation $E(q^2) = 0$ defines a proper face of the cone and E cannot be an interior point. The space $U_q := \ker M_E$ does therefore not depend on the choice of the interior point E. The "good" case where we can recover $\langle q_1, \ldots, q_m \rangle$ is precisely when it is equal to U_q, i.e. if M_E has precisely m zero Eigenvalues.

3.4.15 Remark: Numerical experiments obtained from sampling Gaussian random trace-free quadratics $q_1, \dots, q_m$ in small dimensions n suggest that for $k = 2$ and $m \geq n + 1$, the Sum-of-Squares cone might never be pointed on a typical set of quadratics. This can be seen as some indication that the given analysis could be almost tight. Note that it is not at all unexpected that one finds typical instances for the case $m = n - 1$. However, that case is special since the Bertini theorem then only guarantees that $(q_1, \dots, q_m)$ form a radical ideal but not necessarily a prime ideal. The condition on the real variety would need to be sharpened for that case and also injectivity of the multiplication map from Proposition 3.4.9 is not necessarily given when $V(q_1, \dots, q_m)$ has several irreducible components, whence it is not clear whether the equivalence from Proposition 3.4.9 holds true. However, note that for $(n, m) = (7, 6)$ and a natural parameter distribution, the probability that the cone is pointed appears to be around 98.5%: This value was obtained from sampling 200 instances, each of which consisted of 6 iid trace-free quadratics in 7 variables. The data may be found under the name `pointedsos-uc-200series-1.csv` in `appendices/pointedsos/data` in the appendix [88]. For another 200 instance series `pointedsos-oc-200series-2.csv` with $(n, m) = (7, 7)$, the empirical probability to encounter a pointed sos instance was merely 2%. This is a drop of 96.5%! For $(n, m) = (7, 8)$, yet another series `pointedsos-oc-800series-3.csv` of even *800* instances found *none* with a pointed sos cone. This suggests to believe that there exist none, but it might also be that there is just another very sharp probability drop from $m = n$ to $m = n+1$. I encourage the reader to look into the Julia notebook at `appendices/pointedsos/code` [88] to form an opinion on their own. A trace free quadratic is sampled by choosing some symmetric matrix A whose upper triagonal has iid Gaussian random coefficients and then subtracting $\mathrm{tr}(A)I_n$. This sampling procedure avoids accidentally choosing a form which is positive or negative definite. I wish to thank Greg Blekherman and João Gouveia for hinting me towards trace-free matrices. `MOSEK` [66] was used to solve the SDPs.

Extraction from third and second powers

3.4.16 Proposition: Let $k, m, n \in \mathbb{N}_0$ such that $q_1, \dots, q_m \in S^k(\mathbb{R}^n)$ is a regular sequence of k-forms, $I := (q_1, \dots, q_m)$ is prime and $V(I)$ has dense real points. Then from input $\sum_{i=1}^m q_i^2$, we can compute $u_1, \dots, u_m \in S^k(\mathbb{R}^n)$ which span the space $\langle q_1, \dots, q_m \rangle$ and such that the evaluation map

$$\mathbb{R}[X_1, \dots, X_m] \to \mathbb{R}[q_1, \dots, q_m], X_1 \mapsto u_1, \dots, X_m \mapsto u_m \tag{3.54}$$

is an isomorphism of graded $\mathbb{R}$-algebras.

Proof. By Proposition 3.4.13, we may compute a basis $u_1, \dots, u_m$ for the space $\langle q_1, \dots, q_m \rangle$. The evaluation map from (3.54) then indeed has $\mathbb{R}[u_1, \dots, u_m] = \mathbb{R}[q_1, \dots, q_m]$ as its image. The kernel of this map is trivial, as any algebraic relation between $u_1, \dots, u_m$ would yield an algebraic relation between $q_1, \dots, q_m$, of which there are none by assumption. □

The main theorem is formulated for general $q_1, \dots, q_m$, but the nondegeneracy conditions are the same as in Proposition 3.4.16. For general forms, also the condition on the real variety simplifies via Proposition 3.4.3 and Proposition 3.4.8:

3.4.17 Theorem: Let $k, m, n \in \mathbb{N}$ such that $m \leq n-2$. There exists an efficient algorithm that solves the following problem using only Linear Algebra and Semidefinite Programming: If $q_1, \ldots, q_m \in S^k(\mathbb{R}^n)$ are general k-forms jointly vanishing on a nonzero real point, then compute $\{q_1, \ldots, q_m\}$ from input $\sum_{i=1}^m q_i^3$ and $\sum_{i=1}^m q_i^2$.

Algorithmic Proof. Compute a basis $u_1, \ldots, u_m$ of the space $\langle q_1, \ldots, q_m \rangle$ from input $\sum_{i=1}^m q_i^2$ using Semidefinite Programming as outlined in Proposition 3.4.13. Let φ denote the corresponding evaluation isomorphism that we then obtain from Proposition 3.4.16. By solving a linear system, we are now able to evaluate its inverse isomorphism φ^{-1} in any given point. (For that, we just need to represent a form $p \in \mathbb{R}[q_1, \ldots, q_m]$ of degree kd, where $d \in \mathbb{N}_0$, in the basis of products $(u^\alpha)_{|\alpha|=d}$. Note this is a basis for the degree-kd component of $\mathbb{R}[q_1, \ldots, q_m]$, since the latter is a polynomial algebra by Proposition 3.4.16. Denote $\ell_i := \varphi^{-1}(q_i)$ for each $i \in \{1, \ldots, m\}$. Then

$$\varphi^{-1}(\sum_{i=1}^m q_i^d) = \sum_{i=1}^m \ell_i^d \tag{3.55}$$

for each $d \in \mathbb{N}_0$. Thus, we may compute $\sum_{i=1}^m \ell_i^3$ and $\sum_{i=1}^m \ell_i^2$ from the input. By classical results on Waring decomposition, we can obtain $\{\ell_1, \ldots, \ell_m\}$ from that. We can even do so using only Linear Algebra and Semidefinite Programming, cf. Theorem 2.4.8. But then $\varphi(\{\ell_1, \ldots, \ell_m\}) = \{q_1, \ldots, q_m\}$ is the output we wanted. □

Discussing assumptions: PSD and typical quadratics

The condition of Theorem 3.4.17 is satisfied on a Euclidean open subset of the m-tuples of real forms $S^k(\mathbb{R}^n)$, as long as $m \leq n-2$. This can e.g. be seen from the Poincaré-Miranda theorem:

3.4.18 Theorem: (Poincaré-Miranda, cf. [32, Introduction]) Write $\mathcal{H}$ for the parallelepiped spanned by linearly independent vectors $v_1, \ldots, v_n \in \mathbb{R}^n$ and let $f\colon \mathcal{H} \to \mathbb{R}^n, x \mapsto (f_1(x), \ldots, f_n(x))$ a continuous function. Denote

$$\mathcal{H}_i^1 := \{\sum_{j=1}^m \lambda_j v_j \mid \lambda_j \in [0,1], \lambda_i = 1\}$$
$$\mathcal{H}_i^0 := \{\sum_{j=1}^m \lambda_j v_j \mid \lambda_j \in [0,1], \lambda_i = 0\}$$

for $i \in \{1, \ldots, n\}$. Note these are the facets of $\mathcal{H}$. Assume that for each $i \in \{1, \ldots, n\}$, $f_i \leq 0$ on $\mathcal{H}_i^0$, but $f_i \geq 0$ on $\mathcal{H}_i^1$. Then f has a zero on $\mathcal{H}$.

By Theorem 3.4.18, it clearly suffices to find, say, a rectangle $\mathcal{H}$ and quadratics $q_1, \ldots, q_m, q_{m+1}, \ldots, q_n$ such that for each $i \leq m$, $q_i < 0$ on $\mathcal{H}_i^0$ but $q_i > 0$ on $\mathcal{H}_i^1$, as then the condition of Theorem 3.4.18 will be satisfied in a neighbourhood $\mathcal{U}_1 \times \ldots \times \mathcal{U}_n \times \{q_{m+1}\} \times \ldots \times \{q_n\}$ of $(q_1, \ldots, q_m)$. By introducing some new variable Y, one can take e.g. $m = n$, $q_{n+1} := 0$ and $q_i = X_i Y - \frac{1}{2}Y^2 \in$

$\mathbb{R}[X_1, \dots, X_n, Y]$, for $i \in \{1, \dots, n\}$ where Y is another unknown, and consider the rectangle $\mathcal{H} = [0,1]^n \times \{1\}$ in the affine plane where "$Y = 1$". This corresponds to choosing the basis $e_1 + e_{n+1}, \dots, e_n + e_{n+1}, e_{n+1}$ in Theorem 3.4.18. Note $\mathcal{H}_i^a = [0,1]^{i-1} \times \{a\} \times [0,1]^{n-i} \times \{1\}$ for $a \in \{0,1\}$. We have $q_i = -\frac{1}{2} < 0$ on $\mathcal{H}_i^0$ and $q_i > \frac{1}{2}$. Thus, the condition $V_{\mathbb{R}}(q_1, \dots, q_m) \neq \emptyset$ is a typical property.

However, the condition is never satisfied if *any* of the forms is positive definite (which implies that k is even), as already mentioned in Section 3.4. Fortunately, we can perform an almost operation on the power sums that effectively shift the forms $(q_1, \dots, q_m)$ by some fixed quadratic p. This allows to shift any bounded neighbourhood $\mathcal{U}_1 \times \ldots \times \mathcal{U}_m$ into the cone of psd forms. Thus, we obtain a generalization of Theorem 3.4.17 that succeeds on a typical set of *positive definite* forms.

3.4.19 Lemma: Let $k, m, n \in \mathbb{N}$ and $q_1, \dots, q_m, p \in S^k(\mathbb{R}^n)$. From input p, $\sum_{i=1}^m q_i$, m and $\sum_{i=1}^m q_i^2$, we can compute $\sum_{i=1}^m (q_i + p)^2$.

Proof. $\sum_{i=1}^m q_i^2 + 2p \sum_{i=1}^m q_i + mp^2 = \sum_{i=1}^m (q_i + p)^2$. □

This motivates the following definition:

3.4.20 Definition: Let $k, m, n \in \mathbb{N}$. We say that $q_1, \dots, q_m \in S^k(\mathbb{R}^n)$ have a *common (real) level set*, if there exists $\lambda \in \mathbb{R}^\times$ such that the real affine variety $V_{\mathbb{R}}(q_1 - \lambda, \dots, q_m - \lambda) \subseteq \mathbb{R}^n$ is nonempty or if $V_{\mathbb{R}}(q_1, \dots, q_m) \subseteq \mathbb{R}^n$ contains a nonzero point.

Using Lemma 3.4.19 with $p = \lambda Y^2$ for some new variable Y and $\lambda \in \mathbb{R}$, we obtain

3.4.21 Corollary: (Cf. 3.4.17, 3.4.19, 3.4.20) Let $k, m, n \in \mathbb{N}_0$ such that $m \leq n - 2$. There exists an efficient algorithm for the following problem: If $q_1, \dots, q_m \in S^k(\mathbb{R}^n)$ are general k-forms having a common level set, then, from input $\sum_{i=1}^m q_i^3$ and $\sum_{i=1}^m q_i^2$, compute $\{q_1, \dots, q_m\}$ using only Linear Algebra and Semidefinite Programming.

Proof. The only thing left to show is that for general $q_1, \dots, q_m \in S^k(\mathbb{R}^n)$, $I = (q_1 + \lambda Y, \dots, q_m + \lambda Y)$ is a prime ideal and there are no algebraic relations in between $q_1 + \lambda Y, \dots, q_m + \lambda Y$. Both are clear: There are even no algebraic relations in between $Y, q_1, \dots, q_m$, which is a regular sequence. Thus also $q_1 + \lambda Y, \dots, q_m + \lambda Y$ is a regular sequence and forms a complete intersection, whence I is a prime ideal by Proposition 3.4.3. □

Component extraction from fourth powers

As an aside, let me mention a curious alternative algorithm that does *not* rely on a reduction to tensor decomposition. Under genericity assumptions, this algorithm is worse than the one from Theorem 3.4.17 in almost every relevant aspect: It needs e.g. fourth powers as input as opposed to just second and third powers in 3.4.17. Also, with this approach one has to solve a degree-$4k$ Sums-of-Squares program after computing the space $\langle q_1, \dots, q_m \rangle$ whereas 3.4.17 can recover the space with a degree-$2k$ program, then reduce to a Waring decomposition problem and solve that with a degree-2 Sums-of-Squares program. What makes it interesting is that it does not need the assumptions that $q_1, \dots, q_m$ have no syzygies! It still needs the space $\langle q_1, \dots, q_m \rangle$ as input and has some assumptions on the real variety of $q_1, \ldots, q_m$: essentially that none of $q_1, \ldots, q_m$ is redundant to describe their real variety.

The following SDP is the cornerstone of the algorithm: For $k, m, n \in \mathbb{N}$, $h, q_1, \ldots, q_m \in S^k(\mathbb{R}^n)$, define

$$\begin{aligned} (\mathrm{EXT}_{q,h}) \qquad \text{maximize} \qquad & E[h \sum_{i=1}^m q_i] \\ \text{subject to} \qquad & E \in \mathrm{SOS}_{4k}(\sum_{i=1}^m q_i^2 = 1)^* \\ & E \in \mathrm{SOS}_{4k}(\sum_{i=1}^m q_i^4 = 1)^* \\ & E[1] = 1 \end{aligned} \tag{3.56}$$

3.4.22 Theorem: Let $k, m, n \in \mathbb{N}_0$, $q_1, \ldots, q_m \in S^k(\mathbb{R}^n)$ with $V_\mathbb{R}(q_2, \ldots, q_m) \not\subseteq V_\mathbb{R}(q_1) \subseteq \mathbb{R}^n$. Then for all $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ with $\lambda_1 > \lambda_2 \geq \ldots \geq \lambda_m$ and $h = \lambda_1 q_1 + \ldots + \lambda_m q_m$, any optimal solution E^* of the optimization problem $(\mathrm{EXT}_{q,h})$ from 3.56 satisfies $E^*[q_1^2] = 1$ and $E^*[q_2^2] = \ldots = E^*[q_m^2] = 0$.

Proof. Let $h = \lambda_1 q_1 + \ldots + \lambda_m q_m$ for some real values $\lambda_1 > \lambda_2 \geq \ldots \geq \lambda_m$. Let E optimal for $(\mathrm{EXT}_{q,h})$. First, observe that by Cauchy-Schwarz

$$E[\sum_{\substack{i,j=1 \\ i \neq j}}^m \lambda_j q_i q_j]^2 \leq (\sum_{\substack{i,j=1 \\ i \neq j}}^m \lambda_j^2) \cdot E[\sum_{\substack{i,j=1 \\ i \neq j}}^m q_i^2 q_j^2] = 0$$

where the last equality is true since

$$E[\sum_{\substack{i,j=1 \\ i \neq j}}^m q_i^2 q_j^2] = E[\sum_{i=1}^m q_i^2 \sum_{\substack{j=1 \\ j \neq i}}^m q_j^2] = E[\sum_{i=1}^m q_i^2(1 - q_i^2)] = 1 - E[\sum_{i=1}^m q_i^4] = 0$$

Next, note that any feasible solution E of $(\mathrm{EXT}_{q,h})$ satisfies

$$E[h \sum_{i=1}^m q_i] = E[\sum_{i,j=1}^m \lambda_j q_i q_j] = \sum_{j=1}^m \lambda_j E[q_j^2] \leq \lambda_1 \sum_{i=1}^m E[q_i^2] = \lambda_1$$

This gives an upper bound for the optimal value of $(\mathrm{EXT}_{q,h})$. Thus, if some feasible solution attains it, every optimal solution has to. Clearly, this upper bound is attained by some feasible solution E if and only if $E[q_1^2] = 1$ and $E[q_2^2] = \ldots = E[q_m^2] = 0$. Thus, it suffices to show there exists a feasible solution with the latter property. Pick some $x \in V_\mathbb{R}(q_2, \ldots, q_m) \setminus V_\mathbb{R}(q_1)$. In particular x is nonzero. Since $q_2, \ldots, q_m$ are forms, we may wlog. rescale x such that $q_1(x) = 1$. Then the evaluation $E_x \colon \mathbb{R}[X] \to \mathbb{R}, p \mapsto p(x)$ is feasible for $(\mathrm{EXT}_{q,h})$, as

(a) $\sum_{i=1}^m q_i(x)^2 = q_1(x)^2 = 1$

(b) $\sum_{i=1}^m q_i(x)^4 = q_1(x)^4 = 1$

(c) $E_x[1] = 1$

Furthermore, $E_x[h\sum_{i=1}^m q_i] = \sum_{j=1}^m \lambda_j E_x[q_j^2] = \lambda_1$. □

3.4.23 THEOREM: Let $k, m, n \in \mathbb{N}_0$, $q_1, \ldots, q_m \in S^k(\mathbb{R}^n)$ such that for each $2 \leq i \leq m$ it holds $V_{\mathbb{R}}(q_i, \ldots, q_m) \not\subseteq V_{\mathbb{R}}(q_{i-1}) \subseteq \mathbb{R}^n$, and the same holds true for any permutation of $q_1, \ldots, q_m$. Then, given a basis for $\langle q_1, \ldots, q_m\rangle$ (cf. 3.4.13), we may compute $\{q_1, \ldots, q_m\}$ by solving m instances of (EXT) and Linear Algebra.

