Received: 05 July 2018 | Revised: 06 November 2019 | Accepted: 18 February 2020

DOI: 10.1002/rsa.20926

RESEARCH ARTICLE

WILEY

# The real tau-conjecture is true on average

**Irénée Briquel**[1] | **Peter Bürgisser**[2]

[1]Laboratoire ETIS UMR8051, CY Cergy Paris Université, Paris, France
[2]Institute of Mathematics, Technische Universität Berlin, Berlin, Germany

**Correspondence**
Peter Bürgisser, Institute of Mathematics, Technische Universität Berlin, Berlin, Germany.
Email: pbuerg@math.tu-berlin.de

**Funding information**
This research was supported by the DFG grant (Deutsche Forschungsgemeinschaft), BU 1371/2-2.; ERC under the European's Horizon 2020 Research and Innovation Programme (European Research Council), 787840. (P.B.)

**Abstract**

Koiran's real $\tau$-conjecture claims that the number of real zeros of a structured polynomial given as a sum of $m$ products of $k$ real sparse polynomials, each with at most $t$ monomials, is bounded by a polynomial in $mkt$. This conjecture has a major consequence in complexity theory since it would lead to super-polynomial lower bounds for the arithmetic circuit size of the permanent. We confirm the conjecture in a probabilistic sense by proving that if the coefficients involved in the description of $f$ are independent standard Gaussian random variables, then the expected number of real zeros of $f$ is $\mathcal{O}(mk^2t)$.

**KEYWORDS**

complexity theory, depth four arithmetic circuits, Descartes rule, sparsity, tau-conjecture, zeros of random polynomials

## 1 | INTRODUCTION

We study the number of real zeros of real univariate polynomials. A polynomial $f$ is called $t$-sparse if it has at most $t$ monomials. Descartes rule states that a $t$-sparse polynomial $f$ has at most $t-1$ positive real zeros, no matter what is the degree of $f$. Therefore, a product $f_1 \cdots f_k$ of $k$ many $t$-sparse polynomials $f_j$ can have at most $k(t-1)$ positive real zeros. What can we say about the number of zeros of a sum of $m$ many products? So we consider real univariate polynomials $F$ of the following structure

$$F = \sum_{i=1}^{m} \prod_{j=1}^{k_i} f_{ij}, \tag{1.1}$$

where all $f_{ij}$ are $t$-sparse. In other words, $F$ is given by a depth four arithmetic circuit with the structure $\Sigma\Pi\Sigma\Pi$, where the parameters $m$, $k := \max_i k_i$, and $t$ bound the fan-in at the different levels except at the lowest (since we do not require a bound on the degrees of the $f_{ij}$).

The following conjecture was put forward by Koiran [12].

**Conjecture 1** (Real $\tau$-conjecture)**.** *The number of real zeros of a polynomial $F$ of the form (1.1) is bounded by a polynomial in $m$, $k$, and $t$.*

Koiran [12] proved that the real $\tau$-conjecture implies a major conjecture in complexity theory, namely the separation of complexity classes $\mathrm{VP}^0 \neq \mathrm{VNP}^0$ over $\mathbb{C}$. In Tavenas' PhD thesis [17] it is shown that the real $\tau$-conjecture also implies that $\mathrm{VP} \neq \mathrm{VNP}$ over $\mathbb{C}$. Tavenas also shows that a seemingly much weaker upper bound on the number of real zeros of $F$ is sufficient to deduce $\mathrm{VP} \neq \mathrm{VNP}$: in fact, an upper bound polynomial in $m, t, 2^{\max_i k_i}$ is sufficient [17, §2.1, Cor. 3.23]. In other words, the real $\tau$-conjecture implies that the permanent of $n$ by $n$ matrices requires arithmetic circuits of superpolynomial size. For known upper bounds on the number of real zeros of polynomials of the form $F$, we refer to [13] and the references given there.

The motivation behind Conjecture 1 is Shub and Smale's $\tau$-conjecture [15] asserting that the number of integer zeros of a polynomial computed by an arithmetic circuit is polynomially bounded by the size of the circuit. If true, it gives a superpolynomial lower bound on the circuit complexity of the permanent polynomial [4]. Moreover, it also entails the separation $\mathrm{P}_{\mathbb{C}} \neq \mathrm{NP}_{\mathbb{C}}$ in the Blum-Shub-Smale model [3, 15]. One drawback of the $\tau$-conjecture is that, by referring to integer zeros, it leads to number theory, which is notorious for its hard problems. The $\tau$-conjecture is false when we replace "integer zeros" by "real zeros." Koiran's observation is that when restricting to depth four circuits, the conjecture may be true and we can still derive lower bounds for general circuits. We refer to Hrubes [9] for statements equivalent to the real $\tau$-conjecture that are related to complex zero counting. A $\tau$-conjecture for the Newton polygons of bivariate polynomials, having the same strong complexity theoretic implications, has been formulated by Koiran et al. in [14]. Hrubes [10] recently showed that the real $\tau$-conjecture implies this conjecture on Newton polytopes.

In this work, we prove that the real $\tau$-conjecture is true for random polynomials. More specifically, let $k_1, \ldots, k_m$ and $t$ be positive integers and for $1 \leq i \leq m$ and $1 \leq j \leq k_i$ we fix supports $S_{ij} \subseteq \mathbb{N}$ with $|S_{ij}| \leq t$ for the $t$-sparse polynomials $f_{ij}$. We choose the coefficients $u_{ijs}$ of the polynomials

$$f_{ij}(x) = \sum_{s \in S_{ij}} u_{ijs} x^s$$

as independent standard Gaussian random variables. The resulting $F$ given by (1.1) is a structured random polynomial and we investigate the random variable defined as the number of real zeros of $F$.

Our main result states that the expectation of the number of real zeros of $F$ is polynomially bounded in $m$, $k := \max_i k_i$, and $t$. In fact, we get an at most quadratic bound in the number of parameters!

**Theorem 1.1.** *The expectation of the number of real zeros of a polynomial $F$ of the form (1.1) is bounded as $\mathcal{O}(mk^2t)$ if the coefficient $u_{ijs}$ are independent and standard Gaussian. Thus the real $\tau$-conjecture is true on average.*

Our result can be interpreted in two ways: on the one hand, it supports the real $\tau$-conjecture since we show it is true on average; on the other hand it says that for finding a counterexample to the real $\tau$-conjecture, it is not sufficient to look at generic examples.

We do not think the assumption of Gaussian distributions is relevant. In fact, we have a partial result confirming this (Theorem 6.3). If we assume the coefficients $u_{ijs}$ are independent random variables

whose distributions have densities satisfying some mild assumptions, then the expected number of real zeros of $F$ in $[0, 1]$ is bounded by a polynomial in $k_1 + \cdots + k_m$ and $t$, provided $0 \in S_{ij}$ for all $i, j$. The latter condition means that all the $f_{ij}$ almost surely have a nonzero constant coefficient.

The main proof technique is the Rice formula from the theory of random fields, which has to be analyzed very carefully in order to achieve the good upper bounds. (In fact, we rely on a "Rice inequality," which requires less assumptions.) An interesting intermediate step of the proof is to express the expected number of real zeros of the random structured $F$ from (1.1) in terms of the expected number of real zeros of random linear combinations $R(x) = \sum_{i=1}^{m} u_i q_i(x) x^{d_i}$ of certain weight functions $q_i(x)x^{d_i}$. The deterministic functions $q_i(x)$ are obtained by multiplying and dividing sparse sums of squares in a way reflecting the build-up of the arithmetic circuit forming $F$; see (6.11). The randomness comes from independent coefficients $u_i$, whose distribution is the one of a product of $k_i$ standard Gaussians.

It would be interesting to strengthen our result by concentration statements, showing that it is very unlikely that a random $F$ of the above structure can have many real zeros.

### 1.1 | Outline of paper

Section 2 provides hands-on information on how to deal with conditional expectations, which is mainly basic calculus. In Section 3 we outline the idea of the Rice formula and state a weak version of it (Theorem 3.2), which requires only few technical assumptions. In Section 4 we prepare the ground by proving general estimates on conditional expectations of random linear combinations. Section 5 develops general results of independent interest on the expected number of real zeros of random linear combinations $\sum_{i=1}^{m} w_i(x)u_i$ of weight functions $w_i$, for independent random coefficients $u_i$ having densities satisfying some mild assumptions. We upper bound this in terms of quantities $\mathrm{LV}(w_i)$, for which we coined the name *logarithmic variations*, and which are crucial for achieving good estimations (see Definition 5.6). Finally, combining everything, we provide the proof of the main results in Section 6.

## 2 | PRELIMINARIES

We provide some background on conditional expectations in a general continuous setting, relying on some results from calculus related to the coarea formula. Then we discuss some specific properties pertaining to the distribution of products of Gaussian random variables.

### 2.1 | Conditional expectations

We fix a smooth function $f : \mathbb{R}^N \to \mathbb{R}$ with the property that $\{u \in \mathbb{R}^N : \nabla f(u) = 0\}$ has measure zero. In most of our applications, $f$ will be a nonconstant polynomial function, which satisfies this property. By Sard's theorem, almost all $a \in \mathbb{R}$ are regular values of $f$. For those $a$, the fiber $f^{-1}(a)$ is a smooth hypersurface in $\mathbb{R}^N$.

Suppose we are further given a probability distribution on $\mathbb{R}^N$ with the density $\rho$. To analyze its pushforward measure with respect to $f$, we define for a regular value $a \in \mathbb{R}$

$$\rho_f(a) := \int_{f^{-1}(a)} \frac{\rho}{\|\nabla f\|}\, df^{-1}(a) \in [0, \infty]; \tag{2.1}$$

here $df^{-1}(a)$ denotes the volume element of the hypersurface $f^{-1}(a)$. The coarea formula is a crucial tool going back to Federer [7], see [6, Thm. III.5.2, p. 138] for a comprehensive account. We only need its smooth version [6, p. 159]; see also [8, Appendix] for a short and self-contained proof. The smooth coarea formula implies that $\rho_f$ defined in (2.1) is a probability density on $\mathbb{R}$, namely the density of the random variable $f(a)$. More precisely, $\rho_f$ is the *pushforward measure* with respect to $f$ of the measure on $\mathbb{R}^N$ with density $\rho$.

Let us point out the following simple rule, which we will use all the time: for $\lambda \in \mathbb{R}^*$

$$\rho_{\lambda f}(\lambda a) = \frac{1}{|\lambda|}\rho_f(a). \tag{2.2}$$

We view now $u \in \mathbb{R}^N$ as a random variable with the density $\rho$. Let $a \in \mathbb{R}$ be a regular value of $f$ such that $\rho_f(a) > 0$. We want to define a conditional probability measure on the hypersurface $H := f^{-1}(a)$ that captures the idea that we constrain $u$ to lie in $H$. We do this by defining the *conditional density* for $u \in H$ as

$$\rho_H(u) := \frac{1}{\rho_f(a)} \frac{\rho(u)}{\|\nabla f(u)\|}.$$

Note that we indeed have $\int_H \rho_H \, dH = 1$ by construction, where $dH$ denotes the volume measure of $H$. (As a warning, let us point out that in general, $\rho_H$ does not only depend on $H$, but also on the representation of $H$ by the function $f$.) Using the conditional density, we can define the *conditional expectation*

$$\mathbb{E}(Z \mid f = a) := \int_H Z \rho_H dH \in [0, \infty]$$

of a nonnegative measurable function $Z\colon \mathbb{R}^N \to [0, \infty]$. (This quantity is only defined for regular values $a$ such that $\rho_f(a) > 0$.) In our application, we will always use the following equivalent formula

$$\mathbb{E}(Z \mid f = a)\, \rho_f(a) = \int_H Z \frac{\rho}{\|\nabla f\|} \, dH, \tag{2.3}$$

which is valid for all regular values $a$ of $f$, when interpreting the left-hand side as 0 if $\rho_f(a) = 0$. Thus by Sard's theorem, the equation makes sense for almost all $a \in \mathbb{R}$.

After defining all these notions, we summarize our discussion by stating the following important fact, which is an immediate consequence of the smooth coarea formula (cf. [6, p. 159] or [8, Appendix]).

