Writing Positive Polynomials as Sums of (Few) Squares

Olivier Benoist (Université de Strasbourg, CNRS, France)

In 1927, Artin proved that a real polynomial that is positive semidefinite is a sum of squares of rational functions, thus solving Hilbert's 17th problem. We review Artin's Theorem and its posterity, browsing through basic examples, classical results and recent developments. We focus on a question first considered by Pfister: can one write a positive semidefinite polynomial as a sum of few squares?

1 Hilbert's 17th problem

A real polynomial $f \in \mathbb{R}[X_1, \ldots, X_n]$ is said to be *positive semidefinite* if $f(x_1, \ldots, x_n) \geq 0$ for all $x_1, \ldots, x_n \in \mathbb{R}$.

Artin's Theorem. Can one explain the positivity of such a polynomial by writing it as a sum of squares? The question (popularised by Hilbert as the 17th of his famous list of open problems that he proposed on the occasion of the 1900 International Congress of Mathematicians) was solved by Artin [1]:

Theorem 1.1 (Artin, 1927). Let $f \in \mathbb{R}[X_1, \ldots, X_n]$ be positive semidefinite. Then, f is a sum of squares in $\mathbb{R}(X_1, \ldots, X_n)$.

Artin's proof of Theorem 1.1 was influential, fostering the development of real algebra. In collaboration with Schreier, and with Hilbert's 17th problem in mind, he had developed the theory of ordered fields [2]. A consequence of their work is that an element of a field K may be written as a sum of squares in K if and only if it is nonnegative with respect to all the orderings of K that are compatible with the field structure. It remains to show that if f is negative with respect to some ordering of $\mathbb{R}(X_1, \ldots, X_n)$, its evaluation at some point $(x_1, \ldots, x_n) \in \mathbb{R}^n$ is also negative. This "specialisation argument" is at the heart of the proof.

Sums of squares of polynomials. It would seem more natural to look for an expression of f as a sum of squares of polynomials, but this is too much to ask! It was Minkowski who convinced Hilbert, during his doctoral dissertation in 1885, that such a statement would be too strong. Three years later, in a seminal paper [14], Hilbert was able to show, by abstract means, the existence of counterexamples. Surprisingly, the first explicit counterexample appeared only much later, in 1967, and almost by chance. The polynomial

$$1 + X_1^2 X_2^4 + X_1^4 X_2^2 - 3X_1^2 X_2^2, \qquad (1)$$

introduced by Motzkin [21] for other purposes, was recognised by Taussky–Todd to be the first concrete example of a positive semidefinite polynomial that is not a sum of squares of polynomials. It is positive semidefinite as a consequence

of the arithmetic mean–geometric mean inequality and it satisfies the conclusion of Theorem 1.1 since it coincides with

$$\frac{(X_1^3 X_2 + X_2^3 X_1 - 2X_1 X_2)^2 (1 + X_1^2 + X_2^2) + (X_1^2 - X_2^2)^2}{(X_1^2 + X_2^2)^2}$$

but an elementary analysis of the low degree terms in a hypothetical expression of (1) as a sum of squares of polynomials quickly leads to a contradiction.

Sums of few squares. How many squares are needed in Theorem 1.1? A beautiful answer, surprisingly dependent only on the number of variables, was obtained by Pfister [22].

Theorem 1.2 (Pfister, 1967). Let $f \in \mathbb{R}[X_1, \ldots, X_n]$ be positive semidefinite. Then, f is a sum of 2^n squares in $\mathbb{R}(X_1, \ldots, X_n)$.

Pfister's work is completely independent of Artin's. Indeed, what Pfister really proves is that any sum of squares in $\mathbb{R}(X_1, \ldots, X_n)$ is in fact a sum of 2^n squares. It is only in combination with Theorem 1.1 that Theorem 1.2 is obtained. His result stemmed from important progress in the algebraic theory of quadratic forms: the discovery of the so-called Pfister forms (which enjoy marvellous algebraic properties).

In three variables, Theorem 1.2 had previously been obtained by Ax. It is while reading Ax's manuscript that Pfister realised one could replace the cohomological tools of Ax by the use of Pfister forms, yielding a result in arbitrary dimension. It may not be a coincidence that Pfister forms later turned out to be the key to a cohomological classification of quadratic forms over fields, culminating in Voevodsky's proof of the Milnor conjecture [32].

We refer to [23] for a nice exposition of Theorems 1.1 and 1.2. Whether the 2^n bound in Theorem 1.2 may be improved or not (Question 2.6 below) is the main topic of this survey.

2 Polynomials of low degree or in few variables

Let us illustrate the theorems of Artin and Pfister, starting from basic cases. Let $\mathbb{R}[X_1, \ldots, X_n]_d$ be the space of polynomials of degree d. We consider a positive semidefinite polynomial $f \in \mathbb{R}[X_1, \ldots, X_n]_d$. Of course, since an odd degree polynomial changes sign, the degree d of f must be even.