Algorithmic Proof. Compute an orthonormal basis $u_1, \ldots, u_m$ of the space $U = \langle q_1, \ldots, q_m\rangle$. The algorithm to extract one q_i is as follows: Choose some random $h \in U$ by sampling the coefficients $\mu = (\mu_1, \ldots, \mu_m) \in \mathbb{R}^m$ w.r.t. $u_1, \ldots, u_m$ independently at random. We may take e.g. iid. standard Gaussian normal distributed coefficients. We may also represent $h = \lambda_1 q_1 + \ldots + \lambda_m q_m$ for some $\lambda = (\lambda_1, \ldots, \lambda_m)$. The coefficient vectors λ and μ are related via an unknown linear map $\lambda = C\mu$. Thus $\lambda \sim \mathcal{N}(0, C^TC)$, provided $\mu \sim \mathcal{N}(0, I_m)$. Thus clearly there exists $i \in \{1, \ldots, m\}$ such that λ_i is strictly greater than all other entries of λ. Wlog $i = 1$. (nb: Assuming non-terrible conditioning of the vectors $q_1, \ldots, q_m$, and thus the matrix C, this argument is numerically stable, i.e. we can assume that after reordering, $\lambda_1 > \lambda_2 \geq \ldots \geq \lambda_m$ where λ_1 is significantly bigger than λ_2). Now, solve $(\mathrm{EXT}_{q,h})$. By the assumptions and 3.4.22, $(\mathrm{EXT}_{q,h})$ has an optimal solution. By 3.4.22 we thus obtain a solution E with $E[q_1^2] = 1$ and $E[q_2^2] = \ldots = E[q_m^2] = 0$. In particular, by Cauchy-Schwarz it holds for $i \geq 2$ that $E[q_i]^2 \leq E[1]E[q_i^2] = 0$. We conclude $1 = E[\sum_{i=1}^m q_i] = E[q_1]$. Hence for each $g \in U$, $E[g]$ is the coefficient of q_1 in the basis representation of g w.r.t $q_1, \ldots, q_m$. It follows $\ker E|_U = \langle q_2, \ldots, q_m\rangle$ and therefore we may compute a multiple of q_1 by computing the orthogonal complement of $\ker E|_U$ in U and then searching for the unique point g therein with $E[g] = 1$. Now that we obtained q_1, we may repeat the above procedure by choosing some fresh random $h \in U$ and solving $(\mathrm{EXT}_{q,h})$ with the added constraint $E[q_1^4] = 0$ (nb: this also implies $E[q_1^2] = E[q_1] = 0$ by Cauchy-Schwarz). □

"Of course it is happening inside your head, Harry, but why on earth should that mean that it is not real?"

— Albus P. W. B. Dumbledore, [81].

4 Riemannian Approximation on Parameterized Varieties

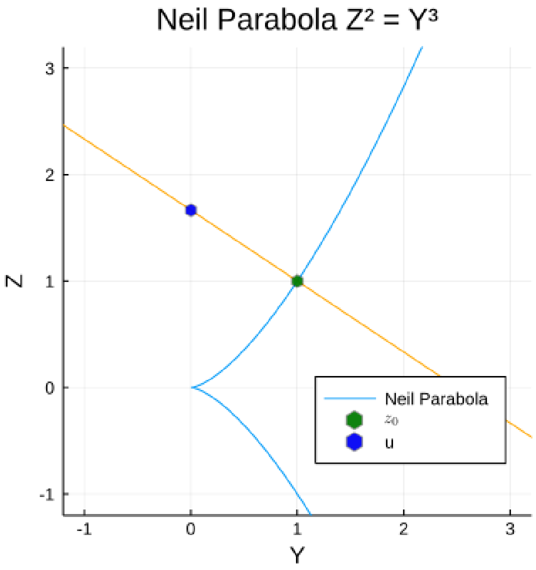

Figure 4.1: The Neil parabola, given by the parametrization $\mathbb{R} \to \mathbb{R}^2, x \mapsto (x^2, x^3)$, some point $u \in \mathbb{R}^2$ outside the parabola and its projection z_0 to the nearest point on the parabola. The orange line in direction $u - z_0$ is normal to the parabola at z_0.

An extremely fundamental algorithmic primitive in Algebraic Statistics is projection to an algebraic variety. Sample-based procedures frequently produce objects in a vector space that lie only *very close* to an interesting algebraic subvariety. Most prominently, this is the case when estimating the moments of a random vector from samples: In Chapter 3, we saw an application where we fitted a rank-2 Gaussian mixture to empirical data, cf. Figure 3.1. Now, this empirical data was of course not generated according to a mixture of two Gaussians, but even if it were, we would not recover the *exact* Gaussian mixture moments as the empirical moments generated from any finite amount of samples.

Moment-based algorithms for estimation problems typically want to make use of the specific algebraic properties of the moment variety (in the particular example above, the second secant of the Gaussian moment variety). But in statistical applications, a moment-based algorithm will likely be fed with inputs that are not actually elements of the variety in question. In some cases, small noise might propagate in an uncontrolled manner and cause unexpected wrong outputs. This is very annoying when designing an algorithm, since the authors of a paper are typically caught in the dilemma to *either* make a clean-looking analysis that understands the noise-free geometry of the problem *or* to take the noise propagation into account and document it in each step. E.g. Garg, Kayal and Saha's [36] algorithm for powers of forms is an example for the first type: The method is inspired from lower bound techniques in circuit complexity, which do not worry about noise at all. Now, of course one can always get lucky if one designs an algorithm with the noiseless situation in mind, but in that case, the authors noticed that it can handle noise of magnitude only inverse exponential in the dimension. On the other hand, many seminal algorithms designed by theoretical computer scientists suffer in the aesthetic category from a very diligent and consequent noise propagation analysis ([65], [59], [37], [38], [5]). [1]

Having to do a noise analysis can obstruct the view on the underlying problem. Thus the question is whether we can somehow separate noise stability from the algorithm for the inverse problem. It would help a lot if there was some *general purpose* "denoising" algorithm that could locally perform the projection to a polynomially-parameterized algebraic variety, given only the outer point and the parametrization as input. This type of *semi-local* projection towards a parameterized variety I will call *Riemannian Approximation* (RA), following Breiding and Vannieuwenhoven [16]. I call it semi-local because it is only local in the image space of the parametrization, but not in the parameter space. Since many estimation algorithms for high-dimensional problems with provable robustness guarantees are based on the Sum-of-Squares method, there typically is a low-degree Sum-of-Squares certificate behind the robustness. Therefore, it should *in principle* not be too wishful to think that in those situations, a Sum-of-Squares relaxation for Riemannian Approximation produces the correct projection, and this might even be proven via essentially the same robustness certificate, if we are lucky. E.g. the analysis of an algorithm of Liu and Moitra for robust estimation of the optimal parameters of a Gaussian mixture model produces a very explicit Sum-of-Squares identity which can be seen as a Sum-of-Squares proof of Lipschitz continuity of the local inverse [59, Equation (1)].

The connection between semi-local projection and inverse problems still

[1] Dear computer scientists: This is a joke. I have the deepest respect for your work.

holds true outside the Sum-of-Squares framework, in the sense that with an algorithm to compute approximate parameters of the inverse problem from noisy input, one may derive an algorithm for RA: First, compute approximations of the optimal parameters, then sharpen them via local Riemannian Optimization to obtain the parameters of the projection. In that sense, RA is a problem we necessarily have to solve on the way when trying to robustly solve the inverse problem. As a caveat, note that this assumes we are talking about *exact* algorithms for the inverse problem, i.e. such algorithms where the error of the output goes to zero if the error of the input does.

Noise stability is also important on the more theoretical side: Identifiability is a nice and very important first property of a statistical model, but *a priori*, it only gives the qualitative guarantee that *exact* moments would *uniquely* determine the parameters.[2] Of course, we would prefer to have a guarantee that *approximate* moments *approximately* determine the parameters. For general parameters, results of that kind can automatically be obtained via the inverse function theorem, provided generic identifiability is known. If anything more than a qualitative statement is needed, there is typically the need to perform a *condition number analysis.* E.g. Breiding and Vannieuwenhoven recently did this for several problems of statistical interest [14], [15], [16], [7].

There are two subtypes of noise stability for parametrized problems: Stability with respect to *outer* noise and with respect to *inner* noise. Given a point $z = s(x)$ from the image of a polynomial map s which is contained in a vector space U, *outer noise* is a small-norm vector $\eta \in U$ that perturbs z, which typically has some genericity properties, e.g. it is often drawn from a continuous random variable. In many statistical applications involving empirical data, the norm $\varepsilon := \|\eta\|$ of the outer noise can be assumed to be arbitrarily small if sufficiently many samples are given. *Inner noise*, on the other hand, is small-norm vector v that perturbs the parameters x.

Noise stability analysis is by itself not necessarily ugly. What makes it messy and hard to read is typically the *propagation* of noise, as in any step of an algorithm, one has to calculate how the algorithmic operations change the noise. Whenever an algorithm for Riemannian Approximation is available for the problem in question, one does not need to worry about noise propagation: Given a problem with true parameters $z = s(x)$, RA can take a noise-perturbed input $s(x) + \eta$ and compute a nearby point $z' = s(x + v)$ of the variety. Thus, one can say that RA converts outer noise to inner noise. It is then possible to run an algorithm without noise stability guarantees on input z'. The quality of the output of such an algorithm will only depend on the *condition* of the associated problem in a neighbourhood of (x, z). Condition analysis can be done separately from the algorithm and with *non-algorithmic* tools. Thus RA is a machinery which might provide a separation of *condition and computation.*

This chapter will focus on the degrees of local exactness certificates, rather than concrete, quantitative robustness guarantees. The degree gives a coarser notion of robustness since by itself it makes no distinction whether local projection is tractable in a large area around the image set or in a very narrow one. For some of the aforementioned cases, the quantitative aspect is very important, e.g. whether the range is exponentially small in the dimension or inverse

[2] As we will see, identifiability *does* actually have some qualitative implications regarding parameter stability in a neighbourhood.

polynomially small would make a big difference for [36].

Outline

The subsequent section will start by collecting some basic geometric properties of tubular neighbourhoods and projections. Afterwards, one will introduce a variant of Lasserre's hierarchy for Riemannian approximation, which will turn out to converge finitely with exactness degree at most $4dD$, if d is the degree of s and D is the degree at which $I(\operatorname{im} s)$ is generated. This finite convergence result motivates the introduction of the *Riemannian Approximation degree*, short *RA-degree*. It may be seen as a local, real algebraic counterpart to the *Euclidean Distance* (ED) degree introduced by Draisma, Horobeţ, Ottaviani, Sturmfels and Thomas [29], with the aim to give a better measurement of complexity for the local projection problem and the estimation problems associated with it. Finally, we will run the semidefinite relaxation on a few concrete varieties to compute RA-degrees e.g. of low-rank tensors.

Related work

Breiding and Vannieuwenhoven [16] studied Riemannian Approximation from the theoretical side and computed condition numbers. With the aforementioned Euclidean Distance degree, Draisma, Horobeţ, Ottaviani, Sturmfels and Thomas [29] introduced a purely algebraic measure for the *global* complexity of projection to a variety. Cifuentes, Agarwal, Parillo and Thomas [25] examined SDP relaxations for a loosely connected problem, namely to minimize the distance to some projection of a variety given by quadratic polynomials, if an input close to the projected variety is given. This is a different kind of problem, but the only one I found which also has a "semi-local spirit" to it, since the input is a point very close to the projected variety and not to the variety itself (e.g. the problem is local in the image of the projection, but global in its kernel). While [25] have no parametrization, another key difference is that the equations are restricted to quadratic polynomials, where in this work, the ideal of $I(\operatorname{im} s)$ can be generated by polynomials of higher degree. In the quadratic setting, one does not need to introduce an explicit constraint for semi-locality, since the Hessian of the Lagrangian is constant.

4.1. Tubular Neighbourhoods

Throughout, let W, U be affine real inner product spaces, thus endowed with the Euclidean topology, and let $s\colon W \to U$ be a polynomial map. Their norms and inner products are denoted with $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$, respectively. Which space is meant will usually be clear from the context. Otherwise, one may clarify by writing e.g. $\|\cdot\|_U$ and $\langle\cdot,\cdot\rangle_U$. Varieties in this chapter are by default affine and still complex, despite the real setting. Thus we see them as subvarieties of $U_{\mathbb{C}} := \mathbb{C} \otimes U \supseteq U$ and $W_{\mathbb{C}} := \mathbb{C} \otimes W \supseteq W$, respectively. Let us consider the *projection problem*

$$(\pi_{s,u}) \qquad \begin{aligned} \text{minimize} \quad & \|z-u\|^2 \\ \text{s. t.} \quad & z \in \overline{\operatorname{im} s} \end{aligned} \tag{4.1}$$

in a *semi-local* setting, where we restrict the parameter u to be from a so-called *tubular neighbourhood* of $\operatorname{im} s$. To define what a tubular neighbourhood is, let us briefly recall notions of *tangent* and *normal bundles* in the smooth setting.

The subsequent exposition largely follows [16, 2.2.], which in turn refers to [55]. Some parts are adapted to the special case of "algebraic manifolds". Recall that any smooth real subvariety of a real affine space such as U is a smooth embedded (second-countable) manifold in the sense of [55]. If the variety is not smooth, then it is still locally a manifold around each smooth point.

4.1.1 Definition: (Tangent and normal bundle, cf. [16]) Let $M \subseteq U$ the set of real points of some smooth subvariety $M_{\mathbb{C}}$ of $U_{\mathbb{C}}$. Then for $x \in M$, we define the $\mathbb{R}$-vector space $T_xM := T_xM_{\mathbb{C}} \cap U$. The "disjoint union"

$$TM = \coprod_{x\in M} T_xM := \{(x,v) \mid x \in M, v \in T_xM\} \tag{4.2}$$

of tangent spaces is called the *tangent bundle* of M. The *normal bundle* is defined as

$$NM = \coprod_{x\in M} N_xM := \{(x,\nu) \mid x \in M, \nu \perp T_xM\} \tag{4.3}$$

where the orthogonal complement $N_xM := (T_xM)^{\perp}$ is taken with respect to the inner product of U. Both TM and NM are C^{∞}-submanifolds of $U \times U$ and real varieties.

4.1.2 Remark: A real variety $V_{\mathbb{R}} \subseteq U$ is the set of real points of some complex variety $V \subseteq U_{\mathbb{C}}$. Note that in principle one needs to be careful, as whether a point in a real variety $V_{\mathbb{R}}$ is smooth depends on the choice of the complex variety V. Some real varieties might be the set of real points of more than one complex variety: Indeed, e.g. the real variety $\{A \in \mathbb{R}^{n\times n} \mid A^TA = I_n\}$ of orthogonal matrices is canonically the set of real points of $\{A \in \mathbb{C}^{n\times n} \mid A^TA = I_n\}$, but also of $\{A \in \mathbb{C}^{n\times n} \mid \det(A)^2 = 1 \,\&\, \operatorname{tr}(A^TA) = n\}$. The latter can be shown via the arithmetic-geometric mean inequality on the Eigenvalues of A^TA. The proof is left as an exercise to the reader. However, there is always a canonical complex variety to consider, which is the Zariski closure $\overline{V_{\mathbb{R}}}$ in $U_{\mathbb{C}}$. Every irreducible component of $V_{\mathbb{R}}$ contains a smooth point if and only if $\overline{V_{\mathbb{R}}} = V$. Smoothness thus implicitly ensures that we do not get a problem with different representations of real varieties via complex ones. Note that in the example above, $\{A \in \mathbb{C}^{n\times n} \mid A^T A = I_n\}$ is the Zariski closure of the set of orthogonal matrices.

4.1.3 Proposition and Definition: For any smooth real subvariety M of U and any function $\delta\colon M \to \mathbb{R}_{>0}$, consider the subset

$$NM_\delta := \{(x, v) \in NM \mid x \in M, \|v\| < \delta(x)\} \tag{4.4}$$

The summation map $\varphi\colon NM \to M, (x, \nu) \mapsto x + \nu$ is smooth and there exists a positive, continuous function δ such that $\varphi|_{NM_\delta}$ is a diffeomorphism onto its image. We call $\mathcal{U}$ a *tubular neighbourhood* of M, if there exists such δ such that $\mathcal{U} = \varphi(NM_\delta)$ is the diffeomorphic image of NM_δ.

4.1.4 Definition: For an arbitrary variety V, we say $\mathcal{U}$ is a tubular neighbourhood of V if $\mathcal{U}$ is a tubular neighbourhood of the smooth locus of V.

4.1.5 Remark: If V is a subvariety of U and $\delta\colon V \to \mathbb{R}_{>0}$ is a semialgebraic function, then the tubular neighbourhood defined by δ is a semialgebraic subset of U: Indeed, by real quantifier elimination it suffices to show that $N\mathcal{S}_\delta$ is a semialgebraic subset of $U \times U$, where $\mathcal{S}$ denotes the smooth locus of V. That a pair $(x, v) \in \mathcal{S} \times U$ lies in the normal bundle means that v is a linear combination of gradients of the equations of V evaluated at x, which is also a semialgebraic condition by quantifier elimination. Finally, $\|v\| < \delta(x)$ is a semialgebraic condition by assumption.

For parameterized varieties, a stronger notion of smoothness is often very useful, which takes the parametrization into account.

4.1.6 Definition: (Parameter-smoothness) A point $z \in \operatorname{im} s$ is called *s-smooth* (or, in abusive ignorance of the dependence on the choice of s, *parameter-smooth*), if there exists $x \in s^{-1}(\{z\})$ such that $\operatorname{rank} Js(x) = \dim T_z V$.

Note that any s-smooth point of $\operatorname{im} s$ is automatically a smooth point of $\operatorname{im} s$, since $\operatorname{rank} Js(x)$ gives a lower bound for the dimension of $\operatorname{im} s$ while $\dim T_z V$ gives an upper bound. In particular, both quantities can only be equal if in fact $\operatorname{rank} Js(x) = \dim \operatorname{im} s = \dim T_z V$. We therefore call $\mathcal{U}$ a *tubular neighbourhood for* s, if $\mathcal{U}$ is a tubular neighbourhood of the locus of s-smooth points of $\operatorname{im} s$. Abusively, we can call such $\mathcal{U}$ a tubular neighbourhood for $\operatorname{im} s$, but note that it clashes with Definition 4.1.4, since for parametrized varieties, we do not want just smooth points, but s-smooth points. Let us recall a few basic properties about distance and projection in a tubular neighbourhood.

