THE AVERAGE NUMBER OF CRITICAL RANK-ONE APPROXIMATIONS TO A TENSOR

JAN DRAISMA AND EMIL HOROBEŢ

Abstract. Motivated by the many potential applications of low-rank multi-way tensor approximations, we set out to count the rank-one tensors that are critical points of the distance function to a general tensor v. As this count depends on v, we average over v drawn from a Gaussian distribution, and find a formula that relates this average to problems in random matrix theory.

1. Introduction

Low-rank approximation of *matrices* via singular value decomposition is among the most important algebraic tools for solving approximation problems in data compression, signal processing, computer vision, etc. Low-rank approximation for *tensors* has the same application potential, but raises substantial mathematical and computational challenges. To begin with, tensor rank and many related problems are NP-hard [Hås90, HL13], although in low degrees (symmetric) tensor decomposition has been approached computationally in [BCMT10, OO13] by greatly generalising classical techniques due to Sylvester and contemporaries. Furthermore, tensors of bounded rank do not form a closed subset, so that a best low-rank approximation of a tensor on the boundary does not exist [dSL08]. This latter problem does not occur for tensors of rank at most one, which *do* form a closed set, and where the best rank-one approximation does exist under a suitable genericity assumption [FO12].

In spite of these mathematical difficulties, much application-oriented research revolves around algorithms for computing low-rank approximations [BW08, BW09, CGLM08, DL08a, DL08b, DLN08, Lim05, IAvHdL11]. Typically, these algorithms are of a local nature and would get into problems near non-minimal critical points of the distance function to be minimised. This motivates our study into the question of *how many* critical points one should expect in the easiest nontrivial setting, namely that of approximation by rank-one tensors. This number should be thought of as a measure of the complexity of finding the closest rank-one approximation. The corresponding complex count, which is the topic of [FO12] and with which we will compare our results, measures the degree of an algebraic field extension needed to write down the critical points as algebraic functions of the tensor to be minimised. We will treat both ordinary tensors and symmetric tensors.

Ordinary tensors. To formulate our problem and results, let $n_1, \ldots, n_p$ be natural numbers and let $X \subset V := \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_p}$ be the variety of rank-one p-way tensors, i.e., those that can be expressed as $x_1 \otimes x_2 \otimes \cdots \otimes x_p$ for vectors $x_i \in \mathbb{R}^{n_i}$, $i = 1, \ldots, p$.

JD is supported by a Vidi grant from the Netherlands Organisation for Scientific Research (NWO).

EH is supported by an NWO free competition grant.

Given a general tensor $v \in V := \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_p}$, one would like to compute $x \in X$ that minimizes the squared Euclidean distance

$$d_v(x) = \sum_{i_1,\ldots,i_p} (v_{i_1,\ldots,i_p} - x_{i_1,\ldots,i_p})^2$$

from v. For the matrix case, where $p = 2$, this minimizer is $\sigma x_1 x_2^T$ where σ is the largest singular value of v and x_1, x_2 are the corresponding left and right singular vectors. Indeed, all critical points of d_v are of this form, with σ running through all singular values of v. For $p > 2$, several algorithms have been proposed for rank-one approximation (see, e.g., [BWG07, DDV00]). These algorithms have a local nature and experience difficulties near critical points of d_v. This is one of our motivations for *counting* these critical points—the main goal of this paper.

In [FO12], a general formula is found for the number of *complex* critical points of d_v on $X_{\mathbb{C}}$. In this case the x_i can have complex coefficients and the expression d_v is copied verbatim, i.e., without inserting complex conjugates. This means that $d_v(x)$ does not really measure a (squared) distance—e.g., it can be zero even for $x \neq v$—but on the positive side the number of critical points of d_v on $X_{\mathbb{C}}$ is constant for v away from some hypersurface (which in particular has measure zero) and this constant is the top Chern class of some very explicit vector bundle [FO12]. For more information on this hypersurface, see [DHO$^+$16, Section 7] and [Hor15]. Explicit equations for these hypersurfaces are not known, even in our setting.

Over the *real* numbers, which we consider, the number of critical points of d_v can jump as v passes through (the real locus of) the same hypersurface. Typically, it jumps by 2, as two real cricital points come together and continue as a complex-conjugate pair of critical points. To arrive at a single number, we therefore impose a probability distribution on our data space V with density function ω (soon specialized to a standard multivariate Gaussian), and we ask: what is the *expected number* of critical points of d_v when v is drawn from the given probability distribution? In other words, we want to compute

$$\int_{\mathbb{R}^{n_1}\otimes\cdots\otimes\mathbb{R}^{n_p}} \#\{\text{real critical points of } d_v \text{ on } X\}\omega(v)dv.$$

This formula is complicated for two different reasons. First, given a point $v \in V$, the value of the integrand at v is not easy to compute. Second, the integral is over a space of dimension $N := \prod_i n_i$, which is rather large even for small values of the n_i. The main result of this paper is the following formula for the above integral, in the Gaussian case, in terms of an integral over a space of much smaller dimension quadratic in the number $n := \sum_i n_i$.

Theorem 1.1. *Suppose that $v \in V$ is drawn from the (standard) multivariate Gaussian distribution with (mean zero and) density function*

$$\omega(v) := \frac{1}{(2\pi)^{N/2}} e^{-(\sum_\alpha v_\alpha^2)/2},$$

where the multi-index α runs over $\{1,\ldots,n_1\} \times \cdots \times \{1,\ldots,n_p\}$. Then the expected number of critical points of d_v on X equals

$$\frac{(2\pi)^{p/2}}{2^{n/2}} \frac{1}{\prod_{i=1}^p \Gamma\left(\frac{n_i}{2}\right)} \int_{W_1} |\det C(w_1)| \, \mathrm{d}\mu_{W_1}.$$

Here W_1 is a space of dimension $1+\sum_{i<j}(n_i-1)(n_j-1)$ with coordinates $w_0 \in \mathbb{R}$ and $C_{ij} \in \mathbb{R}^{(n_i-1)\times(n_j-1)}$ with $i<j$, $C(w_1)$ is the symmetric $(n-p)\times(n-p)$-matrix of block shape

$$\begin{bmatrix} w_0 I_{n_1-1} & C_{1,2} & \cdots & C_{1,p} \\ C_{1,2}^T & w_0 I_{n_2-1} & \cdots & C_{2,p} \\ \vdots & \vdots & & \vdots \\ C_{1,p}^T & C_{2,p}^T & \cdots & w_0 I_{n_p-1} \end{bmatrix},$$

and μ_{W_1} makes w_0 and the $\sum_{i<j}(n_i-1)\cdot(n_j-1)$ matrix entries of the $C_{i,j}$ into independent, standard normally distributed variables. Moreover, Γ is Euler's gamma function.

Not only the dimension of the integral has dropped considerably, but also the integrand can be evaluated easily. The following example illustrates the case where all n_i are equal to 2.

Example 1.2. Suppose that all n_i are equal to 2. Then the matrix $C(w_1)$ becomes

$$C(w_1) = \begin{bmatrix} w_0 & w_{12} & \cdots & w_{1p} \\ w_{12} & w_0 & \cdots & w_{2p} \\ \vdots & \vdots & & \vdots \\ w_{1p} & w_{2p} & \cdots & w_0 \end{bmatrix}$$

where the distinct entries are independent scalar variables $\sim \mathcal{N}(0,1)$. The expected number of critical points of d_v on X equals

$$\frac{(2\pi)^{p/2}}{2^{2p/2}}\frac{1}{\Gamma(\frac{1}{1})^p}\mathbb{E}(|\det(C(w_1))|) = \left(\frac{\pi}{2}\right)^{p/2}\mathbb{E}(|\det(C(w_1))|),$$

where the latter factor is the expected absolute value of the determinant of $C(w_1)$. For $p=2$ that expected value of $|w_0^2 - w_{12}^2|$ can be computed symbolically and equals $4/\pi$. Thus the expression above then reduces to 2, which is just the number of singular values of a 2×2-matrix. For higher p we do not know a closed form expression for $\mathbb{E}(|\det(C(w_1))|)$, but we will present some numerical approximations in Section 5. $\diamondsuit$

In Section 3 we prove Theorem 1.1, and in Section 5 we list some numerically computed values. These values lead to the following intriguing *stabilization conjecture.*

Conjecture 1.3. *Suppose that $n_p - 1 > \sum_{i=1}^{p-1} n_i - 1$. Then, in the Gaussian setting of Theorem 1.1, the expected number of critical points of d_v on X does not decrease if we replace n_p by $n_p - 1$.*

For $p=2$ this follows from the statement that the number of singular values of a sufficiently general $n_1 \times n_2$-matrix with $n_1 < n_2$ equals n_1, which in fact remains the *same* when replacing n_2 by $n_2 - 1$. For arbitrary p the statement is true over $\mathbb{C}$ as shown in [FO12], again with equality, but the proof is not bijective. Instead, it uses vector bundles and Chern classes, techniques that do not carry over to our setting. It would be very interesting to find a direct geometric argument that *does* explain our experimental findings over the reals, as well.