· **d = 2**. A degree 2 polynomial $f \in \mathbb{R}[X_1, \ldots, X_n]$ may be homogenised to a quadratic form $F \in \mathbb{R}[X_0, X_1, \ldots, X_n]$ that is positive semidefinite if f is. Diagonalising the quadratic form F displays it as a sum of $n + 1$ squares of linear forms. Dehomogenising, we see that f is a sum of $n + 1$ squares of polynomials.

As soon as $n \geq 2$, this is a significant improvement on the 2^n bound of Theorem 1.2! Since it is not very impressive to write a quadratic form as a sum of squares, this particular case should not be viewed as representative of the general situation.

· **n = 1**. A polynomial $f \in \mathbb{R}[X]$ in one variable may be factored as a product of irreducible real polynomials:

$$f = \lambda \prod_i (X - a_i) \prod_j (X^2 + 2b_j X + c_j).$$

If f is positive semidefinite, the multiplicity of a_i as a root of f has to be even, and letting $X \to \infty$ implies that $\lambda \geq 0$. Completing the square $X^2 + 2b_j X + c_j = (X + b_j)^2 + (c_j - b_j^2)$ shows that f may be written as a product of sums of two squares of polynomials. The classical identity

$$(A^2 + B^2)(C^2 + D^2) = (AC + BD)^2 + (AD - BC)^2 \quad (2)$$

then implies that f is the sum of two squares of polynomials. We have recovered Theorems 1.1 and 1.2 in a stronger form: it was enough to consider sums of squares of polynomials!

The use of identity (2) is not innocent. In some sense, the contribution of Pfister in his proof of Theorem 1.2 was precisely to find a systematic way to produce identities analogous to (2) in more variables.

· **n = 2 and d = 4**. Degree 4 polynomials in two variables, or, as classical geometers say, *ternary quartics*, are particularly interesting. They have been studied in detail by Hilbert [14], who proved:

Theorem 2.1 (Hilbert, 1888)**.** Let $f \in \mathbb{R}[X_1, X_2]$ be positive semidefinite of degree 4. Then, f is a sum of 3 squares in $\mathbb{R}[X_1, X_2]$.

Not only is f a sum of squares of polynomials (rather than mere rational functions) but the 2^n bound of Theorem 1.2 may also be improved!

In contrast with the $d = 2$ and the $n = 1$ cases above, Theorem 2.1 is a non-trivial result. Hilbert's proof (one of the first applications of topology to algebra) is beautiful. His idea is to start with a ternary quartic that is obviously a sum of three squares of polynomials, such as $f_0 = 1 + X_1^4 + X_2^4$, to carefully choose a path from f_0 to f in the space of positive semidefinite ternary quartics and to deform the representation of f_0 as a sum of three squares to one for f.

We refer to [26] or [31] for modern accounts of Hilbert's proof and to [28] for recent developments.

· **n = 2 and d ≥ 6**. The behaviour of positive semidefinite polynomials in two variables stabilises from degree 6 onward. Hilbert proved in [15] that they are sums of squares in $\mathbb{R}(X_1, X_2)$ and Landau [18, p. 282], analysing Hilbert's proof, noticed that only 4 squares are needed.

Theorem 2.2 (Hilbert, 1893)**.** Let $f \in \mathbb{R}[X_1, X_2]$ be positive semidefinite. Then, f is a sum of 4 squares in $\mathbb{R}(X_1, X_2)$.

Theorem 2.2 is a particular case of Theorem 1.2 in two variables: Pfister's theorem is nothing but a higher-dimensional generalisation of Hilbert's classical result. Hilbert's

argument, an elaboration of his proof of Theorem 2.1, is more difficult than Pfister's. It is also more precise. For instance, it allows one to control the denominators in an expression of f as a sum of 4 squares: if f has degree d, the denominators may be chosen to have degree $\leq \left\lfloor \frac{(d-2)^2}{8} \right\rfloor$.

We have already seen that, even in degree 6, one cannot expect to improve on Theorem 2.2 by requiring that f is a sum of squares of polynomials: Motzkin's polynomial (1) is a counterexample. It is natural to ask whether one could strengthen Theorem 2.2 by writing f as a sum of 3 squares. Again, the answer is negative when $d \geq 6$ and the first known counterexample, discovered by Cassels, Ellison and Pfister [7], was . . . Motzkin's polynomial!

Theorem 2.3 (Cassels-Ellison-Pfister, 1971)**.** Motzkin's polynomial (1) is not a sum of 3 squares in $\mathbb{R}(X_1, X_2)$.

Although it is elementary to verify that Motzkin's polynomial is not a sum of squares in $\mathbb{R}[X_1, X_2]$, showing that it is not a sum of 3 squares in $\mathbb{R}(X_1, X_2)$ requires a little bit of algebraic geometry. In [7], the authors use the precise form of (1) to produce an elliptic surface whose properties control the potential of writing Motzkin's polynomial as a sum of 3 squares and they study it in detail.