4.1.7 Remark: Let $\mathcal{U}$ a tubular neighbourhood of $\operatorname{im} s$.

(a) The intersection $\mathcal{U} \cap \operatorname{im} s$ is the s-smooth locus of $\operatorname{im} s$.

(b) The squared Euclidean distance function $\operatorname{dist}_s^2\colon \mathcal{U} \to \mathbb{R}_{\geq 0}, u \mapsto \operatorname{dist}(u, \operatorname{im} s)^2$ is well-defined and smooth, i.e. C^∞. Well-definedness entails that the minimum in (4.1) is truly attained on a point of $\operatorname{im} s$ rather than just on $\overline{\operatorname{im} s}$.

(c) For every point $z \in \mathcal{U} \cap \operatorname{im} s$, every $\nu \in U$ which is normal to $\operatorname{im} s$ at z and such that $z + \nu \in \mathcal{U}$, it holds that $\operatorname{dist}_s(z + \nu) = \|\nu\|$ and z is the unique optimal solution of (4.1) for $u = z + \nu$.

With *Riemannian Optimization*, there exists a theory which can, given a parametrization s of a manifold M and an initial parameter x, search for a local optimizer of some function f (with sufficiently nice properties) on the manifold. Methods such as *Riemannian Gradient descent* can then track the intial parameter towards a parameter $\hat{x}$ such that $s(\hat{x})$ is a local optimizer of f. Sato's book [82] gives a modern exposition of the theory. However, these methods are not applicable if no initial parameter is known. Finding an initial parameter can be very far from a trivial matter: In fact, for many of the estimation problems mentioned in this chapter's introduction, it is the main computational challenge and a source of computational hardness. One prime example is Waring decomposition: E.g. for some space W, consider the map $s\colon (W^\vee)^m \to S^3(W), (\ell_1, \ldots, \ell_m) \mapsto \sum_{i=1}^m \ell_i^3$. Even without noise, it is highly nontrivial to obtain the parameters, given some point $z = \sum_{i=1}^m \ell_i^3$, despite Theorem 2.4.7 showing that they are generically unique in a neighbourhood (and s thus gives a local diffeomorphism). For sufficiently low rank, e.g. for $m \leq n$, the algorithm of Theorem 2.4.8 might be used to obtain the parameters, but makes no statement on noise. There exist robust variants, e.g. due to Anandkumar, Ge, Hsu, Kakade and Telgarsky [3] that can tolerate some outer noise.

What can go wrong outside of tubular neighbourhoods

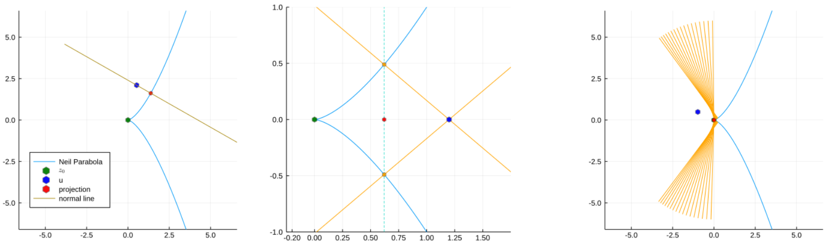

Figure 4.2: Projection in a neighbourhood of the singular point $(0,0)$. Three cases can occur. Left: Some blue points u near the origin get projected to a *unique smooth* point of the parabola (one-to-one situation). Middle: Points on the ray $\mathbb{R}_{\geq 0}(1,0)$ do not have a unique projection, even though such points may be arbitrarily close to the parabola (one-to-many situation). The true nearest points are marked in orange. Right: In a full-dimensional area to the left of the y-axis, all points get projected to the origin (many-to-one situation). The green dot denotes the starting point $(0,0)$, the blue dot a small random ε-perturbation u and the red dot marks the projection back to the variety, computed via $(L\pi)_{s,u,3,2\varepsilon}$ with $s(x) = (x^2, x^3)$ for $x \in \mathbb{R}$. The orange lines are normals in a neighbourhood of $(0,0)$.

The restriction to tubular neighbourhoods of the smooth locus is integral to the subsequent algorithmic approach. Outside a tubular neighbourhood, a unique nearest point need not exist. In fact, the rightmost graphic in Figure 4.2 shows a ray of points that have no unique projection to the Neil parabola. Globally, such points may exist without the presence of a singularity. (Take e.g. the midpoint of a circle, which has no unique projection to the circle). But for the Neil parabola, there exist points *arbitrary close* to it that have no unique projection. This is *one* phenomenon that may occur around a singularity. Another is depicted in the leftmost graphic from Figure 4.2: There can be large typical areas[3] that all get projected to the same singular point. While in principle there is no obvious argument why this case should be computationally hard, it is better to restrict attention to the parameter-smooth locus and understand it first.

4.2. Preliminaries

Real Tangent Spaces and Local Optimality Certificates on Varieties

4.2.1 Motivation: In classical optimization theory, *local optimality certificates* play an important role. For a polynomial $f \in \mathbb{R}[Z]$, the gradient ∇f will vanish at every local extreme point z. If furthermore the Hessian $\operatorname{Hess} f$ of f is positive (resp: negative) definite at such a point z, then the Hessian matrix at this point together with the information that the gradient vanishes is commonly called a *second order certificate* that f has a local minimum (resp: maximum) at z. Things are slightly different when optimizing over a smooth submanifold of $\mathbb{R}^n$: The classical Euclidean gradient then needs to be only orthogonal to the tangent space and a reasonable second order local minimality certificate can only require the Hessian to be positive definite *after* restriction to the tangent space. Riemannian Optimization therefore introduces Riemannian gradients etc., which have the disadvantage that they do not fit nicely into the syntactical framework of ideals and polynomial-based proof systems. In this section we will, among other things, establish that this is not a problem: Whenever f is any polynomial that is extremal on a smooth point of the variety V, we can add a polynomial $g \in I(V)$ vanishing on V such that $f + g$ has a Euclidean second order certificate of optimality. Getting the gradient to vanish is simple via Lagrange multipliers. For the second order part, we use a small trick.

4.2.2 Notation: Throughout, let U a real, affine space with fixed basis $\mathcal{B}$, which is included in some complex vector space $U_{\mathbb{C}} \supseteq U$ of complex dimension equal to the real dimension of U (thus $U_{\mathbb{C}} \cong \mathbb{C} \otimes U$ as real vector spaces). Let $I \subseteq \mathbb{R}[U]$ an ideal, $V(I) = \{z \in U_{\mathbb{C}} \mid \forall f \in I : f(z) = 0\}$ the *variety* (or *complex zero set*) of I and $V_{\mathbb{R}}(I) = V(I) \cap U$ the *real part* of its variety. We will always assume that I is a *radical ideal*, i.e. that I consists of all polynomials vanishing on $V(I)$. We endow U with some inner product $\langle \cdot, \cdot \rangle$ and denote by $M^{\perp} = \{y \in U \mid \forall x \in M : \langle x, y \rangle = 0\}$ the *orthogonal complement* of the set $M \subseteq U$.

While the applications will need arbitrary spaces, many of the technical preliminaries are more easily proven in coordinates. Thus let us introduce some basis for W with respect to which $W \cong \mathbb{R}^n$ and choose variables $X = (X_1, \ldots, X_n)$. Likewise, let us identify $U \cong \mathbb{R}^N$ if needed for some $n, N \in \mathbb{N}_0$, with variables $Z = (Z_1, \ldots, Z_N)$

Throughout this chapter, we will be working with varieties that are also given as the image set of a polynomial map s. On general points of the image set, the tangent space will admit two descriptions: One as the kernel of the Jacobian matrix given by a tuple of defining inequalities, which is the natural

[3] I.e. areas of nonempty Euclidean interior.

definition that was already introduced in Chapter 2. One other as the image set of the Jacobian Js. The latter does not always give the full tangent space, but does so on a dense subset of the parameters, which the following lemma will assert.

4.2.3 PROPOSITION: Let $s\colon W \to U$ a polynomial map. Let $x \in U$ and $z := s(x) \in \operatorname{im} s$. Then

$$\operatorname{im} Js(x) \subseteq T_z \operatorname{im} s \tag{4.5}$$

with equality on a Zariski dense open subset of W. Equality holds precisely if x is an s-smooth point.

Proof. In Proposition 2.2.12, we already saw that the complex extension, i.e. $s_{\mathbb{C}}\colon W_{\mathbb{C}} \to U_{\mathbb{C}}$ induces a map of complex tangent spaces $T_{s_{\mathbb{C}}}(x)\colon W \to T_z \operatorname{im} s_{\mathbb{C}}$ which is generically surjective and in coordinates given by the Jacobian. This readily shows the claim (4.5) for the real spaces, too. Since the image of s is dense in the image of $s_{\mathbb{C}}$, equality holds on an ($\mathbb{R}$-)Zariski dense subset of U. But by definition, equality also holds when x is an s-smooth point, cf. Definition 4.1.6. In particular, s-smoothness is a generic property. □

4.2.4 LEMMA: Let the ideal $I \subseteq \mathbb{R}[U]$ be generated by some polynomials of degree at most $D \in \mathbb{N}_0$. Then there is a Sum-of-Squares polynomial $p \in I$ of degree at most $2D$ such that

(a) $\{z \in U \mid p(z) = 0\} = V_{\mathbb{R}}(I)$

(b) $\nabla p(z) = 0$ for all $z \in V_{\mathbb{R}}(I)$.

(c) In any $z \in V_{\mathbb{R}}(I)$, the directional derivatives $(\partial_h)^2 p(z)$ for $h \in U$ are non-negative and vanish precisely if h lies in the tangent space $T_z V_{\mathbb{R}}(I)$.

(d) The restriction of $\operatorname{Hess} p(z)\colon U \times U \to \mathbb{R}$ to $N_z V_{\mathbb{R}}(I) \times N_z V_{\mathbb{R}}(I)$ is positive definite for each smooth $z \in V_{\mathbb{R}}(I)$.

Proof. Fix $M \in \mathbb{N}_0$ and generators $p_1, \ldots, p_M \in \mathbb{R}[U]$ of I. Set $p := \sum_{i=1}^M p_i^2$. Clearly $p \geq 0$ on U with $\{z \in U \mid p(z) = 0\} = V_{\mathbb{R}}(I)$. Therefore, p has a local minimum in each point $z \in V_{\mathbb{R}}(I)$. We deduce $\nabla p(z) = 0$ for all $z \in V_{\mathbb{R}}(I)$. This shows (a) and (b).

Ad (c): For $z \in V_{\mathbb{R}}(I)$, a simple calculation using $p_i(z) = 0$ shows that

$$\partial_h^2 p(z) = \sum_{i=1}^m (p_i \partial_h^2 p_i + (\partial_h p_i)^2)(z) = \sum_{i=1}^m ((\partial_h p_i)(z))^2 \geq 0 \tag{4.6}$$

It is thus immediate that if $\partial_h^2 p(z) = 0$, then $(\partial_h p_i)(z) = 0$ for $i \in \{1, \ldots, m\}$ and thus $(\partial_h f)(z) = 0$ for all $f \in I$. This implies $h \in T_z V_{\mathbb{R}}(I)$. If, on the other hand, $h \in T_z V_{\mathbb{R}}(I)$, then we can deduce $\partial_h^2 p(z) = 0$ by reading the above argument backwards.

Ad (d): The claim is that $(\partial_h^2)p(z) = \operatorname{Hess} p(z)\langle h, h\rangle > 0$ for all real $h \in N_z V_{\mathbb{R}}(I)^{\perp} \setminus \{0\} \cap U$. Since $(\partial_h^2)p(z) \geq 0$ by (c), it remains to show $(\partial_h^2)p(z) \neq 0$. However, by (c) this is even true for $h \notin T_z V_{\mathbb{R}}(I)$. □

4.2.5 Proposition: Let $s\colon \mathbb{R}^n \to \mathbb{R}^N$ a polynomial morphism, $u \in \mathbb{R}^N$ and $z \in \operatorname{im} s$ a minimizer of $\|Z - u\|^2$ on $\operatorname{im} s$. If z is an s-smooth point, then $(z-u) \in N_z \operatorname{im} s$.

Proof. Choose $x \in s^{-1}(\{z\})$ and set $f\colon \mathbb{R}^N \to \mathbb{R}, z \mapsto \|z-u\|^2$. Using that x is a critical point of $f \circ s$ yields

$$0 = \nabla(f \circ s)(x) = \nabla f(s(x))^T Js(x) = (z-u)^T Js(x)$$

Hence $(z-u) \in (T_z \operatorname{im} s)^\perp$ by s-smoothness. □

4.2.6 Lemma: Let $s\colon \mathbb{R}^n \to \mathbb{R}^N$ a polynomial map. Let $x \in \mathbb{R}^n$, $z = s(x) \in \operatorname{im} s$ and $f\colon \mathbb{R}^n \to \mathbb{R}$ such that $\nabla f(z) = 0$. Then for all $v \in \mathbb{R}^n$,

$$\operatorname{Hess}(f \circ s)(x)\langle v, v\rangle = \operatorname{Hess} f(z)\langle Js(x)v, Js(x)v\rangle$$

In particular, if $\operatorname{Hess}(f \circ s)(x)$ is positive definite, then $\operatorname{Hess} f(z)$ is positive definite when restricted to $\operatorname{im} Js(x)$.

Proof. Deriving $f \circ s$ yields

$$\begin{aligned} &\operatorname{Hess}(f \circ s)(x)\langle v, v\rangle \\ =&\operatorname{Hess} f(s(x))\langle Js(x)v, Js(x)v\rangle + \nabla f(s(x))^T (\operatorname{Hess} s_i(x))_{i \in \{1,\ldots,N\}} \\ =&\operatorname{Hess} f(s(x))\langle Js(x)v, Js(x)v\rangle \end{aligned}$$

□

4.2.7 Lemma: Let $N \in \mathbb{N}_0$ and M a symmetric bilinear form on $\mathbb{R}^N$ such that M is positive definite when restricted to the subspace U. If P is a psd bilinear form on $\mathbb{R}^N$ such that $\ker P = U$, then $M + \mu P$ will be positive definite for all sufficiently large $\mu \in \mathbb{R}$.

Proof. Choosing an Eigenbasis of P yields a decomposition $\mathbb{R}^N = U \oplus W$, where $W = \bigoplus_{\lambda > 0}\{w \in \mathbb{R}^N \mid Pw = \lambda w\}$ is the sum of all Eigenspaces of P corresponding to nonzero Eigenvalues and $U = \ker P$. Representing both P and M w.r.t such a basis, we get a block representation of P as

$$\begin{pmatrix} L & 0 \\ 0 & 0 \end{pmatrix}$$

where $L \in \mathbb{R}^{\dim W \times \dim W}$ is positive definite and of M as

$$\begin{pmatrix} A & B \\ B^T & C \end{pmatrix}$$

where $A \in \mathbb{R}^{\dim W \times \dim W}, B \in \mathbb{R}^{\dim W \times \dim V}, C \in \mathbb{R}^{\dim V \times \dim V}$ and C is positive definite. Consider the matrix

$$M(\mu) = \begin{pmatrix} A(\mu) & B \\ B^T & C \end{pmatrix}$$

with $A(\mu) := A + \mu L$. The Schur complement of the C-block is $A(\mu) - BC^{-1}B^T$. Since L is positive definite, $A - BC^{-1}B^T + \mu L$ will be positive definite for all sufficiently large $\mu \in \mathbb{R}_{>0}$. Since positive definiteness of both C and its Schur complement implies that $M(\mu)$ is positive definite, we conclude that $M(\mu)$ is positive definite for all sufficiently large $\mu \in \mathbb{R}_{>0}$. □

4.2.8 Proposition: Let $V_{\mathbb{R}}$ a real variety. If $z \in V_{\mathbb{R}}$ is smooth, then for the real vector space $T_z V_{\mathbb{R}}$ we have

$$N_z V_{\mathbb{R}} = \{\nabla q(z) \mid q \in I(V_{\mathbb{R}})\}$$

Furthermore, if I is generated by polynomials of degree at most $d \in \mathbb{N}_0$, then

$$N_z V_{\mathbb{R}} = \{\nabla q(z) \mid q \in I, \deg(q) \leq d\}$$

The same holds true under the weaker assumption that I contains codimension-many polynomials of degree at most $d \in \mathbb{N}_0$ with gradients that are linearly independent as vectors of polynomials.

Proof. By definition, the $T_z V_{\mathbb{R}}$ consists precisely of the vectors in $\mathbb{R}^N$ which are orthogonal to all vectors of the kind $\nabla q(z) \in \mathbb{R}^N$, where $q \in I \subseteq \mathbb{R}[Z]$. Hence

$$N_z V_{\mathbb{R}} = \{\nabla q(z) \mid q \in I\}$$

Suppose now I is generated by the tuple $f \in (\mathbb{R}[Z]_{\leq d})^r$ for some $r, d \in \mathbb{N}_0$. Let $q \in I$. Choose $p_1, \dots, p_r \in \mathbb{R}[Z]$ such that $q = q_1 f_1 + \ldots + q_r f_r$. Note that for each $i \in \{1, \dots, r\}$, $(\nabla q_i f_i)(z) = (\nabla q_i)(z) f_i(z) + q_i(z)(\nabla f_i)(z) = q_i(z)(\nabla f_i)(z)$, since $f_i(z) = 0$. Thus $\nabla q(z) \in \langle \nabla f_1(z), \dots, \nabla f_r(z) \rangle$ and therefore $\nabla q(z)$ is also the gradient of some polynomial in the space $\langle f_1, \dots, f_r \rangle$. Now, let $k :=$ codim $V_{\mathbb{R}}$. It is easy to see that f does not need to be a tuple that generates I. Indeed, if the gradients of $f_1, \dots, f_r$ contain a k-element subset which is linearly independent as a set of vectors of polynomials, then this subset gives linearly independent vectors when evaluated on a general point $z \in V_{\mathbb{R}}$. Thus for dimension reasons, they span the complement of the tangent space. □

The following proposition shows that we can certify the minimality of $f_u(z) = \|z - u\|^2$ on $\operatorname{im} s$ in a point $z \in \operatorname{im} s$ by finding a second order optimality certificate for some polynomial $\tilde{f}$ which coincides with f_u as a function on $\overline{\operatorname{im} s}$.

