On the measure of anchored Gaussian simplices, with applications to multivariate medians

DAVY PAINDAVEINE[1,2]

[1]*ECARES and Département de Mathématique, Université libre de Bruxelles, Avenue F.D. Roosevelt, 50, ECARES, CP114/04, B-1050, Brussels, Belgium. E-mail: dpaindav@ulb.ac.be*
[2]*Toulouse School of Economics, Université Toulouse Capitole, 1, Esplanade de l'Université, 31080 Toulouse Cedex 06, France*

We consider anchored Gaussian ℓ-simplices in the d-dimensional Euclidean space, that is, simplices with one fixed vertex $y \in \mathbb{R}^d$ and the remaining vertices $X_1, \ldots, X_\ell$ randomly sampled from the d-variate standard normal distribution. We determine the distribution of the measure of such simplices for any d, any ℓ, and any anchor point y, which is of interest, e.g., when studying the asymptotic behaviour of U-statistics based on such simplex measures. We provide two proofs of the results. The first one is short but is not self-contained as it crucially relies on a technical result for non-central Wishart distributions. The second one is a simple and self-contained proof, that also provides some geometric insight on the results. Quite nicely, variations on this second argument reveal intriguing distributional identities on products of central and non-central chi-square distributions with Beta-distributed non-centrality parameters. We independently establish these distributional identities by making use of Mellin transforms. Beyond the aforementioned use to study the asymptotic behaviour of some U-statistics, our results do find natural applications in the context of robust location estimation, as we illustrate by considering a class of simplex-based multivariate medians that contains the celebrated spatial median and Oja median as special cases. Throughout, our results are confirmed by numerical experiments.

Keywords: Distributional identities; Mellin transforms; Multivariate medians; Oja median; Random simplices; Spatial median; Stochastic geometry

1. Introduction

There is an abundant literature studying the measure of random simplices; see, among many others Miles [27], Ruben [43], Ruben and Miles [44], Anderson [2], Mathai [26], Pivovarov [42], Grote, Kabluchko and Thäle [14], and the references therein. Throughout the paper, Simpl$(x_1, \ldots, x_{\ell+1})$, with $\ell \leq d$, will stand for the simplex with vertices $x_1, \ldots, x_{\ell+1} \in \mathbb{R}^d$, that is, the convex hull of these $\ell + 1$ locations in $\mathbb{R}^d$, and $m_\ell(\text{Simpl}(x_1, \ldots, x_{\ell+1}))$ will denote its ℓ-measure (length for $\ell = 1$, area for $\ell = 2$, etc.) Most of the focus in the literature has been on the ℓ-measure

$$W_{\ell+1,d} := m_\ell(\text{Simpl}(X_1, \ldots, X_{\ell+1})), \qquad \ell \in \{1, \ldots, d\}, \tag{1.1}$$

of random simplices based on a random sample of size $\ell + 1$ from some given distribution over $\mathbb{R}^d$ and on the ℓ-measure

$$V_{\ell,d} := m_\ell(\text{Simpl}(X_1, \ldots, X_\ell, 0)), \qquad \ell \in \{1, \ldots, d\}, \tag{1.2}$$

of simplices "anchored" at the origin of $\mathbb{R}^d$. In the standard normal case, for instance, the distributions of $W_{\ell+1,d}$ and $V_{\ell,d}$ are known (see Corollary 2.1 below) and, remarkably, these are such that $W_{\ell+1,d} =_{\mathcal{D}} \sqrt{\ell + 1} \, V_{\ell,d}$ (throughout, $=_{\mathcal{D}}$ will denote equality in distribution); see, e.g., Grote, Kabluchko and Thäle [14]. The corresponding derivations are quite technical and in particular do not provide geometric insight for this relation between both random ℓ-measures. Such insight could be

obtained by exploiting the connection (Baryshnikov and Vitale [5]) between random projections of a regular simplex onto a fixed subspace and random Gaussian samples in that subspace, but this approach would thus require quite involved results from stochastic geometry (see also Kabluchko and Zaporozhets [20] and Paouris, Pivovarov and Valettas [40]). In this paper, we will rather obtain such geometric insight through simple, independent, derivations. This will actually be achieved as a corollary of our interest in the measure of *anchored* Gaussian simplices. More precisely, we will consider the ℓ-measure

$$W_{\ell,d}(y) := m_\ell(\mathrm{Simpl}(X_1, \ldots, X_\ell, y)), \qquad \ell \in \{1, \ldots, d\}, \tag{1.3}$$

of random simplices anchored at a given $y \in \mathbb{R}^d$, where $X_1, \ldots, X_\ell$ are mutually independent standard normal d-vectors. In particular, we will establish the following result.