One is left to wonder how frequent the sums of 3 squares are. What does the subset of $\mathbb{R}[X_1, X_2]_d$ consisting of polynomials that can be written as sums of 3 squares in $\mathbb{R}(X_1, X_2)$ look like? The first result in this direction, due to Colliot-Thélène [9, 4.3], indicates that they are quite scarce.

Theorem 2.4 (Colliot-Thélène, 1993)**.** If $d \geq 6$, the degree d polynomials that are sums of 3 squares in $\mathbb{R}(X_1, X_2)$ form a meagre subset of measure 0 of $\mathbb{R}[X_1, X_2]_d$.

Hence, sums of 3 squares are negligible both from the topological (*meagre* means a countable union of nowhere dense subsets) and measure theory points of view (cf. Section 4 for an account of the proof).

Despite Theorem 2.4, sums of 3 squares turn out to be dense in the set of positive semidefinite polynomials [3].

Theorem 2.5 (2017)**.** Any degree d positive semidefinite polynomial $f \in \mathbb{R}[X_1, X_2]$ may be approximated by degree d polynomials that are sums of 3 squares in $\mathbb{R}(X_1, X_2)$.

The picture to have in mind is the following. The set of polynomials that may be written as sums of 3 squares of rational functions whose denominators have degree $\leq N$ is a closed subset of $\mathbb{R}[X_1, X_2]_d$. Taking the union on all integers N, we get a countable union of closed subsets and it is only this union that one may hope to be dense. In other words, when approximating a polynomial that is not itself a sum of 3 squares, the degrees of the denominators must grow to infinity. The author is unaware of a constructive approach to Theorem 2.5. In particular, can one write Motzkin's polynomial (1) explicitly as a limit of sums of 3 squares?

· **n ≥ 3 and d ≥ 4.**. In at least 3 variables (and degree ≥ 4), no further general result expressing a positive semidefinite polynomial as a sum of squares of polynomials holds true, as discovered by Hilbert [14]. It is hard to resist writing down a beautiful example, due to Lax and Lax [19], of a degree 4

positive semidefinite polynomial in 3 variables that is not a sum of squares of polynomials:

$$\sum_{i=1}^{5} \prod_{j\neq i} (X_i - X_j).$$

Its five variables are a smokescreen: it only depends on the four homogeneous variables $X_1 - X_2$, $X_2 - X_3$, $X_3 - X_4$ and $X_4 - X_5$, giving rise, after dehomogenisation, to a polynomial in three variables. A survey by Reznick [25] contains many more examples.

In contrast, whether the bound 2^n in Pfister's Theorem 1.2 is optimal or not remains completely mysterious.

Question 2.6 (Pfister). Does there exist a positive semidefinite polynomial $f \in \mathbb{R}[X_1, \ldots, X_n]$ that is not a sum of $2^n - 1$ squares in $\mathbb{R}(X_1, \ldots, X_n)$?

This question was raised by Pfister immediately upon proving Theorem 1.2 and in general is still open today. It is arguably the most important problem of the subject. Defining the *Pythagoras number* $p(K)$ of a field K to be the smallest $p \in \mathbb{N}$ such that every sum of squares in K is actually a sum of p squares (or $+\infty$ if no such integer exists), Question 2.6 may be reformulated as follows.

Question 2.7. Is $p(\mathbb{R}(X_1, \ldots, X_n))$ equal to 2^n?

To see the equivalence, one can reduce to studying polynomials by multiplying a rational function by the square of its denominator and use Artin's Theorem 1.1, according to which the polynomials that are sums of squares in $\mathbb{R}(X_1, \ldots, X_n)$ are exactly those that are positive semidefinite.

We have already seen that Questions 2.6 and 2.7 have positive answers when $n = 1$ (as $1 + X_1^2$ is not a square) and when $n = 2$ (by Cassels-Ellison-Pfister's Theorem 2.3). When $n \geq 3$, the state of the art is the inequality

$$n + 2 \leq p(\mathbb{R}(X_1, \ldots, X_n)) \leq 2^n,$$

where the upper bound is Pfister's Theorem 1.2 and the lower bound is an easy consequence of the Cassels-Ellison-Pfister Theorem [23, p.97].

By analogy with Hilbert's Theorem 2.1, one may expect to obtain better bounds if one restricts to low degree polynomials. This point of view was taken in [4], yielding the following result.

Theorem 2.8 (2016). Let $f \in \mathbb{R}[X_1, \ldots, X_n]_d$ be positive semidefinite. If $n \geq 2$ and $d \leq 2n$, f is a sum of $2^n - 1$ squares in $\mathbb{R}(X_1, \ldots, X_n)$, except maybe if $n \geq 7$ is odd and $d = 2n$.