4.2.9 Proposition: Let $s\colon \mathbb{R}^n \to \mathbb{R}^N$ and $z \in \operatorname{im} s$ be an s-smooth point of $\operatorname{im} s$. Let $u \in \mathbb{R}^N$ such that $u - z \in N_z \operatorname{im} s$, or equivalently: z is a locally optimal solution of

$$\begin{array}{llll} (\pi_{s,u}) & \text{minimize} & \|y - u\|^2 & \\ & \text{s. t.} & y \in \operatorname{im} s & \end{array} \tag{4.7}$$

Write $f_u := \|Z - u\|^2$. Then there is $q \in I(\operatorname{im} s)$ such that $\nabla(f_u + q)(z) = 0$ and $\operatorname{Hess}(f_u + q)(z) \succ 0$. Furthermore, if $I(\operatorname{im} s)$ is generated by some polynomials of degree at most $D \in \mathbb{N}$, then one can choose q such that $\deg(q) \leq 2D$.

Proof. Since z is locally a minimizer of f on $\operatorname{im} s$ we have $\nabla f(z) \in N_z \operatorname{im} s$ by Proposition 4.2.5. By Proposition 4.2.8, we find $q \in I(\operatorname{im} s)$ with $\nabla q(z) = -\nabla f(z)$, and $\deg(q) \leq D$, if $I(\operatorname{im} s)$ is generated by polynomials of degree at most D. Then $\tilde{f} := f + q$ satisfies $\tilde{f}(z) = f(z)$ and $\nabla \tilde{f}(z) = 0$, but $\operatorname{Hess} \tilde{f}(z) = 2I_N + \operatorname{Hess} q(z)$ might not be positive definite. However, by Lemma 4.2.6 we know that $\operatorname{Hess} \tilde{f}(z)\langle h, h\rangle > 0$ for each tangent direction $h \in T_z \operatorname{im} s \setminus \{0\}$. Choose p as in Lemma 4.2.4 of degree at most $2D$. Then $P := \operatorname{Hess} p(z)$ satisfies $U := \ker P = T_z \operatorname{im} s$ and $M := \operatorname{Hess} \tilde{f}(z)$ is pd when restricted to U. Thus we can apply Lemma 4.2.7 to get some $\lambda \in \mathbb{R}$ such that $M + \lambda P = \operatorname{Hess}(f + q + \lambda p)(z)$ is positive definite. Since $q + \lambda p \in I \cap \mathbb{R}[X]_{\leq 2D}$ and $\nabla p(z) = 0$, this shows the claim. □

4.3. A Sum-of-Squares Relaxation for Riemannian Approximation with finite convergence

4.3.1 Definition: For real vector spaces W, U, a polynomial morphism $s\colon W \to U$, $u \in U$ and $k \in 2(\mathbb{N}_{\geq \deg(s)}), \varepsilon \in \mathbb{R}_{>0}$ we define the semi-local Lasserre relaxation $(L\pi_{s,u,k,\varepsilon})$ of $(\pi_{s,u})$ as the Sum-of-Squares optimization problem

$$\begin{array}{lll} (L\pi_{s,u,k,\varepsilon}) & \text{minimize} & E[\|s-u\|^2] \\ & \text{s. t.} & E \in \mathrm{SOS}_k^*(\|s-u\|^2 \leq \varepsilon^2) \\ & & E[1] = 1 \end{array} \tag{4.8}$$

4.3.2 Remark: The conic dual of $(L\pi)_{s,u,k,\varepsilon}$ from Definition 4.3.1 is

$$\begin{array}{ll} \text{maximize} & \mu \\ \text{s.t.} & \|s-u\|^2 - \mu \in \mathrm{SOS}_{2k}(\|s-u\|^2 \leq \varepsilon^2) \end{array} \tag{4.9}$$

It is often useful to identify U with $\mathbb{R}^N$ and W with $\mathbb{R}^n$, where $n, N \in \mathbb{N}$, via choosing variables $X = (X_1, \dots, X_n)$ for W and $Z = (Z_1, \dots, Z_N)$ for U. Then we can identify the polynomial map s with the vector of polynomials $(s_1, \dots, s_N) \in \mathbb{R}[X]^N$ describing it and write down expressions such as $\|s\|^2 = \sum_{i=1}^N s_i^2 \in \mathbb{R}[X]$ etc. This convention is used whenever needed throughout the entire chapter and its usage will be mentioned by saying that we choose variables.

4.3.3 Definition: Let $s\colon W \to U$ a polynomial map between real affine spaces. We call a functional $L \in \mathbb{R}[W]^*$ *s-atomic* (with respect to $z \in \operatorname{im} s$), if for any $f \in \mathbb{R}[U]$,

$$L(f \circ s) = f(z) \tag{4.10}$$

More generally, a functional on a subspace of $\mathbb{R}[W]$ is called s-atomic if it is the restriction of an s-atomic functional on $\mathbb{R}[W]$.

Note that for $U = W$ in the previous definition, the id_W-atomic functionals are just the point evaluations (and thus the ring homomorphisms) on $\mathbb{R}[W]$.

4.3.4 Proposition: Let W, U be real affine spaces, $s\colon W \to U$, $d = \deg(s)$ and $u \in U$.

(a) Let $\varepsilon \in \mathbb{R}_{>0}$ be fixed. Then the sequence

$$(L\pi_{s,u,2d,\varepsilon}), (L\pi_{s,u,2d+2,\varepsilon}), (L\pi_{s,u,2d+4,\varepsilon}), \dots$$

defines a hierarchy of Sum-of-Squares relaxations for $\pi_{s,u}$ with semidecreasing optimal values. By convention, we understand the optimal value of an infeasible minimization problem to be ∞.

(b) If $\varepsilon \in \mathbb{R}_{>0}$ is such that $\operatorname{dist}(u, \operatorname{im} s) \leq \varepsilon$, then for any $k \in \mathbb{N}_0$, $(L\pi_{s,u,k,\varepsilon})$ is feasible.

(c) Let now $k \in 2(\mathbb{N}_{\geq d})$ be fixed and $\varepsilon_1, \varepsilon_2, \dots$ a semidecreasing sequence in $\mathbb{R}_{\geq \operatorname{dist}(u, \operatorname{im} s)^2}$. Then

$$(L\pi_{s,u,k,\varepsilon_1}), (L\pi_{s,u,k,\varepsilon_2}), (L\pi_{s,u,k,\varepsilon_3}), \dots$$

is a hierarchy of Sum-of-Squares relaxations for $\pi_{s,u}$ with semidecreasing optimal values.

4.3.5 Theorem: Let W, U real inner product spaces, $s\colon W \to U$ a polynomial map, $u \in U$ general, $k \in \mathbb{N}_0$ and $\varepsilon \in \mathbb{R}_{>0}$. Let $d \in \mathbb{N}$ such that s is given by polynomials of degree at most d and let $D \in \mathbb{N}$ such that $I(\operatorname{im} s)$ is generated in degree at most D. If ε is sufficiently small, then for all $k \geq 2dD$ and all solutions E of $(L\pi_{s,u,k,\varepsilon})$ such that

$$E[\|s-u\|^2] \leq \operatorname{dist}(u, \operatorname{im} s)^2, \tag{4.11}$$

equality holds in Equation (4.11) and $E[s]$ is the orthogonal projection of u to $\operatorname{im} s$. Furthermore, E is then s-atomic.

Proof. Choose a vector of variables Z for U and denote $f_u := \|Z-u\|^2 \in \mathbb{R}[Z]$. Let $z \in U$ be the unique orthogonal projection of u to $\operatorname{im} s$. Let E some feasible solution with $E[\|s-u\|^2] \leq f_u(z) = \operatorname{dist}(u, \operatorname{im} s)^2$. By Proposition 4.2.9, we may choose $g \in \mathbb{R}[U]$ such that $g - f \in I(\operatorname{im} s)$, $\nabla g(z) = 0$, $\operatorname{Hess} g(z) \succ 0$ and $\deg(g) \leq 2D$. The polynomial map $s\colon W \to U$ induces a map of polynomial rings $\hat{s}\colon \mathbb{R}[U] \to \mathbb{R}[W], f \mapsto f(s)$, via which any functional on $\mathbb{R}[W]_{\leq 2dD}$ pulls back to a functional on $\mathbb{R}[U]_{\leq 2D}$. Denote by $\tilde{E}\colon \mathbb{R}[U] \to \mathbb{R}, f \mapsto E[f(s)]$ the pullback of E. From Lemma 2.3.11, we obtain that $\tilde{E}$ is an optimal solution of the SOS program

$$\begin{aligned} \text{minimize} \quad & L[g] \\ \text{s.t.} \quad & L \in \mathrm{SOS}^*_{2D}(f_u \leq \varepsilon^2) \end{aligned}$$

in variables Z. Thus, from Lemma 2.3.11 we also obtain that $\tilde{E} = \delta_z$. In particular, $E[s] = \tilde{E}[Z] = z$ and E is s-atomic. □

4.3.6 Remark: (a) Recall that if $\operatorname{dist}(u, \operatorname{im} s) \leq \varepsilon$, $(L\pi_{s,u,k,\varepsilon})$ is always feasible independent of the degree k, and there is a solution satisfying (4.11). Indeed, simply take a point $z \in \operatorname{im} s$ with $\|z-u\| \leq \varepsilon$, choose a preimage $x \in W$ such that $z = s(x)$ and then observe that the Dirac functional δ_x is feasible for $(L\pi_{s,u,k,\varepsilon})$. In particular, if in addition the conditions of Theorem 4.3.5 are met, then any optimal solution of $(L\pi_{s,u,k,\varepsilon})$ is s-atomic.

(b) If the conditions of Theorem 4.3.5 are satisfied, then a solution of $(L\pi_{s,u,k,\varepsilon})$ satisfying (4.11) exists if and only if $\operatorname{dist}(u, \operatorname{im} s) \leq \varepsilon$: The "if" direction is covered by (a). For the "only if" direction, take some feasible E satisfying (4.11). By Theorem 4.3.5, E is then s-atomic. Thus it holds that $z := E[s] \in \operatorname{im} s$ and therefore $\|z-u\|^2 \leq \varepsilon^2$.

(c) The condition that $I(\operatorname{im} s)$ is generated in degree at most D may be replaced by the possibly weaker condition that there exist $\operatorname{codim}(W)$-many polynomials in $I(\operatorname{im} s) \subseteq K[U]$ of degree at most D whose gradients are linearly independent vectors of polynomials.

4.3.7 Remark: While generators of $I(\operatorname{im} s)$ give a degree bound for the RA degree, it is worth noting that the RA degree might be smaller by an arbitrarily large amount. E.g. in Section 4.5, we will see an explicit family of polynomial maps $s_n\colon (\mathbb{R}^n)^{n-1} \to \mathbb{R}^{n\times n}$, apparently of constant RA degree 4, such that even set-theoretic equations for $\operatorname{im} s$ have degree at least n. Intuitively, it is good to have a hierarchy rather than relying on equations since there might be non-obvious reasons for exactness of $(L\pi)$ that have nothing to do with the equations. Note that whenever we know ideal generators $g_1, \ldots, g_l$, $(l \in \mathbb{N}_0)$ of $I(\operatorname{im} s)$ explicitly, and these are easy to evaluate, the optimization problem can be solved much faster via Lagrangian optimization: To an input $u \in U$ sufficiently close to some s-smooth point of $\operatorname{im} s$, simply compute a local optimizer of the Lagrangian: $\mathcal{L}(z, \mu) := \|z-u\|^2 + \sum_{i=1}^m \mu_i g_i(z)$. Any critical point z will satisfy $0 = \nabla_z \mathcal{L}(z, \mu) = 2(z-u) + \sum_{i=1}^m \mu_i \nabla g_i(z)$ and $0 = \partial_{\lambda_i} \mathcal{L}(z, \mu) = g_i(z)$ for each $i \in \{1, \ldots, l\}$. The latter conditions guarantee that $z \in \operatorname{im} s$. The first condition implies $z - u \in N_z \operatorname{im} s$. Locally, this characterizes the orthogonal projection.

Lipschitz bounds for the local inverse

We have seen that equations for $\operatorname{im} s$ give a degree bound for convergence of $(L\pi)_s$. Another method are proofs of the kind

$$\|X - x\|^2 \preceq C \cdot \|s - s(x)\|^2 \tag{4.12}$$

where $x \in W$ and $C \in \mathbb{R}$. An identity such as (4.12) can only exist if s is an injective map and it is a Sum-of-Squares proof of a Lipschitz constant for the inverse of s. Any proof showing that $\|X - x\|^2 \leq \mathcal{O}(\varepsilon^2)$ will imply a similar local exactness guarantee, since we can simply use a Taylor expansion on W rather than U: Indeed, notice that for $f_u := \|Z - u\|^2$ and $\nu := u - s(x)$, the first derivative of the polynomial $f_u \circ s$ is

$$Jf_u(s(x))Js(x) = (s(x) - u)^T Js(x) \tag{4.13}$$

which vanishes if $s(x)$ is the projection of u. The Hessian at $v \in \mathbb{R}^n$ is calculated as was already done in the proof of Lemma 4.2.6:

$$\begin{aligned} &\operatorname{Hess}(f_u \circ s)(x)\langle v, v\rangle \\ &= 2\langle Js(x)v, Js(x)v\rangle + (\nu^T (\operatorname{Hess} s_i(x))_{i \in \{1,\ldots,N\}})\langle v, v\rangle \end{aligned} \tag{4.14}$$

The left addend in the second line is a positive definite form in v (for general $x \in W$, assuming the map s is nondegenerate, as otherwise an identity such as (4.12) cannot exist) and the right addend goes to zero as $\|\nu\| = \|u - z\| \to 0$. Thus semi-locally the matrix will be positive definite and then Lemma 2.3.11 applied to the space W and the form $f_u \circ s$ (rather than U and some form $g \in I(\operatorname{im} s) + f_u$ as in Theorem 4.3.5) yields that any optimal solution E of $(L\pi)_{s,u,k,\mathcal{O}(\varepsilon)}$, where k is the degree of the proof for (4.12), satisfies $E[\|X - x\|^2] = 0$ and is thus an atomic pseudo-expectation supported on x. Therefore, $E[\|s - s(x)\|^2] = 0$ and E is obviously also s-atomic. Note that this framework is very limited, since it requires a “certifiably injective” map s.

4.4. The Riemannian Approximation degree

4.4.1 Definition: Let $s\colon W \to U$ a polynomial map and $z \in \operatorname{im} s$. We say that $(L\pi)_s$ is *semi-locally exact* in degree $k \in 2(\mathbb{N}_{\geq \deg s})$ at z, if there exists $\varepsilon > 0$ such that $(L\pi)_{s,u,k,2\varepsilon}$ is exact for each $u \in \mathcal{B}_\varepsilon(z)$ with $u - z \in N_z \operatorname{im} s$.

4.4.2 Definition: Let $s\colon W \to U$ a polynomial map and $z \in \operatorname{im} s$ general. We call the minimum $k \in \mathbb{N}_{\geq \deg s}$ such that $(L\pi)_s$ is *semi-locally exact* in degree k the *Riemannian Approximation (RA) degree* of s at z.

By Theorem 4.3.5, the Riemannian Approximation degree is well-defined and bounded by a finite number not depending on k.

4.4.3 Remark: If $(L\pi)_s$ is semi-locally exact in degree $k \in 2(\mathbb{N}_{\geq \deg s})$ at the general point $z \in \operatorname{im} s$, then in particular for all sufficiently small-length normal vectors ν at z, $u := z + \nu$, $f_u := \|Z - u\|^2$ and all $E \in \mathrm{SOS}_k(\|s-u\|^2 \leq \varepsilon^2)^*$, it holds that $E[f_u(s)] - f_u(z) \geq 0$. Thus $f_u(s) - f_u(z) \in \mathrm{SOS}_k(\|s-u\|^2 \leq \varepsilon^2)$ for all sufficiently small ε. Therefore, for each small $\varepsilon > 0$ there exist SOS polynomials $r_\varepsilon, t_\varepsilon$ of degrees at most k and $k - 2d$, respectively, such that

$$\forall u \in \mathcal{B}_\varepsilon(z), u - z \in N_z\colon \quad f_u(s) - f_u(z) = r_\varepsilon + t_\varepsilon(\varepsilon^2 - \|s-u\|^2) \tag{4.15}$$

But semi-local exactness also implies that $E[\|s-z\|^2] = 0$ for all E which lie in the face of $\mathrm{SOS}_k^*(f_u(s) \leq \varepsilon^2)$ exposed by the hyperplane

$$\mathcal{H} := \{E \mid E[f_u(s)] = f_u(z)\} \tag{4.16}$$

of optimal value. In an equivalent conic dual formulation, this means that there exists $\mu \in \mathbb{R}$ such that for all $u \in \mathcal{B}_\varepsilon(z)$ with $u - z \in N_z$,

$$-\|s-z\|^2 \in \mathrm{SOS}_k(\|s-u\|^2 \leq \varepsilon^2) + \mu(f_u(s) - f_u(z)) \tag{4.17}$$

Group actions

4.4.4 Remark: Consider the setting where U is a subspace of the polynomial ring $\mathbb{R}[W]$ and $\operatorname{im} s$ is invariant under some subgroup G of $\mathrm{GL}(W)$, which acts on $\mathbb{R}[W]$ by substitution. The RA degree is constant on orbits of such actions: Choose variables $X = (X_1, \dots, X_n)$ for the space W. For z, u, ε, ν as in Remark 4.4.3, if an equation of the kind (4.15) holds, then it also holds that

$$\|s(gX) - u(gX)\|^2 - \|\nu(gX)\|^2 = r_\varepsilon(gX) + t_\varepsilon(gX)(\varepsilon^2 - \|s(gX) - u(gX)\|^2)$$

Thus $\|\nu(gX)\|^2$ is the optimal value of $(L\pi)_{s,gu,k,\varepsilon}$, provided that δ_{gz} is still a feasible solution, i.e. $\|g(z-u)\| \leq \varepsilon$. Similarly, we obtain a certificate analogous to (4.17): $-\|gs - gz\|^2 \in \mathrm{SOS}_k(\|gs - gu\|^2 \leq \varepsilon^2) + \mu(f_{gu}(gs) - f_{gu}(gz))$ Thus the RA degree is constant on orbits.