Example 1.4. Alternatively, one could try and prove the conjecture directly from the integral formula in Theorem 1.1. The smallest open case is when $p = 3$ and $(n_1, n_2, n_3) = (2, 2, 4)$, and here the conjecture says that

$$\frac{\sqrt{\pi}}{2\sqrt{2}} \int_{\mathbb{R}} \int_{\mathbb{R}^7} \left| \det \begin{pmatrix} w_0 & w_{12} & w_{13} & w_{14} & w_{15} \\ w_{12} & w_0 & w_{23} & w_{23} & w_{25} \\ w_{13} & w_{23} & w_0 & 0 & 0 \\ w_{14} & w_{24} & 0 & w_0 & 0 \\ w_{15} & w_{25} & 0 & 0 & w_0 \end{pmatrix} \right| e^{-\frac{w_0^2 + \sum w_{ij}^2}{2}} \mathrm{d}w_0 \mathrm{d}w_{ij}$$

$$\leqslant \int_{\mathbb{R}} \int_{\mathbb{R}^5} \left| \det \begin{pmatrix} w_0 & w_{12} & w_{13} & w_{14} \\ w_{12} & w_0 & w_{23} & w_{24} \\ w_{13} & w_{23} & w_0 & 0 \\ w_{14} & w_{24} & 0 & w_0 \end{pmatrix} \right| e^{-\frac{w_0^2 + \sum w_{ij}^2}{2}} \mathrm{d}w_0 \mathrm{d}w_{ij}.$$

The determinant in the first integral is approximately w_0 times a determinant like in the second integral, but we do not know how to turn this observation into a proof of this integral inequality. $\diamondsuit$

Symmetric tensors. In the second part of this paper, we discuss *symmetric tensors*. There we consider the space $V = S^p\mathbb{R}^n$ of homogeneous polynomials of degree p in the standard basis $e_1, \ldots, e_n$ of $\mathbb{R}^n$, and X is the subvariety of V consisting of all polynomials that are of the form $\pm u^p$ with $u \in \mathbb{R}^n$. We equip V with the *Bombieri norm*, in which the monomials in the e_i form an orthogonal basis with squared norms

$$||e_1^{\alpha_1} \cdots e_n^{\alpha_n}||^2 = \frac{\alpha_1! \cdots \alpha_n!}{p!}.$$

Our result on the average number of critical points of d_v on X is as follows.

Theorem 1.5. *When $v \in S^p\mathbb{R}^n$ is drawn from the standard Gaussian distribution relative to the Bombieri norm, then the expected number of critical points of d_v on the variety of (plus or minus) pure p-th powers equals*

$$\frac{1}{2^{(n^2+3n-2)/4} \prod_{i=1}^n \Gamma(i/2)} \int_{\lambda_2 \leqslant \ldots \leqslant \lambda_n} \int_{-\infty}^{\infty} \left(\prod_{i=2}^n |\sqrt{p} w_0 - \sqrt{p-1}\lambda_i| \right)$$
$$\cdot \left(\prod_{i<j} (\lambda_j - \lambda_i) \right) e^{-w_0^2/2 - \sum_{i=2}^n \lambda_i^2/4} \mathrm{d}w_0 \mathrm{d}\lambda_2 \cdots \mathrm{d}\lambda_n.$$

Here the dimension reduction is even more dramatic: from an integral over a space of dimension $\binom{p+n-1}{p}$ to an integral over a polyhedral cone of dimension n. In this case, the corresponding complex count is already known from [CS13]: it is the geometric series $1 + (p-1) + \cdots + (p-1)^{n-1}$.

Example 1.6. For $p = 2$ the integral above evaluates to n (see Subsection 4.8 for a direct computation). Indeed, for $p = 2$ the symmetric tensor v is a symmetric matrix, and the critical points of d_v on the manifold of rank-one symmetric matrices are those of the form λuu^T, with u a norm-1 eigenvector of v with eigenvalue λ.

For $n = 2$ it turns out that the above integral can also be evaluated in closed form, with value $\sqrt{3p-2}$; a different proof of this fact appeared in [DHO$^+$16]. For $n = 3$ we provide a closed formula in Section 5. In all of these cases, the average count is an algebraic number. We do not know if this persists for larger values of n. $\diamondsuit$

Outline. The remainder of this paper is organized as follows. First, in Section 2 we explain a double counting strategy for computing the quantity of interest. This strategy is then applied to ordinary tensors in Section 3 and to symmetric tensors in Section 4. We conclude with some (symbolically or numerically) computed values in Section 5.

Acknowledgments. This paper fits in the research programme laid out in [DHO+16], which asks for *Euclidean distance degrees* of algebraic varieties arising in applications. We thank the authors of that paper, as well as our Eindhoven colleague Rob Eggermont, for several stimulating discussions on the topic of this paper.

2. Double counting

Suppose that we have equipped $V = \mathbb{R}^N$ with an inner product $(.|.)$ and that we have a smooth manifold $X \subseteq V$. Assume that we have a probability density ω on $V = \mathbb{R}^N$ and that we want to count the average number of critical points $x \in X$ of the function $d_v(x) := (v - x|v - x)$ when v is drawn according to that density. Let Crit denote the set

$$\text{Crit} := \{(v, x) \mid v - x \perp T_x X\} \subseteq V \times X$$

of pairs $(v, x) \in X \times V$ for which x is a critical point of d_v. For fixed $x \in X$ the $v \in V$ with $(v, x) \in \text{Crit}$ form an affine space, namely, $x + (T_x X)^\perp$. In particular, Crit is a manifold of dimension N. On the other hand, for fixed $v \in V$, the $x \in X$ for which $(v, x) \in \text{Crit}$ are what we want to count. Let $\pi_V : \text{Crit} \to V$ be the first projection. Then (the absolute value of) the pull-back $|\pi_V^* \omega \mathrm{d}v|$ is a pseudo volume form on Crit, and we have

$$\int_V \#(\pi_V^{-1}(v))\omega(v)\mathrm{d}v = \int_{\text{Crit}} 1|\pi_V^* \omega \mathrm{d}v|.$$

Now suppose that we have a smooth $1:1$ parameterization $\varphi : \mathbb{R}^N \to \text{Crit}$ (perhaps defined outside some set of measure zero). Then the latter integral is just

$$\int_{\mathbb{R}^N} |\det J_w(\pi_V \circ \varphi)|\omega(\pi_V(\varphi(w)))\mathrm{d}w,$$

where $J_w(\pi_V \circ \varphi)$ is the Jacobian of $\pi_V \circ \varphi$ at the point w. We will see that if X is the manifold of rank-one tensors or rank-one symmetric tensors, then Crit (or in fact, a slight variant of it) has a particularly friendly parameterization, and we will use the latter expression to compute the expected number of critical points of d_v. In a more general setting, this double-counting approach is discussed in [DHO+16].

3. Ordinary tensors

3.1. **Set-up.** Let $V_1, \ldots, V_p$ be real vector spaces of dimensions $n_1 \leqslant \ldots \leqslant n_p$ equipped with positive definite inner products $(.|.)$. Equip $V := \bigotimes_{i=1}^p V_i$, a vector space of dimension $N := n_1 \cdots n_p$, with the induced inner product and associated norm, also denoted $(.|.)$. Given a tensor $v \in V$, we want to count the number of critical points of the function

$$d_v : x \mapsto ||v - x||^2 = (v|v) - 2(v|x) + (x|x)$$

on the manifold $X \subseteq V$ of non-zero rank-one tensors $x = x_1 \otimes \cdots \otimes x_p$. The following well-known lemma (see for instance [FO12]) characterizes which x are critical for a given $v \in V$. In its statement we extend the notation $(v|u)$ to the setting where u is a tensor in $\bigotimes_{i \in I} V_i$ for some subset $I \subseteq \{1, \ldots, p\}$, to stand for the tensor in $\bigotimes_{i \notin I} V_i$ obtained by contracting v with u using the inner products.

Lemma 3.1. *The non-zero rank-one tensor $x = x_1 \otimes \cdots \otimes x_p$ is a critical point of d_v if and only if for all $i = 1, \ldots, p$ we have*

$$(v|x_1 \otimes \cdots \otimes \hat{x_i} \otimes \cdots \otimes x_p) = \left(\prod_{j \neq i} (x_j|x_j) \right) x_i.$$

In words: pairing v with the tensor product of the x_j with $j \neq i$ gives a well-defined scalar multiple of x_i, and this should hold for all i.