**Proposition 2.1.** *Let $f\colon \mathbb{R}^N \to \mathbb{R}$ be a smooth function such that $\{u \in \mathbb{R}^N : \nabla f(u) = 0\}$ has measure zero. Moreover, let $\rho$ be a probability density on $\mathbb{R}^N$ and $Z\colon \mathbb{R}^N \to [0, \infty]$ be measurable. Then we have*

$$\mathbb{E}(Z) = \int_{\mathbb{R}} \mathbb{E}(Z \mid f = a)\, \rho_f(a)\, da.$$

We next discuss how to compute the right-hand side in concrete situations. As a first step, we express the volume element of the hypersurface $H$ in local coordinates. If $\partial_{u_1} f \neq 0$, then by the implicit function theorem, we can locally express $u_1$ as a function of $u_2, \dots, u_N$. The following lemma is well known. For the understanding of the following, it is helpful provide the proof.

**Lemma 2.2.** *We have*

$$dH = \frac{\|\nabla f\|}{|\partial_{u_1} f|}\, du_2 \cdots du_N.$$

*Proof.* Generally, if we parametrize $H$ by $u = \psi(t_1, \dots, t_{N-1})$, using local coordinates $t_1, \dots, t_{N-1}$, it is well known that the volume element of $H$ is given by $dH = \sqrt{\det((D\psi)^T D\psi)}\, dt_1 \cdots t_{N-1}$. In our situation, we locally write $u_1 = h(u_2, \dots, u_N)$ and use the parametrization $\psi(u_2, \dots, u_N) := (h(u_2, \dots, u_N), u_2, \dots, u_N)$ of $H$. A straightforward calculation shows $(D\psi)^T D\psi = I + \nabla h(\nabla h)^T$. Moreover, $\det(I + \nabla h(\nabla h)^T) = 1 + \|\nabla h\|^2$. (In order to see this, use the orthogonal matrix $S \in O(N)$ such that $S\nabla h = (0, \dots, 0, \|\nabla h\|)$.) Hence the volume element of $H$ satisfies

$$dH = \sqrt{1 + \|\nabla h\|^2}\, du_2 \cdots du_N.$$

By implicit differentiation we get $\partial_{u_i} h = -\partial_{u_i} f / \partial_{u_1} f$. Hence,

$$1 + \|\nabla h\|^2 = \frac{\|\nabla f\|^2}{(\partial_{u_1} f)^2},$$

and the assertion follows. ∎

Assume now that $H$ is parametrized when $(u_2, \dots, u_N)$ runs over (an open dense subset of) $\mathbb{R}^{N-1}$. Then, due to Lemma 2.2, we can express the pushforward density $\rho_f$ as follows:

$$\rho_f(a) = \int_{\mathbb{R}^{N-1}} \frac{\rho}{|\partial_{u_1} f|}\, du_2 \cdots du_N. \tag{2.4}$$

Moreover, Formula (2.3) reads as

$$\mathbb{E}\,(Z \mid f = a)\, \rho_f(a) = \int_{\mathbb{R}^{N-1}} Z\, \frac{\rho}{|\partial_{u_1} f|}\, du_2 \cdots du_N. \tag{2.5}$$

**Example 2.3.** Consider the linear function $f(u) = \sum_{i=1}^N w_i u_i$ for a nonzero $w \in \mathbb{R}^N$. Then $H = f^{-1}(a)$ is a hyperplane and $\nabla f = w$. We have by definition

$$\rho_f(a) = \frac{1}{\|w\|} \int_H \rho\, dH, \quad \rho_H(u) = \left( \int_H \rho dH \right)^{-1} \rho(u).$$

If $w_1 = \partial_{u_1} f \neq 0$, Formula (2.5) gives

$$\mathbb{E}\,(Z \mid f = a)\, \rho_f(a) = \frac{1}{|w_1|} \int_{\mathbb{R}^{N-1}} Z\rho\, du_2 \cdots du_N.$$

In the special case $f(u) = u_N$, we retrieve the known notion of the marginal distribution $\rho_{u_N}(a) = \int_{\mathbb{R}^{N-1}} \rho(u_1, \dots, u_{N-1}, a)\, du_1 \cdots du_{N-1}$, and the conditional density of $Z$ satisfies

$$\mathbb{E}\,(Z \mid u_N = a)\, \rho_{u_N}(a) = \int_{\mathbb{R}^{N-1}} Z(u_1, \dots, u_{N-1}, a) \rho(u_1, \dots, u_{N-1}, a)\, du_1 \cdots du_{N-1}. \tag{2.6}$$

**Example 2.4.** Consider the product function $f(y) = y_1 \cdot \ldots \cdot y_k$, and for nonzero $a \in \mathbb{R}$ the smooth hypersurface

$$C_a := \{y \in \mathbb{R}^k : y_1 \cdot \ldots \cdot y_k = a\}.$$

If $\rho$ is the joint density of $y \in \mathbb{R}^k$, then the pushforward density $\rho_f$ of the product $f(y)$ satisfies, by (2.1) and Lemma 2.2, that

$$\rho_f(a) = \int_{C_a} \frac{\rho}{\|\nabla f\|}\, dC_a = \int_{\mathbb{R}^{k-1}} \frac{\rho}{\partial_{y_1} f}\, dy_2 \cdots dy_k = \int_{\mathbb{R}^{k-1}} \rho\, \frac{dy_2}{|y_2|} \cdots \frac{dy_k}{|y_k|}, \tag{2.7}$$

since $\partial_{y_1} f = y_2 \cdots y_k$. We also note that $\|\nabla f(y)\| = |a|(\sum_{i=1}^k y_i^{-2})^{\frac{1}{2}}$. Moreover, (2.3) combined with Lemma 2.2, reads as

$$\mathbb{E}\,(Z \mid f = a)\, \rho_f(a) = \int_H Z \frac{\rho}{\|\nabla f\|}\, dC_a = \int_{\mathbb{R}^{k-1}} Z\rho\, \frac{dy_2}{|y_2|} \cdots \frac{dy_k}{|y_k|}. \tag{2.8}$$

## 2.2 | Products of Gaussians

In the sequel, we denote by $\varpi_k$ the density of the product $y_1 \cdot \ldots \cdot y_k$ of independent standard Gaussian distributed random variables $y_1, \ldots, y_k$; see [16]. According to (2.7) we have for $a \in \mathbb{R}^*$

$$\varpi_k(a) = \int_{(y_2,\ldots,y_k)\in\mathbb{R}^{k-1}} \varphi\Big(\frac{a}{y_2 \cdot \ldots \cdot y_k}\Big)\varphi(y_2) \cdot \ldots \cdot \varphi(y_k)\, \frac{dy_2}{|y_2|} \cdots \frac{dy_k}{|y_k|}, \tag{2.9}$$

where $\varphi(y) = (2\pi)^{-\frac{1}{2}} e^{-\frac{y^2}{2}}$ denotes the density of the standard Gaussian distribution.

More generally, if $y_i \sim \mathcal{N}(0, \sigma_i^2)$ are independent centered Gaussians with variance $\sigma_i^2$, then we may write $y_i = \sigma_i \tilde{y}_i$ with $\tilde{y}_i \sim \mathcal{N}(0,1)$. The density $\rho_f$ of the product $f(y) = y_1 \cdot \ldots \cdot y_k = \sigma_1 \cdot \ldots \cdot \sigma_k \tilde{y}_1 \cdot \ldots \cdot \tilde{y}_k$ then can be expressed via (2.2) as

$$\rho_f(a) = \frac{1}{\sigma_1 \cdot \ldots \cdot \sigma_k}\, \varpi_k\left(\frac{|a|}{\sigma_1 \cdot \ldots \cdot \sigma_k}\right). \tag{2.10}$$

It is easy to see that the density $\varpi_k$ of the product of $k$ standard Gaussians is unbounded for $k \geq 2$: we have $\lim_{a\to 0} \varpi_k(a) = \infty$, which causes some technical problems. However, the following lemma states that the growth of $\varpi_k$ for $a \to 0$ is slow, which will be needed for the proof of Theorem 1.1: more specifically, for guaranteeing the assumption (4.1) so that Proposition 5.5 can be applied to the random linear combination $R(x) = \sum_{i=1}^m u_i q_i(x) x^{d_i}$, where the coefficients $u_i$ are independent random variables with the distribution $\varpi_{k_i}$.

**Lemma 2.5.** *(1) $\varpi_k$ is monotonically decreasing on $(0,\infty)$ and $\varpi_k(-a) = \varpi_k(a)$.*
*(2) For $0 < \delta \leq \frac{1}{2}$ and $a \in \mathbb{R}^*$ we have $\varpi_2(a) \leq |a|^{\delta-1}$.*
*(3) For $a \in \mathbb{R}^*$ we have $\varpi_k(a) \leq e\, |a|^{\frac{1}{2k}-1}$.*

*Proof.* (1) Taking the derivative in (2.9) we obtain, using symmetry, that

$$\varpi_k'(a) = 2^{k-1} \int_{(y_2,\ldots,y_k)\in\mathbb{R}_+^{k-1}} \varphi'\left(\frac{a}{y_2 \cdot \ldots \cdot y_k}\right) \varphi(y_2) \cdot \ldots \cdot \varphi(y_k)\, \frac{dy_2}{y_2^2} \cdots \frac{dy_k}{y_k^2}.$$

Since $\varphi'(y) \le 0$ for $y \ge 0$, we see that $\varpi_k'(a) \le 0$ for $a > 0$. It is clear that $\varpi_k(-a) = \varpi_k(a)$.

(2) By (2.9) we have

$$\varpi_2(a) = \int_{\mathbb{R}} \varphi\left(\frac{a}{y}\right) \varphi(y) \frac{dy}{|y|} = \frac{2}{2\pi} \int_0^\infty \frac{1}{y} e^{-\frac{a^2}{2y^2}} e^{-\frac{y^2}{2}}\, dy,$$

which we bound as follows:

$$\varpi_2(a) \;\le\; \frac{1}{\pi} \int_0^1 \frac{1}{y} e^{-\frac{a^2}{2y^2}}\, dy + \frac{1}{\pi} \int_1^\infty e^{-\frac{y^2}{2}}\, dy \;\le\; \frac{1}{\pi} \int_0^1 \frac{1}{y} e^{-\frac{a^2}{2y^2}}\, dy + \frac{1}{\sqrt{2\pi}}. \tag{2.11}$$

Let $0 < \delta \le \frac{1}{2}$. Since $2x^{\frac{1-\delta}{2}} e^{-x} \le 1$ for $x \ge 0$, we obtain $e^{-\frac{a^2}{2y^2}} \le 2^{-\frac{1}{2}} |a|^{\delta-1} y^{1-\delta}$ for all $y > 0$. Integrating, we obtain

$$\int_0^1 \frac{1}{y} e^{-\frac{a^2}{2y^2}}\, dy \;\le\; 2^{-\frac{1}{2}} |a|^{\delta-1} \int_0^1 y^{-\delta}\, dy \;=\; 2^{-\frac{1}{2}} \frac{|a|^{\delta-1}}{1-\delta} \;\le\; 2^{\frac{1}{2}} |a|^{\delta-1}.$$

Altogether, we get from (2.11) for $|a| \le 1$,

$$\varpi_2(a) \;\le\; \frac{\sqrt{2}}{\pi} |a|^{\delta-1} + \frac{1}{\sqrt{2\pi}} \;\le\; \left(\frac{\sqrt{2}}{\pi} + \frac{1}{\sqrt{2\pi}}\right) |a|^{\delta-1} \;<\; |a|^{\delta-1}.$$

One can check that $a^{1-\delta} \varpi_2(a) \le a \varpi_2(a) < 1$ for $a \ge 1$. The assertion follows.