This improvement on Theorem 1.2 seems incremental but is already new for degree 4 polynomials in three variables. In this case, it shows that a positive semidefinite polynomial is a sum of 7 squares. It is not known if this can be improved.

The hypothesis that $n \geq 2$ cannot be dispensed with, as attested by the polynomial $1 + X_1^2$. It is, however, likely that the result continues to hold when $n \geq 7$ is odd and $d = 2n$.

It may be expected that the degree range $d \leq 2n$ appearing in Theorem 2.8 is the correct one, in the sense that, from degree $d = 2n + 2$ onward, there would exist positive semidefinite polynomials that are not sums of $2^n - 1$ squares, hence

giving a positive answer to Question 2.6. We will give a geometric interpretation for this value of the degree at the end of Section 4.

3 A rich legacy

Hilbert's 17th problem has triggered developments in many other directions. A few will be listed here, without any attempt to be exhaustive.

Arithmetic base fields. What if the coefficients of f belong to a smaller field, say the field $\mathbb{Q}$ of rational numbers? Then, it was already known to Artin [1] that f is a sum of squares in $\mathbb{Q}(X_1, \ldots, X_n)$. On the other hand, obtaining a bound à la Pfister on the number of squares involved is much harder. The best result to date is the following arithmetic geometry masterpiece.

Theorem 3.1 (Jannsen, 2016). Let $f \in \mathbb{Q}[X_1, \ldots, X_n]$ be positive semidefinite. If $n \geq 2$, f is a sum of 2^{n+1} squares in $\mathbb{Q}(X_1, \ldots, X_n)$.

This theorem was found to follow from two outstanding conjectures by Colliot-Thélène and Jannsen [10]: the Milnor conjecture established by Voevodsky [32] and Kato's cohomological local-global principle eventually settled by Jannsen in [16].

The hypothesis $n \geq 2$ is necessary. When $n = 0$, the optimal statement is Euler's precursor of Lagrange's Theorem, according to which a non-negative rational number is a sum of 4 squares of rational numbers [12]. When $n = 1$, Pourchet [24] has proved that a positive semidefinite polynomial $f \in \mathbb{Q}[X]$ is a sum of 5 squares and his result is the best possible. When $n \geq 2$, it is not known whether Jannsen's bound is optimal. In the terminology introduced in Section 2, is the Pythagoras number $p(\mathbb{Q}(X_1, \ldots, X_n))$ equal to 2^{n+1} for $n \geq 2$?

Effectivity. Artin's proof of Theorem 1.1, relying on Zorn's lemma, is not constructive. The search for effective proofs was initiated by Kreisel, allowing one to derive bounds on the degrees of the rational functions involved. The history of this line of thought is explained in Delzell's survey [11]. Lombardi, Perrucci and Roy [20] have recently obtained the following theorem.

Theorem 3.2 (Lombardi, Perrucci, Roy, 2014). A positive semidefinite polynomial $f \in \mathbb{R}[X_1, \ldots, X_n]$ may be written as a sum of squares of rational functions of degree $\leq 2^{2^{2^{d^{4^n}}}}$.

The reader should not be intimidated by the formidable bound: it is a tremendous improvement on previous results!

Positivstellensätze. What can be said if the polynomial f is only known to be nonnegative on some domain $\Omega \subset \mathbb{R}^n$? The following theorem, due to Stengle [30] but close to independent earlier work of Krivine [17], is the prototype of such a result: a *Positivstellensatz*.

Theorem 3.3 (Krivine, Stengle, 1974). Let $f \in \mathbb{R}[X_1, \ldots, X_n]$ be positive on $\Omega := \{x \in \mathbb{R}^n \mid g_1(x) \geq 0, \ldots, g_k(x) \geq 0\}$,

where $g_1, \ldots, g_k \in \mathbb{R}[X_1, \ldots, X_n]$. Then, f belongs to the sub-semiring of $\mathbb{R}(X_1, \ldots, X_n)$ generated by the g_i and the squares.

Given the many counterexamples to the variant of Hilbert's 17$^{\text{th}}$ problem involving polynomials instead of rational functions, it came as a surprise when Schmüdgen [29] obtained a Positivstellensatz without denominators, at the expense of a compactness hypothesis.

Theorem 3.4 (Schmüdgen, 1991). Under the assumptions of Theorem 3.3, if Ω is compact, f belongs to the sub-semiring of $\mathbb{R}[X_1, \ldots, X_n]$ generated by the g_i and the squares.

The reader will find these statements and many more in a survey by Scheiderer [27].

The cone of sums of squares. What positive semidefinite polynomials $f \in \mathbb{R}[X_1, \ldots, X_n]$ are sums of squares of polynomials? We have already seen in Section 2 that some positive semidefinite polynomials are not (unless $n = 1$ or $d = 2$ or $n = 2$ and $d = 4$) but which ones? More precisely, can one describe a set of linear inequalities defining the closed convex cone $\Sigma_{n,d} \subset \mathbb{R}[X_1, \ldots, X_n]_d$ consisting of polynomials that are sums of squares of polynomials?