Proof. The tangent space at x to the manifold of rank-one tensors is $\sum_{i=1}^{p} x_1 \otimes \cdots \otimes V_i \otimes \cdots \otimes x_p$. Fixing i and $y \in V_i$, the derivative of d_v in the direction $x_1 \otimes \cdots \otimes y \otimes \cdots \otimes x_p$ is

$$-2(v - x_1 \otimes \cdots \otimes x_p | x_1 \otimes \cdots \otimes y \otimes \cdots \otimes x_p).$$

Equating this to zero for all y yields that

$$(v|x_1 \otimes \cdots \otimes \hat{x_i} \otimes \cdots \otimes x_p) = (x_1 \otimes \cdots \otimes x_p | x_1 \otimes \cdots \otimes \hat{x_i} \otimes \cdots \otimes x_p) = \left(\prod_{j \neq i} (x_j|x_j) \right) x_i,$$

as claimed. □

The lemma can also be read as follows: a rank-one tensor $x_1 \otimes \cdots \otimes x_p$ is critical for d_v if and only if first, for each i the contraction $(v|x_1 \otimes \cdots \otimes \hat{x_i} \otimes \cdots \otimes x_p)$ is some scalar multiple of x_i, and second, $(v|x_1 \otimes \cdots \otimes x_p)$ equals $\prod_j (x_j|x_j)$. From this description it is clear that if $x_1 \otimes \cdots \otimes x_p$ merely satisfies the first condition, then some scalar multiple of it is critical for d_v. Also, if a rank-one tensor u is critical for d_v, then tu is critical for d_{tv} for all $t \in \mathbb{R}$. These considerations give rise to the following definition and proposition.

Definition 3.2. Define Crit to be the subset of $V \times (\mathbb{P}V_1 \times \cdots \times \mathbb{P}V_p)$ consisting of points $(v, ([u_1], \ldots, [u_p]))$ for which all 2×2-determinants of the $\dim V_i \times 2$-matrix $[(v|u_1 \otimes \cdots \otimes \hat{u_i} \otimes \cdots \otimes u_p) \quad | \quad u_i]$ vanish, for each $i = 1, \ldots, p$.

Proposition 3.3. *The projection* Crit $\to \prod_i \mathbb{P}V_i$ *is a smooth sub-bundle of the trivial bundle $V \times \prod_i \mathbb{P}V_i$ over $\prod_i \mathbb{P}V_i$ of rank $N - (n_1 + \cdots + n_p) + p$, while the fiber of the projection $\pi_V :$* Crit $\to V$ *over a tensor v counts the number of critical points of d_v in the manifold of non-zero rank-one tensors.*

Proof. The second statement is clear from the above. For the first observe that the fiber above $\mathbf{u} = ([u_1], \ldots, [u_p])$ equals $W_{\mathbf{u}} \times \{([u_1], \ldots, [u_p])\}$ where

$$W_{\mathbf{u}} = \left(\bigoplus_{i=1}^{p} u_1 \otimes \cdots \otimes (u_i)^{\perp} \otimes \cdots \otimes u_p \right)^{\perp} \subseteq V.$$

This space varies smoothly with $\mathbf{u}$ and has codimension $\sum_i (n_i - 1)$, whence the dimension formula. □

We want to compute the average fiber size of the projection Crit $\to V$. Here *average* depends on the choice of a measure on V, and we take the Gaussian measure $\frac{1}{(2\pi)^{N/2}} e^{-||v||^2/2} \mathrm{d}v$, where $\mathrm{d}v$ stands for ordinary Lebesgue measure obtained from identifying V with $\mathbb{R}^N$ by a linear map that relates $(.|.)$ to the standard inner product on $\mathbb{R}^N$.

3.2. **Parameterizing** Crit**.** To apply the double counting strategy from Section 2, we introduce a convenient parameterization of Crit. Fix norm-1 vectors $e_i \in V_i,\ i = 1, \ldots, p$, write $\mathbf{e} = (e_1, \ldots, e_p)$ and $[\mathbf{e}] := ([e_1], \ldots, [e_p])$, and define

$$W := W_{[\mathbf{e}]} = \left(\bigoplus_{i=1}^{p} e_1 \otimes \cdots \otimes (e_i)^{\perp} \otimes \cdots \otimes e_p \right)^{\perp}.$$

We parameterize (an open subset of) $\mathbb{P}V_i$ by the map $e_i^{\perp} \to \mathbb{P}V_i, u_i \mapsto [e_i + u_i]$. Write $U := \prod_{i=1}^{p} (e_i^{\perp})$. For $\mathbf{u} = (u_1, \ldots, u_p) \in U$ let $R_{\mathbf{u}}$ denote a linear isomorphism $W \to W_{[\mathbf{e}+\mathbf{u}]}$, to be chosen later, but at least smoothly varying with $\mathbf{u}$ and perhaps defined outside some subvariety of positive codimension.

Next define

$$\varphi : W \times U \to V,\ (w, \mathbf{u}) \mapsto R_{\mathbf{u}} w.$$

Then we have the following fundamental identity

$$\frac{1}{(2\pi)^{N/2}} \int_V (\#\pi_V^{-1}(v)) \cdot e^{-\frac{||v||^2}{2}} \mathrm{d}v = \frac{1}{(2\pi)^{N/2}} \int_{W \times U} |\det J_{(w,\mathbf{u})}\varphi| e^{-\frac{||R_{\mathbf{u}} w||^2}{2}} \mathrm{d}\mathbf{u}\ \mathrm{d}w,$$

where $J_{(w,\mathbf{u})}\varphi$ is the Jacobian of φ at $(w, \mathbf{u})$, whose determinant is measured relative to the volume form on V coming from the inner product and the volume form on $W \times U$ coming from the inner products of the factors, which are interpreted perpendicular to each other. The left-hand side is our desired quantity, and our goal is to show that the right-hand side reduces to the formula in Theorem 1.1.

We choose $R_{\mathbf{u}}$ to be the tensor product $R_{u_1} \otimes \cdots \otimes R_{u_p}$, where R_{u_i} is the element of $\mathrm{SO}(V_i)$ determined by the conditions that it maps e_i to a positive scalar multiple of $e_i + u_i$ and that it restricts to the identity on $\langle e_i, u_i \rangle^{\perp}$; this map is unique for non-zero $u_i \in e_i^{\perp}$. Indeed, we have

$$\begin{aligned} R_{u_i} &= \left(I - e_i e_i^T - \frac{u_i}{||u_i||} \frac{u_i^T}{||u_i||} \right) + \left(\frac{e_i + u_i}{\sqrt{1 + ||u_i||^2}} e_i^T + \frac{u_i - ||u_i||^2 e_i}{||u_i|| \sqrt{1 + ||u_i||^2}} \frac{u_i^T}{||u_i||} \right) \\ &= \left(I - e_i e_i^T - \frac{u_i u_i^T}{||u_i||^2} \right) + \left(\frac{e_i + u_i}{\sqrt{1 + ||u_i||^2}} e_i^T + \frac{u_i - ||u_i||^2 e_i}{\sqrt{1 + ||u_i||^2}} \frac{u_i^T}{||u_i||^2} \right) \end{aligned}$$

where the first term is the orthogonal projection to $\langle e_i, u_i \rangle^{\perp}$ and the second term is projection onto the plane $\langle e_i, u_i \rangle$ followed by a suitable rotation there. Two important remarks concerning symmetries are in order. First, by construction of R_{u_i} we have

$$R_{u_i}^{-1} = R_{-u_i}. \tag{1}$$

Second, for any element $g \in \mathrm{SO}(e_i^{\perp}) \subseteq \mathrm{SO}(V_i)$ we have

$$R_{g u_i} = g \circ R_{u_i} \circ g^{-1}. \tag{2}$$

We now compute the derivative at u_i of the map $e_i^\perp \to \mathrm{SO}(V_i), u \mapsto R_u$ in the direction $v_i \in e_i^\perp$. First, when v_i is perpendicular to both e_i and u_i, this derivative equals

$$(3) \qquad \frac{\partial R_{u_i}}{\partial v_i} = \frac{1}{\sqrt{1+||u_i||^2}}(v_i e_i^T - e_i v_i^T) - \frac{\sqrt{1+||u_i||^2}-1}{||u_i||^2\sqrt{1+||u_i||^2}}(u_i v_i^T + v_i u_i^T).$$

Second, when v_i equals u_i, the derivative equals

$$(4) \qquad \frac{\partial R_{u_i}}{\partial u_i} = \frac{1}{(1+||u_i||^2)^{3/2}}(-u_i u_i^T + u_i e_i^T - e_i u_i^T - ||u_i||^2 e_i e_i^T).$$

For now, fix $(w, \mathbf{u}) \in W \times U$. On the subspace $T_w W = W$ of $T_{(w,\mathbf{u})} W \times U$ the Jacobian of φ is just the map $W \to V, w \mapsto R_{\mathbf{u}} w$. Hence relative to the orthogonal decompositions $V = W^\perp \oplus W$ and $U \times W$, we have a block decomposition

$$R_{\mathbf{u}}^{-1} J_{(w,\mathbf{u})}\varphi = \begin{bmatrix} A_{(w,\mathbf{u})} & 0 \\ * & I_W \end{bmatrix}$$

for a suitable matrix $A_{(w,\mathbf{u})}$. Note that this matrix has size $(n-p)\times(n-p)$, which is the size of the determinant in Theorem 1.1. As $R_{\mathbf{u}}$ is orthogonal with determinant 1, we have $\det J_{(w,\mathbf{u})}\varphi = \det A_{(w,\mathbf{u})}$ and $||R_{\mathbf{u}} w|| = ||w||$. This yields the following proposition.