Theorem 1.1. *For any positive integer d, any $\ell \in \{1, \ldots, d\}$, and any $y \in \mathbb{R}^d$,*

$$W_{\ell,d}(y) =_{\mathcal{D}} \frac{1}{\ell!} Q_d(\ell \|y\|^2) \prod_{i=2}^{\ell} Q_{d+1-i}, \tag{1.4}$$

where $Q_d^2(\ell\|y\|^2)$ is non-central chi-square with d degrees of freedom and non-centrality parameter $\ell\|y\|^2$ (throughout, $\|x\| = \sqrt{x'x}$ is the Euclidean norm of x), Q_m^2 is chi-square with m degrees of freedom, and where the factors in (1.4) are mutually independent.

Such anchored random simplices have receive little attention (if any) in the literature, beyond the very particular case obtained with $y = 0$ in (1.2). This is quite surprising since the distribution of the ℓ-measure of such anchored simplices naturally arises, e.g., when studying the asymptotic distribution of U-statistics based on simplex measures (incidentally, U-statistics were also used in Paouris, Pivovarov and Zinn [41] when deriving a central limit theorem for the volume of the projection of a high-dimensional cube onto a random subspace of fixed dimension). To elaborate on this, let $X_1, \ldots, X_n$ be a random sample from the d-variate standard normal distribution and consider the U-statistic

$$U_{n,\ell,r} = \frac{1}{\binom{n}{\ell+1}} \sum_{1 \le i_1 < \cdots < i_\ell \le n} m_\ell^r(\mathrm{Simpl}(X_{i_1}, \ldots, X_{i_{\ell+1}})), \tag{1.5}$$

where r is a positive real number. This is a U-statistic involving a symmetric kernel with a finite variance, so that its asymptotic behaviour can in principle be derived using the standard theory of U-statistics. In particular, the consistency result in page 190 of Serfling [46] entails that U_n converges almost surely to $\mathrm{E}[W_{\ell+1,d}^r]$, a value that can be obtained from the known distribution of $W_{\ell+1,d}$ in (1.1). More importantly, writing $\to_{\mathcal{D}}$ for weak convergence, the classical U-statistics theory then also yields that, as n diverges to infinity,

$$\sqrt{n}\big(U_{n,\ell,r} - \mathrm{E}[W_{\ell+1,d}^r]\big) = \frac{\ell+1}{\sqrt{n}} \sum_{i=1}^{n} h_{\ell,r}(X_i) + o_{\mathrm{P}}(1) \to_{\mathcal{D}} \mathcal{N}\big(0, \mathrm{E}[h_{\ell,r}^2(X_1)]\big),$$

with $h_{\ell,r}(y) := \mathrm{E}[m_\ell^r(\mathrm{Simpl}(X_1, \ldots, X_\ell, y))] - \mathrm{E}[W_{\ell+1,d}^r]$; see, e.g., Theorem 12.3 in van der Vaart [48]. Clearly, understanding the asymptotic behaviour of U_n thus requires the study of anchored random simplices we present in this work. In particular, our results will entail that
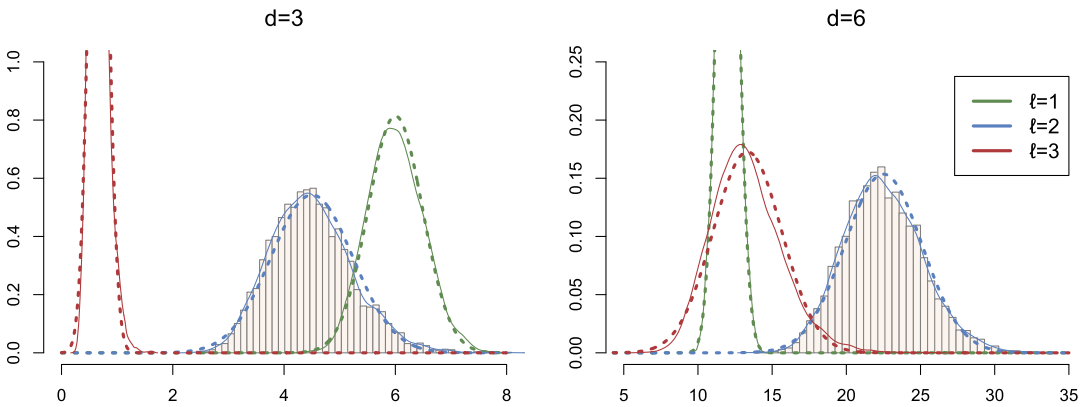
Figure 1. (Left:) Kernel density estimates (solid lines), for each $\ell \in \{1, 2, 3\}$, of the values of $U_{n,\ell,2}$ in (1.5) obtained from 5 000 mutually independent random samples of size $n = 100$ from the $(d = 4)$-variate standard normal distribution (kernel density estimates are obtained by using the R command `density` with default parameter values). For $\ell = 2$, a histogram of raw values is provided. The respective asymptotic densities, obtained from (1.6), are also plotted (dashed lines). (Right:) The corresponding results in dimension $d = 10$.