Semicontinuity

4.4.5 Definition: Let $s\colon W \to U$ a polynomial map and $\mathcal{S}$ the s-smooth locus of $\operatorname{im} s$. Then we denote by $\operatorname{radeg}_s\colon \mathcal{S} \to \mathbb{N}$ the function which maps s-smooth points of $\operatorname{im} s$ to their RA degree.

The goal of this section is to show that the function radeg_s is lower semicontinuous, i.e. for any $k \in \mathbb{N}$, the set of points in $\mathcal{S}$ of RA degree at most k forms a closed (semialgebraic) subset of $\mathcal{S}$. This is shown in Proposition 4.4.7. We will need a technical lemma beforehand.

4.4.6 Lemma: Let $k \in \mathbb{N}$ and $\mathcal{A}_p = \{g \geq 0 \mid g \in \mathcal{A}_{p,\geq}\}$ a system of polynomial inequalities on a real affine space U which continuously depend on the parameters p that are constrained to some compact set $\mathcal{M}$. Assume that the values of $f_p \in \mathrm{SOS}_k(\mathcal{A}_p)$ are bounded on the semialgebraic set

$$\mathcal{S}_{\mathcal{A}_p} := \{x \in U \mid x \text{ satisfies } \mathcal{A}_p\} \tag{4.18}$$

and that there exists a set of nonempty interior $\mathcal{S}$ contained in the interior of all $\mathcal{S}_{\mathcal{A}_p}$. Then there exists a constant $C \in \mathbb{R}_{>0}$ such that for any $p \in \mathcal{M}$ and any Sum-of-Squares representation,

$$f_p = \sum_{g \in \mathcal{A}_{\geq,p}} s_g g \tag{4.19}$$

with Sum-of-Squares polynomials $s_g \in \mathrm{SOS}_{k-\deg(g)}(U)$, the coefficients of the s_g are bounded in absolute value by C.

Proof. Wlog let $\mathcal{S}$ be compact, otherwise replace with a compact subset of nonempty interior. Fix $d \in \mathbb{N}$ and choose variables $X = (X_1, \dots, X_n)$ for U, with $n := \dim(U)$. For $\beta \in \mathbb{N}_0^n$ with $|\beta| \leq d$ and $x \in \mathcal{S}$, both the coefficient map $c_\beta \colon \sum_{|\alpha| \leq d} p_\alpha X^\alpha \mapsto p_\beta$, and the evaluation map $\delta_x \colon p \mapsto p(x)$, are linear functionals on $\mathbb{R}[X]_d$. Since by assumption, $\mathcal{S}$ has nonempty interior, no nonzero polynomial can vanish on all points of $\mathcal{S}$. Thus $\{\delta_x \mid x \in \mathcal{S}^\circ\}$ span $\mathbb{R}[X]_d^\vee$ by duality. For $N := \dim(U)$, we may therefore choose $x_1, \dots, x_N \in \mathcal{S}$ such that $\delta_{x_1}, \dots, \delta_{x_N}$ form a basis. By assumption, all f_p are bounded on $\mathcal{S}$ by some constant C', which wlog does not depend on p by compactness of $\mathcal{M}$. Thus for each representation of the kind (4.19), all addends $s_g(x_i)g(x_i)$ are bounded by C', too, for $i \in \{1, \dots, N\}$, since they are all nonnegative. We have $g(x_i) > 0$ due to $x_i \in \mathcal{S}^\circ$. Therefore the $s_g(x_i)$ are bounded by

$$\frac{C'}{\min\{g(x_i) \mid g \in \mathcal{A}_{p,\geq}\}} \tag{4.20}$$

where the minimum exists by compactness of $\mathcal{M}$ and continuous dependency of the inequalities on p. The minimum has to be positive, since the x_i are interior points of $\mathcal{S}_{\mathcal{A}}$ and thus no defining inequality of $\mathcal{S}_{\mathcal{A}_p}$ can vanish on them. Since the evaluations form a basis, we may represent each coefficient map c_β as a linear combination $c_\beta = \sum_{i=1}^N \lambda_i \delta_{x_i}$, where $\lambda_1, \dots, \lambda_N \in \mathbb{R}$ are constants not depending on p but only on the choice of $x_1, \dots, x_N$. Since $\delta_{x_i}(s_g)$ are bounded, so are the $c_\beta(s_g)$. □

4.4.7 Proposition: Let $s \colon W \to U$ a polynomial map. Then radeg_s is a semialgebraic function on U and lower semicontinuous, i.e. for any sequence $(z_n)_{n \in \mathbb{N}}$ in the domain converging to an s-smooth point z of $\operatorname{im} s$, it holds that

$$\mathrm{radeg}(z) \leq \liminf_{n \to \infty} \mathrm{radeg}(z_n) \tag{4.21}$$

Proof. Let $\mathcal{S}$ denote the s-smooth locus of $\operatorname{im} s$. By Theorem 4.3.5, radeg_s attains only finitely many values. The graph of radeg_s is thus the union of the finitely many nonempty fibers $\mathcal{N}_d := \{(z,d) \mid z \in \mathcal{S}, \operatorname{radeg}_s(z) = d\}$, where $d \in \mathbb{N}_0$. For each $d \in \mathbb{N}_0$, it holds that $\mathcal{N}_d \setminus (\bigcup_{1 \le k < d} \mathcal{N}_k)$ is semialgebraic, since it is given by those points $z \in \mathcal{S}$ which satisfy the formulae from (4.15) and (4.17). Thus the graph of radeg_s is a semialgebraic set.

As for semicontinuity, let $\varepsilon > 0$ and $z \in \operatorname{im} s$ be an s-smooth point. Let $\nu \in \mathbb{N}_z \operatorname{im} s$ a normal of length $\delta \in (0, \varepsilon)$. Assume there is a sequence of s-smooth points $(z_n)_{n \in \mathbb{N}}$ in $\operatorname{im} s$ converging to z. Then one may also choose a sequence of length-δ normals $(\nu_n)_{n \in \mathbb{N}}$ such that $\nu_n \to \nu$ as $n \to \infty$. Assume further that for each $n \in \mathbb{N}$ there exist certificates of degree at most $k \in \mathbb{N}$ of the kind (4.15) and (4.17) for (z_n, ν_n) and some fixed $\varepsilon > 0$ *not depending* on n. It suffices to show that the coefficients of such representations are necessarily bounded, for then there exists a convergent subsequence whose limit gives representations of the type (4.15) and (4.17) for (z, ν). To this end, let us apply Lemma 4.4.6: For the semialgebraic sets

$$S_n := \{x \in W \mid \|z_n + \nu_n - s(x)\| \le \varepsilon\} \tag{4.22}$$

there exists $N \in \mathbb{N}_0$ such that $\bigcap_{n \ge N} S_n$ has nonempty interior: Indeed, choose N such that for all $n \ge N$, $\|z_n - z\| \le \varepsilon/4$ and $\|\nu_n - \nu\| \le \varepsilon/4$. Then $\|z_n + \nu_n - s(x)\|$ is upper-bounded by $\|z + \nu - s(x)\| + \varepsilon/2$. Thus any point x satisfying $\|z + \nu - s(x)\| < \varepsilon/2$ will lie in the interior of all S_n. One may clearly even choose an open ball $\mathcal{B}$ in the (open) preimage of $\{y \in U \mid \|z + \nu - y\| < \varepsilon/2\}$ under s whose size does not depend on n. On the closure of the ball $\mathcal{B}$, the polynomials $\|s - z_n - \nu_n\|^2 - \|\nu_n\|^2$ and $-\|s - z_n\|^2$ from (4.15) and (4.17), respectively, are bounded. Thus the condition of Lemma 4.4.6 is fulfilled. Therefore, Sum-of-Squares certificates for their nonnegativity have bounded coefficients and we may thus pass to a convergent subsequence which gives a proof that $(L\pi_{s,z+\nu,k,\varepsilon})$ is exact, too. □

Inner products and linear transformations

4.4.8 Proposition: The RA degree does not depend on the choice of the inner product on U.

Proof. Let $U = \mathbb{R}^N$ for some $N \in \mathbb{N}$ and $H \in S\mathbb{R}^{N \times N}$ a positive definite matrix. Let us endow $\mathbb{R}^N$ both with the standard inner product $\langle \cdot, \cdot \rangle$ and with $\langle x, y \rangle_H := x^T H y$. Fix some s-smooth point $z \in \operatorname{im} s$. First, notice that for any ν which is normal at z w.r.t. $\langle \cdot, \cdot \rangle$, it holds that $H^{-1}\nu$ is normal at z w.r.t. $\langle \cdot, \cdot \rangle_H$. Indeed, for any tangent vector h at z it holds,

$$0 = \langle h, \nu \rangle = \langle h, H^{-1}\nu \rangle_H \tag{4.23}$$

Assume there exists an identity such as (4.15), i.e. there is $z \in \operatorname{im} s$ and a normal $H^{-1}\nu$ with respect to $\langle \cdot, \cdot \rangle_H$, $u := z + H^{-1}\nu$ and $\varepsilon > 0$ such that

$$\|s - z - H^{-1}\nu\|_H^2 - \|H^{-1}\nu\|_H^2 = r_\varepsilon + t_\varepsilon(\varepsilon^2 - \|s - z - H^{-1}\nu\|_H^2) \tag{4.24}$$

The left hand side can be bounded:

$$\begin{aligned} & \|s-z\|_H^2 - 2\langle s-z, H^{-1}\nu\rangle_H \qquad (4.25)\\ & \|s-z\|_H^2 - 2\langle s-z, \nu\rangle \\ \preceq\ & \lambda_{\max}(H)\|s-z\|^2 - 2\langle s-z,\nu\rangle \\ =\ & \lambda_{\max}(H)(\|s-z\|^2 - 2\langle s-z, \frac{1}{\lambda_{\max}(H)}\nu\rangle) \\ =\ & \lambda_{\max}(H)(\|s-z-\tilde{\nu}\|^2 - \|\tilde{\nu}\|^2) \end{aligned}$$

Here, $\lambda_{\max}(H)$ denotes the largest Eigenvalue of H and $\tilde{\nu} := \frac{1}{\lambda_{\max}(H)}\nu$. On the right hand side, note that, $\varepsilon^2 - \|s-z-H^{-1}\nu\|_H^2 \succeq \varepsilon^2 - \lambda_{\max}(H)\|s-z-H^{-1}\nu\|^2$. Furthermore, by the Sum-of-Squares triangle inequality,

$$\begin{aligned} \|s-z-H^{-1}\nu\|^2 &\preceq 2\|s-z-\nu\|^2 + 2\|\nu - H^{-1}\nu\|^2 \\ &\preceq 2\|s-z-\nu\|^2 + 4\|\nu\|^2 + 4\|H^{-1}\nu\|^2 \\ &\preceq 2\|s-z-\nu\|^2 + 4\left(1+\frac{1}{\lambda_{\min}(H)}\right)\|\nu\|^2 \end{aligned}$$

Thus if $8\left(1+\frac{1}{\lambda_{\min}(H)}\right)\|\nu\|^2 < \varepsilon^2$,

$$\varepsilon^2 - \lambda_{\max}(H)\|s-z-H^{-1}\nu\|^2 \succeq \frac{1}{2}\varepsilon^2 - 2\lambda_{\max}(H)\|s-z-\nu\|^2 \qquad (4.26)$$

Thus we found an analogous equation of the type (4.15) for the norm $\|\cdot\|$. The same argument works backwards, i.e. from $\|\cdot\|$ to $\|\cdot\|_H$, and a similar argument works for equations of the type (4.17). □

4.4.9 Proposition: The RA degree is constant under linear transformations $s \mapsto As$, where $A \in \mathrm{GL}(U)$. That is, $\mathrm{radeg}_{As}(Az) = \mathrm{radeg}_s(z)$ for all $z \in \operatorname{im} s$.

Proof. Let $A \in \mathrm{GL}(U), z \in \operatorname{im} s$. To show: $\mathrm{radeg}_{As}(Az) = \mathrm{radeg}_s(z)$. Assume for $k \geq 2\deg(s)$ there exist certificates of exactness of the kind (4.15) and (4.17) for s around z in all normal direction $\nu \in N_z \operatorname{im} s$. Then for all tangent vectors $h \in T_z \operatorname{im} s$, it holds that $0 = \langle \nu, h\rangle = \langle (A^{-1})^T\nu, Ah\rangle$. We have that $T_z \operatorname{im} s \to T_{Az} \operatorname{im} As, h \mapsto Ah$ is an isomorphism of vector spaces. It follows that $N_z \operatorname{im} s \to N_{Az} \operatorname{im} As, \nu \mapsto A^{-1}h$ is an isomorphism, too. Thus, we have to check whether for all $\nu \in N_z \operatorname{im} s$ there exists $\varepsilon > 0$ such that,

$$\|A(s-z) - A^{-1}\nu\|^2 - \|A^{-1}\nu\|^2 \in \mathrm{SOS}_k(\|A(s-z) - A^{-1}\nu\|^2 \leq \varepsilon^2) \qquad (4.27)$$

However, by Proposition 4.4.8, we may change the inner product at will. Thus, take the positive definite matrix $H := (A^{-1})^T A^{-1}$ and replace $\|\cdot\|^2$ in (4.27) by $\|\cdot\|_H = \|A^{-1}\cdot\|$. Since $\|A(s-z) - A^{-1}\nu\|_H^2 = \|s-z-A^{-2}\nu\|$ is lower-bounded by $\|A^{-2}\nu\|^2$ via a Sum-of-Squares certificate from $\mathrm{SOS}_k(\|s-z-A^{-2}\nu\|^2 \leq \varepsilon^2)$ by assumption (due to $\mathrm{radeg}_s(z) \leq k$), the first certificate transfers. The same argument works for a certificate of the type (4.17). Thus $\mathrm{radeg}_{As}(Az) \leq \mathrm{radeg}_s(z)$. Since A is invertible, the argument works forwards as backwards. We conclude $\mathrm{radeg}_{As}(Az) = \mathrm{radeg}_s(z)$. □

The following observation is trivial:

4.4.10 PROPOSITION: The RA degree does not change under shifts, i.e. for any vector $y \in U$ and any polynomial map $s\colon W \to U$, it holds that

$$\forall z \in \operatorname{im} s\colon \operatorname{radeg}_{s+y}(z+y) = \operatorname{radeg}_s(z) \tag{4.28}$$

In particular, s and $s+y$ have the same RA degrees.

Dimension independent RA degrees

Some polynomial maps inherently only make sense on a certain vector space of fixed dimension, with fixed coordinates. E.g. I do not know of a higher-dimensional analogue of the Neil parabola, cf. Figure 4.1. Other maps, such as the *Veronese embedding* $\operatorname{ver}_d\colon U \to S^d(U), x \mapsto x^{\otimes d}$ exist on arbitrary affine spaces U. For such maps, it is reasonable to ask the question how the RA degree scales with the dimension of U. It seems that for a large class of such "coordinate agnostic" maps, called *morphisms of (finite degree) polynomial functors*, the RA degree will eventually become a constant as the dimension of U grows. In fact, one has reason to believe that for this class of maps, even the upper bound from Theorem 4.3.5 is bounded by a constant, i.e. in other words, that the degrees of generating equations for the image set are bounded by some universal constant. This is a direct consequence of recent results for finite-degree polynomial functors, e.g. Noetherianity due to Draisma [28] and (algorithmic) implicitization due to Blatter, Draisma and Ventura [10].

For the precise definition of finite-degree polynomial functors on the category $\mathbf{Vec}_K$ of finite-dimensional vector spaces over some field K and their morphisms, let me refer to [28] and [10] in order to not break the scope. The most important thing to know is that a polynomial functor P maps any finite-dimensional vector space U to a finite-dimensional vector space $P(U)$ and for each linear map $f\colon U \to W$ also induces a map $P(f)\colon P(U) \to P(W)$ that is given by a polynomial of finite degree $d \in \mathbb{N}_0$ not depending on U. In addition, some compatibility conditions are required for P to be a functor, and polynomial functors can also be seen as functors $\mathbf{Vec}_K \to \mathbf{Top}$ into the class of topological spaces, by endowing each $P(U)$ with the Zariski topology. This allows to talk about zero sets of equations as closed subspaces of the functor P. Schur functors are examples of polynomial functors, e.g. the symmetric powers $U \mapsto S^d(U)$ for $d \in \mathbb{N}_0$ or the alternating powers $U \mapsto \bigwedge^d U$ and tensor powers $U \mapsto U^{\otimes d}$. Instead of formally defining morphisms of polynomial functors, let me provide a few examples and a non-examples. For fixed $k, d \in \mathbb{N}_0$, the powers-of-forms variety $V_{k,d}(U)$ from Notation 2.4.1 and Chapter 3 is the image of the morphism $S^k(U) \mapsto S^{kd}(U), q \mapsto q^d$. The map $U^\vee \times S^2(U) \to S^6(U), (\ell, q) \mapsto \ell^6 + 15q\ell^4 + 45q^2\ell^2 + 15q^3$ from (3.27) is a morphism of polynomial functors whose image closure is the degree-6 Gaussian moment variety $\mathrm{GM}_6(U)$. Generally, for any variety which is the image of a morphism $s(U) : P(U) \to Q(U)$ of polynomial functors P and Q and any constant $m \in \mathbb{N}$, the m-th secant variety is the closure of the image of $t(U)\colon P(U)^m \to Q(U), (x_1, \dots, x_m) \mapsto \sum_{i=1}^m s(U)(x_i)$.