Proposition 3.4. *The expected number of critical tank-one approximations to a standard Gaussian tensor in V is*

$$I := \frac{1}{(2\pi)^{N/2}} \int_W \int_U |\det A_{(w,\mathbf{u})}| e^{-\frac{||w||^2}{2}} \mathrm{d}\mathbf{u}\, \mathrm{d}w.$$

For later use, consider the function $F : U \to \mathbb{R}$ defined as

$$F(u) = \frac{1}{(2\pi)^{N/2}} \int_W |\det A_{(w,\mathbf{u})}| e^{-\frac{||w||^2}{2}} \mathrm{d}w.$$

From (2) and the fact that the Gaussian density on W is orthogonally invariant, it follows that F is invariant under the group $\prod_{i=1}^p \mathrm{SO}(e_i^\perp)$. In particular, its value depends only on the tuple $(||u_1||, \ldots, ||u_p||) =: (t_1, \ldots, t_p)$. This will be used in the following subsection.

3.3. **The shape of $A_{(w,\mathbf{u})}$.** Recall that $U = \prod_{i=1}^p (e_i^\perp)$. Correspondingly, the columns of the matrix $A_{(w,\mathbf{u})}$ come in p blocks, one for each $e_i^\perp$. The i-th block records the $W^\perp$-components of the vectors $\left(R_{\mathbf{u}}^{-1} \frac{\partial R_{\mathbf{u}}}{\partial \mathbf{v}_i}\right) w$, where $\mathbf{v}_i = (0, \ldots, v_i, \ldots, 0)$ and v_i runs through an orthonormal basis $e_i^{(1)}, \ldots, e_i^{(n_i-1)}$ of $e_i^\perp$. We have

$$(5) \qquad R_{\mathbf{u}}^{-1} \frac{\partial R_{\mathbf{u}}}{\partial \mathbf{v}_i} = \mathrm{Id} \otimes \cdots \otimes R_{u_i}^{-1} \frac{\partial R_{u_i}}{\partial v_i} \otimes \cdots \otimes \mathrm{Id}.$$

Furthermore, if v_i is also perpendicular to u_i, then by 3 and 1

$$(6) \qquad R_{u_i}^{-1} \frac{\partial R_{u_i}}{\partial v_i} = \frac{1}{\sqrt{1+||u_i||^2}}(v_i e_i^T - e_i v_i^T) + \frac{1-\sqrt{1+||u_i||^2}}{||u_i||^2\sqrt{1+||u_i||^2}}(v_i u_i^T - u_i v_i^T).$$

On the other hand, when v_i is parallel to u_i, then

$$(7) \qquad R_{u_i}^{-1} \frac{\partial R_{u_i}}{\partial v_i} = \frac{1}{1+||u_i||^2}(v_i e_i^T - e_i v_i^T).$$

This is derived from (1) and (4), keeping in mind the fact that here v_i needs not be *equal* to u_i, but merely *parallel* to it. Note that both matrices are skew-symmetric. This is no coincidence: the directional derivative $\partial R_{u_i}/\partial v_i$ lies in the tangent space to $\mathrm{SO}(V_i)$ at u_i, and left multiplying by $R_{u_i}^{-1}$ maps these elements into the Lie algebra of $\mathrm{SO}(V_i)$, which consists of skew symmetric matrices.

We decompose the space W as

$$W = \left(\bigoplus_{i=1}^{p} e_1 \otimes \cdots \otimes (e_i)^\perp \otimes \cdots \otimes e_p\right)^\perp = \mathbb{R} \cdot e_1 \otimes e_2 \otimes \cdots \otimes e_p$$
$$\oplus \left(\bigoplus_{1 \leqslant i < j \leqslant p} e_1 \otimes \cdots \otimes e_i^\perp \otimes \cdots \otimes e_j^\perp \otimes \cdots \otimes e_p\right) \oplus W' =: W_0 \oplus W',$$

where W' contains the summands that contain at least three $e_i^\perp$-s as factors. From (5) it follows that $R_{\mathbf{u}}^{-1}\frac{\partial R_{\mathbf{u}}}{\partial \mathbf{v}_i} W' \subseteq W$. So for a general w we use the parameters

$$w = w_0 \cdot e_1 \otimes \cdots \otimes e_p + \sum_{1 \leqslant i < j \leqslant p} \sum_{1 \leqslant a \leqslant n_i - 1} \sum_{1 \leqslant b \leqslant n_j - 1} w_{i,j}^{a,b} e_1 \otimes \cdots \otimes e_i^{(a)} \otimes \cdots \otimes e_j^{(b)} \otimes \cdots \otimes e_p + w',$$

where w_0 and $w_{i,j}^{a,b}$ are real numbers, and where $w' \in W'$ will not contribute to $A_{(w,\mathbf{u})}$. We also write $w_1 = (w_0, (w_{i,j}^{a,b}))$ for the components of w that do contribute.

As a further simplification, we take each u_i equal to a scalar $t_i \geqslant 0$ times the first basis vector $e_i^{(1)}$ of $e_i^\perp$. This is justified by the observation that the function F is invariant under the group $\prod_i \mathrm{SO}(e_i^\perp)$. Thus we want to determine $A_{\left(w,(t_1 e_1^{(1)}, t_2 e_2^{(1)}, \ldots, t_p e_p^{(1)})\right)}$. This matrix has a natural block structure $(B_{i,j})_{1 \leqslant i,j \leqslant p}$, where $B_{i,j}$ is the part of the Jacobian containing the $e_1 \otimes \cdots \otimes e_i^\perp \otimes \cdots \otimes e_p$-coordinates of $\left(R_{\mathbf{u}}^{-1} \frac{\partial R_{\mathbf{u}}}{\partial \mathbf{v}_j}\right) w$ with $\mathbf{v}_j = (0, \ldots, v_j, \ldots, 0)$.

Fixing $i < j$, the matrix $B_{i,j}$ is of type $(n_i - 1) \times (n_j - 1)$, where the (a,b)-th element is the $e_1 \otimes \cdots \otimes e_i^{(a)} \otimes \cdots \otimes e_p$-coordinate of

$$\left(R_{u_j}^{-1} \frac{\partial R_{u_j}}{\partial e_j^{(b)}}\right) w.$$

First, if $b \neq 1$, then we have a directional derivative in a direction perpendicular to $u_j = t_j e_j^{(1)}$. Applying formula 6 for the directions $e_j^{(b)}$ yields

$$B_{i,j}(a,b) = \frac{-w_{i,j}^{a,b}}{\sqrt{1 + t_j^2}}.$$

Second, if $b = 1$, then we consider directional derivatives parallel to u_j, so applying formula 7 for direction $e_j^{(1)}$, we get

$$B_{i,j}(a,1) = \frac{-w_{i,j}^{a,1}}{1 + t_j^2}.$$

Putting all together, the matrix $B_{i,j}$ is as follows

$$B_{i,j} = \left(\frac{1}{1 + t_j^2} C_{i,j}^1, \frac{1}{\sqrt{1 + t_j^2}} C_{i,j}^2, \ldots, \frac{1}{\sqrt{1 + t_j^2}} C_{i,j}^{n_j - 1}\right),$$

where $C_{i,j}^b = \left(-w_{i,j}^{a,b}\right)_{1 \leqslant a \leqslant n_i - 1}$ are column vectors for all $1 \leqslant b \leqslant n_j - 1$. Denote the matrix consisting of these column vectors by $C_{i,j}$. Doing the same calculations but now for the matrix $B_{j,i}$, and writing $C_{j,i} = C_{i,j}^T$, we find that

$$B_{j,i} = \left(\frac{1}{1+t_i^2} C_{j,i}^1, \frac{1}{\sqrt{1+t_i^2}} C_{j,i}^2, \ldots, \frac{1}{\sqrt{1+t_i^2}} C_{j,i}^{n_i - 1} \right).$$

The only remaining case is when $i = j$, and then similar calculations yield that $B_{j,j} = \frac{1}{(1+t_j^2)^{\frac{n_j}{2}}} w_0 I_{n_j - 1}$. We summarize the content of this subsection as follows.

Proposition 3.5. *For $(w, \mathbf{u}) \in W \times U$ with $\mathbf{u} = (t_1 e_1^{(1)}, \ldots, t_p e_p^{(1)})$ we have*

$$\det A_{(w,\mathbf{u})} = \prod_{k=1}^{p} \frac{1}{(1+t_k^2)^{\frac{n_k}{2}}} \det \begin{pmatrix} C_1 & C_{1,2} & \cdots & C_{1,p} \\ C_{1,2}^T & C_2 & \cdots & C_{2,p} \\ \vdots & \vdots & & \vdots \\ C_{1,p}^T & C_{2,p}^T & \cdots & C_p \end{pmatrix},$$

where $C_{i,j} = \left(-w_{i,j}^{a,b}\right)_{a,b}$ and $C_j = w_0 I_{n_j - 1}$ for all $1 \leqslant i < j \leqslant p$.

For further reference we denote the above matrix $(C_{i,j})_{1 \leqslant i,j \leqslant p}$ by $C(w_1)$.

3.4. **The value of I.** We are now in a position to prove our formula for the expected number of critical rank-one approximations to a Gaussian tensor v.