$$\sqrt{n}\left(U_{n,\ell,2} - \frac{(\ell+1)}{\ell!}\binom{d}{\ell}\right) = \frac{\ell+1}{\sqrt{n}(\ell-1)!d}\binom{d}{\ell}\sum_{i=1}^{n}(\|X_i\|^2 - d) + o_P(1)$$

$$\to_{\mathcal{D}} \mathcal{N}\left(0, \frac{2(\ell+1)^2}{((\ell-1)!)^2 d}\binom{d}{\ell}^2\right) \tag{1.6}$$

as n diverges to infinity (the asymptotic distribution of $U_{n,\ell,1}$ can be obtained explicitly from our results, too, but formulae are more complicated). To illustrate the asymptotic normality result in (1.6), we generated, for each $d \in \{4, 6\}$, a collection of 5 000 mutually independent random samples of size $n = 100$ from the d-variate standard normal distribution. For both values of d and for each $\ell \in \{1, 2, 3\}$, Figure 1 plots kernel density estimates of the resulting 5 000 values of $U_{n,\ell,2}$, as well as the theoretical Gaussian fits resulting from (1.6). Obviously, this numerical exercise fully supports our asymptotic results.

The previous example shows that anchored simplices naturally arise when considering some U-statistics. In this paper, we will primarily show the relevance of our results in a much more involved framework associated with robust location estimation. In the univariate case $d = 1$, the most robust location functional is the median μ_P of the probability measure P at hand, that can be characterized as the location $\mu \in \mathbb{R}$ minimizing the L_1 expected loss $E_P[|X_1 - \mu|] = E_P[m_1(\text{Simpl}(X_1, \mu))]$, where X_1 has distribution P. This may be naturally extended to an arbitrary dimension d by defining the ℓ-median of a probability measure P over $\mathbb{R}^d$ as

$$\mu_{\ell,P} := \arg\min_{\mu \in \mathbb{R}^d} E_P[m_\ell(\text{Simpl}(X_1, \ldots, X_\ell, \mu))],$$

where $X_1, \ldots, X_\ell$ form a random sample from P. This defines a collection of multivariate medians indexed by $\ell \in \{1, \ldots, d\}$, that is related to the one introduced in Vitale [49] by considering "intrinsic volumes" of suitable zonotopes. To the best of the author's knowledge, the properties of these

ℓ-medians have not been not investigated outside the particular cases $\ell = 1$ and $\ell = d$, for which these medians reduce to the celebrated *spatial median* (Brown [8], Chaudhuri [9], Möttönen, Nordhausen and Oja [30], Magyar and Tyler [24]) and *Oja median* (Oja [35], Hettmansperger, Möttönen and Oja [16], Ollila, Oja and Hettmansperger [39]), respectively. We will study the asymptotic behaviour of the sample version of these ℓ-medians, and, as we will show, our results on anchored simplices will be key to obtain in the Gaussian case explicit expressions of the influence function and asymptotic variance of these sample multivariate medians.

Above, we mainly focused on the aspects that motivated our study. We now provide the outline of the paper, which gives us the opportunity to be more specific about our contribution. In Section 2, we provide a first proof of Theorem 1.1 that relies on a technical result for non-central Wishart matrices. We also show how this theorem allows us to recover results for unanchored simplices available in the literature. In Section 3, we give an alternative, simple and self-contained, proof, that provides some geometric insights on the results. We also show that variations on this alternative proof reveal intriguing distributional identities on products of central and non-central chi-square distributions with Beta-distributed non-centrality parameters. In Section 4, we use Mellin transforms to establish independently these distributional identities. In Section 5, we define the multivariate medians described above and derive the asymptotic distribution of their sample versions. We then exploit our results on anchored Gaussian simplices to obtain explicit expressions of the influence function and asymptotic variance of these sample medians in the Gaussian case. We briefly discuss perspectives for future research in Section 6. While some of our proofs, that are part of our contribution, are to be found in the main body of the paper, proofs of some auxiliary results as well as the lengthy proofs dedicated to multivariate medians are given in a technical appendix.