(3) The case $k = 1$ follows from (1). Suppose now $k \ge 2$. We have by (2.9)

$$\begin{aligned}
\varpi_k(a) &= \int_{(y_2,\dots,y_k)\in\mathbb{R}^{k-1}} \varphi\left(\frac{a}{y_2 \cdot \ldots \cdot y_k}\right) \varphi(y_2) \cdot \ldots \cdot \varphi(y_k) \frac{dy_2}{|y_2|} \cdots \frac{dy_k}{|y_k|} \\
&= \int_{(y_3,\dots,y_k)\in\mathbb{R}^{k-2}} \left[\int_{y_2\in\mathbb{R}} \varphi\left(\frac{a}{y_2 \cdot \ldots \cdot y_k}\right) \varphi(y_2) \frac{dy_2}{|y_2|}\right] \varphi(y_3) \cdot \ldots \cdot \varphi(y_k) \frac{dy_3}{|y_3|} \cdots \frac{dy_k}{|y_k|} \\
&= \int_{(y_3,\dots,y_k)\in\mathbb{R}^{k-2}} \varpi_2\left(\frac{a}{y_3 \cdot \ldots \cdot y_k}\right) \varphi(y_3) \cdot \ldots \cdot \varphi(y_k) \frac{dy_3}{|y_3|} \cdots \frac{dy_k}{|y_k|}.
\end{aligned}$$

By item (2) we can bound this by

$$\varpi_k(a) \;\le\; |a|^{\delta-1} \left(\int_{y\in\mathbb{R}} |y|^{-\delta} \varphi(y)\, dy\right)^{k-2} = |a|^{\delta-1} \left(\mathbb{E}\, |y|^{-\delta}\right)^{k-2}.$$

Is well known that

$$\mathbb{E}\, |y|^{-\delta} = \frac{1}{\sqrt{\pi}} 2^{-\frac{\delta}{2}} \Gamma\left(\frac{1-\delta}{2}\right) \;\le\; \frac{1}{\sqrt{\pi}} \Gamma\left(\frac{1-\delta}{2}\right).$$

The Taylor expansion $\frac{1}{\sqrt{\pi}} \Gamma(\frac{1-\delta}{2}) = 1 + 0.9819\ldots \cdot \delta + O(\delta^2)$ gives the growth for small $\delta$: it is straightforward to verify that $\frac{1}{\sqrt{\pi}} \Gamma(\frac{1-\delta}{2}) \le 1 + 2\delta$ for $0 < \delta \le \frac{1}{2}$. Setting $\delta = 1/(2k)$, we obtain $\left(\mathbb{E}\, |y|^{-\delta}\right)^{k-2} \le (1 + 2\delta)^k = (1 + \frac{1}{k})^k < e$ and assertion follows. ∎

## 3 | THE RICE FORMULA

### 3.1 | Outline

The Rice formula is a major tool in the theory of random fields. It gives a concise integral expression for the expected number of zeros of random functions. For comprehensive treatments we refer to [1,2].

We are going to apply this formula in the following special situation. Let $\mathbb{R}[X]_{\leq D}$ denote the finite dimensional space of polynomials of degree at most $D$ in the single variable $X$. We study a family of structured polynomials given by a parametrization $\mathbb{R}^N \to \mathbb{R}[X]_{\leq D}, u \mapsto F_u(X)$, where $F_u(X)$ is a polynomial function in the parameter $u$ and the variable $X$. In our case of interest, it is the parametrization of polynomials by arithmetic circuits of depth four in terms of their parameters.

Here is a rough outline of the method. We fix a probability density on the space $\mathbb{R}^N$ of parameters. Its pushforward measure on $\mathbb{R}[X]_{\leq D}$ defines a class of random polynomial functions $F : \mathbb{R} \to \mathbb{R}$. (It is common to notationally drop the dependence on the parameter $u$.) The number $\#\{x \in [0,1] : F(x) = 0\}$ of real zeros of $F$ then becomes a random variable, whose expectation we wish to analyze. For this, let us assume that for almost all $x \in \mathbb{R}$, the real random variable $F(x)$ has a density, denoted by $\rho_{F(x)}$. Moreover, we assume that the conditional expectation $\mathbb{E}\left(|F'(x)| \mid F(x) = 0\right)$ is well defined for almost all $x \in \mathbb{R}$. The *Rice formula* states that, under some technical assumptions,

$$\mathbb{E}\left(\#\{x \in [0,1] : F(x) = 0\}\right) = \int_0^1 \mathbb{E}\left(|F'(x)| \mid F(x) = 0\right) \rho_{F(x)}(0)\, dx.$$

While the idea behind this formula can be easily explained (e.g., see [2, §3.1]), the rigorous justification can be quite hard, especially in case of non-Gaussian distributions that we encounter in our work; compare [2, Thm. 3.4]). For this reason, we will rely on a weaker version of the Rice formula, tailored to our situation, that only claims the inequality $\leq$ above, but has the advantage of requiring less assumptions. This is the topic of the next subsection. Let us emphasize that we do not attempt to state this weaker version of the Rice formula in the greatest generality possible.

### 3.2 | A Rice inequality

Let $\mathbb{R}^N \times I \to \mathbb{R}$, $(u,x) \mapsto F_u(x)$ be a polynomial function, where $I$ is a compact interval. We think of $F$ as a parametrization of structured polynomial functions in the variable $x$ in terms of the parameters $u_1, \ldots, u_N$. We assume that for all $x \in I$, the polynomial function

$$F(x) : \mathbb{R}^N \to \mathbb{R},\ u \mapsto F_u(x)$$

is not constant and thus $\{u \in \mathbb{R}^N : \nabla F(x)(u) = 0\}$ has measure zero.

**Example 3.1.** (1) Fix integers $0 = d_1 < d_2 < \cdots < d_t$. Then $F_u(x) := u_1 + u_2 x^{d_2} + \cdots + u_t x^{d_t}$ parametrizes sparse polynomials with support $\{d_1, \ldots, d_t\}$. Note that for all $x \in \mathbb{R}$, $F(x)$ is a nonconstant linear function (of the argument $u$). In particular, $F(x)$ does not have singular values.

(2) Fix integers $0 = d_1 < d_2 < \cdots < d_t$ and $0 = e_1 < e_2 < \cdots < e_t$. The family $F_{u,v}(x) := (u_1 + u_2 x^{d_2} + \cdots + u_t x^{d_t})(v_1 + v_2 x^{e_2} + \cdots + v_t x^{e_t})$ parameterizes products of two sparse polynomials with supports given by $\{d_1, \ldots, d_t\}$ and $\{e_1, \ldots, e_t\}$. The set of singular points of $F(x)$ consists of the pairs $(u,v)$ such that $u_1 + u_2 x^{d_2} + \cdots + u_t x^{d_t} = 0, v_1 + v_2 x^{e_2} + \cdots + v_t x^{e_t} = 0$. Thus, for all $x \in \mathbb{R}$, $F(x)$ is surjective and 0 is its only singular value. (We generalize this example in Lemma 6.1.)

Following Section 2.1, if a probability distribution with a density $\rho$ is given on the space $\mathbb{R}^N$ of parameters, for all $x \in I$, $F(x)$ becomes a random variable with a well-defined density $\rho_{F(x)}$.

The following "Rice inequality" is the version of Rice's formula that we apply in this paper. It is essentially Azaïs and Wschebor [2, Exercise 3.9, p. 69]. We state it in a way that makes the method convenient to apply in our setting. We provide the proof for lack of a suitable reference.

**Theorem 3.2.** *Let $\mathbb{R}^N \times [x_0, x_1] \to \mathbb{R}$, $(u, x) \mapsto F_u(x)$ be a smooth function such that, for all $x \in [x_0, x_1]$, $\{u \in \mathbb{R}^N : \nabla F(x)(u) = 0\}$ has measure zero. Moreover, we assume that, for almost all $u \in \mathbb{R}^N$, the function $[x_0, x_1] \to \mathbb{R}$ has only finitely many zeros. Furthermore, let a probability density $\rho$ be given on $\mathbb{R}^N$. We assume there exists an integrable function $g : [x_0, x_1] \to [0, \infty]$ and $\varepsilon > 0$ such that for all $x \in [x_0, x_1]$ and almost all $a \in (-\varepsilon, \varepsilon)$ we have*

$$\mathbb{E}\left(|F'(x)| \mid F(x) = a\right) \rho_{F(x)}(a) \;\le\; g(x).$$

*Then, for a random $u$ with the density $\rho$, we can bound the expected number of zeros of the random function $x \mapsto F_u(x)$ in the interval $[x_0, x_1]$ as follows:*

$$\mathbb{E}\left(\#\{x \in [x_0, x_1] : F(x) = 0\}\right) \;\le\; \int_{x_0}^{x_1} g(x)\, dx.$$

The starting point for the proof of Theorem 3.2 is Kac's counting formula [11, Lemma 1 and Remark 1]. A turning point of function is a point where its derivative changes sign.

**Lemma 3.3.** *A $C^1$ function $f : [x_0, x_1] \to \mathbb{R}$ with only finitely many turning points satisfies*

$$N(f) := \#\{x \in (x_0, x_1) : f(x) = 0\} \;\le\; \lim_{\delta \to 0} \frac{1}{2\delta} \int_{x_0}^{x_1} \mathbb{1}_{\{|f(x)| < \delta\}}\, |f'(x)|\, dx.$$

*In fact, for sufficiently small $\delta > 0$, the right-hand side equals $N(f) + \eta$, where $\eta = 0, \frac{1}{2}, 1$ according to as none, one, or both of the numbers $x_0, x_1$ are zeros of $f$.*

*Proof of Theorem 3.2.* In the setting of this theorem, we apply Lemma 3.3 to $f = F_u$ for a random $u \in \mathbb{R}^N$. Taking expectations over $u$ and using Fatou's lemma, we obtain (for convenience, we drop the index $u$)

$$\mathbb{E}\left(N(F)\right) \;\le\; \liminf_{\delta \to 0} \mathbb{E}\left(\frac{1}{2\delta} \int_{x_0}^{x_1} \mathbb{1}_{\{|F(x)| < \delta\}}\, |F'(x)|\, dx\right).$$

Due to Tonelli's lemma (nonnegative integrands), we can interchange the integral over $x$ and the expectation. We obtain

$$\mathbb{E}\left(N(F)\right) \;\le\; \liminf_{\delta \to 0} \int_{x_0}^{x_1} J_\delta(x)\, dx,$$

where we have put

$$J_\delta(x) := \frac{1}{2\delta}\, \mathbb{E}\left(\mathbb{1}_{\{|F(x)| < \delta\}}\, |F'(x)|\right).$$

Proposition 2.1 gives for $x \in (x_0, x_1)$,

$$J_\delta(x) = \frac{1}{2\delta} \int_{-\delta}^{\delta} \mathbb{E}\left(|F'(x)| \mid |F(x)| = a\right) \rho_{F(x)}(x)\, da.$$

By assumption, the integrand is upper bounded by $g(x)$ for almost all $a \in (-\varepsilon, \varepsilon)$, hence we obtain $J_\delta(x) \leq g(x)$ for $\delta < \varepsilon$. Therefore,

$$\mathbb{E}(N(F)) \leq \liminf_{\delta \to 0} \int_{x_0}^{x_1} J_\delta(x)\, dx \leq \int_{x_0}^{x_1} g(x)\, dx.$$

Finally, $\mathbb{E}(\#\{x \in (x_0, x_1) : f(x) = 0\}) = \mathbb{E}(N(F))$ since $F(x_0) = 0$ and $F(x_1) = 0$ happens with probability zero. ∎

## 4 | CONDITIONAL EXPECTATIONS OF RANDOM LINEAR COMBINATIONS

Throughout, we assume that $u_1, \ldots, u_m$ are independent real random variables having the densities $\varphi_1, \ldots, \varphi_m$, respectively. We fix real weights $w_1, \ldots, w_m$, not all being zero, and study the random variable

$$f := w_1 u_1 + \cdots + w_m u_m.$$

We shall study bounds for the quantity $\mathbb{E}(|u_i| \mid f = a)\, \rho_f(a)$. Since $\nabla f = w \neq 0$, there is no singular value of $f$.

We begin with a simple bound on the density $\rho_f$ of $f$. It is only useful if the densities $\varphi_i$ are bounded (which is not the case for $\varphi = \varpi_k$).