A full answer has been obtained by Blekherman [5] in the first two cases where $\Sigma_{n,d}$ does not coincide with the set of positive semidefinite polynomials: ternary sextics ($n = 2$ and $d = 6$) and quaternary quartics ($n = 3$ and $d = 4$). Surprisingly, the required linear inequalities are precisely those that were introduced by Hilbert [14] to show the existence of positive semidefinite polynomials that are not sums of squares of polynomials.

4 Sums of fewer squares

The theorems that express positive semidefinite polynomials as a sum of fewer squares than predicted by Pfister's Theorem 1.2, as well as those that show that it is impossible, use tools from algebraic geometry. More precisely, they rely on the study of *algebraic cycles*, that is, of the algebraic subvarieties of a fixed algebraic variety.

The first indication of such a link was Cassels, Ellison and Pfister's proof of Theorem 2.3. Its main step is the computation of the Mordell-Weil group of an elliptic curve over a function field [7, §7], a problem essentially equivalent to the determination of all algebraic curves lying on an elliptic surface.

The relation between sums of squares and algebraic cycles is much more transparent in Colliot-Thélène's proof of Theorem 2.4 [9]. We devote the greater part of this section to explaining its principle. The way algebraic cycles enter Theorems 2.5 and 2.8 is similar and we will comment on these proofs too. Our main goal is to understand how algebraic geometry governs the dependence of the properties of a positive semidefinite polynomial on its degree.

Sums of 3 squares in $\mathbb{R}[X_1, X_2]$. Fix a degree d positive semidefinite polynomial $f \in \mathbb{R}[X_1, X_2]_d$. We want to understand under which conditions f is a sum of 3 squares in $\mathbb{R}(X_1, X_2)$ and to deduce, following Colliot-Thélène [9], that this is a rare phenomenon when $d \geq 6$.

We reformulate this property. Let Y be a square root of $-f$ and consider the quadratic field extension $K := \mathbb{R}(X_1, X_2)[Y]$ of $\mathbb{R}(X_1, X_2)$.

Lemma 4.1. That f is a sum of 3 squares in $\mathbb{R}(X_1, X_2)$ is equivalent to -1 being a sum of 2 squares in K.

Proof. This is elementary and we only verify the direct implication, which is the one we actually use. If $f = a^2 + b^2 + c^2$ in $\mathbb{R}(X_1, X_2)$, dividing out by $-f = Y^2$ yields an identity of the form $-1 = r^2 + s^2 + t^2$ in K. Applying (2) cleverly, one gets

$$-1 = \left(\frac{rs + t}{1 + r^2}\right)^2 + \left(\frac{s - rt}{1 + r^2}\right)^2. \qquad \square$$

We are reduced to understanding when -1 is a sum of 2 squares in K. To do so, we introduce the geometric incarnation of K, that is, the algebraic surface S defined as a set by

$$\{(x_1, x_2, y) \in \mathbb{C}^3 \mid y^2 + f(x_1, x_2) = 0\}. \qquad (3)$$

To be more precise, what we will really denote by S is the compactification of (3) obtained by adding "points at infinity". We will also assume that f has been chosen so that S has no singularities. Elements of K may be viewed as functions on S (which may not be defined everywhere: there may be poles). For this reason, K is called the *function field* of S.

The proof of Theorem 2.4 depends on the knowledge of algebraic curves on S, that is, of the subsets $C \subset S$ of complex dimension 1 that are defined by polynomial equations. There are obvious algebraic curves on S, defined by a single polynomial equation $g \in \mathbb{C}[X_1, X_2, Y]$:

$$C = \{(x_1, x_2, y) \in S \mid g(x_1, x_2, y) = 0\}. \qquad (4)$$

There may, however, be more! This happens, for instance, if the restriction of f to some complex line in $\mathbb{C}^2$ is the square of a polynomial. Indeed, suppose that the line is defined, say, by the equation $X_2 = 0$ and that $f(X_1, 0) = h(X_1)^2$ for some $h \in \mathbb{C}[X_1]$. Then, the algebraic curve $C = \{(x_1, x_2, y) \in S \mid x_2 = 0\}$ splits into two algebraic curves in S:

$$\begin{cases} C_+ &= \{(x_1, x_2, y) \in S \mid x_2 = 0 \text{ and } y = +ih(x_1)\}, \\ C_- &= \{(x_1, x_2, y) \in S \mid x_2 = 0 \text{ and } y = -ih(x_1)\}, \end{cases}$$

which are not individually of the form (4). Moreover, one can check that, if $d \leq 4$, there always exist such lines in $\mathbb{C}^2$, so that not all curves in S are of the form (4). When $d \geq 6$, the situation is completely different.