2. A proof based on non-central Wishart matrices

In this section, we first prove Theorem 1.1 by making use of a technical result on non-central Wishart matrices, and we then explore consequences of this theorem. Before proceeding, we introduce some notation and recall some basic geometric facts. For $x_1, \ldots, x_\ell \in \mathbb{R}^d$, with $\ell \le d$, consider the parallelotope $\mathrm{Parall}(x_1, \ldots, x_\ell, 0) := \{\sum_{i=1}^{\ell} \lambda_i x_i : \lambda_1, \ldots, \lambda_\ell \in [0, 1]\}$. In the full-dimensional case $\ell = d$, it is well-known that $m_d(\mathrm{Parall}(x_1, \ldots, x_d, 0)) = |\det(x_1 \ldots x_d)|$. Measures of lower-dimensional simplices are then obtained as follows. Denoting as $x_{1:\ell}$ the $d \times \ell$ matrix $(x_1 \ldots x_\ell)$ and letting $\{v_{\ell+1}, \ldots, v_d\}$ be an arbitrary orthonormal basis of the orthogonal complement of the vector space spanned by $x_1, \ldots, x_\ell$,

$$m_\ell(\mathrm{Parall}(x_1, \ldots, x_\ell, 0)) = m_d(\mathrm{Parall}(x_1, \ldots, x_\ell, v_{\ell+1}, \ldots, v_d, 0)) = |\det(x_{1:\ell} \, v_{\ell+1} \, \ldots \, v_d)|$$

$$= \sqrt{\det((x_{1:\ell} \, v_{\ell+1} \, \ldots \, v_d)'(x_{1:\ell} \, v_{\ell+1} \, \ldots \, v_d))} = \sqrt{\det(x'_{1:\ell} x_{1:\ell})}. \tag{2.1}$$

Alternatively, denoting as $\pi_\perp(z; z_1, \ldots, z_k)$ the orthogonal projection of z onto the orthogonal complement of the vector space spanned by $z_1, \ldots, z_k$, orthogonalization also provides

$$m_\ell(\mathrm{Parall}(x_1, \ldots, x_\ell, 0)) = m_d(\mathrm{Parall}(x_1, \ldots, x_\ell, v_{\ell+1}, \ldots, v_d, 0)) = |\det(x_1 \, \ldots \, x_\ell \, v_{\ell+1} \, \ldots \, v_d)|$$

$$= |\det(x_1 \, \pi_\perp(x_2; x_1) \ldots \pi_\perp(x_\ell; x_1, \ldots, x_{\ell-1}) \, v_{\ell+1} \ldots v_d)| = \|x_1\| \prod_{i=2}^{\ell} \|\pi_\perp(x_i; x_1, \ldots, x_{i-1})\|. \tag{2.2}$$

Finally, an obvious combinatorial argument yields that the ℓ-measure of the corresponding simplex satisfies $m_\ell(\mathrm{Simpl}(x_1, \ldots, x_\ell, 0)) = (\ell!)^{-1} m_\ell(\mathrm{Parall}(x_1, \ldots, x_\ell, 0))$, hence can obtained from the above formulae.

We can now provide a first proof of Theorem 1.1.

Proof of Theorem 1.1. Let O be an arbitrary $d \times d$ orthogonal matrix such that $Oy = \|y\|e_{1d}$, where e_{1k} is the first vector of the canonical basis of $\mathbb{R}^k$. Since the common distribution of the X_i's is spherically symmetric about the origin in $\mathbb{R}^d$ (in the sense that $PX_1 =_{\mathcal{D}} X_1$ for any $d \times d$ orthogonal matrix P), we obtain

$$W_{\ell,d}(y) = m_\ell(\text{Simpl}(X_1 - y, \ldots, X_\ell - y, 0))$$

$$= m_\ell(\text{Simpl}(OX_1 - \|y\|e_{1d}, \ldots, OX_\ell - \|y\|e_{1d}, 0))$$

$$=_{\mathcal{D}} m_\ell(\text{Simpl}(X_1 - \|y\|e_{1d}, \ldots, X_\ell - \|y\|e_{1d}, 0)),$$

so that (2.1) yields

$$(\ell!)^2 W_{\ell,d}^2(y) =_{\mathcal{D}} \det(X_y' X_y), \quad \text{with } X_y := (X_1 - \|y\|e_{1d} \ \ldots \ X_\ell - \|y\|e_{1d}).$$