**Lemma 4.1.** *Suppose that $\|\varphi_i\|_\infty \leq A$ for all $i$. Then $\|\rho_f\|_\infty \leq \frac{A}{\max_i |w_i|}$. In particular, we have $\|\rho_f\|_\infty \leq A$ if $w_i = 1$ for some $i$.*

*Proof.* For $a \in \mathbb{R}$ we have by (2.4)

$$\rho_f(a) = \frac{1}{|w_1|} \int_{\mathbb{R}^{k-1}} \varphi_1(w_1^{-1}(a - w_2 u_2 - \cdots - w_m u_m))\varphi_2(u_2) \cdot \ldots \cdot \varphi_m(u_m) du_2 \cdots du_m,$$

which we can bound as

$$\rho_f(a) \leq \frac{A}{|w_1|} \cdot \int_{\mathbb{R}^{k-1}} \varphi_2(u_2) \cdots \varphi_m(u_m) du_2 \cdots du_m = \frac{A}{|w_1|}.$$

Since the same argument works for $w_i$, this finishes the proof. ∎

**Definition 4.2.** We call a probability density $\varphi$ on $\mathbb{R}$ *convenient* if $\varphi$ is monotonically decreasing on $(0, \infty)$ and symmetric around the origin, that is, $\varphi(-u) = \varphi(u)$ for all $u \in \mathbb{R}$. Moreover, we require

$$\mathbb{E}_\varphi := \int_{\mathbb{R}} |u| \varphi(u)\, du < \infty.$$

Clearly, a distribution with a convenient density $\varphi$ is centered: $\int_{\mathbb{R}} u\varphi(u)\,du = 0$. The densities $\varpi_k$ of the products of independent Gaussian random variables provide examples of convenient densities (see Section 2.2). Note that $\mathbb{E}_{\varpi_k} = (\mathbb{E}_{\varphi})^k \leq 1$ with $\varphi$ denoting the density of the standard Gaussian distribution.

**Lemma 4.3.** *Let $\varphi$ and $\psi$ be densities on $\mathbb{R}$ and assume that $\varphi$ is convenient. Then:*

*(1)* $|u|\varphi(u) \leq \frac{1}{2}$.
*(2)* $\int_{\mathbb{R}} |u|\varphi(u)\psi(u)du \leq 1$.

*Proof.* (1) We have $u\varphi(u) \leq \int_0^u \varphi(t)\,dt \leq \int_0^\infty \varphi(t)\,dt = \frac{1}{2}$, for $u > 0$.

(2) By Fubini,

$$\int_0^\infty u\varphi(u)\psi(u)du = \int_0^\infty \left(\int_0^u dt\right)\varphi(u)\psi(u)du$$
$$= \int_{0\leq t\leq u} \varphi(u)\psi(u)dt\,du = \int_0^\infty \int_t^\infty \varphi(u)\psi(u)dudt.$$

Now we use that $\varphi$ is monotonically decreasing on $(0,\infty)$ to upper bound this by

$$\int_0^\infty \varphi(t)\int_t^\infty \psi(u)du\,dt \leq \int_0^\infty \varphi(t)dt = \frac{1}{2}.$$

The assertion follows by the symmetry of $\varphi$. ∎

**Proposition 4.4.** *Consider $f = w_1u_1 + \cdots + w_mu_m$, where $(w_1,\ldots,w_m) \neq 0$. If the density $\varphi_i$ of $u_i$ is convenient, then we have for any $a \in \mathbb{R}$*

$$|w_i|\,\mathbb{E}\,(|u_i|\,|f = a)\,\rho_f(a) \leq 1.$$

*Proof.* We begin with a general observation. Let $v_1$ and $v_2$ be independent random variables with the densities $\psi_1$ and $\psi_2$ and assume $\psi_1$ to be convenient. Consider the sum $g(v_1,v_2) := v_1 + v_2$. By (2.5) we have for $a \in \mathbb{R}$,

$$\mathbb{E}\,(|v_1|\,|\,g(v_1,v_2) = a)\,\rho_g(a) = \int_{\mathbb{R}} |v_1|\psi_1(v_1)\psi_2(a-v_1)\,dv_1$$

and Lemma 4.3(2) implies $\mathbb{E}\,(|v_1|\,|\,v_1+v_2=a)\,\rho_g(a) \leq 1$. Applying this observation to $v_1 := w_iu_i$ and $v_2 := \sum_{j\neq i} w_ju_j$ yields the assertion. ∎

We provide now another bound on the conditional expectation, which is better for small weights $w_i$. For this we need a stronger assumption on the densities. We will have to deal with unbounded densities, namely with the density $\varpi_k$ of the product of $k \geq 2$ standard Gaussian random variables. Lemma 2.5 will allow us to apply the following result to these densities.

**Proposition 4.5.** *Suppose $u_i$ has a convenient density $\varphi_i$ with $\mathbb{E}_{\varphi_i} \leq B$, for $i = 2,\ldots,m$. Furthermore, assume the density $\varphi_1$ of $u_1$ satisfies*

$$\forall u\ \varphi_1(u) \leq C\,|u|^{\delta-1} \tag{4.1}$$

*for some constants $C > 0$ and $0 < \delta \leq 1$. Then, for all $w_2, \ldots, w_m \in \mathbb{R}$, the random linear combination $f := u_1 + w_2 u_2 + \cdots + w_m u_m$ satisfies for $i \geq 2$ and all $a \in \mathbb{R}$,*

$$\mathbb{E}\,(|u_i| \mid f = a)\,\rho_f(a) \;\leq\; C\left(\delta^{-1} + B\right)|w_i|^{\delta-1}.$$

*Proof.* Using the symmetry of $\varphi_i$, we can assume w.l.o.g. that all the weights $w_i$ are positive. We first provide the proof in the case $m = 2$. Let $f = u_1 + w u_2$ with $w > 0$. By (2.5) we have

$$I := \mathbb{E}\,(|u_2| \mid f = a)\,\rho_f(a) = \int_{\mathbb{R}} |u_2|\varphi_1(a - w u_2)\varphi_2(u_2) du_2. \tag{4.2}$$

By assumption, we have $\varphi_1(a - w u_2) \leq C|a - w u_2|^{\delta-1}$ for all $u_2 \in \mathbb{R}$. Using this, we obtain

$$I \;\leq\; C\int_{\mathbb{R}} |a - w u_2|^{\delta-1}|u_2|\varphi_2(u_2)\,du_2 \;\leq\; C|w|^{\delta-1}\int_{\mathbb{R}} \left|\frac{a}{w} - u_2\right|^{\delta-1}|u_2|\varphi_2(u_2)\,du_2.$$

We bound this integral by splitting according to whether $\left|\frac{a}{w} - u_2\right|$ is smaller or larger than one. Using that $|u_2|\varphi_2(u_2) \leq \frac{1}{2}$, which holds since $\varphi_2$ is convenient (see Lemma 4.3(1)), we get

$$\begin{aligned}\int_{\mathbb{R}} \left|\frac{a}{w} - u_2\right|^{\delta-1}|u_2|\varphi_2(u_2)\,du_2 \;&\leq\; \frac{1}{2}\int_{|u_2 - a/w| \leq 1} \left|\frac{a}{w} - u_2\right|^{\delta-1} du_2 + \int_{|u_2 - a/w| \geq 1} |u_2|\varphi_2(u_2)\,du_2 \\ &\leq\; \frac{1}{2}\int_{-1}^{1} |x|^{\delta-1}\,dx \;+\; \mathbb{E}_{\varphi_2} \leq \frac{1}{\delta} \;+\; B.\end{aligned}$$

We have thus shown that $\mathbb{E}\,(|u_2| \mid f = a)\,\rho_f(a) \leq C'|w|^{\delta-1}$, where $C' := C(\delta^{-1} + B)$, settling the case $m = 2$.

We now turn to the general case $m \geq 2$. Let $f := u_1 + w_2 u_2 + \cdots + w_m u_m$ and w.l.o.g. $i = 2$. As for (4.2),

$$\begin{aligned}&\mathbb{E}\,(|u_2| \mid f = a)\;\rho_f(a) = \\ &\int_{\mathbb{R}^{m-1}}\int_{\mathbb{R}} |u_2|\varphi_1((a - w_3 u_3 - \cdots - w_m u_m) - w_2 u_2)\varphi_2(u_2)\,du_2\;\varphi_3(u_3)\cdots\varphi_m(u_m)du_3\cdots du_m.\end{aligned}$$

We bound the inner integral using the case $m = 2$ and obtain,

$$\mathbb{E}\,(|u_2| \mid f = a)\;\rho_f(a) \;\leq\; C'|w_2|^{\delta-1}\int_{\mathbb{R}^{m-1}} \varphi_3(u_3)\cdots\varphi_m(u_m)du_3\cdots du_m = C'|w_2|^{\delta-1},$$

which finishes the proof. ∎

## 5 | RANDOM LINEAR COMBINATIONS OF FUNCTIONS

Throughout this section we fix analytic functions $w_1, \ldots, w_m : [x_0, x_1] \to \mathbb{R}$ and study for $u \in \mathbb{R}^m$ their linear combination

$$F(x) := \sum_{i=1}^{m} w_i(x)u_i.$$

We assume that $w_1, \ldots, w_m$ do not have a common zero in $[x_0, x_1]$. Note that $\nabla F(x) = (w_1(x), \ldots, w_m(x)) \neq 0$ for all $x$.

**Lemma 5.1.** *The set of $u \in \mathbb{R}^m$ such that $\sum_{i=1}^m w_i(x)u_i$ has finitely many zeros is of measure zero.*

*Proof.* W.l.o.g. we can assume that $w_1, \dots, w_k$ is a basis of the span of $w_1, \dots, w_m$, where $k \geq 1$. If we write $w_i = \sum_{j=1}^k \lambda_{ij} w_j$, for $i > k$ with $\lambda_{ij} \in \mathbb{R}$, then $F(x) = \sum_{i=1}^m w_i(x)u_i = \sum_{j=1}^k w_j(x)v_j$, where $v_j = u_j + \sum_{i=k+1}^m \lambda_{ij}u_i$. If the analytic function $F(x)$ has infinitely many zeros in $[x_0, x_1]$, then it must vanish identically and thus $v_j = u_j + \sum_{i=k+1}^m \lambda_{ij}u_i = 0$ for all $j \leq k$. Since the set of $u \in \mathbb{R}^m$ satisfying these conditions lie in a lower dimensional subspace, the assertion follows. ∎

We note that any family of polynomials without common zeros in $[x_0, x_1]$ satisfies the above assumptions. For instance, we can take the family of monomials $w_i(x) = x^{d_i}$ with $d_1 = 0 \leq d_2 \leq \cdots \leq d_m$, which amounts to studying the random fewnomial $F(x) = \sum_{i=1}^m u_i x^{d_i}$.

We assume now that the $u_1, \dots, u_m$ are independent real random variables with the densities $\varphi_1, \dots, \varphi_m$ and consider random linear combination $F(x)$. (Notationally, we again drop the dependence on $u$.) Our goal is to bound the expected number of real zeros of $F$ via the Rice inequality.

We begin with a simple estimation.

**Proposition 5.2.** *Suppose $A, B$ are constants such that*

$$\forall i \quad \|\varphi_i\|_\infty \leq A, \quad \mathbb{E}_{\varphi_i} \leq B.$$

*Then $F(x) := u_1 + \sum_{i=2}^m w_i(x)u_i$ satisfies for all $x \in [x_0, x_1]$ and $a \in \mathbb{R}$,*

$$\mathbb{E}\left(|F'(x)| \mid F(x) = a\right) \rho_{F(x)}(a) \leq AB \sum_{i=2}^m |w_i'(x)|.$$

*Moreover, we have*

$$\mathbb{E}\,\#\{x \in [x_0, x_1] : F(x) = 0\} \leq AB \sum_{i=2}^m \int_{x_0}^{x_1} |w_i'(x)| dx.$$

*Proof.* We have $F'(x) = \sum_{i=2}^m w_i'(x)u_i$, hence $|F'(x)| \leq \sum_{i=2}^m |w_i'(x)| \cdot |u_i|$. If we put $w_j := w_j(x)$ for fixed $x$, we have

$$\mathbb{E}\left(|F'(x)| \mid F(x) = a\right) \leq \sum_{i=2}^m |w_i'(x)| \cdot \mathbb{E}\left(|u_i| \mid \sum_{j=1}^m w_j u_j = a\right).$$

Then $u_2$ and $v := u_1 + w_3u_3 + \cdots + w_mu_m$ are independent random variables and $F(x) = w_2u_2 + v$. Let $\vartheta$ denote the density of $v$. By Lemma 4.1 we have $\|\vartheta\|_\infty \leq A$. Hence, by (2.5),

$$\mathbb{E}\left(|u_2| \mid F(x) = a\right) \rho_{F(x)}(a) = \int_{\mathbb{R}} |u_2| \varphi_2(u_2) \vartheta(a - w_2u_2)\, du_2 \leq A \cdot \mathbb{E}_{\varphi_2} \leq AB.$$

The same bound holds for all $u_i$ with $i \geq 2$ and the first assertion follows.