Theorem 4.2 (Noether–Lefschetz). If $d \geq 6$ and f is very general, all algebraic curves $C \subset S$ are of the form (4).

Here, *very general* is the algebraic geometers' way to indicate a generic behaviour: it means that the statement holds for all f except maybe for those that belong to a countable union of algebraic subvarieties of the parameter space $\mathbb{C}[X_1, X_2]_d$. It implies that the set of those f for which the statement fails is meagre and of measure 0.

We may now conclude the proof of Theorem 2.4. Suppose that $-1 = u^2 + v^2$ is a sum of two squares in K. Then, it can be claimed that the algebraic curve

$$\Gamma = \{(x_1, x_2, y) \in S \mid u = 0 \text{ and } v = i\} \subset S$$

is not of the form (4). By the Noether–Lefschetz Theorem 4.2, this can only happen for very particular choices of f, and Theorem 2.4 is proven.

Let us give a hint for the purely topological argument proving the above claim. Remember that our polynomial f has real coefficients. Consequently, the complex conjugation

$$\sigma : (x_1, x_2, y) \mapsto (\overline{x_1}, \overline{x_2}, \overline{y})$$

induces an involution of S. Triangulating the curve Γ yields a class $[\Gamma] \in H_2(S, \mathbb{Z})$ in the homology of S that is called the fundamental class of Γ (indeed, as a complex curve, Γ is topologically a surface). One can verify that this homology class cannot be realised by a 2-cycle that is invariant under the action of the complex conjugation, whereas the fundamental classes of all the algebraic curves of the form (4) can!

Why degree 6? The above proof indicates a reason why positive semidefinite polynomials in two variables exhibit different behaviours when $d \leq 4$ (Theorem 2.1) and $d \geq 6$ (Theorem 2.4). This is due to the Noether–Lefschetz Theorem! When $d \leq 4$, the associated surface S has a rich geometry and contains plenty of algebraic curves but when $d \geq 6$, a typical S contains only obvious algebraic curves.

Still, our understanding is not yet complete: why is the Noether–Lefschetz Theorem only valid in degree ≥ 6? Since it will also be important in our discussion of Theorem 2.5, we explain this now. The main tool is *Hodge theory*.

We need to understand when the surface S contains unexpected algebraic curves that are not of the form (4). To do so, we fix a homology class $\gamma \in H_2(S, \mathbb{Z})$ and we consider the question: when is γ the fundamental class of an algebraic curve $\Gamma \subset S$ or, rather, a linear combination with integral coefficients of such classes?

A necessary condition is that if ω is a holomorphic 2-form on S (for every $s \in S$, ω_s is an alternating $\mathbb{C}$-bilinear form on the tangent space of S at s varying holomorphically with s), the integral $\int_\gamma \omega$ needs to vanish. This is a simple dimension argument: ω vanishes in restriction to algebraic curves on S because there are no non-zero alternating $\mathbb{C}$-bilinear forms on a one-dimensional $\mathbb{C}$-vector space. This condition also turns out to be sufficient. Denoting by $\Omega^2(S)$ the space of holomorphic 2-forms on S, this is the famous Lefschetz $(1, 1)$ Theorem.

Theorem 4.3 (Lefschetz $(1, 1)$). *A class $\gamma \in H_2(S, \mathbb{Z})$ is a linear combination of classes of algebraic curves on S if and only if $\int_\gamma \omega = 0$ for every $\omega \in \Omega^2(S)$.*

If $d \leq 4$, one can compute that $\Omega^2(S) = 0$. Consequently, Theorem 4.3 predicts the existence of many algebraic curves on S, in particular of curves not of the shape (4).

On the other hand, if $d \geq 6$, one can check that $\Omega^2(S) \neq 0$. The Lefschetz $(1, 1)$ Theorem then gives non-trivial obstructions to the existence of algebraic curves on S and one can verify that, for most values of f, these obstructions prevent the existence of any curve not of the form (4). This proves Theorem 4.2.

This completely explains why the properties of positive semidefinite polynomials $f \in \mathbb{R}[X_1, X_2]_d$ change when $d \geq 6$.

It is the influence of the geometry of the associated surface S that carries non-zero holomorphic 2-forms if and only if $d \geq 6$.

Density. Now that we have understood why there are few semidefinite polynomials that are sums of 3 squares (in degree ≥ 6), let us explain why these are dense in the set of positive semidefinite polynomials (Theorem 2.5). Recall that we have associated to a degree d polynomial $f \in \mathbb{C}[X_1, X_2]_d$ an algebraic surface S defined by (3):

$$y^2 + f(x_1, x_2) = 0,$$

and explained that if f is a sum of 3 squares in $\mathbb{R}(X_1, X_2)$, the surface S carries more algebraic curves than expected. The archetype of the density result we need has essentially been obtained by Ciliberto, Harris, Miranda and Green [8].