Proof of Theorem 1.1. Combine Propositions 3.4 and 3.5 into the expression

$$I = \frac{1}{(2\pi)^{\frac{N}{2}}} \prod_{k=1}^{p} \mathrm{Vol}(S^{n_k - 2}) \int_W \int_0^\infty \cdots \int_0^\infty \prod_{i=1}^{p} \frac{t_i^{n_i - 2}}{(1+t_i^2)^{\frac{n_i}{2}}} \left|\det C(w_1)\right| e^{-\frac{||w||^2}{2}} \mathrm{d}t_1 \cdots \mathrm{d}t_p \mathrm{d}w.$$

Here the factors $t_i^{n_i - 2}$ and the volumes of the sphere account for the fact that F is orthogonally invariant and $\mathrm{d}u_i = t_i^{n_i - 2} d_t d_S$, where d_S is the surface element of the $(n_i - 2)$-dimensional unit sphere in $e_i^\perp$. Now recall that

$$\int_0^\infty \frac{t^{n_i - 2}}{(1+t^2)^{\frac{n_i}{2}}} \mathrm{d}t = \frac{\sqrt{\pi}}{2} \frac{\Gamma(\frac{n_i - 1}{2})}{\Gamma(\frac{n_i}{2})},$$

and that the volume of the $(n-2)$-sphere is

$$\mathrm{Vol}(S^{n_i - 2}) = \frac{2\pi^{\frac{n_i - 1}{2}}}{\Gamma(\frac{n_i - 1}{2})}.$$

Plugging in the above two formulas, we obtain

$$I = \frac{\sqrt{\pi}^n}{\sqrt{2\pi}^N} \frac{1}{\prod_{i=1}^{p} \Gamma\left(\frac{n_i}{2}\right)} \int_W \left|\det C(w)\right| e^{-\frac{||w||^2}{2}} \mathrm{d}w.$$

Now the integral splits as an integral over W_1 and one over W':

$$\int_W |\det C(w)|\, e^{-\frac{||w||^2}{2}} \mathrm{d} = \int_{W'} e^{-\frac{||w'||^2}{2}} \mathrm{d}w' \int_{W_1} |\det C(w_1)|\, e^{-\frac{||w_1||^2}{2}} \mathrm{d}w_1$$

$$= \sqrt{2\pi}^{\dim W} \left(\frac{1}{\sqrt{2\pi}^{\dim W_1}} \int_{W_1} |\det C(w_1)|\, e^{-\frac{||w_1||^2}{2}} \mathrm{d}w_1 \right)$$

$$= \sqrt{2\pi}^{N-(n-p)} \mathbb{E}(|\det C(w_1)|)$$

where w_1 is drawn from a standard Gaussian distribution on W_1. Inserting this in the expression for I yields the expression for I in Theorem 1.1. □

3.5. **The matrix case.** In this section we perform a sanity check, namely, we show that our formula in Theorem 1.1 gives the correct answer for the case $p = 2$ and $n_1 = n_2 = n$—which is n, the number of singular values of any sufficiently general matrix. In this special case we compute

$$J := \int_W |\det C(w)|\, \mathrm{d}\mu_W = \int_{-\infty}^{\infty} \int_{\mathrm{M}_{n-1}} \left| \det \begin{pmatrix} w_0 I_{n-1} & B \\ B^T & w_0 I_{n-1} \end{pmatrix} \right| e^{\frac{||w_0^2||}{2}} \mathrm{d}\mu_B \mathrm{d}w_0 =$$

$$= \int_{-\infty}^{\infty} \int_{\mathrm{M}_{n-1}} \left| \det(w_0^2 I_{n-1} - BB^T) \right| e^{\frac{||w_0||^2}{2}} \mathrm{d}\mu_B \mathrm{d}w_0,$$

where $B \in \mathrm{M}_{n-1}(\mathbb{R})$ is a real $(n-1) \times (n-1)$ matrix. The matrix $A := BB^T$ is a symmetric positive definite matrix and since the entries of B are independent and normally distributed, A is drawn from the Wishart distribution with density $W(A)$ on the cone of real symmetric positive definite matrices [Rou07, Section 2.1]. Denote this space by Sym_{n-1}. So the integral we want to calculate is

$$J = \int_{-\infty}^{\infty} \int_{\mathrm{Sym}_{n-1}} \left| \det(w_0^2 I_{n-1} - A) \right| e^{\frac{||w_0||^2}{2}} \mathrm{d}W(A) \mathrm{d}w_0.$$

Now by [Rou07, Part 2.2.1] the joint probability density of the eigenvalues λ_j of A on the orthant $\lambda_j > 0$ is

$$\frac{1}{Z(n-1)} \prod_{j=1}^{n-1} \frac{e^{\frac{-\lambda_j}{2}}}{\sqrt{\lambda_j}} \prod_{1 \leqslant j < k < n-1} |\lambda_k - \lambda_j|, \tag{8}$$

where the normalizing constant is

$$Z(n-1) = \sqrt{2}^{(n-1)^2} \left(\frac{2}{\sqrt{\pi}} \right)^{n-1} \prod_{j=1}^{n-1} \Gamma\left(1 + \frac{j}{2}\right) \Gamma\left(\frac{n-j}{2}\right).$$

Using this fact we obtain

$$J = \frac{1}{Z(n-1)} \int_{\mathbb{R}} \int_{\lambda > 0} \prod_{j=1}^{n-1} \frac{e^{\frac{-\lambda_j}{2}}}{\sqrt{\lambda_j}} \prod_{1 \leqslant j < k < n-1} |\lambda_k - \lambda_j| \prod_{j=1}^{n-1} |w_0^2 - \lambda_j| e^{\frac{||w_0||^2}{2}} \mathrm{d}\lambda \mathrm{d}w_0.$$

Now making the change of variables $w_0^2 = \lambda_n$, so that
$$J = 2\frac{Z(n)}{Z(n-1)}.$$
Plugging in the remaining normalizing constants we find that the expected number of critical rank-one approximations to an $n \times n$-matrix is
$$I = \frac{\sqrt{\pi}^{2n}}{\sqrt{2\pi}^{n^2}}\Gamma\left(\frac{n}{2}\right)^{-2} 2\frac{Z(n)}{Z(n-1)} = n.$$

4. Symmetric tensors

4.1. **Set-up.** Now we turn our attention from arbitrary tensors to symmetric tensors, or, equivalently, homogeneous polynomials. For this, consider $\mathbb{R}^n$ with the standard orthonormal basis $e_1, e_2, \ldots, e_n$ and let $V = S^p\mathbb{R}^n$ be the space of homogeneous polynomials of degree p in n variables $e_1, e_2, \ldots, e_n$. Recall that, up to a positive scalar, V has a unique inner product that is preserved by the orthogonal group O_n in its natural action on polynomials in $e_1, \ldots, e_n$. This inner product, sometimes called the *Bombieri inner product*, makes the monomials $e^\sigma := \prod_i e_i^{\alpha_i}$ (with $\sigma \in \mathbb{Z}^n_{\geqslant 0}$ and $\sum_i \sigma_i = p$, which we will abbreviate to $\sigma \vdash p$) into an orthogonal basis with square norms
$$(e^\sigma | e^\sigma) = \frac{\sigma_1! \cdots \sigma_n!}{p!} =: \binom{p}{\sigma}^{-1}.$$
The scaling ensures that that the squared norm of a pure power $(t_1e_1 + \ldots + t_ne_n)^p$ equals $(\sum_i t_i^2)^p$. The scaled monomials
$$f_\sigma := \sqrt{\binom{p}{\sigma}} e^\sigma$$
form an orthonormal basis of V, and we equip V with the standard Gaussian distribution relative to this orthonormal basis.

Now our variety X can be defined by the parameterization
$$\psi : \mathbb{R}^n \to S^p\mathbb{R}^n,$$
$$t \mapsto t^p = \sum_{\sigma \vdash p} t_1^{\sigma_1} \cdots t_n^{\sigma_n} \sqrt{\binom{p}{\sigma}} f_\sigma.$$
In fact, if p is odd, then this parameterization is one-to-one, and $X = \operatorname{im}\psi$. If p is even, then this parameterization is two-to-one, and $X = \operatorname{im}\psi \cup (-\operatorname{im}\psi)$.

Definition 4.1. Define Crit to be the subset of $V \times X$ consisting of all pairs of (v, x) such that $v - x \perp T_xX$.