For such maps of polynomial functors, there exists a universal degree bound D such that for each $n \in \mathbb{N}_0$, the closure of the image set of the morphism $s(K^n)$ is cut out set-theoretically by equations of degree D. This is formalized in the following theorem.

4.4.11 Theorem: [10, Theorem 1.5.1] Given a morphism $s\colon P \to Q$ of polynomial functors P, Q, there exists a(n affine) vector space U such that for any other space U', the equations for $\operatorname{im} s(U)$ pull back to *set-theoretic* defining equations for $\operatorname{im} s(U')$.

However, there is a catch: A priori the theorem just makes a statement about set-theoretic equations. While for Theorem 4.3.5, we do not strictly need generators of the vanishing ideal, it is necessary to have equations which give the tangent space at a general point in the image set. After some private conversation with the authors of [10], I am mostly convinced that getting tangent space defining equations, and thus a universal constant bounding the RA degree, is possible with their methods.

Note that there are important classes of coordinate-agnostic maps that are not morphisms of finite-degree polynomial functors. E.g. this is the case for functors whose parameters also depend on the dimension. Take e.g. varieties of low-rank matrices, provided the rank is more than a constant: For instance, the image of the map

$$U^{2m} \to U \otimes U, (x_1, y_1, \dots, x_m, y_m) \mapsto \sum_{i=1}^{m} x_i \otimes y_i \tag{4.29}$$

is for any m described by equations of degree $m+1$, but of course the number of ranks m for which this map is interesting increases with m. Plugging in $m = \dim(U) - 1$ will thus not yield a polynomial functor on the left side, and it is for the better, since the degree of equations for the image of (4.29) depends of course on m. In fact, it is classically known that the ideal is generated by the $(m+1) \times (m+1)$ minors. Recent work of Draisma, Kahle and Wiersig [30, Theorem 3.1] even shows that high-degree equations for the image cannot be simple due to sparsity, since all nonzero polynomials in the ideal of the image of (4.29) have at least $(m+1)!$ terms. This motivates the choice of examples in the next section.

4.5. Denoising explicit varieties

The optimization program $(L\pi_{s,u,k,\varepsilon})$ from 4.3.1 may be implemented on a computer and then solved. This section collects some numerical experiments for interesting varieties. All code was written in `Julia` [8]. Modeling was done with the `JuMP` [31] package and the `PolyJuMP` [56] extension. The SDPs were solved with `MOSEK` [66]. Code and results are to be found in [88, `appendices/ra`]. The `data` subfolder contains raw data of the experiments, organized in `.csv` files. The `code` subfolder contains Jupyter/Julia notebooks that were used, e.g. to generate the data.

Semilocal projection instances can in principle be generated in the style of a forward-analysis: Sample some random $x \in W$, compute $s(x)$ and some unit length normal ν at $s(x)$. Choose $\varepsilon > 0$ and let $u := s(x) + \varepsilon\nu$. Then, solve $(L\pi_{s,u,k,2\varepsilon})$ for some degree k and check whether for the computed optimal solution E, $E[s]$ equals $s(x)$.[4] However, in our two specific examples, Remark 4.4.4 justifies to look at *one specific* point z in the image of s instead.

Low-rank psd matrices

As a first example, let us examine the local projection to the set $\mathcal{P}_m(\mathbb{R}^n)$ of rank $m \in \{1, \dots, n-1\}$ psd matrices in $S\mathbb{R}^{n\times n}$, endowed with the Frobenius norm $\|\cdot\|_F \colon A \mapsto \sqrt{\sum_{i,j=1}^n A_{ij}^2}$. As a parametrization, we may take

$$s \colon (\mathbb{R}^n)^m \to S\mathbb{R}^{n\times n}, (x_1, \dots, x_m) \mapsto \sum_{i=1}^m x_i x_i^T \tag{4.30}$$

Note that the Zariski closure of $\operatorname{im} s$ in $S\mathbb{R}^{n\times n}$ is the real variety of matrices $\mathcal{R}_m(\mathbb{R}^n)$ of rank at most m . Thus for a general matrix $M \in S\mathbb{R}^{n\times n}$ sufficiently close to a general point of $\operatorname{im} s$, the projection of M to $\operatorname{im} s$ is equal to the projection of M to $\mathcal{R}_m(\mathbb{R}^n)$. Projections to $\mathcal{R}_m(\mathbb{R}^n)$, even globally, are characterized by the *Eckart-Young-Mirski* theorem: Since $\|\cdot\|_F$ is invariant under both the left and right action of the orthogonal group, one may equivalently first diagonalize $M \approx \operatorname{diag}(\lambda_1, \dots, \lambda_n)$, where $\lambda_1 > \dots > \lambda_n \in \mathbb{R}$ are the ordered Eigenvalues of M, and then project $\operatorname{diag}(\lambda_1, \dots, \lambda_n)$ to $\mathcal{R}_m(\mathbb{R}^n)$. The Eckart-Young-Mirski theorem then states that $\operatorname{diag}(\lambda_1, \dots, \lambda_m, 0, \dots, 0)$ is the (unique) projection, and undoing the diagonalization gets the unique orthogonal projection of M to $\mathcal{R}_m(\mathbb{R}^n)$. In particular, there exists an efficient spectral algorithm that computes the best low-rank approximation of M in polynomial time in the dimension. Therefore, one would hope that quantities measuring the complexity of projection problems should deem this particular problem easy.

The *Euclidean distance (ED) degree* (due to [29], cf. the introduction of this chapter) is such a measure and an invariant of the variety which counts the critical points of projection. However, there are a total of $\binom{n}{m}$ critical points of the Frobenius distance function that can be obtained by choosing any m Eigenvalues of M instead of just the m largest ones, cf. [29, 2.3]. Thus $\binom{n}{m}$ is the ED degree of the variety of symmetric rank-m matrices. For $m = \lceil n/2 \rceil$,

[4] It is possible to tune the ε-parameter of the relaxation in a different way, but I did most experiments with a constant factor of two between the actual distance and the distance parameter.

the ED degree of $\mathcal{R}_m(\mathbb{R}^n)$ grows at least exponentially in n. As a measure of complexity, the ED degree thus overestimates the difficulty of $\mathcal{R}_m(\mathbb{R}^n)$ by quite a bit.

n	m	solver time (s)	$\mathbb{P}(L\pi_{s,z+\frac{\varepsilon}{2}\nu,2,\varepsilon}$ exact)	$\|E[s]-z\|^2$	No. samples
2	1	0.01	1	1.86E-09	2000
3	1	0.01	1	5.31E-10	200
3	2	0.04	1	1.64E-10	1000
4	1	0.01	1	1.67E-10	100
4	2	0.24	1	1.82E-11	100
4	3	7.45	1	2.22E-12	34
5	1	0.02	1	1.47E-10	20
5	2	1.58	1	9.63E-12	10
5	3	64.94	1	8.27E-14	2
5	4	1042.67	1	1.33E-12	2
6	1	0.17	1	2.68E-11	3
6	2	7.57	1	7.55E-12	3
6	3	381.55	1	4.24E-11	3
7	1	0.10	1	3.39E-12	3
7	2	29.63	1	2.51E-10	3

Table 4.1: Numerical experiments indicate that for $m < n$ the Riemannian Approximation degree is always 4, independent of the dimension. All experiments were done with $k = 4$, $z = \sum_{i=1}^{m} e_i e_i^T$ and the ball constraint parameter set to $\varepsilon = 0.4$. The starting points u all satisfied $\|z - u\| = 0.2$ and were generated by uniformly sampling a unit length normal $\nu \in N_z \operatorname{im} s$. The probability in the fourth column refers to the choice of ν. From $n \geq 6$ onwards, some values were skipped because of large solver times and/or memory requirements. Raw data of the experiments may be found in [88, `appendices/ra/data/d2-semilocal`].

On the contrary, the RA degree appears to be constantly 4 for all $1 \leq m < n$: Numerical evidence for this is collected in Table 4.1. An intuitive explanation for this would be that the existence of a spectral algorithm for a problem makes it plausible that a canonical SDP relaxation might succeed. However, the Eckart-Young-Mirski theorem implies that the problem is *globally* easy. Thus we should repeat the same experiment without the ball constraint and see if we observe a change in the outcome. Numerical experiments indicate that this is not the case. Apparently our example problem was “too easy”. Since the table without the ball constraint shows essentially the same outcome as the one with, the experimental results are deferred to [88, `appendices/ra/data/d2-global`].

Undercomplete Waring rank

Thus let us turn to 3-tensors. Any element of the space $S^3(\mathbb{R}^n)$ may be represented by a unique, totally symmetric 3-tensor $t \in \mathbb{R}^{n\times n\times n}$, for which there exists a canonical generalization of the Frobenius norm, $\|t\|_F := \sqrt{\sum_{i,j,k=1}^{m} t_{ijk}^2}$. Endowing $S^3(\mathbb{R}^n)$ with that, we may consider projections to the secants of the Veronese variety $V_{1,3}(\mathbb{R}^n)$. For $m \in \mathbb{N}$, the m-th real affine secant variety is the

Zariski closure in $S^3(\mathbb{R}^n)$ of the image of

$$s\colon ((\mathbb{R}^n)^\vee)^m \to S^3(\mathbb{R}^n), (\ell_1, \ldots, \ell_m) \mapsto \sum_{i=1}^m \ell_i^3 \tag{4.31}$$

As in the previous section, for $m \leq n$ it suffices to look at $z = \sum_{i=1}^m e_i^{\otimes 3}$ by Remark 4.4.4. Table 4.2 shows that the map s from (4.31) has minimal RA degree 6 for $m = 1$ and $n \in \{2, 3, 4\}$. On the other side, for $m = 2$ and $n \in \{3, 4\}$, the map s does not have minimal RA degree.

n	m	solver time (s)	$\mathbb{P}(L\pi_{s,z+\frac{\varepsilon}{2}\nu,3,\varepsilon}$ exact)	$\Vert E[s] - z\Vert^2$	No. samples
2	1	0.04	1	2.95E-10	20
3	1	0.02	1	1.57E-09	20
3	2	10.00	0	1.96E-04	20
3	3	1400.80	0	4.00E-04	3
4	1	1.76	1	4.08E-09	10
4	2	230.90	0	4.14E-05	1

Table 4.2: Numerical experiments indicate that for $m > 1$, the Riemannian Approximation degree of Waring rank m forms is not minimal. All experiments were done with $k = 3$, $z = \sum_{i=1}^m e_i^{\otimes 3}$ and the ball constraint parameter set to $\varepsilon = 0.04$. The starting points u all satisfied $\Vert z - u\Vert = 0.02$ and were generated by uniformly sampling a unit length normal $\nu \in N_z \operatorname{im} s$. The probability in the fourth column refers to the choice of ν. Raw data of the experiments may be found in [88, `appendices/ra/data/d3-semilocal-k3`].

Let us speculate about potential explanations. Note that there is evidence that some normal directions might be easier than others: The Frobenius norm does not change under flattenings, e.g. if we rearrange the entries of some $t \in \mathbb{R}^{n\times n\times n}$ to a matrix $M_t \in \mathbb{R}^{n\times n^2}$ with $(M_t)_{i,(j,k)} = t_{ijk}$. Any element of $\sigma_m(V_{1,3}(\mathbb{R}^n))$ has matrix rank at most m when flattened. Now, observe that our specific z is *orthogonally decomposable*, i.e. the rank-one components are orthogonal. Thus the matrix flattening $M_z = \sum_{i=1}^m x_i \otimes (x_i x_i^T)$ has 1 and 0 as m-fold and $(n - m)$-fold singular values, respectively. For $m < n$, there exists a normal which is a power of a linear form: Indeed, take $y \in \langle e_1, \ldots, e_m\rangle^\perp$ of unit length and note that $\nu := y^{\otimes 3} \in S^3(\mathbb{R}^n)$ is normal to $\operatorname{im} s$ at z. For any $\varepsilon < 1$, it holds that M_z is the orthogonal projection of $M_{z+\varepsilon\nu}$ to the set of matrices of rank at most m by the Eckart-Young-Mirski theorem (for non-symmetric matrices). Of course, the previous section was about symmetric matrices, but if the reader believes that the RA degree of non-square matrices should be "easy", too, then it seems plausible that there exists at least one "easy" normal direction.

The question remains at what degrees we can get rid of more normal directions. Further numerics, to be found in [88, `ra/data/d3-semilocal-k4`], indicate that degree 8 is still not sufficient to denoise a rank-2 secant of $V_{1,3}$. There exist Sum-of-Squares based noise stable algorithms for *orthogonal* tensor decomposition, e.g. due to Ma, Shi and Steurer [60]. At the same time, the variety of orthogonal tensors has a much simpler structure. E.g., the (complex) Zariski closure of the set of real orthogonally decomposable tensors is cut out up to irreducible component by Robeva's degree-2 equations (originally in [79, Lemma 3.7], but perhaps [9, Proposition 4.2.1] states more directly that we get the tangent space). This was already mentioned in Remark 2.5.17 and both [9, Proposition 4.2.1] and [60] can e.g. be separately used to see that semi-local projection is tractable for orthogonally decomposable tensors. Perhaps surprisingly, for $m \leq n$, the general element t of $\sigma_m(V_{1,3})$ can be transformed into an orthogonally decomposable one via a procedure called *whitening*. This method uses the pseudoinverse of some maximum rank psd matrix in the matrix space defined by t and did already occur earlier in this thesis, namely in the second proof of Theorem 2.4.8. It is not at all clear whether the Sum-of-Squares method captures whitening in low-degrees. The whitening transform is also generally not orthogonal, unless it is trivial. The degree bound that Robeva's equations give together with Theorem 4.3.5 would be $2dD = 2 \cdot 3 \cdot 2 = 12$, but of course Robeva's equations do not lie in the vanishing ideal of $\operatorname{im} s$, but we only have that after pulling them back via s, they lie in the ideal of the variety $A \subseteq ((\mathbb{R}^n)^\vee)^m$ of orthogonal systems of m linear forms. Note that whenever an element $t \in \sigma_m(V_{1,3})$ is orthogonally decomposable (and smooth in both varieties), $N_t\sigma_m(V_{1,3})$ is contained in the normal space to the variety of orthogonally decomposable tensors. This does not imply exactness of $(L\pi)$ in degree 12, but it makes it somewhat plausible that something could happen at that threshold. Unfortunately, solving degree 12 Sum-of-Squares programs is out of computational budget, at least for me.

"I don't see how he can ever finish (...)"

— Alice's Adventures in Wonderland, [18]

5 Singular Matrix Spaces

Part of this chapter, predominantly the introduction, intersects with a previously published article, see [92].

Motivation

This last chapter is dedicated to the study of some curious objects that have been present all throughout this thesis, while staying modestly in the back. It is therefore time that we direct some attention to: *matrix spaces.* Let us briefly discuss how ubiquitous these were in the previous chapters: E.g. all three algorithms that were given for Theorem 2.4.8 have in common that they work with the matrix space $\langle \ell_1^{\otimes 2}, \dots, \ell_m^{\otimes 2} \rangle$, or, equivalently, the space $\langle \ell_1^2, \dots, \ell_m^2 \rangle$ of quadratic forms.

Every Tensor $T \in \mathbb{R}^n \otimes \mathbb{R}^n \otimes \mathbb{R}^m$ defines a space of matrices, by expanding $T = \sum_{i=1}^m A_i \otimes e_i$ for some $A_i \in \mathbb{R}^n \otimes \mathbb{R}^n \cong \mathbb{R}^{n \times n}$ and taking the space $\langle A_1, \dots, A_m \rangle$ spanned by the *slices* $A_1, \dots, A_m$. Conversely, from every matrix space, we obtain an orbit of 3-tensors under the action of $\mathrm{GL}(\mathbb{R}^m)$ whose elements $\sum_{i=1}^m A_i \otimes e_i$ correspond to the bases $A_1, \dots, A_m$ of the space. Sum-of-Squares optimization is essentially about finding the psd matrices in a given matrix space.

Another problem is to check whether a given matrix space contains *only singular* matrices. That is, given a tuple of matrices $A = (A_1, \dots, A_m)$, decide with a *deterministic* algorithm whether

$$0 \overset{?}{=} \det\left(\sum_{i=1}^m X_i A_i\right) \in \mathbb{R}[X_1, \dots, X_m] \tag{5.1}$$

is the zero polynomial in variables $X_1, \dots, X_m$. This problem, called *determinant identity testing* (DIT) is somewhat fundamental to circuit complexity, cf. e.g. [17] and [92, Introduction] and also to derandomization, cf. [47]. It is astonishing how easy the problem becomes once the requirement of a *deterministic* algorithm is dropped. Indeed, from the Schwartz-Zippel Lemma ([84],

[93]), it follows the existence of an efficient *probabilistic* algorithm: Evaluate $A(X) := \sum_{i=1}^{m} X_i A_i$ in a few random points x from a hypercube S^m, where S is a set with more than n elements. Compute $\det(A(x))$ for these points with standard methods of Linear Algebra. The Schwartz-Zippel lemma asserts that a nonzero n-variate degree-d polynomial has at most $d|S|^{n-1}$ roots in S^n. Thus, choosing e.g. $|S| = 2n$, $\det(A(X))$ will not vanish on a uniformly random $x \in S^n$ with probability at least $\frac{1}{2}$, unless $\det(A(X)) = 0$. Therefore checking on a few independently random points x efficiently distinguishes between singular matrix tuples and nonsingular matrix tuples with reliably high probabilty.