Pick then an arbitrary $\ell \times \ell$ orthogonal matrix U satisfying $U'1_\ell = -\sqrt{\ell}e_{1\ell}$; throughout, $1_k = (1, \ldots, 1)' \in \mathbb{R}^k$. We then have that $Z := X_y U = (X_1 \ldots X_\ell)U - \|y\|e_{1d}1_\ell'U =_{\mathcal{D}} (X_1 \ldots X_\ell) + \sqrt{\ell}\|y\|e_{1d}e_{1\ell}'$ is (d, ℓ)-variate matrix normal with mean $\sqrt{\ell}\|y\|e_{1d}e_{1\ell}'$ and covariance matrix $I_{d\ell}$. Recall that if the random $\ell \times d$ matrix T is normal with mean M and covariance $I_d \otimes \Sigma$, then $S = T'T$ is said to follow the non-central d-variate Wishart distribution with ℓ degrees of freedom, covariance matrix Σ and non-centrality parameter $\Omega = \Sigma^{-1}M'M$, which is denoted as $S \sim W_d(\ell, \Sigma, \Omega)$; see, e.g., Section 10.3 from Muirhead [32]. By definition, we thus have $Z'Z \sim \mathcal{W}_\ell(d, I_\ell, \ell\|y\|^2 e_{1\ell}e_{1\ell}')$. Therefore,

$$(\ell!)^2 W_{\ell,d}^2(y) =_{\mathcal{D}} \det\left((X_y U)' X_y U\right) = \det(Z'Z) =_{\mathcal{D}} Q_d^2(\ell\|y\|^2) \prod_{i=2}^{\ell} Q_{d+1-i}^2,$$

where we used Theorem 10.3.8 from Muirhead [32]. The result is proved. $\qquad \square$

For $\ell = 1$, it is trivial that $W_{\ell,d}^2(y) = \|X_1 - y\|^2 = \sum_{r=1}^{d}\{(X_1 - y)_r\}^2$ is non-central chi-square with d degrees of freedom and non-centrality parameter $\|y\|^2$, which is compatible with Theorem 1.1. The result for $\ell = d$ was obtained in Möttönen *et al.* [29] through an argument that also exploits results on non-central Wishart matrices (more precisely, this argument relies on Theorem 4.3 from Dahel and Giri [10]). To the best of our knowledge, the result is original for the other values of ℓ.

Theorem 1.1 allows us to derive the distribution of the ℓ-measures $W_{\ell+1,d}$ and $V_{\ell,d}$ of unanchored simplices and simplices anchored at the origin; see (1.1)–(1.2). We have the following result.

Corollary 2.1. *When based on mutually independent standard normal d-vectors,*

$$(i) \ W_{\ell+1,d} =_{\mathcal{D}} \frac{\sqrt{\ell+1}}{\ell!} \prod_{i=1}^{\ell} Q_{d+1-i} \quad and \quad (ii) \ V_{\ell,d} =_{\mathcal{D}} \frac{1}{\ell!} \prod_{i=1}^{\ell} Q_{d+1-i}$$

(so that $W_{\ell+1,d} =_{\mathcal{D}} \sqrt{\ell+1}\, V_{\ell,d}$), where Q_m^2 is chi-square with m degrees of freedom and the Q_m's are mutually independent.

The proof requires the following result (see the appendix for a proof).

Lemma 2.1. *Let d be a positive integer and λ be a non-negative real number. Let Q be a random variable that is chi-square with d degrees of freedom and R be a random variable that, conditional on Q, is non-central chi-square with d degrees of freedom and non-centrality parameter λQ (i.e., $Q \sim \chi_d^2$ and $R|[Q=q] \sim \chi_d^2(\lambda q)$ for any $q > 0$). Then $R =_{\mathcal{D}} \sqrt{\lambda + 1}\, Q$.*

Proof of Corollary 2.1. (i) Theorem 1.1 yields that, conditional on $X_{\ell+1}$,

$$W_{\ell+1,d}|X_{\ell+1} =_{\mathcal{D}} \frac{1}{\ell!} Q_d(\ell \|X_{\ell+1}\|^2) \prod_{i=2}^{\ell} Q_{d+1-i}.$$

Since $\|X_{\ell+1}\|^2 \sim \chi_d^2$, the result then follows from Lemma 2.1. (ii) The result is obtained by taking $y = 0$ in Theorem 1.1. $\qquad\qquad\square$

Let us consider some particular cases. Of course, we have $V_{1,d} = \|X_1\| \sim \chi_d$, whereas $W_{2,d} = \|X_1 - X_2\| \sim \sqrt{2}\chi_d$ (since $(X_1 - X_2)/\sqrt{2}$ is d-variate standard normal); both results are in line with Corollary 2.1. It is remarkable that, irrespective of ℓ and d, the distribution of $W_{\ell+1,d}$ and $V_{\ell,d}$ only differ by a scale factor (depending on ℓ only). The proof above, however, fails to provide any insight on why this holds true and on the value of the proportionality factor. This provides a motivation to derive the simple proof we give in Section 3 below.