By the assumptions on the functions $w_i$ made at the beginning of Section 5, we can apply Theorem 3.2 and the second assertion follows. ∎

The following corollary is of independent interest.

**Corollary 5.3.** *In the situation of Proposition 5.2, if all functions $w_i$ are monotonically increasing, then*

$$\mathbb{E}\,\#\{x \in [x_0, x_1] : F(x) = 0\} \;\leq\; AB \sum_{i=2}^{m} (w_i(x_1) - w_i(x_0)).$$

In particular, in the case of monomials $w_i(x) = x^{d_i}$ with $d_1 = 0 < d_2 < \cdots < d_m$, the random fewnomial $F(x) = \sum_{i=1}^{m} u_i x^{d_i}$ satisfies $\mathbb{E}\,\#\{x \in [0,1] : F(x) = 0\} \leq AB\,(m-1)$, which can be seen as a probabilistic version of Descartes rule.

*Remark* 5.4. Better bounds can be obtained for particular probability distributions of the coefficients $u_i$. For instance, one can show that $\mathbb{E}\,\#\{x \in \mathbb{R} : F(x) = 0\} = \mathcal{O}(\sqrt{m}\log m)$ for the random sparse polynomial $F(x) = \sum_{i=1}^{m} u_i x^{d_i}$ with independent standard Gaussian coefficients; see [5].

Following Proposition 4.5, we now provide an estimation, which is better for small values of $w_i(x)$. It is relevant that this does not require the density $\varphi_i$ to be bounded. This estimation can be applied to the distributions of products of independent Gaussians, which will be of importance for the proof of the main result.

**Proposition 5.5.** *Suppose $u_i$ has a convenient density $\varphi_i$ with $\mathbb{E}_{\varphi_i} \leq B$, for $i = 2, \ldots, m$. Furthermore, assume there are $C \geq 1$ and $0 < \delta \leq 1$ such that the density $\varphi_1$ of $u_1$ satisfies $\varphi_1(u) \leq C\,|u|^{\delta-1}$ for all $u$. Then, for all $w_2, \ldots, w_m \in \mathbb{R}$, the random linear combination $F(x) := u_1 + \sum_{i=2}^{m} w_i(x) u_i$ satisfies for all $x \in [x_0, x_1]$ and all $a \in \mathbb{R}$:*

$$\mathbb{E}\,\left(|F'(x)| \mid F(x) = a\right) \rho_{F(x)}(a) \;\leq\; C(\delta^{-1} + B) \sum_{i=2}^{m} \frac{|w_i'(x)|}{\max\{|w_i(x)|, |w_i(x)|^{1-\delta}\}}.$$

*Proof.* Put $C' := C(\delta^{-1} + B)$. Proposition 4.5 gives for $i \geq 2$, $a \in \mathbb{R}$, and $x \in [x_0, x_1]$,

$$\mathbb{E}\,(|u_i| \mid F(x) = a)\,\rho_{F(x)}(a) \;\leq\; \frac{C'}{|w_i(x)|^{1-\delta}}.$$

On the other hand, Proposition 4.4 gives

$$\mathbb{E}\,(|u_i| \mid F(x) = a)\,\rho_{F(x)}(a) \;\leq\; \frac{1}{|w_i(x)|}.$$

Therefore, since $C' \geq C \geq 1$,

$$\mathbb{E}\,(|u_i| \mid F(x) = a)\,\rho_{F(x)}(a) \;\leq\; C'\,\frac{1}{\max\{|w_i(x)|, |w_i(x)|^{1-\delta}\}}.$$

The assertion follows now with $|F'(x)| \leq \sum_{i=2}^{m} |w_i'(x)||u_i|$. ∎

In order to make effective use of Proposition 5.5 for certain structured weight functions having product form, we introduce the following notion, related to the total variation $\int_0^1 |q'(x)|dx$ of a function $q$.

**Definition 5.6.** The *logarithmic variation* of a function $q : [x_0, x_1] \to (0, \infty)$ is defined as

$$\mathrm{LV}(q) := \int_{x_0}^{x_1} \left|\frac{d}{dx} \ln q(x)\right| dx = \int_{x_0}^{x_1} \frac{|q'(x)|}{q(x)}\, dx.$$

The logarithmic variation has the following basic properties, whose proof is obvious.

**Lemma 5.7.** *(1) If $q$ is monotonically increasing, then* $\mathrm{LV}(q) = \ln q(x_1) - \ln q(x_0)$.
*(2)* $LV(q_1 \cdot q_2) \le \mathrm{LV}(q_1) + \mathrm{LV}(q_2)$.
*(3)* $\mathrm{LV}(q^r) = |r|\,\mathrm{LV}(q)$ *for* $r \in \mathbb{R}$.

For reasons to become clear in the next section, we assign to a finite subset $S \subseteq \mathbb{N}$ the sparse sum of squares with "support" $S$ defined as the polynomial

$$\alpha_S(x) := \sum_{s \in S} x^{2s}.$$

We will assume $0 \in S$, hence $\alpha_S(x) \ge 1$ for all $x \in \mathbb{R}$ and $\alpha_S(0) = 1$. Moreover, $\alpha_S(1) = |S|$.

Assume now we have a family of subsets $S_i \subseteq \mathbb{N}$ satisfying $0 \in S_i$ and $|S_i| \le t$, for $1 \le i \le \ell$. We choose $1 \le k \le \ell$ and define the function

$$q(x) := \left( \frac{\alpha_{S_1}(x) \cdot \ldots \cdot \alpha_{S_k}(x)}{\alpha_{S_{k+1}}(x) \cdot \ldots \cdot \alpha_{S_\ell}(x)} \right)^{\frac{1}{2}}.$$

**Proposition 5.8.** *Let $d \in \mathbb{N}$ and $0 < \delta \le 1$. The function $w\colon [0,1] \to [0,\infty)$, $x \mapsto q(x)x^d$ satisfies* $\mathrm{LV}(q) \le \frac{1}{2}\ell \ln t$. *Moreover,*

$$\int_0^1 \frac{|w'(x)|}{\max\{w(x), w(x)^{1-\delta}\}}\, dx \;\le\; 2\mathrm{LV}(q) + kt + \frac{1}{\delta}.$$

*Proof.* 1. By Lemma 5.7, we have $\mathrm{LV}(\alpha_{S_i}) \le \ln t$, since $\alpha_{S_i}$ is monotonically increasing Moreover, $\alpha_{S_i}(0) = 1$, and $\alpha_{S_i}(1) \le t$. Again using Lemma 5.7, we get $\mathrm{LV}(q) \le \frac{1}{2}\sum_{i=1}^{\ell} \mathrm{LV}(\alpha_{S_i}) \le \frac{1}{2}\ell \ln t$, showing the first assertion.
2. We will choose $\varepsilon = \varepsilon(k,t,d) \in (0,1)$ and bound

$$\int_0^1 \frac{|w'(x)|}{\max\{w(x), w(x)^{1-\delta}\}}\, dx \;\le\; \int_0^\varepsilon \frac{|w'(x)|}{w(x)^{1-\delta}}\, dx + \int_\varepsilon^1 \frac{|w'(x)|}{w(x)}\, dx.$$

For bounding the left-hand integral, we take logarithmic derivatives to get from $w(x) = q(x)x^d$

$$\frac{w'(x)}{w(x)} = \frac{q'(x)}{q(x)} + \frac{d}{x}, \tag{5.1}$$

and hence

$$\frac{w'(x)}{w(x)^{1-\delta}} = \frac{q'(x)}{q(x)} w(x)^\delta + \frac{d}{x} w(x)^\delta = \frac{q'(x)}{q(x)} q(x)^\delta x^{d\delta} + d\, q(x)^\delta x^{d\delta - 1}.$$

For $0 \le x \le \varepsilon$ we have $\alpha_{S_i}(x) \le 1 + \varepsilon^2(t-1) \le 1 + \varepsilon^2 t$, and hence $q(x) \le (1+\varepsilon^2 t)^{\frac{k}{2}}$. We can therefore bound

$$\frac{|w'(x)|}{w(x)^{1-\delta}} \;\le\; (1+\varepsilon^2 t)^{\frac{k\delta}{2}}\, \varepsilon^{d\delta} \frac{|q'(x)|}{q(x)} + \frac{1}{\delta}(1+\varepsilon^2 t)^{\frac{k\delta}{2}} \frac{d}{dx} x^{d\delta}.$$

Integrating over $[0, \varepsilon]$, we obtain

$$\begin{aligned}\int_0^\varepsilon \frac{|w'(x)|}{|w(x)|^{1-\delta}}\,dx &\le (1+\varepsilon^2 t)^{\frac{k\delta}{2}}\,\varepsilon^{d\delta}\int_0^\varepsilon \frac{|q'(x)|}{q(x)}\,dx + \frac{1}{\delta}(1+\varepsilon^2 t)^{\frac{k\delta}{2}}\varepsilon^{d\delta}\\ &\le \left((1+\varepsilon^2 t)^{\frac{k}{2}}\,\varepsilon^{d}\right)^{\delta}\left(\mathrm{LV}(q)+\frac{1}{\delta}\right).\end{aligned}$$

We now choose $\varepsilon := e^{-\frac{kt}{d}}$. Then $\varepsilon^d = e^{-kt}$ and

$$(1+\varepsilon^2 t)^{\frac{k}{2}}\varepsilon^d \le (1+t)^{\frac{k}{2}}\varepsilon^d \le e^{\frac{kt}{2}}\varepsilon^d = e^{-\frac{kt}{2}} \le 1.$$

With this choice of $\varepsilon$, we therefore have

$$\int_0^\varepsilon \frac{|w'(x)|}{|w(x)|^{1-\delta}}\,dx \le \mathrm{LV}(q)+\frac{1}{\delta}.$$

We next bound the integral over $[\varepsilon, 1]$, again using (5.1),

$$\int_\varepsilon^1 \frac{|w'(x)|}{w(x)}\,dx \le \int_\varepsilon^1 \frac{|q'(x)|}{q(x)}\,dx + d\int_\varepsilon^1 \frac{dx}{x} \le \mathrm{LV}(q) + d\ln\frac{1}{\varepsilon} = \mathrm{LV}(q)+kt,$$

where we used $d\ln\frac{1}{\varepsilon} = kt$ by our choice of $\varepsilon$. Altogether, we obtain

$$\int_0^1 \frac{|w'(x)|}{\max\{w(x), w(x)^{1-\delta}\}}\,dx \le \mathrm{LV}(q)+\frac{1}{\delta}+\mathrm{LV}(q)+kt \le 2\mathrm{LV}(q)+kt+\frac{1}{\delta}$$

completing the proof. ∎

## 6 | SUM OF PRODUCTS OF SPARSE POLYNOMIALS

Let us first fix some notation. We assign to a finite subset $S \subseteq \mathbb{Z}$ of exponents and a collection of coefficients $u_s$, for $s \in S$, the Laurent polynomial

$$f_S(x) := \sum_{s\in S} u_s x^s.$$

Note that $f_S(x^{-1}) = f_{-S}(x)$ and $f_{d+S}(x) = x^d f_S(x)$ for $d \in \mathbb{Z}$. This allows to achieve a normalization by shifting exponents: let $d$ be the minimum of $S$ and put $S' := S - d$. Then $S' \subseteq \mathbb{N}$ and $0 \in S'$. Since $f_S(x) = x^d f_{S'}(x)$, the functions $f_S$ and $f_{S'}$ have the same number of nonzero roots.