There are many interesting geometrical questions associated with matrix spaces that are quite poorly understood. Specifically, (DIT) is about the membership of $(n \times n)$-matrices $A_1, \ldots, A_m$ in the (complex) variety $\mathrm{SING}_{\mathbb{C},n,m}$ given by the equations

$$\forall \lambda \in \mathbb{Q}^m \colon \ \det(\sum_{i=1}^{m} \lambda_i A_i) = 0 \tag{5.2}$$

As this variety is invariant under a $\mathrm{GL}_n \times \mathrm{GL}_n \times \mathrm{GL}_m$ action defined by left-right-back multiplication on the matrix tuple $(A_1, \ldots, A_m)$, people studied invariant theoretical aspects, e.g. Makam and Wigderson found out that this variety is not a null cone [61], [62] and argued that this can be seen as evidence for the difficulty of (DIT). Perhaps surprisingly, the equations from (5.2) (when seen with the entries of the $A_1, \ldots, A_m$ as unknowns) will in general not define a radical ideal. In [92], Vill and Michałek and I found polynomials that vanish on $\mathrm{SING}_{\mathbb{C},n,m}$ but do not lie in the ideal generated by the forms from (5.2). These additional equations arise from the 3-tensor structure of the matrix tuple $(A_1, \ldots, A_m)$.

From a more philosophical standpoint, people questioned how essential randomness is as a resource to algorithms and how exactly complexity classes for randomized algorithms such as BPP (bounded-error probabilistic polynomial time) relate to the complexity classes for deterministic machines. Results were established which showed connections to hardness results in circuit complexity (e.g. that certain one-way functions whose inverses had high circuit complexity could be used as a source of pseudorandomness), cf. e.g. the references in [47, Introduction]. On the converse side, a celebrated result of Kabanets and Impagliazzo [47] showed that any subexponential time deterministic algorithm for (DIT) would either show superexponential circuit lower bounds for the permanent (and thus VP≠VNP, cf. [91], [61]) or that NEXP is computable by polynomially-sized boolean circuits. *Derandomization* is, vaguely spoken, the attempt to replace random decisions in algorithms by deterministic ones. In the spirit of derandomization, it appears natural to replace sampling by optimization and the samples from a probability distribution μ by functionals $\mathbb{E}_\mu$ that represent the distribution itself. For tractability, of course one has to make significant compromises. One possible compromise is to replace expectation operators of measures by pseudo-expectation operators, cf. Proposition and Definition 2.3.2. The first hope is that a Sum-of-Squares relaxation of sufficiently high degree can provably solve (DIT), even if it gives no improvement on the complexity of naively expanding the determinant. It will turn out that this hope is justified, as we will see in Theorem 5.2.3.

This approach will yield an alternative quantification of the complexity of singular matrix tuples: Not via equations, but via certain Sum-of-Squares inequalities and hierarchies of Sum-of-Squares feasibility problems. There are some very modest results obtained: I characterize the singular matrix spaces detected by the first level of the hierarchy in Proposition 5.2.2 and show that there is a "last" level that detects all elements of $\mathrm{SING}_{\mathbb{C},n,m}$ (Theorem 5.2.3, which can be interpreted as a finite convergence result). Most interesting problems are left open, e.g: Which are the singular matrix spaces that get detected by the *second* level of the hierachy? These could give a class of still relatively easy singular matrix spaces that is not as trivial as that of the first level. The semidefinite framework also suggests to understand the relation between (DIT) and some natural associated quadratic optimization problem, which is is to find the maximum over $x \in \mathbb{S}^{m-1}$ over all minimum singular values of $A(x)$.

5.1. Matrix Sum-of-Squares Cones

5.1.1 DEFINITION: Let $n, k \in \mathbb{N}$ and $A, B \in \mathbb{R}[X]^{n\times n}$ symmetric matrix polynomials. As an ad-hoc notation for this chapter, we define $\mathrm{SOS}^*_k(A \succeq 0)$ to be the convex cone

$$\{E \in \mathrm{SOS}^*_k(X) \mid \forall p \in \mathbb{R}[X]^n : (\deg(p^T A p) \leq k \implies E[p^T A p] \geq 0)\}$$

and

$$\mathrm{SOS}^*_k(A = 0) := \{E \in \mathrm{SOS}^*_k(X) \mid \forall q \in \mathbb{R}[X] : (\deg(qA) \leq k \implies E[qA] = 0)\}$$

Furthermore, $\mathrm{SOS}^*_k(A \succeq B) := \mathrm{SOS}^*_k(A - B \succeq 0)$ and for several (matrix) in:equality constraints $c_1, \dots, c_m, m \in \mathbb{N}_0$ we define

$$\mathrm{SOS}^*_k(c_1, \dots, c_m) := \mathrm{SOS}^*_k(c_1) \cap \dots \cap \mathrm{SOS}^*_k(c_m)$$

In all of the above, the *degree* of a matrix polynomial is to be understood as the maximum degree of its entries.

5.1.2 REMARK: Let n and A as in Definition 5.1.1. For any $k \in \mathbb{N}$, $\mathrm{SOS}^*_k(A \succeq 0)$ is a spectrahedral cone. Seeing this requires a moderately long argument, which is generously left to the reader. From the notation, one might be tempted to think that $\mathrm{SOS}^*_k(A \succeq 0)$ should be seen as the class of square-definite functionals "respecting" the formal matrix inequality $A \succeq 0$, similar in spirit to what we did in Section 2.3. However, note that in the non-matrix case, the Positivstellensätze discussed in Section 2.3 prove that Lasserre's semidefinite approximations of the cone of nonnegative polynomials on a set are much more than a heuristic, and have in fact some convergence properties as $k \to \infty$. These Positivstellensätze exist for matrix polynomials, too, [27] but they require slightly more than just conjugation by vectors of polynomials. E.g. one needs to allow conjugation by matrix polynomials, too. For our limited purposes, vector conjugation will turn out to be enough, so it is reasonable to take the simpler option, since it yields an SDP with less variables.

5.2. Singular Matrix Tuples

Throughout, let $n \in \mathbb{N}, m \in \mathbb{N}_0$. For any field K, denote by

$$\mathrm{SING}_{K,n,m} = \{A \in (K^{n\times n})^m \mid \forall \lambda_1, \ldots, \lambda_m \in K : \det(\sum_{i=1}^{m} \lambda_i A_i) = 0\}$$

the variety of singular matrix tuples. If no field is specified, we assume $K = \mathbb{R}$ and thus write $\mathrm{SING}_{n,m} := \mathrm{SING}_{\mathbb{R},n,m}$.

5.2.1 Definition: Each matrix tuple A defines a unique linear matrix polynomial

$$A(X) := A_1X_1 + \ldots + A_mX_m \in \mathbb{R}[X]$$

in variables $X = (X_1, \ldots, X_m)$. The notation $A(X)$ might be considered slight abuse of notation, as for $x \in K^m$ we also denote $A(x)$ for the evaluation of $A(X)$ in x. $A \in (K^{n\times n})^m$ is called a *singular matrix tuple*, if $A \in \mathrm{SING}_{K,n,m}$. A matrix polynomial $B \in K[X]^{n\times n}$ is called (completely) *singular*, if for all $x \in K^n$, $B(x)$ is singular.

Over the real field, testing singularity of a matrix tuple is captured by the polynomial optimization problem

$$\begin{array}{lll} (\mathrm{SING})_A & \text{maximize} & \lambda \\ & \text{subject to} & \lambda \in \mathbb{R}, x \in \mathbb{R}^n \\ & & A(x)^T A(x) \succeq \lambda I_n \\ & & \|x\|^2 = 1 \end{array} \tag{5.3}$$

Clearly, for each $A \in (\mathbb{R}^{n\times n})^m$, the optimal value λ^* of $(\mathrm{SING})_A$ is nonnegative. A is a singular matrix tuple if and only if $\lambda^* = 0$. It is natural to consider Sum-of-Squares Relaxations for the problem, e.g. the degree-2 relaxation

$$\begin{array}{lll} (\mathrm{LSING})_A & \text{maximize} & \lambda \\ & \text{subject to} & \lambda \in \mathbb{R}, E \in \mathbb{R}[X]^*_{\leq 2} \\ & & E[X^T X] = 1 \\ & & E[A(X)^T A(X)] \succeq \lambda I_n \\ & & E[1] = 1 \end{array} \tag{5.4}$$

As it turns out, this relaxation detects only a small but very explicit subset of $\mathrm{SING}_{\mathbb{R},n,m}$.

5.2.2 Proposition: Let $A \in (\mathbb{R}^{n\times n})^m$. Then $(\mathrm{LSING})_A$ has a feasible solution (E, λ) with $\lambda > 0$ if and only if

$$\bigcap_{x\in\mathbb{R}^n} \ker A(x) = \ker A_1 \cap \ldots \cap \ker A_m = \{0\}$$

Proof. First, let (E, λ) a feasible solution with $\lambda > 0$. Let $v \in \bigcap_{x\in\mathbb{R}^n} \ker A(x)$. This means that $A(X)v = 0$ is the vector of zero polynomials. Now, as $E[A(X)^T A(X)] \succeq \lambda I_n$, we have

$$0 = E[v^T A(X)^T A(X)v] = v^T E[A(X)^T A(X)]v \geq \lambda v^T v$$

Since $\lambda > 0$, this implies $v^T v = 0$ and therefore $v = 0$.

For the other direction, let $\ker A_1 \cap \ldots \cap \ker A_m = \{0\}$. Choose $E \in \mathbb{R}[X]^*_{\leq 2}$ such that

$$\begin{aligned} E[XX^T] &= \frac{1}{n} I_n \\ E[X] &= 0 \\ E[1] &= 1 \end{aligned}$$

It is a short exercise to verify that this is square-definite and satisfies the constraint $E[X^T X] = 1$. Writing out

$$A(X)^T A(X) = \sum_{i=1}^{m} X_i^2 A_i^T A_i + \sum_{\substack{i,j=1 \\ i<j}}^{m} X_i X_j (A_i^T A_j + A_j^T A_i)$$

we obtain

$$E[A(X)^T A(X)] = \frac{1}{n} \sum_{i=1}^{m} A_i^T A_i$$

Clearly, $\ker(A_1^T A_1 + \ldots + A_m^T A_m) = \ker A_1 \cap \ldots \cap \ker A_m = \{0\}$. Therefore, choosing λ as the minimum eigenvalue of $A_1^T A_1 + \ldots + A_m^T A_m$, which must be strictly positive, we obtain a pair (E, λ) that is feasible for $(\mathrm{LSING})_A$ as $E[A(X)^T A(X)] - \lambda I_n = (A_1^T A_1 + \ldots + A_m^T A_m) - \lambda I_n$ is positive semidefinite by construction. □

From a practical perspective, this Proposition 5.2.2 is not very interesting, as detecting whether $A_1, \ldots, A_m$ have a common kernel element may also be done with Linear Algebra. In light of this, it should be apparent that it is reasonable to look at “higher order” relaxations of $(\mathrm{SING})_A$. It is not completely apparent how such a relaxation should look like, but we will provide one by ad-hoc improvisation. Note that while our target is to detect membership in $\mathrm{SING}_{\mathbb{R},n,m}$, the optimization problem $(\mathrm{SING})_A$ actually computes $\max\{\sigma_{\min}(A(x))^2 \mid x \in \mathbb{S}^{n-1}\}$, where for a matrix $M \in \mathbb{R}^{n \times n}$, $\sigma_{\min}(M)$ denotes the minimum singular value of M. It is not clear whether computing the optimal value and checking whether it is nonzero are problems of similar computational complexity. As an intuitive argument against, note that there is an obvious syntactic certificate that $A^T A$ has nonnegative Eigenvalues at each point, but even if the optimal value is positive, a certificate of the kind $A^T A - \lambda I_n = BB^T$ for some matrix polynomial B need not necessarily exist. For the higher order relaxations, it actually makes it easier for our purpose to see them as feasibility problems rather than optimization problems.

For $k \in \mathbb{N}_0$ and $\varepsilon \in \mathbb{R}_{>0}$, define the feasibility problem

$$\begin{array}{lll} (\mathrm{LSING})_{A,k,\varepsilon} & \text{find} & E \in \mathrm{SOS}_k^*(X^T X = 1, A(X)^T A(X) \succeq \varepsilon I_n) \\ & \text{subject to} & E[1] = 1 \end{array} \tag{5.5}$$

This optimization problem can be solved by Semidefinite Programming. We claim that for sufficiently large relaxation parameter k, $(\mathrm{LSING})_{A,k,\varepsilon}$ can distinguish between the case where $A \in \mathrm{SING}_{\mathbb{R},n,m}$ and the case where

$$\max\{\sigma_{\min}(A(x))^2 \mid x \in \mathbb{S}^{n-1}\} \geq \varepsilon$$

5.2.3 Theorem: Let $A \in (\mathbb{R}[X]^{n\times n})^m$ a linear matrix polynomial and $k \geq 2n$. Then, the following are equivalent:

(a) There exists some $\varepsilon \in \mathbb{R}_{>0}$ and $E \in \mathrm{SOS}_k^*(X^TX = 1, A(X)^TA(X) \succeq \varepsilon I_n)$ with $E[1] = 1$.

(b) $A \notin \mathrm{SING}_{n,m}$.

Proof. **(a)** $\Longrightarrow$ **(b)**: Let $\varepsilon \in \mathbb{R}_{>0}$ and $E \in \mathrm{SOS}_k^*(X^TX = 1, A(X)^TA(X) \succeq \varepsilon I_n)$. To the contrary, assume $A \in \mathrm{SING}_{n,m}$ and thus $\det(A(X)) = 0$. Clearly, for $P := \mathrm{adj}(A(X))$, which is a matrix of rank $n-1$, we have $A(X)P = \det(A(X))I_n = 0$, as $A \in \mathrm{SING}_{n,m}$. By assumption on E, it holds that for every column p of P:

$$0 = E[\det(A(X))^2] = E[p^TA(X)^TA(X)p] \geq \varepsilon E[p^Tp]$$

Therefore, $E[p^Tp] = 0$. But p^Tp is a sum of squares of $(n-1)\times(n-1)$ minors of $A(X)$. As p was an arbitrary column of the adjugate matrix of $A(X)$, this means that for all $(n-1)\times(n-1)$ minors m of $A(X)$, $E[m^2] = 0$. We will construct a set of vectors $\mathcal{Q}$ with the properties that

(a) For $p \in \mathcal{Q}$, all entries of p are either $(n-2)\times(n-2)$ minors of A or 0.

(b) All $(n-2)\times(n-2)$ minors of $A(X)$ occur as entries of some element of $\mathcal{Q}$.

(c) For $p \in \mathcal{Q}$, all entries of $A(X)p$ are $(n-1)\times(n-1)$ minors of A or 0.

Let us first see how the existence of such a set will yield the claim. By the third property, we will get that $0 = E[p^TA(X)^TA(X)p] \geq \varepsilon E[p^Tp]$ for each $p \in \mathcal{Q}$. By the two other properties, we can inductively continue this process and eventually deduce that $E[A(X)_{ij}^2] = 0$ for each $i,j \in \{1,\dots,n\}$. This is a clear contradiction, as then $0 = E[\mathrm{tr}(A(X)^TA(X))] \geq E[\varepsilon \mathrm{tr}(I_n)] = n\varepsilon$. As for the set $\mathcal{Q}$, look at any $(n-1)\times(n-1)$ submatrix B of $A(X)$. By Laplace expansion, we may for each $i \in \{1,\dots,n-1\}$ write $\det(B)$ as a sum $\det(B) = \sum_{j=1}^{n-1} B_{ij}m_{ij}$, where, up to sign, m_{ij} are $(n-2)\times(n-2)$ minors of B and thus also of $A(X)$. Let

$$p_{B,i} := \begin{pmatrix} m_{i1} \\ \vdots \\ m_{i(n-1)} \end{pmatrix}, \quad (B\ (n-1)\text{-sized submatrix of } A(X), i \in \{1,\dots,n-1\})$$

Now, observe that $Bp_{B,i}$ can have at most one nonzero entry: By construction, clearly we have that $(Bp_{B,i})_i = \det(B)$. But for $k \in \{1,\dots,n-1\}$ with $k \neq i$, we must have $(Bp_{B,i})_k = 0$: Indeed, it holds that

$$(Bp_{B,i})_k = \sum_{j=1}^{m} B_{kj}m_{ij} = \pm\det(\tilde{B})$$

where $\tilde{B}$ arises from B by replacing the i-th row with the k-th row of B. But $\tilde{B}$ has two equal rows and thus $\det(\tilde{B}) = 0$. Now, remember that as an $(n-1)\times(n-1)$ submatrix of $A(X)$, B was obtained from $A(X)$ by deleting some row, say the ℓ-th row for some $\ell \in \{1,\dots,n\}$ and some column, say $k \in \{1,\dots,n\}$. After padding with one zero entry at the k-th position, we can obtain a vector $q_{B,i}$ from $p_{B,i}$ such that $A(X)q_{B,i} = \pm\det(B)e_i \pm \det(\tilde{A})e_\ell$, for some linear matrix polynomial $\tilde{A}$, where e_i, e_ℓ denote the i-th and ℓ-th standard basis vectors, respectively: (Here wlog $i < \ell$, otherwise it is $A(X)q_{B,i} = \pm\det(B)e_{i+1} \pm \det(\tilde{A})e_\ell$). The new entry $\pm\det(\tilde{A})e_\ell$ is the result of multiplying the ℓ-th row of $A(X)$ with $q_{B,i}$. By Laplace expansion, we see that it is also the determinant of some matrix $\tilde{A}$ which is obtained from $A(X)$ by deleting the k-th column and i-th row. Thus it is in fact another $(n-1)\times(n-1)$ minor of $A(X)$. We conclude that

$$\mathcal{Q} := \{q_{B,i} \mid B\ (n-1)\text{-sized submatrix of } A(X), i \in \{1,\dots,n-1\}\}$$

is a set as desired.