Before providing this alternative proof, we illustrate numerically the results of this section. For various combinations of d and ℓ, we generated 100 000 mutually independent random samples $X_1, \ldots, X_{\ell+1}$ from the d-variate standard normal distribution. The top panels of Figure 2 provide kernel density estimates of the resulting 100 000 values of $W_\ell(e_{1d})$ (measures of ℓ-simplices anchored at the first vector of the canonical basis of $\mathbb{R}^d$), whereas the bottom panels do the same for $W_{\ell+1,d}$ (measures of unanchored ℓ-simplices). In both cases, we also plot kernel density estimates obtained from random variables generated according to the corresponding distributions in Theorem 1.1 and Corollary 2.1(i). For both types of simplices and for all combinations of d and ℓ, Figure 2 clearly confirms our theoretical results.

3. A self-contained proof

We now present a simple proof of Theorem 1.1 that will not require technical results for non-central Wishart matrices. The following preliminary result is key to this alternative proof.

Lemma 3.1. *Let d be a positive integer and $\ell \in \{1, \ldots, d\}$. Let $X_1, \ldots, X_\ell$ be mutually independent standard normal d-vectors. Let $A = (A_1 \ldots A_\ell)$ and $B = (B_1 \ldots B_\ell)$ be $d \times \ell$ deterministic matrices sharing the same singular values. Then, $m_\ell(\mathrm{Parall}(X_1 + A_1, \ldots, X_\ell + A_\ell, 0)) =_{\mathcal{D}} m_\ell(\mathrm{Parall}(X_1 + B_1, \ldots, X_\ell + B_\ell, 0))$.*