Let now $k_1, \dots, k_m$ and $t$ be positive integers and fix supports $S_{ij} \subseteq \mathbb{Z}$ for $1 \le i \le m$ and $1 \le j \le k_i$ such that $|S_{ij}| \le t$. We study the number of nonzero real roots of the sum of products $\sum_{i=1}^m f_{i1}\cdot\ldots\cdot f_{ik_i}$, where $f_{ij} := f_{S_{ij}}$.

By shifting exponents, we assume without loss of generality

$$\forall i,j \quad S_{ij} \subseteq \mathbb{N}, \quad 0 \in S_{ij} \quad \text{and} \quad |S_{ij}| \le t,$$

and consider

$$F(x) := \sum_{i=1}^{m} f_{i1}(x) \cdot \ldots \cdot f_{ik_i}(x) x^{d_i}. \tag{6.1}$$

where we allow for a degree pattern $0 = d_1 \leq d_2 \leq \cdots \leq d_m$ consisting of natural numbers $d_i$.

The probabilistic setting is as follows. For each $i, j$ and $s \in S_{ij}$ we fix a convenient probability density $\varphi_{ijs}$ on $\mathbb{R}$ and assume that there are constants $A, B$ such that

$$\forall i, j, s \quad \|\varphi_{ijs}\|_\infty \leq A, \quad \mathbb{E}_{\varphi_{ijs}} \leq B.$$

We suppose that we have random univariate polynomials

$$f_{ij}(x) = \sum_{s \in S_{ij}} u_{ijs} x^s \tag{6.2}$$

with independent real coefficients $u_{ijs}$ having the convenient density $\varphi_{ijs}$. The goal is to study the expected number of real zeros of the resulting random polynomial $F$.

We assign to the support $S_{ij}$ the following generating functions

$$\alpha_{ij}(x) := \sum_{s \in S_{ij}} x^{2s}, \quad \beta_{ij}(x) := \sum_{s \in S_{ij}} x^s. \tag{6.3}$$

Note that $\mathbb{E}\,(f_{ij}(x)^2) = \alpha_{ij}(x)$ if $\mathbb{E}\,(u_{ijs}^2) = 1$, since $\mathbb{E}\,(u_{ijs}) = 0$.

The next lemma makes sure we can apply Theorem 3.2 in the above setting.

**Lemma 6.1.** *Let $N := \sum_{i=1}^{m} \sum_{j=1}^{k_i} |S_{ij}|$ denote the number of parameters. For $x \in \mathbb{R}$ consider the polynomial map $F(x) : \mathbb{R}^N \to \mathbb{R}$ sending a system $u = (u_{ijs}) \in \mathbb{R}^N$ of coefficients to $F(x)$, as defined in (6.1). For all $x \in \mathbb{R}$ we have:*

*(a) $F(x)$ is surjective and thus nonconstant.*
*(b) All nonzero $a \in \mathbb{R}$ are regular values of $F(x)$.*
*(c) 0 is a singular value of $F(x)$ unless $k_1 = \cdots = k_m = 1$.*

*The conditional density $\rho_{F(x)}(a)$ is defined at every nonzero $a \in \mathbb{R}$. However, it is undefined at $a = 0$, unless $k_1 = \cdots = k_m = 1$.*

*Proof.* We fix $x \in \mathbb{R}$.

(a) After specializing $u_{ijs} := 0$ for $s \neq 0$, $F(x)$ becomes the function mapping $(u_{ij0})$ to $\sum_{i=1}^{m} u_{i10} \cdot \ldots \cdot u_{ik_i0}$, which clearly is a surjective function.
(b) The $f_{ij}(x)$ are linear functions in disjoints sets of variables and all have a nonzero coefficient. Therefore, their gradients, viewed as vectors in $\mathbb{R}^N$, are linearly independent. Suppose now $u = (u_{ijs}) \in \mathbb{R}^N$ is a singular point of $F(x)$. We have (dropping the argument $u$)

$$\nabla F(x) = \sum_{i=1}^{m} f_{i,1} \cdot \ldots \cdot f_{i,j-1} \nabla f_{i,j} f_{i,j+1} \cdot \ldots \cdot f_{i,k_i}.$$

Since the $\nabla f_{i,j}$ are linearly independent, we must have $f_{i,1} \cdot \ldots \cdot f_{i,j-1} \nabla f_{i,j} f_{i,j+1} \cdot \ldots \cdot f_{i,k_i} = 0$ for all $i, j$. This means that for all $i$ there are different $j$ and $j'$ such that $f_{ij}(x) = 0$ and $f_{ij'}(x) = 0$. In particular, we have

$F(x)(u) = 0$ for such $u$ and hence 0 is the only possible singular value of $F(x)$. If $k_i > 1$ for some $i$, then $u = 0$ is a singular point of $F(x)$ and thus 0 is a singular value.
(c) This follows from the reasoning in (b). ∎

For applying Theorem 3.2, the main work consists now in exhibiting a "small" integrable function $g(x)$ that upper bounds the conditional expectations. We embark on this next.

### 6.1 | Products of sparse polynomials

We analyze here the case $m = 1$ of one product

$$g(x) := f_1(x) \cdot \ldots \cdot f_k(x)$$

of random $t$-sparse polynomials $f_j(x) = \sum_{s \in S_j} u_{js} x^s$, where for convenience, we drop the index $i = 1$. In particular, we write $\beta_j(x) := \sum_{s \in S_j} x^s$. So we assume $0 \in S_j$ and $|S_j| \leq t$ for all $j$.

By Lemma 6.1, every nonzero $a \in \mathbb{R}$ is a regular value of the map $g(x) \colon \mathbb{R}^N \to \mathbb{R}$, thus the conditional density $\rho_{g(x)}(a)$ is well defined and so are the conditional expectations with respect to the condition $g(x) = a$, provided $\rho_{g(x)}(a) > 0$.

**Lemma 6.2.** *For all $x \in \mathbb{R}$ and all nonzero $a \in \mathbb{R}$ we have*

$$\mathbb{E}\left(\left|\frac{f_j'(x)}{f_j(x)}\right| \mid g(x) = a\right) \rho_{g(x)}(a) \;\leq\; AB\,\frac{\beta_j'(x)}{|a|},$$

$$\mathbb{E}\left(|g'(x)| \mid g(x) = a\right) \rho_{g(x)}(a) \;\leq\; AB \sum_{j=1}^{k} \beta_j'(x).$$

*Proof.* Fix $x \in \mathbb{R}$ and consider the random variables $y_j := f_j(x)$ and $z_j := f_j'(x)$. If $\psi_j(y_j, z_j)$ denotes the joint density of $(y_j, z_j)$, then by the independence of $(y_1, z_1), \ldots, (y_k, z_k)$, the probability density of $(y, z) \in \mathbb{R}^k \times \mathbb{R}^k$ is given by $\psi_1(y_1, z_1) \cdot \ldots \cdot \psi_k(y_k, z_k)$. Note that $g(x) = y_1 \cdot \ldots \cdot y_k$.

We are going to apply some insights from Section 2. Namely, we apply Equation (2.3) to the function $f \colon \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}$, $(y, z) \mapsto y_1 \cdot \ldots \cdot y_k$ and the random variable $Z(y, z) := |\frac{z_1}{y_1}|$. For nonzero $a \in \mathbb{R}$ we consider the hypersurface $C_a := \{y \in \mathbb{R}^k : y_1 \cdots y_k = a\}$ and note that $\|\nabla(y_1 \cdot \ldots \cdot y_k)\| = |a|(\sum_{i=1}^k y_i^{-2})^{\frac{1}{2}}$. We obtain

$$\begin{aligned}
&\mathbb{E}\left(\left|\frac{z_1}{y_1}\right| \;\middle|\; y_1 \cdot \ldots \cdot y_k = a\right) \rho_{g(x)}(a) \\
&\quad = \int_{C_a \times \mathbb{R}^k} \left|\frac{z_1}{y_1}\right| \psi_1(y_1, z_1) \cdot \ldots \cdot \psi_k(y_k, z_k)\, \frac{d(C_a \times \mathbb{R}^k)}{|a| \left(\sum_{i=1}^k y_i^{-2}\right)^{\frac{1}{2}}} \\
&\quad = \int_{y \in C_a} \left[\int_{z \in \mathbb{R}^k} \left|\frac{z_1}{y_1}\right| \psi_1(y_1, z_1) \cdot \ldots \cdot \psi_k(y_k, z_k)\, dz_1 \cdots dz_k\right] \frac{dC_a}{|a| \left(\sum_{i=1}^k y_i^{-2}\right)^{\frac{1}{2}}}.
\end{aligned} \tag{6.4}$$

For fixed $y \in C_a$, the inner integral can be simplified to

$$\int_{z_1\in\mathbb{R}} \left|\frac{z_1}{y_1}\right| \psi_1(z_1,y_1) \left[\int_{(z_2,\dots,z_k)\in\mathbb{R}^{k-1}} \psi_2(y_2,z_2)\cdot\ldots\cdot\psi_k(y_k,z_k)\,dz_2\cdots dz_k\right] dz_1$$
$$= \int_{z_1\in\mathbb{R}} \left|\frac{z_1}{y_1}\right| \psi_1(y_1,z_1)\,dz_1 \cdot \psi_2(y_2)\cdot\ldots\cdot\psi_k(y_k),$$

with the marginal densities $\psi_i$ defined by $\psi_i(y_i) := \int_{\mathbb{R}} \psi_i(y_i,z_i)dz_i$. By (2.6) we have for $y_1 \in \mathbb{R}^*$,

$$\int_{z_1\in\mathbb{R}} \left|\frac{z_1}{y_1}\right| \psi_1(y_1,z_1)\,dz_1 = \mathbb{E}\left(\left|\frac{z_1}{y_1}\right| \;\middle|\; y_1\right)\psi_1(y_1).$$

We thus obtain from (6.4)

$$\mathbb{E}\left(\left|\frac{z_1}{y_1}\right| \;\middle|\; y_1\cdot\ldots\cdot y_k = a\right)\rho_{g(x)}(a)$$
$$= \int_{y\in C_a} \mathbb{E}\left(\left|\frac{z_1}{y_1}\right| \;\middle|\; y_1\right)\psi_1(y_1)\,\psi_2(y_2)\cdot\ldots\cdot\psi_k(y_k)\,\frac{dC_a}{|a|\left(\sum_{i=1}^k y_i^{-2}\right)^{\frac{1}{2}}}. \tag{6.5}$$

Proposition 5.2 applied to the random linear combination $f_j(x) = \sum_{s\in S_j} u_{js}x^s$ implies

$$\mathbb{E}\left(|z_1| \mid y_1\right)\psi_1(y_1) \;\le\; AB\,\beta_1'(x).$$

Here we essentially use that, due to the assumption $0 \in S_j$, the polynomial $f_j(x) = u_{j0} + \ldots$ has a constant term. Using this bound, we get from (6.5),

$$\mathbb{E}\left(\left|\frac{z_1}{y_1}\right| \;\middle|\; y_1\cdot\ldots\cdot y_k = a\right)\rho_{g(x)}(a) \le\; AB\cdot\beta_1'(x)\int_{y\in C_a}\frac{1}{|y_1|}\psi_2(y_2)\cdot\ldots\cdot\psi_k(y_k)\,\frac{dC_a}{|a|\left(\sum_{i=1}^k y_i^{-2}\right)^{\frac{1}{2}}}.$$

Using (2.8), the integral over $C_a$ simplifies to

$$\int_{(y_2,\dots,y_k)\in\mathbb{R}^{k-1}} \frac{|y_2\cdot\ldots\cdot y_k|}{|a|}\cdot\psi_2(y_2)\cdot\ldots\cdot\psi_k(y_k)\frac{dy_2\cdots dy_k}{|y_2|\cdots|y_k|}$$
$$= \frac{1}{|a|}\int_{\mathbb{R}}\psi_2(y_2)dy_2\cdot\ldots\cdot\int_{\mathbb{R}}\psi_k(y_k)dy_k = \frac{1}{|a|}.$$

Therefore, indeed

$$\mathbb{E}\left(\left|\frac{f_1'(x)}{f_1(x)}\right| \;\middle|\; g(x) = a\right)\rho_{g(x)}(a) \;\le\; AB\,\frac{\beta_1'(x)}{|a|}.$$

The same argument works with $f_j$ instead of $f_1$, so that we have proved the first statement.