4.2. **Parameterizing** Crit**.** We derive a convenient parameterization of Crit, as follows. Taking the derivative of ψ at $t \neq 0$, we find that $T_{\pm t^p}X$ both equal $t^{p-1} \cdot \mathbb{R}^n$. In particular, for t any non-zero scalar multiple of e_1, this tangent space is spanned by all monomials that contain at least $(n-1)$ factors e_1. Let W denote the orthogonal complement of this space, which is spanned by all monomials that contain at most $(p-2)$ factors e_1. For $u \in e_1^\perp \setminus \{0\}$, recall from Subsection 3.2 the orthogonal map $R_u \in \mathrm{SO}_n$ that is the identity on $\langle e_1, u\rangle^\perp$ and a rotation sending e_1 to a scalar multiple of $e_1 + u$ on $\langle e_1, u\rangle$. We write S^pR_u for the induced linear map on V, which, in particular, sends e_1^p to $(e_1 + u)^p$. We have the following parameterization of Crit:

$$\begin{aligned} e_1^\perp \times \mathbb{R}e_1^p \times W &\to \text{Crit}, \\ (u, w_0 e_1^p, w) &\mapsto (w_0 S^p R_u e_1^p, w_0 S^p R_u e_1^p + S^p R_u w). \end{aligned}$$

Combining with the projection to V, we obtain the map

$$\varphi : e_1^\perp \times \mathbb{R}e_1^p \times W \to V,\ (u, w_0 e_1^p, w) \mapsto S^p R_u(w_0 e_1^p + w).$$

Following the strategy in Section 2, the expected number of critical points of d_v on X for a Gaussian v equals

$$I := \frac{1}{(2\pi)^{\dim V/2}} \int_{e_1^\perp} \int_{-\infty}^{\infty} \int_W |\det J_{(u,w_0,w)}\varphi| e^{-(w_0^2+||w||^2)/2} \mathrm{d}w \mathrm{d}w_0 \mathrm{d}u,$$

where we have used that $S^p R_u$ preserves the norm, and that $w \perp e_1^p$.

To determine the Jacobian determinant, we observe that $J_{(u,w_0,w)}\varphi$ restricted to $T_{w_0 e_1^p}\mathbb{R}e_1^p \oplus T_w W$ is just the linear map $S^p R_u$. Hence, relative to a block decomposition $V = (W + \mathbb{R}e_1^p)^\perp \oplus \mathbb{R}e_1^p \oplus W$ we find

$$S^p(R_u)^{-1} J_{(u,w_0,w)}\varphi = \left[\begin{array}{c|c|c} A_{(u,w_0,w)} & 0 & 0 \\ \hline * & 1 & 0 \\ \hline * & 0 & I \end{array}\right]$$

for a suitable linear map $A_{(u,w_0,w)} : e_1^\perp \to (W \oplus \mathbb{R}e_1^p)^\perp$.

4.3. **The shape of $A_{(u,w_0,w)}$.** For the computations that follow, we will need only part of our orthonormal basis of V, namely, e_1^p and the vectors

$$\begin{aligned} f_i &:= \sqrt{p} e_1^{p-1} e_i \\ f_{ii} &:= \sqrt{p(p-1)/2} e_1^{p-2} e_i^2 \\ f_{ij} &:= \sqrt{p(p-1)} e_1^{p-2} e_i e_j \end{aligned}$$

where $2 \leqslant i \leqslant n$ in the first two cases and $2 \leqslant i < j \leqslant n$ in the last case. The target space of $A_{(u,w_0,w)}$ has an orthonormal basis $f_2, \ldots, f_n$, while the domain has an orthonormal basis $e_2, \ldots, e_n$. Let a_{kl} be the coefficient of f_k in $A_{(u,w_0,w)} e_l$. To compute a_{kl}, we expand w as

$$w = \sum_{2 \leqslant i \leqslant j} w_{ij} f_{ij} + w' =: w_1 + w'$$

where w' contains the terms with at most $p-3$ factors e_1. We have the identity

$$S^p(R_u)^{-1} \frac{\partial S^p R_u(e_{i_1} \cdots e_{i_p})}{\partial e_l} = \sum_{m=1}^{p} e_{i_1} \cdots (R_u^{-1} \frac{\partial R_u}{\partial e_l} e_{i_m}) \cdots e_{i_p}.$$

For this expression to contain terms that are multiples of some f_k, we need that at least $p-2$ of the i_m are equal to 1. Thus a_{kl} is independent of w', which is why we need only the basis vectors above.

As in the case of ordinary tensors, we make the further simplification that $u = te_2$. Then we have to distinguish two cases: $l = 2$ and $l > 2$. For $l = 2$ formula (7) applies, and we compute modulo $\langle f_2, \ldots, f_n\rangle^\perp$

$$\begin{aligned}
&(S^pR_{te_2})^{-1}\frac{\partial(S^pR_{te_2}(w_0e_1^p + w_1))}{\partial e_2}\\
&= (S^pR_{te_2})^{-1}\frac{\partial(S^pR_{te_2}(w_0e_1^p + \sum_{2\leqslant i} w_{ii}f_{ii} + \sum_{2\leqslant i<j} w_{ij}f_{ij}))}{\partial e_2}\\
&= \frac{1}{1+t^2}(pw_0e_1^{p-1}e_2 - 2w_{22}\sqrt{p(p-1)/2}e_1^{p-1}e_2 - \sum_{2<j} w_{2j}\sqrt{p(p-1)}e_1^{p-1}e_j)\\
&= \frac{1}{1+t^2}((\sqrt{p}w_0 - \sqrt{2(p-1)}w_{22})f_2 - \sum_{2<j}\sqrt{p-1}w_{2j}f_j).
\end{aligned}$$

For $l > 2$ formula (6) applies, but in fact the second term never contributes when we compute modulo $\langle f_2, \ldots, f_n\rangle^\perp$:

$$\begin{aligned}
&(S^pR_{te_2})^{-1}\frac{\partial(S^pR_{te_2}(w_0e_1^p + w_1))}{\partial e_l}\\
&= (S^pR_{te_2})^{-1}\frac{\partial(S^pR_{te_2}(w_0e_1^p + \sum_{2\leqslant i} w_{ii}f_{ii} + \sum_{2\leqslant i<j} w_{ij}f_{ij}))}{\partial e_l}\\
&= \frac{1}{\sqrt{1+t^2}}\Bigg(pw_0e_1^{p-1}e_l - 2w_{ll}\sqrt{p(p-1)/2}e_1^{p-1}e_l\\
&- \sqrt{p(p-1)}(\sum_{2\leqslant i<l} w_{il}e_1^{p-1}e_i + \sum_{l<j} w_{2j}e_1^{p-1}e_j)\Bigg)\\
&= \frac{1}{\sqrt{1+t^2}}\left((\sqrt{p}w_0 - \sqrt{2(p-1)}w_{ll})f_l - \sum_{i\neq l}\sqrt{p-1}w_{il}f_i\right);
\end{aligned}$$

here we use the convention that $w_{il} = w_{li}$ if $i > l$. We have thus proved the following proposition.

Proposition 4.2. *The determinant of $A_{(te_2,w_0,w)}$ equals*

$$\frac{1}{(1+t^2)^{n/2}}\det\left(\sqrt{p}w_0I - \sqrt{p-1}\cdot\begin{bmatrix}\sqrt{2}w_{22} & w_{23} & \cdots & w_{2n}\\ w_{23} & \sqrt{2}w_{33} & \cdots & w_{3n}\\ \vdots & \vdots & & \vdots\\ w_{2n} & w_{3n} & \cdots & \sqrt{2}w_{nn}\end{bmatrix}\right).$$

We denote the $(n-1)\times(n-1)$-matrix by $C(w_1)$.

4.4. **The value of I.** We can now formulate our theorem for symmetric tensors.

Proposition 4.3. *For a standard Gaussian random symmetric tensor $v \in S^p\mathbb{R}^n$ (relative to the Bombieri norm) the expected number of critical points of d_v on the manifold of non-zero symmetric tensors of rank one equals*

$$\frac{\sqrt{\pi}}{2^{(n-1)/2}\Gamma(\frac{n}{2})}\mathbb{E}(|\det(\sqrt{p}w_0I - \sqrt{p-1}C(w_1))|),$$

where w_0 and the entries of w_1 are independent and $\sim\mathcal{N}(0,1)$.

Proof. Combining the results from the previous subsections, we find

$$
\begin{aligned}
I = &\frac{1}{(2\pi)^{\dim V/2}}\mathrm{Vol}(S^{n-2}) \\
&\cdot \int_0^\infty \int_{-\infty}^\infty \int_W |\det(\sqrt{p}w_0 I - \sqrt{p-1}C(w_1))| e^{-\frac{w_0^2+||w||^2}{2}} \frac{t^{n-2}}{(1+t^2)^{n/2}} \mathrm{d}w\mathrm{d}w_0\mathrm{d}t.
\end{aligned}
$$

Here, like in the ordinary tensor case, we have used that the function $F(u)$ in the definition of I is $\mathrm{O}(e_1^\perp)$-invariant. Now plug in

$$
\int_0^\infty \frac{t^{n-2}}{(1+t^2)^{\frac{n}{2}}}\mathrm{d}t = \frac{\sqrt{\pi}}{2}\frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})} \text{ and } \mathrm{Vol}(S^{n-2}) = \frac{2\pi^{\frac{n-1}{2}}}{\Gamma(\frac{n-1}{2})}
$$

to find that I equals

$$
\frac{1}{2^{\dim V/2}\pi^{(\dim V-n)/2}\Gamma(\frac{n}{2})} \cdot \int_{-\infty}^\infty \int_W |\det(\sqrt{p}w_0 I - \sqrt{p-1}C(w_1))| e^{-\frac{w_0^2+||w||^2}{2}} \mathrm{d}w\mathrm{d}w_0.
$$