Proof of Lemma 3.1. Since A and B share the same singular values, there exist $d \times d$ orthogonal matrices O_A, O_B and $\ell \times \ell$ orthogonal matrices U_A, U_B such that $A = O_A D U_A'$ and $B = O_B D U_B'$, where the only non-zero entries in the $d \times \ell$ matrix D are the common singular values D_{rr}, $r = 1, \ldots, \ell$, of A and B (some of these singular values may be zero). Thus, $B = OAU'$, where $O = O_B O_A'$
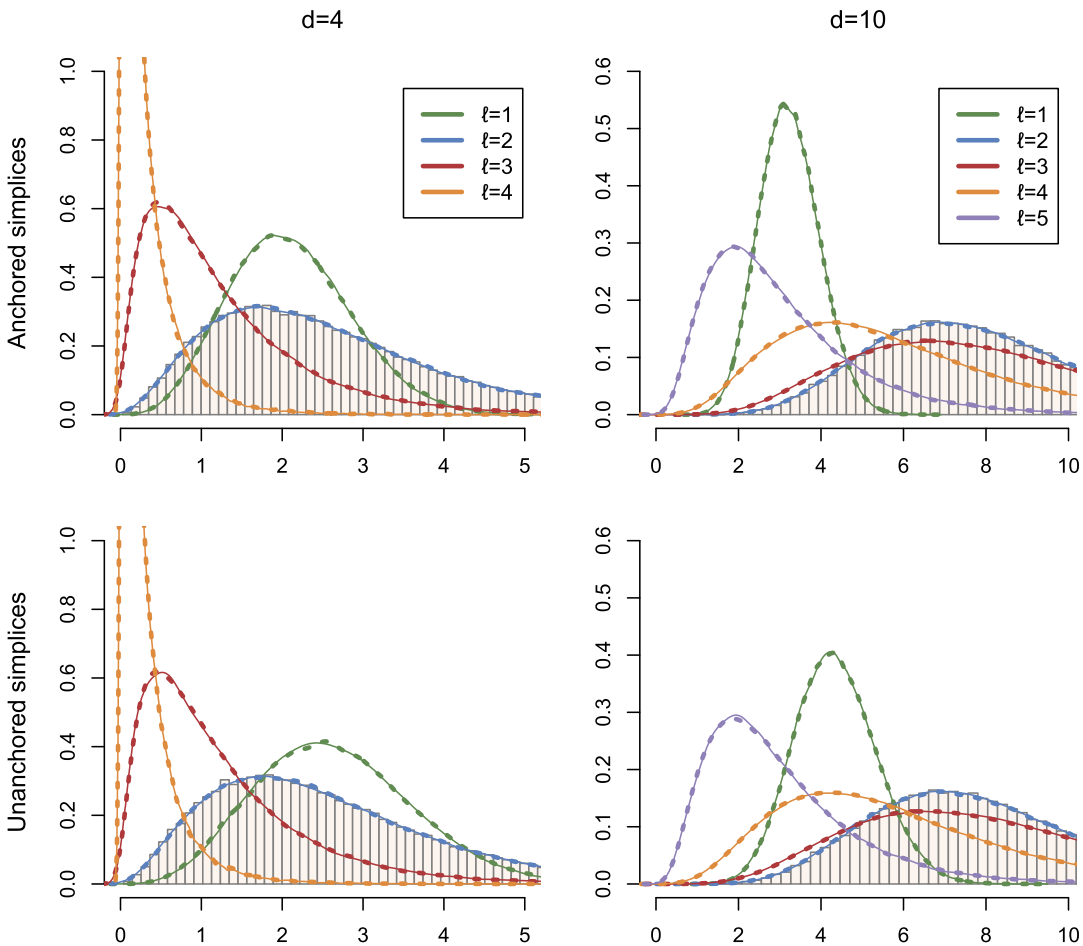
Figure 2. (Top left:) Kernel density estimates (solid lines), for each $\ell \in \{1, 2, 3, 4\}$, of the values of $W_\ell(e_{1d})$ in (1.3) obtained from 100 000 mutually independent random samples $X_1, \ldots, X_\ell$ from the $(d = 4)$-variate standard normal distribution (kernel density estimates are obtained by using the R command `density` with default parameter values). For $\ell = 2$, a histogram of raw values is provided. Kernel density estimates obtained, still for each $\ell \in \{1, 2, 3, 4\}$, from a collection of 100 000 mutually independent random variables generated according to the righthand side of (1.4) are also plotted (dashed lines). (Top right:) The same quantities in dimension $d = 10$, where the value $\ell = 5$ is also considered. (Bottom left and right:) The corresponding results for the measure $W_{\ell+1}$ of unanchored Gaussian simplices, where dashed lines are obtained from the distribution in Corollary 2.1(i).

is $d \times d$ orthogonal and $U = U_B U_A'$ is $\ell \times \ell$ orthogonal. Letting $X = (X_1 \ldots X_\ell)$, (2.1) then yields

$$m_\ell^2(\text{Parall}(X_1 + A_1, \ldots, X_\ell + A_\ell, 0)) = \det((X + A)'(X + A))$$

$$= \det(U(X + A)'O'O(X + A)U') = \det((OXU' + B)'(OXU' + B))$$

$$=_{\mathcal{D}} \det((X + B)'(X + B)) = m_\ell^2(\text{Parall}(X_1 + B_1, \ldots, X_\ell + B_\ell, 0)),$$

which, since these ℓ-measures are non-negative, establishes the result. $\qquad \square$

We can now provide our simple proof of Theorem 1.1.

Proof of Theorem 1.1. Proceeding as in the proof from Section 2, we first write

$$W_{\ell,d}(y) =_{\mathcal{D}} m_\ell(\mathrm{Simpl}(X_1 - \|y\|e_{1d}, \ldots, X_\ell - \|y\|e_{1d}, 0))$$

$$= \frac{1}{\ell!} m_\ell(\mathrm{Parall}(X_1 - \|y\|e_{1d}, \ldots, X_\ell - \|y\|e_{1d}, 0)).$$

Let $A = -\|y\|e_{1d}1'_\ell$ and $B = \sqrt{\ell}\|y\|e_{1d}e'_{1\ell}$. Note that the only non-zero eigenvalue of $A'A = \|y\|^2 1_\ell 1'_\ell$ and $B'B = \ell\|y\|^2 e_{1\ell}e'_{1\ell}$ is $\ell\|y\|^2$ in both cases. Thus, A and B share the same singular values, so that Lemma 3.1 yields

$$W_{\ell,d}(y) =_{\mathcal{D}} \frac{1}{\ell!} m_\ell(\mathrm{Parall}(X_1 + \sqrt{\ell}\|y\|e_{1d}, X_2, \ldots, X_\ell, 0)).$$