In order to show the second statement, taking logarithmic derivatives, we get

$$\frac{g'(x)}{g(x)} = \sum_{j=1}^k \frac{f_j'(x)}{f_j(x)},$$

hence

$$\left|\frac{g'(x)}{g(x)}\right| \;\le\; \sum_{j=1}^{k}\left|\frac{f_j'(x)}{f_j(x)}\right|.$$

Therefore,

$$\mathbb{E}\left(\left|\frac{g'(x)}{g(x)}\right| \;\middle|\; g(x)=a\right) \;\le\; \sum_{j=1}^{k}\mathbb{E}\left(\left|\frac{f_j'(x)}{f_j(x)}\right| \middle| g(x)=a\right)$$

hence

$$\mathbb{E}\left(|g'(x)| \mid g(x)=a\right)\rho_{g(x)}(a) \;\le\; \sum_{j=1}^{k}|a|\,\mathbb{E}\left(\left|\frac{f_j'(x)}{f_j(x)}\right| \;\middle|\; g(x)=a\right)\rho_{g(x)}(a).$$

Inserting here the bound of the first statement yields the second statement. ∎

### 6.2 | Polynomials with nonzero constant coefficient

We deal here with the special case $d_1 = \cdots = d_m = 0$. So we are in the situation where all the $f_{ij}$ almost surely have a nonzero constant coefficient. It turns out that this situation is way easier to analyze than the general case.

The next result shows that the real $\tau$-conjecture is true on average under the assumption $d_1 = \cdots = d_m = 0$, if we only count zeros in $[0, 1]$. It is worthwhile noting that this results holds for any convenient distribution of the coefficients $u_{ijs}$, as long as they are independent.

**Theorem 6.3.** *Under the assumptions from the beginning of Section 6, the random polynomial* $F = \sum_{i=1}^{m} f_{i1}\cdot\ldots\cdot f_{ik_i}$ *satisfies*

$$\mathbb{E}\,\#\{x\in[0,1] : F(x)=0\} \;\le\; AB\,(k_1+\cdots+k_m)(t-1).$$

*Proof.* Lemma 6.1 guarantees that $(u,x)\mapsto g(x)(u)$ satisfies the assumptions of Theorem 3.2.

We are going to show that forall $x\in\mathbb{R}$ and all nonzero $a\in\mathbb{R}$,

$$\mathbb{E}\,(|F'(x)| \mid F(x)=a)\rho_{F(x)}(a) \;\le\; AB\sum_{i=1}^{m}\sum_{j=1}^{k_i}\beta_{ij}'(x), \tag{6.6}$$

where we recall that $\beta_{ij}(x)$ was defined in (6.3). Then, taking into account Lemma 6.1 and $\int_0^1 \beta_{ij}'(x)dx = \beta_{ij}(1)-\beta_{ij}(0)\le t-1$, the assertion will follow by Theorem 3.2.

Towards proving (6.6), we put $g_i(x) := f_{i1}(x)\cdot\ldots\cdot f_{ik_i}(x)$. Then we have $F(x) = g_1(x)+\cdots+g_m(x)$ and hence $|F'(x)|\le\sum_{i=1}^{m}|g_i'(x)|$. Therefore,

$$\mathbb{E}\,(|F'(x)| \mid F(x)=a)\rho_{F(x)}(a) \;\le\; \sum_{i=1}^{m}\mathbb{E}\,(|g_i'(x)| \mid F(x)=a)\rho_{F(x)}(a).$$

Lemma 6.2 gives for nonzero $b\in\mathbb{R}$ that

$$\mathbb{E}\,(|g_i'(x)| \mid g_i(x)=b)\rho_{g_i(x)}(b) \;\le\; AB\sum_{j=1}^{k_i}\beta_{ij}'(x). \tag{6.7}$$

For proving (6.6), it suffices to show that the same bound holds when conditioning on $F(x) = a$, namely

$$\mathbb{E}\left(|g_i'(x)| \mid F(x) = a\right)\rho_{F(x)}(a) \;\le\; AB \sum_{j=1}^{k_i} \beta_{ij}'(x). \tag{6.8}$$

For showing this, we fix $1 \le i \le m$. We put $y_i := g_i(x)$ and $z_i := g_i'(x)$, and denote by $\psi_i(y_i, z_i)$ the joint density of $(y_i, z_i)$. Moreover, we write $\psi_i(y_i) := \int_{\mathbb{R}} \psi_i(y_i, z_i)\, dz_i$ for the first marginal distribution. By construction, the pairs $(y_1, z_1), \dots, (y_m, z_m)$ are independent. We claim that

$$\mathbb{E}\left(|z_1| \mid y_1 + \cdots + y_m = a\right) \rho_{y_1+\cdots+y_m}(a) = \int_{\mathbb{R}} \mathbb{E}\left(|z_1| \mid y_1 = b\right) \psi_1(b)\, \rho_{y_2+\cdots+y_m}(a-b)\, db. \tag{6.9}$$

It remains to prove this claim, since it implies, combined with (6.7), that

$$\begin{aligned}\mathbb{E}\left(|g_1'(x)| \,\Big|\, \sum_{j=1}^{m} g_j(x) = a\right)\rho_{F(x)}(a) &= \int_{b\in\mathbb{R}} \mathbb{E}\left(|g_1'(x)| \,\big|\, g_1(x) = b\right) \rho_{g_1(x)}(b)\, \rho_{\sum_{j\neq 1} g_j(x)}(a-b)\, db \\ &\le\; AB \sum_{j=1}^{k_i} \beta_{1j}'(x) \int_{b\in\mathbb{R}} \rho_{\sum_{j\neq i} g_j(x)}(a-b)\, db = AB \sum_{j=1}^{k_i} \beta_{1j}'(x),\end{aligned}$$

which is (6.8) (for w.l.o.g. $i = 1$).

We deduce now the claim (6.9). By (2.5) we have

$$\begin{aligned}&\mathbb{E}\left(|z_1| \mid y_1 + \cdots + y_m = a\right) \rho_{y_1+\cdots+y_m}(a) \\ &\quad= \int_{\mathbb{R}^{m-1}} \int_{\mathbb{R}^m} |z_1| \psi_1(a - y_2 - \cdots - y_m, z_1) \psi_2(y_2, z_2) \cdots \psi_m(y_m, z_m)\, dz_1 \cdots dz_m\, dy_2 \cdots dy_m \\ &\quad= \int_{\mathbb{R}^{m-1}} \left[\int_{\mathbb{R}} |z_1| \psi_1(a - y_2 - \cdots - y_m, z_1)\, dz_1\right] \psi_2(y_2) \cdots \psi_m(y_m)\, dy_2 \cdots dy_m.\end{aligned} \tag{6.10}$$

For fixed $y_2, \dots, y_m$ and $b = a - y_2 - \cdots - y_m$, the expression in parenthesis equals

$$\mathbb{E}\left(|z_1| \mid y_1 = b\right) \psi_1(b).$$

By applying Proposition 2.1 to the map $T : \mathbb{R}^{m-1} \to \mathbb{R}, (y_2, \dots, y_m) \mapsto a - y_2 - \cdots - y_m$, taking into account Lemma 2.2, we can express the above integral (6.10) as

$$\int_{\mathbb{R}} \mathbb{E}\left(|z_1| \mid y_1 = b\right) \psi_1(b) \left[\int_{T^{-1}(b)} \psi_2(y_2) \cdots \psi_m(y_m)\, \frac{dT^{-1}(b)}{\sqrt{m-1}}\right] db.$$

By definition, the expression in the parenthesis equals the pushforward density $\rho_{y_2+\cdots+y_m}(a-b)$, which shows the claim (6.9) and finishes the proof. ∎

### 6.3 | Proof of main result

We specialize the setting described at the beginning of Section 6 to the case where all the coefficients $u_{ijs}$ are standard Gaussian.

For $1 \le i \le m$ we define the auxiliary analytic weight functions

$$q_i(x) := \prod_{j=1}^{k_i} \left( \frac{\alpha_{ij}(x)}{\alpha_{1j}(x)} \right)^{\frac{1}{2}}, \tag{6.11}$$

and recall that $\alpha_{ij}(x)$ was defined in (6.3). Note that $q_i(x) > 0$ for all $x \in \mathbb{R}$ and $q_1(x) = 1$. We define the analytic weight function $w_i(x) := q_i(x)x^{d_i}$ for $1 \le i \le m$ and note that $w_1(x) = 1$.

We will reduce the problem of counting the expected number of zeros of the structured random polynomial $F(x)$ to the study of the expected number of zeros of random linear combinations

$$R(x) := \sum_{i=1}^{m} u_i q_i(x) x^{d_i} = u_1 + u_2 q_2(x) x^{d_2} + \cdots + u_m q_m(x) x^{d_m},$$

of the weight functions $w_i(x)$, where the coefficients $u_i$ are independent and follow the distribution $\varpi_{k_i}$ of a product of $k_i$ standard Gaussians (cf. Section 2.2). We note that, due to Lemma 5.1, for almost all $u \in \mathbb{R}^m$, the function $R$ has only finitely many zeros in $[x_0, x_1]$. Thus it satisfies the assumptions stated at the beginning of Section 5.

**Proposition 6.4.** *For $x \in \mathbb{R}$ and nonzero $a \in \mathbb{R}$, we have*

$$\begin{aligned} \mathbb{E}\left(|F'(x)| \mid F(x) = a\right) \rho_{F(x)}(a) \le\ & \frac{1}{\sqrt{2\pi}} \sum_{i=1}^{m} \sum_{j=1}^{k_i} \beta'_{ij}(x) + \sum_{i=1}^{m} \left| \frac{q_i'(x)}{q_i(x)} \right| \\ & + \mathbb{E}\left(|R'(x)| \mid R(x) = a\right) \rho_{R(x)}(a). \end{aligned}$$

*Proof.* We write $F(x) = h_1(x) + \cdots + h_m(x)$, where

$$h_i(x) := g_i(x) x^{d_i} \quad \text{and} \quad g_i(x) := f_{i1}(x) \cdot \ldots \cdot f_{ik_i}(x).$$

Note that $h_i'(x) = g_i'(x)x^{d_i} + g_i(x)d_i x^{d_i - 1}$. We bound with the triangle inequality:

$$|F'(x)| \le \sum_{i=1}^{m} \left| g_i'(x) x^{d_i} \right| + \left| \sum_{i=1}^{m} g_i(x) d_i x^{d_i - 1} \right|.$$

Here, it is essential not to upper bound further the right-hand contribution by $\sum_{i=1}^{m} |g_i(x)d_i x^{d_i-1}|$. Continuing, we get

$$\begin{aligned} \mathbb{E}\left(|F'(x)| \mid F(x) = a\right) \rho_{F(x)}(a) \le\ & \sum_{i=1}^{m} \mathbb{E}\left(|g_i'(x)x^{d_i}| \mid F(x) = a\right) \rho_{F(x)}(a) \\ & + \mathbb{E}\left( \left| \sum_{i=2}^{m} g_i(x) d_i x^{d_i - 1} \right| \mid F(x) = a \right) \rho_{F(x)}(a). \end{aligned} \tag{6.12}$$

By the same reasoning as for (6.9), we have for nonzero $a \in \mathbb{R}$

$$\mathbb{E}\left(|g_i'(x)x^{d_i}| \mid F(x) = a\right) \rho_{F(x)}(a) = \int_{\mathbb{R}} \mathbb{E}\left(|g_i'(x)x^{d_i}| \mid h_i(x) = b\right) \rho_{h_i(x)}(b) \rho_{H_i(x)}(a - b)\, db,$$

where $H_i(x) := \sum_{j\neq i} h_j(x)$. Moreover, setting $\tilde{b} := \frac{b}{x^{d_i}}$, we get for nonzero $b \in \mathbb{R}$

$$\mathbb{E}\left(|g_i'(x)x^{d_i}| \mid h_i(x) = b\right)\rho_{h_i(x)}(b) = \mathbb{E}\left(|g_i'(x)x^{d_i}| \mid g_i(x) = \tilde{b}\right)\frac{1}{x^{d_i}}\rho_{g_i(x)}(\tilde{b}).$$

The $x^{d_i}$ cancels and by Lemma 6.2, we have with $AB = \frac{1}{\sqrt{2\pi}}$,

$$\mathbb{E}\left(|g_i'(x)x^{d_i}| \mid h_i(x) = b\right)\rho_{h_i(x)}(b) \leq \frac{1}{\sqrt{2\pi}}\sum_{j=1}^{k_i}\beta_{ij}'(x).$$

Note that the right hand-side does not depend on $b$. Multiplying with $\rho_{H_i(x)}(a-b)$, integrating over $b$ (which does not change anything) and summing over $i$, yields the first contribution in the theorem's upper bound.