Finally, we can factor out the part of the integral concerning w', which lives in a space of dimension $\dim V - 1 - (n-1) - n(n-1)/2 = \dim V - n(n+1)/2$. As a consequence, we need only integrate over the space W_1 where w_1 lives, and have to multiply by a suitable power of 2π:

$$
\begin{aligned}
I =& \frac{1}{2^{n(n+1)/4}\pi^{n(n-1)/4}\Gamma(\frac{n}{2})} \\
&\cdot \int_{-\infty}^\infty \int_{W_1} |\det(\sqrt{p}w_0 I - \sqrt{p-1}C(w_1))| e^{-\frac{w_0^2+||w_1||^2}{2}} \mathrm{d}w_1\mathrm{d}w_0 \\
=& \frac{\sqrt{\pi}}{2^{(n-1)/2}\Gamma(\frac{n}{2})}\mathbb{E}(|\det(\sqrt{p}w_0 I - \sqrt{p-1}C(w_1))|)
\end{aligned}
$$

as desired. □

4.5. **Further dimension reduction.** Since the matrix C from Proposition 4.3 is just $\sqrt{2}$ times a random matrix from the standard Gaussian orthogonal ensemble, and in particular has an orthogonally invariant probability density, we can further reduce the dimension of the integral, as follows.

Proof of Theorem 1.5. First we denote the diagonal entries of C

$$
\tilde{w}_{ii} := \sqrt{2}w_{ii},\ i = 2,\dots,n
$$

Then the joint density function of the random matrix C equals

$$
f_{n-1}(\tilde{w}_{ii}, w_{ij}) := \frac{1}{2^{(n-1)/2}\cdot(2\pi)^{n(n-1)/4}} e^{-(\tilde{w}_{22}^2+\cdots+\tilde{w}_{nn}^2)/4-\sum_{2\leqslant i<j\leqslant n} w_{ij}^2/2}.
$$

This function is invariant under conjugating C with an orthogonal matrix, and as a consequence, the joint density of the ordered tuple $(\lambda_2 \leqslant \ldots \leqslant \lambda_n)$ of eigenvalues of C equals

$$
Z(n-1)f_{n-1}(\Lambda)\prod_{i<j}(\lambda_j - \lambda_i),
$$

(see [Mui82, Theorem 3.2.17][1]). Here Λ is the diagonal matrix with $\lambda_2, \ldots, \lambda_n$ on the diagonal, and

$$Z(n-1) = \frac{\pi^{n(n-1)/4}}{\prod_{i=1}^{n-1} \Gamma(i/2)}.$$

Consequently, we have

$$I = \frac{\sqrt{\pi}}{2^{(n-1)/2}\Gamma(\frac{n}{2})} \int\limits_{\lambda_2 \leqslant \ldots \leqslant \lambda_n} \int\limits_{-\infty}^{\infty} \left(\prod_{i=2}^{n} |\sqrt{p} w_0 - \sqrt{p-1}\lambda_i| \right) \left(\prod_{i<j} (\lambda_j - \lambda_i) \right)$$
$$\cdot Z(n-1) f_{n-1}(\Lambda) \left(\frac{1}{\sqrt{2\pi}} e^{-w_0^2/2} \right) \mathrm{d}w_0 \mathrm{d}\lambda_2 \cdots \mathrm{d}\lambda_n.$$
$$= \frac{1}{2^{(n^2+3n-2)/4} \prod_{i=1}^{n} \Gamma(i/2)} \int\limits_{\lambda_2 \leqslant \ldots \leqslant \lambda_n} \int\limits_{-\infty}^{\infty} \left(\prod_{i=2}^{n} |\sqrt{p} w_0 - \sqrt{p-1}\lambda_i| \right)$$
$$\cdot \left(\prod_{i<j} (\lambda_j - \lambda_i) \right) e^{-w_0^2/2 - \sum_{i=2}^{n} \lambda_i^2/4} \mathrm{d}w_0 \mathrm{d}\lambda_2 \cdots \mathrm{d}\lambda_n,$$

as required. □

4.6. **The cone over the rational normal curve.** In the case where $n = 2$, the integral from Theorem 1.5 is over a 2-dimensional space and can be computed in closed form.

Theorem 4.4. *For $n = 2$ the number of critical points in Theorem 1.5 equals $\sqrt{3p-2}$.*

A slightly different computation yielding this result can be found in [DHO+16].

4.7. **Veronese embeddings of the projective plane.** In the case where $n = 3$, the integral from Theorem 1.5 gives the number of critical points to the cone over the p-th Veronese embedding of the projective plane. In this case the integral can be computed in closed form, using symbolic integration in `Mathematica` we have the following result.

Theorem 4.5. *For $n = 3$ the number of critical points in Theorem 1.5 equals*

$$1 + 4 \cdot \frac{p-1}{3p-2} \sqrt{(3p-2)\cdot(p-1)}.$$

We do not know whether a similar closed formula exists for higher values of n.

4.8. **Symmetric matrices.** In Example 1.6 we saw that the case where $p = 2$ concerns rank-one approximations to symmetric matrices, and that the average number of critical points is n. We now show that the integral above also yields n. Here we have

$$I = \frac{\sqrt{\pi}}{2^{(n-1)/2}\Gamma(\frac{n}{2})} \int\limits_{\lambda_2 \leqslant \ldots \leqslant \lambda_n} \int\limits_{-\infty}^{\infty} \left(\prod_{i=2}^{n} |\sqrt{2} w_0 - \lambda_i| \right) \left(\prod_{i<j} (\lambda_j - \lambda_i) \right)$$
$$\cdot Z(n-1) f_{n-1}(\Lambda) \left(\frac{1}{\sqrt{2\pi}} e^{-w_0^2/2} \right) \mathrm{d}w_0 \mathrm{d}\lambda_2 \cdots \mathrm{d}\lambda_n.$$

[1]The theorem there concerns the positive-definite case, but is true for orthogonally invariant density functions on general symmetric matrices.

Now set $\lambda_1 := \sqrt{2}w_0$. Then the inner integral over λ_1 splits into n integrals, according to the relative position of λ_1 among $\lambda_2 \leqslant \cdots \leqslant \lambda_n$. Moreover, these integrals are all equal. Hence we find

$$\begin{aligned} I = n&\frac{\sqrt{\pi}}{2^{(n-1)/2}\Gamma(\frac{n}{2})} \int_{\lambda_1\leqslant\ldots\leqslant\lambda_n} \left(\prod_{1\leqslant i<j\leqslant n} (\lambda_j - \lambda_i)\right) \\ &\cdot Z(n-1) \cdot \frac{1}{2^{n/2}\cdot(2\pi)^{(n(n-1)+2)/4}} e^{-(\lambda_1^2+\cdots+\lambda_n^2)/4} \mathrm{d}\lambda_1\cdots\mathrm{d}\lambda_n \\ = n&\frac{\sqrt{\pi}}{2^{(n-1)/2}\Gamma(\frac{n}{2})} \int_{\lambda_1\leqslant\ldots\leqslant\lambda_n} \left(\prod_{1\leqslant i<j\leqslant n} (\lambda_j - \lambda_i)\right) \\ &\cdot Z(n-1)\cdot f_n(\mathrm{diag}(\lambda_1,\ldots,\lambda_n))\cdot(2\pi)^{(n-1)/2}\mathrm{d}\lambda_1\cdots\mathrm{d}\lambda_n. \end{aligned}$$

Now, again by [Mui82, Theorem 3.2.17], the integral of $\prod_{1\leqslant i<j\leqslant n}(\lambda_j - \lambda_i)\cdot f_n$ equals $1/Z(n)$. Inserting this into the formula yields $I = n$.

5. Values

In this section we record some values of the expressions in Theorem 1.1 and 1.5.

5.1. **Ordinary tensors.** Below is a table of expected numbers of critical rank-one approximations to a Gaussian tensor, computed from Theorem 1.1. We also include the count over $\mathbb{C}$ from [FO12]. Unfortunately, the dimensions of the integrals from Theorem 1.1 seem to prevent accurate computation numerically, at least with all-purpose software such as `Mathematica`. Instead, we have estimated these integrals as follows: for some initial value I (we took $I = 15$), take 2^I samples of C from the multivariate standard normal distribution, and compute the average absolute determinant. Repeat with a new sample of size 2^I, and compare the absolute difference of the two averages divided by the first estimate. If this relative difference is $< 10^{-4}$, then stop. If not, then group the current 2^{I+1} samples together, sample another 2^{I+1}, and perform the same test. Repeat this process, doubling the sample size in each step, until the relative difference is below 10^{-4}. Finally, multiply the last average by the constant in front of the integral in Theorem 1.1. We have not computed a confidence interval for the estimate thus computed, but repetitions of this procedure suggest that the first three computed digits are correct; we give one more digit below.