Theorem 4.4 (Ciliberto, Harris, Miranda, Green). *The set of $f \in \mathbb{C}[X_1, X_2]_d$ such that the associated surface S contains algebraic curves not of the form (4) is dense in $\mathbb{C}[X_1, X_2]_d$.*

Of course, this cannot imply Theorem 2.5 because it says nothing about density in $\mathbb{R}[X_1, X_2]_d$. Proving Theorem 2.5 requires an adaption over $\mathbb{R}$, carried out in [3], of the techniques of [8].

Let us explain what enters the proof of Theorem 4.4 and of its real variant yielding Theorem 2.5. One has to analyse how the obstructions to the existence of algebraic curves on S that are provided by Theorem 4.3 vary with $f \in \mathbb{C}[X_1, X_2]_d$. This amounts to understanding the variation with f of the integrals $\int_\gamma \omega = 0$, called the *periods* of the surface S. Since the work of Griffiths, this very classical topic has been known as the study of *infinitesimal variations of Hodge structures*. Both [8] and [3] rely on these modern tools.

More variables. To study a positive semidefinite polynomial $f \in \mathbb{R}[X_1, \ldots, X_n]$ in $n \geq 3$ variables, it is still useful to introduce the algebraic variety X defined by the equation

$$y^2 + f(x_1, \ldots, x_n) = 0 \qquad (5)$$

and its function field $K = \mathbb{R}(X_1, \ldots, X_n)[Y]$, where $Y = \sqrt{-f}$. An analogue of Lemma 4.1 holds: f is a sum of $2^n - 1$ squares in $\mathbb{R}(X_1, \ldots, X_n)$ if and only if -1 is a sum of 2^{n-1} squares in K. When $n \geq 3$, this is not as elementary as Lemma 4.1 and relies on Pfister's work on quadratic forms.

Relating the latter property to algebraic cycles on X depends on the far-reaching enhancement of Pfister's work provided by Voevodsky's proof of the Milnor conjecture [32]. This allows a cohomological reformulation: letting the group $G := \mathbb{Z}/2\mathbb{Z}$ act on X by the complex conjugation

$$\sigma : (x_1, \ldots, x_n, y) \mapsto (\overline{x_1}, \ldots, \overline{x_n}, \overline{y}),$$

there exists a class $\alpha \in H_G^n(X, \mathbb{Z}/2\mathbb{Z})$ in G-equivariant cohomology such that -1 is a sum of 2^{n-1} squares in K if and only if α vanishes in the complement of an algebraic subvariety of X. Intuitively, this will happen if and only if X contains many algebraic subvarieties.

Proving Theorem 2.8 requires one to show that if $d \leq 2n$, the variety X contains many algebraic subvarieties in this sense. The Hodge theory arguments are of no use now and [4] relies on other methods, such as Bloch-Ogus theory. Let us explain the origin of the hypothesis $d \leq 2n$ on the degree. As before, it reflects a geometric property of the algebraic variety X. Namely, it ensures that X is *rationally connected*: that there exist enough algebraic maps $\mathbb{P}^1 \to X$ to connect any two points $p, q \in X$. It has been known since Bloch and Srinivas [6] that this geometric information gives strong control on the cohomology of X, which is exactly what is needed.

In contrast, when trying to answer Question 2.6, one has to show that the algebraic variety X may contain only few algebraic subvarieties if the degree d is high enough (maybe if $d \geq 2n + 2$?). When $n = 3$, which is the smallest value for which Question 2.6 is open, the required statement is a variant of a classical question asked by Griffiths and Harris in [13]. To give a flavour of what is needed, we state a slightly different question, closer to the one raised in [13]. Recall that the *degree* of an algebraic curve $C \subset X$ is the cardinality, taking multiplicities into account, of the set

$$\{(x_1, \ldots, x_n, y) \in C \mid x_1 = 0\}.$$

Question 4.5. Let $f \in \mathbb{C}[X_1, X_2, X_3]_d$ and X be defined by (5). If f is very general and $d \geq 10$, are all algebraic curves in X of even degree?

References

[1] E. Artin. Über die Zerlegung definiter Funktionen in Quadrate. *Abh. Math. Sem. Univ. Hamburg*, 5(1):100–115, 1927.

[2] E. Artin and O. Schreier. Algebraische Konstruktion reeller Körper. *Abh. Math. Sem. Univ. Hamburg*, 5(1):85–99, 1927.

[3] O. Benoist. Sums of 3 squares and Noether-Lefschetz loci. arXiv:1706.02053.

[4] O. Benoist. On Hilbert's 17th problem in low degree. *Algebra Number Theory*, 11(4):929–959, 2017.

[5] G. Blekherman. Nonnegative polynomials and sums of squares. *J. Amer. Math. Soc.*, 25(3):617–635, 2012.

[6] S. Bloch and V. Srinivas. Remarks on correspondences and algebraic cycles. *Amer. J. Math.*, 105(5):1235–1253, 1983.