It remains to bound the right-hand contribution in (6.12), for fixed $x \in \mathbb{R}$ and nonzero $a \in \mathbb{R}$. For this, note that $f_{ij}(x)$ is a centered Gaussian random variable having the variance $\alpha_{ij}(x)$ (recall (6.2) and (6.3)). So we may write $f_{ij}(x) = \alpha_{ij}(x)^{\frac{1}{2}}v_{ij}$ with independent standard Gaussian random variables $v_{ij}$. Hence, if we abbreviate $u_i := v_{i1}\cdots v_{ik_i}$ and put $p_i(x) := (\alpha_{i1}(x)\cdot\ldots\cdot\alpha_{ik_i}(x))^{-\frac{1}{2}}$, then

$$g_i(x) = f_{i1}(x)\cdot\ldots\cdot f_{ik_i}(x) = \alpha_{i1}(x)^{\frac{1}{2}}\cdot\ldots\cdot\alpha_{ik_i}(x)^{\frac{1}{2}}v_{i1}\cdot\ldots\cdot v_{ik_i} = p_i(x)^{-1}u_i.$$

By its definition, the random variable $u_i$ has the distribution $\varpi_{k_i}$ (cf. Section 2.2). It is a convenient distribution (cf. Definition 4.2). We also note that $q_i(x) = \frac{p_1(x)}{p_i(x)}$ by (6.11). With these notations, we can write

$$F(x) = \sum_{i=1}^m g_i(x)x^{d_i} = \sum_{i=1}^m \frac{u_i}{p_i(x)}x^{d_i} = \frac{1}{p_1(x)}\sum_{i=1}^m u_iq_i(x)x^{d_i} = \frac{1}{p_1(x)}R(x).$$

Hence $\rho_{F(x)}(a) = p_1(x)\rho_{R(x)}(\zeta)$, where $\zeta := p_1(x)a$. We analyze now the right-hand contribution in (6.12):

$$\mathbb{E}\left(\Big|\sum_{i=2}^m \frac{u_i}{p_i(x)}d_ix^{d_i-1}\Big| \mid F(x) = a\right)\rho_{F(x)}(a) = \mathbb{E}\left(\Big|\sum_{i=2}^m u_iq_i(x)d_ix^{d_i-1}\Big| \mid F(x) = a\right)\rho_{R(x)}(\zeta).$$

Using

$$R'(x) = \sum_{i=1}^m u_iq_i'(x)x^{d_i} + \sum_{i=1}^m u_iq_i(x)d_ix^{d_i-1},$$

we can bound

$$\Big|\sum_{i=2}^m u_iq_i(x)d_ix^{d_i-1}\Big| \leq \sum_{i=1}^m \Big|u_iq_i'(x)x^{d_i}\Big| + |R'(x)|.$$

Therefore,

$$\begin{aligned}\mathbb{E}\left(\Big|\sum_{i=2}^m u_iq_i(x)d_ix^{d_i-1}\Big| \mid R(x) = \zeta\right)\rho_{R(x)}(\zeta) \leq \sum_{i=1}^m \mathbb{E}\left(\Big|u_iq_i'(x)x^{d_i}\Big| \mid R(x) = \zeta\right)\rho_{R(x)}(\zeta)\\ + \mathbb{E}\left(|R'(x)| \mid R(x) = \zeta\right)\rho_{R(x)}(\zeta).\end{aligned}$$

Note that the right-hand contribution equals $\mathbb{E}\left(|R'(x)| \mid R(x)=a\right)\rho_{R(x)}(a)$ as desired. In order to bound the left-hand sum, we can apply Proposition 4.4 since the densities of the $u_i$ are convenient, and we thus obtain

$$|q_i(x)x^{d_i}| \cdot \mathbb{E}\left(|u_i| \mid R(x)=\zeta\right)\rho_{R(x)}(\zeta) \;\leq\; 1.$$

This yields

$$|q_i'(x)x^{d_i}| \cdot \mathbb{E}\left(\left|u_i\right| \mid R(x)=\zeta\right)\rho_{R(x)}(\zeta) \;\leq\; \frac{|q_i'(x)|}{q_i(x)}.$$

Summarizing, we have shown that

$$\mathbb{E}\left(\Big|\sum_{i=2}^{m} g_i(x)d_i x^{d_i-1}\Big| \mid F(x)=a\right)\rho_{F(x)}(a) \;\leq\; \sum_{i=1}^{m}\frac{|q_i'(x)|}{q_i(x)} + \mathbb{E}\left(|R'(x)| \mid R(x)=\zeta\right)\rho_{R(x)}(\zeta),$$

which completes the proof. ∎

We can finally provide the proof of the main result.

*Proof of Theorem 1.1.* The right-hand term in the statement of Proposition 6.4 can be bounded with Proposition 5.5. Indeed, due to Lemma 2.5 we know that $\varpi_{k_1}(a) \;\leq\; e\;|a|^{\frac{1}{2k_1}-1}$ for all $a$. Applying Proposition 5.5 with the parameters $B=1$, $C=e$, and $\delta=(2k_1)^{-1}$ yields

$$\mathbb{E}\left(|R'(x)| \mid R(x)=a\right)\rho_{R(x)}(a) \;\leq\; e(2k_1+1)\sum_{i=2}^{m}\frac{|w_i'(x)|}{\max\{|w_i(x)|,|w_i(x)|^{1-\frac{1}{2k_1}}\}}.$$

Applying Proposition 6.4 implies for $x\in\mathbb{R}$ and $a\in\mathbb{R}^*$, recalling that $w_i(x) := q_i(x)x^{d_i}$,

$$\begin{aligned}\mathbb{E}\left(|F'(x)| \mid F(x)=a\right)\rho_{F(x)}(a) \;&\leq\; \frac{1}{\sqrt{2\pi}}\sum_{i=1}^{m}\sum_{j=1}^{k_i}\beta_{ij}'(x) + \sum_{i=1}^{m}\left|\frac{q_i'(x)}{q_i(x)}\right| \\ &+\; e(2k_1+1)\sum_{i=2}^{m}\frac{|w_i'(x)|}{\max\{|w_i(x)|,|w_i(x)|^{1-\frac{1}{2k_1}}\}} =: g(x). \qquad (6.13)\end{aligned}$$

$$\int_0^1 \frac{|w_i'(x)|}{\max\{w_i(x),w_i(x)^{1-\frac{1}{2k_1}}\}}\,dx \;\leq\; 2\mathrm{LV}(q_i)+k_i t+2k_1.$$

The function $g(x)$ on the right-hand side of (6.13) is integrable:

$$\int_0^1 g(x)\,dx \;\leq\; \frac{1}{\sqrt{2\pi}}\sum_{i=1}^{m}\sum_{j=1}^{k_i}(t-1)+\sum_{i=1}^{m}\mathrm{LV}(q_i)+e(2k_1+1)\sum_{i=2}^{m}(2\mathrm{LV}(q_i)+k_i t+2k_1)<\infty.$$

By Proposition 5.8 we can bound $\mathrm{LV}(q_i) \;\leq\; \frac{1}{2}2k_i\ln t$. Moreover, Theorem 3.2 can be applied (see Lemma 6.1) and states that $\mathbb{E}\left(\#\{x\in[0,1] : F(x)=0\}\right) \leq \int_0^1 g(x)\,dx$. Hence,

$$\begin{aligned}\mathbb{E}\left(\#\{x\in[0,1] : F(x)=0\}\right) \;&\leq\; \frac{1}{\sqrt{2\pi}}(k_1+\cdots+k_m)(t-1)+(k_1+\cdots+k_m)\ln t \\ &+\; e(2k_1+1)\left((k_2+\cdots+k_m)(2\ln t+t)+(m-1)2k_1\right) \\ &= \mathcal{O}(k^2 m t), \qquad (6.14)\end{aligned}$$

where $k$ denotes the maximum of the $k_i$.

The number of zeros of $F$ in $[1,\infty)$ equals the number of zeros $x \in (0,1]$ of $F(x^{-1})$. Moreover, $F(x^{-1})$ has the same structure as $F$ except that the supports $S_{ij}$ are replaced by $-S_{ij}$. Since we can shift the degrees without changing the number of positive zeros, we conclude that $\mathbb{E}(\#\{x \in [1,\infty) : F(x) = 0\})$ is also bounded by (6.14). Therefore, $\mathbb{E}(\#\{x \in \mathbb{R} : F(x) = 0\})$ is upper bounded by four times (6.14). ■

## ACKNOWLEDGMENTS

We thank Pascal Koiran and Mario Kummer for helpful discussions. Peter Bürgisser is grateful to the late Mario Wschebor for introducing him into the wonders of the Rice formula. We thank the anonymous referees whose comments led to an improved presentation.

## REFERENCES

1. R.J. Adler and J.E. Taylor, *Random fields and geometry*. Springer Monographs in Mathematics, Springer, New York, 2007.
2. J.-M. Azaïs and M. Wschebor, *Level sets and extrema of random processes and fields*, John Wiley & Sons, Inc., Hoboken, NJ, 2009.
3. L. Blum, F. Cucker, M. Shub, and S. Smale, *Complexity and real computation*, Springer-Verlag, New York, 1998. With a foreword by Richard M. Karp.
4. P. Bürgisser, *On defining integers and proving arithmetic circuit lower bounds*, Comput. Complexity. **18** (2009), 81-103.
5. P. Bürgisser, A. Erguer, and J. Tonelli-Cueto, *On the number of real zeros of random fewnomials*, SIAM J. Appl. Algebra Geom. **3** (2019), 721-732.
6. I. Chavel, *Riemannian geometry*. Cambridge University Press.
7. H. Federer, *Curvature measures*, Trans. Amer. Math. Soc. **93** (1959), 418-491.
8. R. Howard, *The kinematic formula in Riemannian homogeneous spaces*, Mem. Amer. Math. Soc. **106** (1993), vi+69.
9. P. Hrubes, *On the real $\tau$-conjecture and the distribution of complex roots*, Theory Comput. **9** (2013), 403-411.
10. P. Hrubes, *On the distribution of runners on a circle*, 2019, arXiv:1906.02511.
11. M. Kac, *On the average number of real roots of a random algebraic equation*, Bull. Amer. Math. Soc. **49** (1943), 314-320.
12. P. Koiran, *Shallow circuits with high-powered inputs*. Proceedings of the Second Symposium on Innovations in Computer Science, pp. 309-320, 2011.
13. P. Koiran, N. Portier, and S. Tavenas, *A Wronskian approach to the real $\tau$-conjecture*, J. Symb. Comput. **68** (2015), 195-214.
14. P. Koiran, N. Portier, S. Tavenas, and S. Thomassé, *A $\tau$-conjecture for Newton polygons*, Found. Comput. Math. **15** (2015), 185-197.
15. Michael Shub and Steve Smale, *On the intractability of Hilbert's Nullstellensatz and an algebraic version of "*NP $\neq$ P*?"*, Duke Math. J.. **81** (1995), 47-54. A celebration of John F. Nash, Jr..
16. M.D. Springer and W.E. Thompson, *The distribution of products of beta, gamma and Gaussian random variables*, SIAM J. Appl. Math. **18** (1970), 721-737.
17. Sébastien Tavenas, *Bornes inferieures et superieures dans les circuits arithmetiques*, PhD thesis, Ecole normale supérieure de Lyon, 2014.

**How to cite this article:** Briquel I, Bürgisser P. The real tau-conjecture is true on average. *Random Struct Alg*. 2020;57:279–303. https://doi.org/10.1002/rsa.20926