Tensor format	average count over $\mathbb{R}$	count over $\mathbb{C}$
$n \times m$	$\min(n, m)$	$\min(n, m)$
$2^3 = 2 \times 2 \times 2$	4.287	6
2^4	11.06	24
2^5	31.56	120
2^6	98.82	720
2^7	333.9	5040
2^8	$1.206 \cdot 10^3$	40320
2^9	$4.611 \cdot 10^3$	362880
2^{10}	$1.843 \cdot 10^4$	3628800
$2 \times 2 \times 3$	5.604	8
$2 \times 2 \times 4$	5.556	8
$2 \times 2 \times 5$	5.536	8
$2 \times 3 \times 3$	8.817	15
$2 \times 3 \times 4$	10.39	18
$2 \times 3 \times 5$	10.28	18
$3 \times 3 \times 3$	16.03	37
$3 \times 3 \times 4$	21.28	55
$3 \times 3 \times 5$	23.13	61

Except in some small cases, we do not expect that there exists a closed form expression for $\mathbb{E}(|\det(C)|)$. However, asymptotic results on expected absolute determinants such as those in [TV12] should give asymptotic results for the counts in Theorems 1.1 and 1.5, and it would be interesting to compare these with the count over $\mathbb{C}$.

From [FO12] we know that the count for ordinary tensors stabilizes for $n_p - 1 \geqslant \sum_{i=1}^{p-1}(n_i - 1)$, i.e., beyond the *boundary format* [GKZ94, Chapter 14], where the variety dual to the variety of rank-one tensors ceases to be a hypersurface. We observe a similar behavior experimentally for the average count according to Theorem 1.1, although the count seems to *decrease* slightly rather than to *stabilize*. It would be nice to prove this behavior from our formula, but even better to give a geometric explanation both over $\mathbb{R}$ and over $\mathbb{C}$.

5.2. **Symmetric tensors.** The following table contains the average number of rank-one tensor approximations to $S^p\mathbb{R}^n$ according to Theorem 1.5 (on the left). The integrals here are over a much lower-dimensional domain than in the previous section, and they *can* be evaluated accurately with `Mathematica`. On the right we list the corresponding count over $\mathbb{C}$. By [FO12, Theorem 12] these values are simply $1 + (p-1) + \cdots + (p-1)^{n-1}$.

$p\backslash n$	1	2	3	4
1	1	1	1	1
2	1	2	3	4
3	1	$\sqrt{7}$	$1+4\cdot\frac{2}{7}\cdot\sqrt{7\cdot 2}$	9.3951
4	1	$\sqrt{10}$	$1+4\cdot\frac{3}{10}\cdot\sqrt{10\cdot 3}$	16.254
5	1	$\sqrt{13}$	$1+4\cdot\frac{4}{13}\cdot\sqrt{13\cdot 4}$	24.300
6	1	$\sqrt{16}$	$1+4\cdot\frac{5}{16}\cdot\sqrt{16\cdot 5}$	33.374
7	1	$\sqrt{19}$	$1+4\cdot\frac{6}{19}\cdot\sqrt{19\cdot 6}$	43.370
8	1	$\sqrt{22}$	$1+4\cdot\frac{7}{22}\cdot\sqrt{22\cdot 7}$	54.211
9	1	$\sqrt{25}$	$1+4\cdot\frac{8}{25}\cdot\sqrt{25\cdot 8}$	65.832
10	1	$\sqrt{28}$	$1+4\cdot\frac{9}{28}\cdot\sqrt{28\cdot 9}$	78.185

$p\backslash n$	1	2	3	4
1	1	1	1	1
2	1	2	3	4
3	1	3	7	15
4	1	4	13	40
5	1	5	21	85
6	1	6	31	156
7	1	7	43	259
8	1	8	57	400
9	1	9	73	585
10	1	10	91	820

References

[BCMT10] Jerome Brachat, Pierre Comon, Bernard Mourrain, and Elias Tsigaridas. Symmetric tensor decomposition. *Linear Algebra Appl.*, 433(11–12):1851–1872, 2010.

[BW08] Femke van Belzen and Siep Weiland. Diagonalization and low-rank appromixation of tensors: a singular value decomposition approach. In *Proceedings 18th International Symposium on Mathematical Theory of Networks & Systems (MTNS), 28 July - 1 August 2008, Blacksburg, Virginia*, Blacksburg, Virginia, USA, 2008. MTNS.

[BW09] Femke van Belzen and Siep Weiland. Approximation of nD systems using tensor decompositions. In *Proceedings of the International Workshop on Multidimensional (nD) Systems, June 29th - July 1st, 2009, Thessaloniki, Greece*, pages 1–8, Piscataway, 2009. IEEE Service Center.

[BWG07] F. van Belzen, S. Weiland, and J. de Graaf. Singular value decompositions and low rank approximations of multi-linear functionals. In *Proceedings of the 46th Conference on Decision and Control (CDC 2007) 12-14 December 2007, New Orleans, Louisiana, USA. (pp. 3751-3756)*, Piscataway, New Jersey, USA, 2007. IEEE.

[CGLM08] Pierre Comon, Gene Golub, Lek-Heng Lim, and Bernard Mourrain. Symmetric tensors and symmetric tensor rank. *SIAM J. Matrix Anal. Appl.*, 30(3):254–1279, 2008.

[CS13] Dustin Cartwright and Bernd Sturmfels. The number of eigenvalues of a tensor. *Linear Algebra Appl.*, 438(2):942–952, 2013.

[DDV00] Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. On the best rank-1 and rank-$(R_1, R_2, ..., R_N)$ approximation of higher-order tensors. *SIAM J. Matrix Anal. Appl.*, 21(4):1324–1342, 2000.

[DHO+16] Jan Draisma, Emil Horobet, Giorgio Ottaviani, Bernd Sturmfels, and Rekha R. Thomas. The Euclidean distance degree of an algebraic variety. *Found. Comput. Math.*, 2016. To appear; `arXiv:1309.0049`.

[DL08a] Lieven De Lathauwer. Decompositions of a higher-order tensor in block terms. I: Lemmas for partitioned matrices. *SIAM J. Matrix Anal. Appl.*, 30(3):1022–1032, 2008.

[DL08b] Lieven De Lathauwer. Decompositions of a higher-order tensor in block terms. II: Definitions and uniqueness. *SIAM J. Matrix Anal. Appl.*, 30(3):1033–1066, 2008.

[DLN08] Lieven De Lathauwer and Dimitri Nion. Decompositions of a higher-order tensor in block terms. III: Alternating least squares algorithms. *SIAM J. Matrix Anal. Appl.*, 30(3):1067–1083, 2008.

[dSL08] Vin de Silva and Lek-Heng Lim. Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM J. Matrix Anal. Appl.*, 30(3):1084–1127, 2008.

[FO12] Shmuel Friedland and Giorgio Ottaviani. The number of singular vector tuples and uniqueness of best rank one approximation of tensors. *Found. Comp. Math.*, 2012. To appear, `arXiv:1210.8316`.

[GKZ94] Israel M. Gelfand, Mikhail M. Kapranov, and Andrei V. Zelevinsky. *Discriminants, resultants, and multidimensional determinants*. Mathematics: Theory & Applications. Birkhäuser, Boston, MA, 1994.

[Hås90] Johan Håstad. Tensor rank is NP-complete. *J. Algorithms*, 11(4):644–654, 1990.

[HL13] Christopher J. Hillar and Lek-Heng Lim. Most tensor problems are NP-hard. *J. ACM*, 60(6):39, 2013.

[Hor15] Emil Horobeţ. The data singular and the data isotropic loci for affine cones. 2015. Preprint; `arxiv:1507.02923`.

[IAvHdL11] Mariya Ishteva, P.-A. Absil, Sabine van Huffel, and Lieven de Lathauwer. Best low multilinear rank approximation of higher-order tensors, based on the Riemannian trust-region scheme. *SIAM J. Matrix Anal. Appl.*, 32(1):115–135, 2011.

[Lim05] Lek-Heng Lim. Singular values and eigenvalues of tensors: a variational approach. In *Proceedings of the IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP '05)*, volume 1, pages 129–132, 2005.

[Mui82] Robb J. Muirhead. Aspects of multivariate statistical theory. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley & Sons, Inc. XIX, 673 p. (1982)., 1982.

[OO13] Luke Oeding and Giorgio Ottaviani. Eigenvectors of tensors and algorithms for Waring decomposition. *J. Symb. Comput.*, 54:9–35, 2013.

[Rou07] Alain Rouault. Asymptotic behavior of random determinants in the Laguerre, Gram and Jacobi ensembles. *ALEA, Lat. Am. J. Probab. Math. Stat.*, 3:181–230, 2007.

[TV12] Terence Tao and Van Vu. A central limit theorem for the determinant of a Wigner matrix. *Adv. Math.*, 231(1):74–101, 2012.

Department of Mathematics and Computer Science, Technische Universiteit Eindhoven, P.O. Box 513, 5600 MB Eindhoven, The Netherlands, and Vrije Universiteit Amsterdam and Centrum Wiskunde en Informatica, Amsterdam, The Netherlands.

E-mail address: `j.draisma@tue.nl`

Department of Mathematics and Computer Science, Technische Universiteit Eindhoven, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.

E-mail address: `e.horobet@tue.nl`