[7] J. W. S. Cassels, W. J. Ellison and A. Pfister. On sums of squares and on elliptic curves over function fields. *J. Number Theory*, 3:125–149, 1971.

[8] C. Ciliberto, J. Harris and R. Miranda. General components of the Noether-Lefschetz locus and their density in the space of all surfaces. *Math. Ann.*, 282(4):667–680, 1988.

[9] J.-L. Colliot-Thélène. The Noether-Lefschetz theorem and sums of 4 squares in the rational function field $\mathbf{R}(x, y)$. *Compositio Math.*, 86(2):235–243, 1993.

[10] J.-L. Colliot-Thélène and U. Jannsen. Sommes de carrés dans les corps de fonctions. *C. R. Acad. Sci. Paris Sér. I Math.*, 312(11):759–762, 1991.

[11] C. N. Delzell. Kreisel's unwinding of Artin's proof. In *Kreiseliana*, pages 113–246. A K Peters, Wellesley, MA, 1996.

[12] L. Euler. Demonstratio theorematis Fermatiani omnem numerum sive integrum sive fractum esse summam quatuor pauciorumve quadratorum. *Novi Commentarii academiae scientiarum Petropolitanae*, 5:13–58, 1760.

[13] P. Griffiths and J. Harris. On the Noether-Lefschetz theorem and some remarks on codimension-two cycles. *Math. Ann.*, 271(1):31–51, 1985.

[14] D. Hilbert. Ueber die Darstellung definiter Formen als Summe von Formenquadraten. *Math. Ann.*, 32(3):342–350, 1888.

[15] D. Hilbert. Über ternäre definite Formen. *Acta Math.*, 17(1):169–197, 1893.

[16] U. Jannsen. Hasse principles for higher-dimensional fields. *Ann. of Math. (2)*, 183(1):1–71, 2016.

[17] J.-L. Krivine. Anneaux préordonnés. *J. Analyse Math.*, 12:307–326, 1964.

[18] E. Landau. Über die Darstellung definiter binärer Formen durch Quadrate. *Math. Ann.*, 57(1):53–64, 1903.

[19] A. Lax and P. D. Lax. On sums of squares. *Linear Algebra and Appl.*, 20(1):71–75, 1978.

[20] H. Lombardi, D. Perrucci and M.-F. Roy. An elementary recursive bound for effective Positivstellensatz and Hilbert 17th problem, 2014, to appear in Mem. Amer. Math. Soc.

[21] T. S. Motzkin. The arithmetic-geometric inequality. In *Inequalities (Proc. Sympos. Wright-Patterson Air Force Base, Ohio, 1965)*, pages 205–224. Academic Press, New York, 1967.

[22] A. Pfister. Zur Darstellung definiter Funktionen als Summe von Quadraten. *Invent. Math.*, 4:229–237, 1967.

[23] A. Pfister. *Quadratic forms with applications to algebraic geometry and topology*, Volume 217 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1995.

[24] Y. Pourchet. Sur la représentation en somme de carrés des polynômes à une indéterminée sur un corps de nombres algébriques. *Acta Arith.*, 19:89–104, 1971.

[25] B. Reznick. Some concrete aspects of Hilbert's 17th Problem. In *Real algebraic geometry and ordered structures (Baton Rouge, LA, 1996)*, Volume 253 of *Contemp. Math.*, pages 251–272. Amer. Math. Soc., Providence, RI, 2000.

[26] W. Rudin. Sums of squares of polynomials. *Amer. Math. Monthly*, 107(9):813–821, 2000.

[27] C. Scheiderer. Positivity and sums of squares: a guide to recent results. In *Emerging applications of algebraic geometry*, Volume 149 of *IMA Vol. Math. Appl.*, pages 271–324. Springer, New York, 2009.

[28] C. Scheiderer. Hilbert's theorem on positive ternary quartics: a refined analysis. *J. Algebraic Geom.*, 19(2):285–333, 2010.

[29] K. Schmüdgen. The K-moment problem for compact semialgebraic sets. *Math. Ann.*, 289(2):203–206, 1991.

[30] G. Stengle. A Nullstellensatz and a Positivstellensatz in semialgebraic geometry. *Math. Ann.*, 207:87–97, 1974.

[31] R. G. Swan. Hilbert's theorem on positive ternary quartics. In *Quadratic forms and their applications (Dublin, 1999)*, Volume 272 of *Contemp. Math.*, pages 287–292. Amer. Math. Soc., Providence, RI, 2000.

[32] V. Voevodsky. Motivic cohomology with $\mathbb{Z}/2$-coefficients. *Publ. Math. IHES*, (98):59–104, 2003.

Olivier Benoist [olivier.benoist@unistra.fr] received his PhD in Paris in 2012. He is now a CNRS researcher at IRMA, Strasbourg. His field of interest is algebraic geometry, with a focus on algebraic cycles and real algebraic geometry.