(b) $\Longrightarrow$ (a): Choose $x \in \mathbb{S}^{n-1}$ such that $A(x)$ is nonsingular. Let $E_x\colon \mathbb{R}[X] \to \mathbb{R}, p \mapsto p(x)$ and let ε the smallest Eigenvalue of $A(x)^TA(x)$. Then E_x satisfies $E_x[X^TX] = 1$ and $E_x[A(X)^TA(X)] \succeq \varepsilon I_n$. As E_x is in fact a ring homomorphism, this directly implies that for all $k \in \mathbb{N}_{\geq 2}$,

$$E_x \in \mathrm{SOS}_k^*(X^TX = 1, A(X)^TA(X) \succeq \varepsilon I_n)$$

□

Let us end by formulating an open problem.

5.2.4 Problem: Characterize the set of matrix tuples A such that

$$\bigcap_{\varepsilon\in\mathbb{R}_{>0}} \mathrm{SOS}_{2k}^*(X^TX = 1, A(X)^TA(X) \succeq \varepsilon I_n) = \{0\} \tag{5.6}$$

for $k = 2$.

Theorem 5.2.3 and Proposition 5.2.2 characterize the set of matrix tuples such that (5.6) holds for $k \geq n$ and $k = 1$, respectively. This should just be seen as the very beginning, as both of these results do not give any interesting algorithmic consequences: The high-degree relaxations are computationally intractable whereas the class detected by the first relaxation can also be detected just with Linear Algebra. What remains open and is really interesting is whether e.g. the degree-4 relaxation has a concise and explicit description that is similarly easy to understand.

Bibliography

[1] Amendola, C., Faugere, J.-C., and Sturmfels, B. Moment varieties of Gaussian mixtures. *Journal of Algebraic Statistics 7*, 1 (2016).

[2] Améndola, C., Ranestad, K., and Sturmfels, B. Algebraic Identifiability of Gaussian Mixtures. *International Mathematics Research Notices 2018*, 21 (Nov. 2018), 6556–6580.

[3] Anandkumar, A., Ge, R., Hsu, D., Kakade, S. M., and Telgarsky, M. Tensor decompositions for learning latent variable models. *Journal of Machine Learning Research* (2012).

[4] Arthurs, N., Stenhaug, B., Karayev, S., and Piech, C. Grades are not normal: Improving exam score models using the logit-normal distribution. In *EDM* (2019), C. F. Lynch, A. Merceron, M. Desmarais, and R. Nkambou, Eds., vol. 12 of *Proceedings of the International Conference on Educational Data Mining (EDM)*, International Educational Data Mining Society, pp. 252–257.

[5] Bafna, M., Hsieh, T., Kothari, P., and Xu, J. Polynomial-time power-sum decomposition of polynomials. *2022 IEEE 63st Annual Symposium on Foundations of Computer Science (FOCS)* (2022), (to appear).

[6] Bandeira, A. S., Blum-Smith, B., Kileel, J., Perry, A., Weed, J., and Wein, A. S. Estimation under group actions: recovering orbits from invariants, 2017. arXiv:1712.10163 [math.ST].

[7] Beltrán, C., Breiding, P., and Vannieuwenhoven, N. The average condition number of most tensor rank decomposition problems is infinite. *Foundations of Computational Mathematics* (2022).

[8] Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. Julia: A fresh approach to numerical computing. *SIAM Review 59*, 1 (2017), 65–98.

[9] Biaggi, B., Draisma, J., and Seynnaeve, T. On the quadratic equations for odeco tensors, 2022. arXiv:2206.01521 [math.AG].

[10] Blatter, A., Draisma, J., and Ventura, E. Implicitisation and parameterisation in polynomial functors, 2022. arXiv:2206.01555 [math.AG].

[11] Blekherman, G., Smith, G. G., and Velasco, M. Sharp degree bounds for sum-of-squares certificates on projective curves. *Journal de Mathématiques Pures et Appliquées 129* (2019), 61–86.

[12] Blomenhofer, A. T., Casarotti, A., Oneto, A., and Michałek, M. Identifiability for mixtures of centered Gaussians and sums of powers of quadratics, 2022. (under review). Preprint: arXiv:2204.09356 [math.AG].

[13] Bogačev, V. I. *Measure theory.* Springer, 2007.

[14] Breiding, P., and Vannieuwenhoven, N. The condition number of join decompositions. *SIAM Journal on Matrix Analysis and Applications 39*, 1 (2018), 287–309.

[15] Breiding, P., and Vannieuwenhoven, N. On the average condition number of tensor rank decompositions. *IMA Journal of Numerical Analysis 40*, 3 (2019), 1908–1936.

[16] Breiding, P., and Vannieuwenhoven, N. The condition number of riemannian approximation problems. *SIAM Journal on Optimization 31*, 1 (2021), 1049–1077.

[17] Bürgisser, P. *Completeness and reduction in algebraic complexity theory.* Springer, 2011.

[18] Carroll, L. *Alice's Adventures in Wonderland.* T. Y. Crowell & co, 1893.

[19] Casarotti, A., and Mella, M. From non-defectivity to identifiability. *Journal of the European Mathematical Society* (2022).

[20] Chiantini, L., and Ciliberto, C. Weakly defective varieties. *Transactions of the American Mathematical Society 354*, 1 (2001), 151–178.

[21] Chiantini, L., and Ottaviani, G. On generic identifiability of 3-tensors of small rank. *SIAM Journal on Matrix Analysis and Applications 33*, 3 (2012), 1018–1037.

[22] Chiantini, L., Ottaviani, G., and Vannieuwenhoven, N. An algorithm for generic and low-rank specific identifiability of complex tensors. *SIAM Journal on Matrix Analysis and Applications 35*, 4 (2014), 1265–1287.

[23] Chiantini, L., Ottaviani, G., and Vannieuwenhoven, N. On generic identifiability of symmetric tensors of subgeneric rank. *Transactions of the American Mathematical Society 369*, 6 (2016), 4021–4042.

[24] Cifani, M. G., Cuzzucoli, A., and Moschetti, R. Monodromy of projections of hypersurfaces. *Annali di Matematica Pura ed Applicata (1923 -) 201*, 2 (2021), 637–654.

[25] Cifuentes, D., Agarwal, S., Parrilo, P. A., and Thomas, R. R. On the local stability of semidefinite relaxations. *Mathematical Programming 193*, 2 (2021), 629–663.

[26] di Dio, P. J., and Schmüdgen, K. The multidimensional truncated moment problem: The moment cone. *Journal of Mathematical Analysis and Applications 511*, 1 (2022), 126066.

[27] Dinh, T. H., Ho, T., and Le, C. Positivstellensätze for polynomial matrices. *Positivity 25* (09 2021).

[28] Draisma, J. Topological noetherianity of polynomial functors. *Journal of the American Mathematical Society 32*, 3 (2019), 691–707.

[29] Draisma, J., Horobeţ, E., Ottaviani, G., Sturmfels, B., and Thomas, R. R. The Euclidean distance degree of an algebraic variety. *Foundations of Computational Mathematics 16*, 1 (2015), 99–149.

[30] Draisma, J., Kahle, T., and Wiersig, F. No short polynomials vanish on bounded rank matrices, 2021. arXiv:2112.11764 [math.AC].

[31] Dunning, I., Huchette, J., and Lubin, M. JuMP: A modeling language for mathematical optimization. *SIAM Review 59*, 2 (2017), 295–320.

[32] Fonda, A., and Gidoni, P. Generalizing the Poincaré–Miranda theorem: The avoiding cones condition. *Annali di Matematica Pura ed Applicata 195*, 4 (2015), 1347–1371.

[33] Fröberg, R. An inequality for Hilbert series of graded algebras. *Mathematica Scandinavica 56*, 2 (1985), 117–144.

[34] Fröberg, R., Ottaviani, G., and Shapiro, B. On the Waring problem for polynomial rings. *Proceedings of the National Academy of Sciences 109*, 15 (2012), 5600–5602.

[35] Garg, A., Gurvits, L., Oliveira, R., and Wigderson, A. Operator scaling: Theory and applications. *Foundations of Computational Mathematics 20* (05 2019), 1–68.

[36] Garg, A., Kayal, N., and Saha, C. Learning sums of powers of low-degree polynomials in the non-degenerate case. *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)* (2020).

[37] Ge, R., Huang, Q., and Kakade, S. M. Learning mixtures of Gaussians in high dimensions. *Proceedings of the forty-seventh annual ACM symposium on Theory of Computing* (2015).

[38] Ge, R., and Ma, T. Decomposing Overcomplete 3rd Order Tensors using Sum-of-Squares Algorithms. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2015)* (Dagstuhl, Germany, 2015), N. Garg, K. Jansen, A. Rao, and J. D. P. Rolim, Eds., vol. 40 of *Leibniz International Proceedings in Informatics (LIPIcs)*, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, pp. 829–849.

[39] Geronimi, C. and Jackson, W. and Luske, H. (Dir.s), Walt, D. (Prod.) et. al. Alice in Wonderland, 1951.

[40] Grigoriev, D., and Vorobjov, N. Complexity of Null- and Positivstellensatz proofs. *Annals of Pure and Applied Logic 113*, 1 (2001), 153–160. First St. Petersburg Conference on Days of Logic and Computability.

[41] Hardesty, L. The history of Amazon's recommendation algorithm, Jul 2022. Website: amazon.science/the-history-of-amazons-recommendation-algorithm.

[42] Harris, J. Galois groups of enumerative problems. *Duke Mathematical Journal 46*, 4 (1979).

[43] Harshman, R. Foundations of the parafac procedure: Models and conditions for an “explanatory” multi-modal factor analysis. *UCLA Working Papers in Phonetics 16* (1970).

[44] Hartshorne, R. *Algebraic geometry*, vol. 52. Springer Science & Business Media, 2013.

[45] Hillar, C. J., and Lim, L.-H. Most tensor problems are NP-hard. *Journal of the ACM 60*, 6 (2013), 1–39.

[46] Jaynes, E. T. *Probability Theory: The Logic of Science.* Cambridge University Press, 2003.

[47] Kabanets, V., and Impagliazzo, R. Derandomizing polynomial identity tests means proving circuit lower bounds. *Proceedings of the thirty-fifth ACM symposium on Theory of computing - STOC 03* (2003).

[48] Kaminski, J. Y., Kanel-Belov, A., and Teicher, M. Trisecant lemma for nonequidimensional varieties. *Journal of Mathematical Sciences 149*, 2 (2008), 1087–1097.

[49] Kaminski, J. Y., Kanel-Belov, A., and Teicher, M. Multi-secant lemma. *Israel Journal of Mathematics 177*, 1 (2010), 253–266.

[50] Kleinberg, J., and Sandler, M. Using mixture models for collaborative filtering. *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing - STOC ’04* (2004).

[51] Klenke, A. *Probability theory: A comprehensive course.* Springer, 2020.

[52] Kohn, A. The dangerous myth of grade inflation. *The Chronicle of higher education 49* (2002).

[53] Landsberg, J. M. *Tensors: Geometry and applications.* American Mathematical Society, 2012.

[54] Laurent, M. Sums of squares, moment matrices and optimization over polynomials. *Emerging Applications of Algebraic Geometry* (2008), 157–270.

[55] Lee, J. M. *Introduction to smooth manifolds.* Springer, 2012.

[56] Legat, B., Timme, S., and Deits, R. Juliaalgebra/multivariatepolynomials.jl: v0.3.18, 07 2021.

[57] Leurgans, S. E., Ross, R. T., and Abel, R. B. A decomposition for three-way arrays. *SIAM Journal on Matrix Analysis and Applications 14*, 4 (1993), 1064–1083.

[58] Linden, G., Smith, B., and York, J. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing 7*, 1 (2003), 76–80.

[59] Liu, A., and Moitra, A. Settling the robust learnability of mixtures of Gaussians. *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing* (2021).

[60] Ma, T., Shi, J., and Steurer, D. Polynomial-time tensor decompositions with sum-of-squares. In *FOCS* (2016), I. Dinur, Ed., IEEE Computer Society, pp. 438–446.

[61] Makam, V., and Wigderson, A. Symbolic determinant identity testing (sdit) is not a null cone problem; and the symmetries of algebraic varieties. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)* (2020), IEEE, pp. 881–888.

[62] Makam, V., and Wigderson, A. Singular tuples of matrices is not a null cone (and the symmetries of algebraic varieties). *Journal für die reine und angewandte Mathematik 2021* (08 2021).

[63] McKernan, J. Lecture notes on introduction to algebraic geometry, 2012. math.mit.edu/ mckernan/Teaching/11-12/Spring/18.726.

[64] Minn, M. Curving grades with a normal distribution, 2000-2022. michaelminn.net/tutorials/normal-curve-grading.

[65] Moitra, A., and Valiant, G. Settling the polynomial learnability of mixtures of Gaussians. *2010 IEEE 51st Annual Symposium on Foundations of Computer Science* (2010).

[66] MOSEK-ApS. *Semidefinite Optimization — Mosek Optimizer API for Python*, 2019.

[67] Nenashev, G. A note on Fröberg's conjecture for forms of equal degrees. *Comptes Rendus Mathematique 355*, 3 (2017), 272–276.

[68] Nicklasson, L. On the Hilbert series of ideals generated by generic forms. *Communications in Algebra 45*, 8 (2017), 3390–3395.

[69] O'Donnell, R. SOS is not obviously automatizable, even approximately. In *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)* (Dagstuhl, Germany, 2017), C. H. Papadimitriou, Ed., vol. 67 of *Leibniz International Proceedings in Informatics (LIPIcs)*, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, pp. 59:1–59:10.

[70] O'Donnell, R., and Zhou, Y. Approximability and proof complexity. *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms* (2013).

[71] Pearson, K. Mathematical contributions to the theory of evolution. VII. On the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character 195* (1900), pp. 1–47+405.

[72] Perry, R. P., and Johnson, V. E. Grade inflation: A crisis in college education. *Academe 90*, 1 (2004), 90.

[73] Prestel, A., and Delzell, C. *Mathematical Logic and Model Theory: A Brief Introduction.* Universitext. Springer London, 2011.

[74] Prestel, A., and Delzell, C. N. *Positive Polynomials.* Springer Monographs in Mathematics. Springer Berlin, Heidelberg, 01 2001.

[75] Putinar, M., and Vasilescu, F.-H. Positive polynomials on semi-algebraic sets. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics 328*, 7 (1999), 585–589.

[76] Renegar, J. *A mathematical view of interior-point methods in convex optimization.* Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104), 2001.

[77] Richter, H. Parameterfreie Abschätzung und Realisierung von Erwartungswerten. *Blätter der DGVFM 3*, 2 (1957), 147–162.

[78] Riener, C., and Schweighofer, M. Optimization approaches to quadrature: New characterizations of Gaussian quadrature on the line and quadrature with few nodes on plane algebraic curves, on the plane and in higher dimensions. *Journal of Complexity 45* (2018), 22–54.

[79] Robeva, E. Orthogonal decomposition of symmetric tensors. *SIAM Journal on Matrix Analysis and Applications 37* (09 2014).

[80] Rowling, J. K. *Harry Potter and the Goblet of Fire.* Bloomsbury, 2000.

[81] Rowling, J. K. *Harry Potter and the Deathly Hallows.* Bloomsbury, 2007.

[82] Sato, H. *Riemannian Optimization and Its Applications.* SpringerBriefs in Electrical and Computer Engineering. Springer Cham, 01 2021.

[83] Schmüdgen, K. The k-moment problem for compact semi-algebraic sets. *Mathematische Annalen 289*, 1 (1991), 203–206.

[84] Schwartz, J. Fast probabilistic algorithms for verification of polynomial identities. *Journalof the ACM, 27:701717* (1980).

[85] Schweighofer, M. Real algebraic geometry, positivity and convexity, 2017. math.uni-konstanz.de/ schweigh/17/real-alg-geo-16-17.pdf.

[86] Steurer, D., and Barak, B. Proofs, beliefs, and algorithms through the lens of sum-of-squares. sumofsquares.org.

[87] Taveira Blomenhofer, A. Base case computation for: Sums of third powers of quadratics are generically identifiable up to quadratic rank, April 2022. Available at github.com/a44l/cubes-of-quadratics.

[88] Taveira Blomenhofer, A. PhD Thesis appendix for: Gaussian Mixture Separation and Denoising on Parameterized Varieties, September 2022. Available at github.com/a44l/phd-thesis.

[89] Tolkien, J. R. R. *The fellowship of the ring.* 1954.

[90] Top Hat editorial team. The ultimate guide to grading on a curve, Nov 2021. Author unspecified (editorial team). Website: tophat.com/blog/grading-on-a-curve.

[91] Valiant, L. G. Completeness classes in algebra. In *Proceedings of the eleventh Annual ACM Symposium on Theory of Computing* (New York, NY, USA, 1979), STOC '79, Association for Computing Machinery, pp. 249–261.

[92] Vill, J., Michałek, M., and Blomenhofer, A. T. Ideals of spaces of degenerate matrices. *Linear Algebra and its Applications* (2022).

[93] Zippel, R. *Effective Polynomial Computation.* The Springer International Series in Engineering and Computer Science. Springer US, 1993.