Applying (2.2), we thus have

$$W_{\ell,d}(y) =_{\mathcal{D}} \frac{1}{\ell!} \|X_1 + \sqrt{\ell}\|y\|e_{1d}\| \left(\prod_{i=2}^{\ell} \|\pi_\perp(X_i; X_1 + \sqrt{\ell}\|y\|e_{1d}, X_2, \ldots, X_{i-1})\| \right) \quad (3.1)$$

$$=: \frac{1}{\ell!} T_1 \left(\prod_{i=2}^{\ell} T_i \right),$$

where T_2 is of course associated with the projection onto the orthogonal complement of the vector space spanned by $X_1 + \sqrt{\ell}\|y\|e_{1d}$ only. Now, since $X_1 + \sqrt{\ell}\|y\|e_{1d} \sim \mathcal{N}_d(\sqrt{\ell}\|y\|e_{1d}, I_d)$, we have that $T_1^2 \sim \chi_d^2(\ell\|y\|^2)$. Conditional on X_1, the distribution of T_2^2 is then χ_{d-1}^2 since it is the squared norm of the orthogonal projection of X_2 onto a vector space that has almost surely dimension $d - 1$. Therefore, T_2 and X_1, or equivalently T_2 and T_1, are mutually independent. Repeating this reasoning shows that $T_1^2 \sim \chi_d^2(\ell\|y\|^2)$, $T_i^2 \sim \chi_{d+1-i}^2$, $i = 2, \ldots, \ell$, and that these ℓ random variables are mutually independent. This establishes the result. □

For unanchored simplices, working conditionally on $X_{\ell+1}$ in the argument above shows that

$$W_{\ell+1,d}|X_{\ell+1} =_{\mathcal{D}} \frac{1}{\ell!} \|X_1 + \sqrt{\ell}\|X_{\ell+1}\|e_{1d}\| \left(\prod_{i=2}^{\ell} \|\pi_\perp(X_i; X_1 + \sqrt{\ell}\|X_{\ell+1}\|e_{1d}, X_2, \ldots, X_{i-1})\| \right)$$

$$=: \frac{1}{\ell!} T_1 \left(\prod_{i=2}^{\ell} T_i \right).$$

Spherical symmetry entails that

$$T_1^2 = \|X_1 + \sqrt{\ell}\|X_{\ell+1}\|e_{1d}\|^2 =_{\mathcal{D}} \|X_1 + \sqrt{\ell}X_{\ell+1}\|^2 =_{\mathcal{D}} (\ell+1)\|X_1\|^2 \sim (\ell+1)\chi_d^2,$$

whereas the same conditioning argument as above still yields that the remaining factors $T_i^2 \sim \chi_{d+1-i}^2$, $i = 2, \ldots, \ell$, are mutually independent and independent of T_1. In this simplified proof, thus, we obtain the result in Corollary 2.1(i) without even needing Lemma 2.1. This argument, that relies on orthogonalization, with the only conditionally non-central chi-factor being unconditionally distributed as $\sqrt{\ell+1}\chi_d$, also gives the result a geometric interpretation.

We now discuss some variation on the alternative proof above, that exploits the flexibility in the choice of the matrix B in that proof. Rather than $B = \sqrt{\ell}\|y\|e_{1d}e'_{1\ell}$, we take here $B = \sqrt{\ell}\|y\|e_{1d}e'_{k\ell}$ for some fixed $k \in \{2, \ldots, \ell\}$, where $e_{k\ell}$ denotes the kth vector of the canonical basis of $\mathbb{R}^\ell$. Since this new matrix B still has the same singular values as the matrix A, Lemma 3.1 yields

$$W_{\ell,d}(y) =_{\mathcal{D}} \frac{1}{\ell!} m_\ell(\text{Parall}(X_1, \ldots, X_{k-1}, X_k + \sqrt{\ell}\|y\|e_{1d}, X_{k+1}, \ldots, X_\ell, 0)).$$

Using the same notation as in (3.1), this leads to

$$W_{\ell,d}(y) =_{\mathcal{D}} \frac{1}{\ell!} \|X_1\| \left(\prod_{i=2}^{k-1} \|\pi_\perp(X_i; X_1, \ldots, X_{i-1})\| \right) \|\pi_\perp(X_k + \sqrt{\ell}\|y\|e_{1d}; X_1, \ldots, X_{k-1})\|$$

$$\times \left(\prod_{i=k+1}^{\ell} \|\pi_\perp(X_i; X_1, \ldots, X_{k-1}, X_k + \sqrt{\ell}\|y\|e_{1d}, X_{k+1}, \ldots, X_{i-1})\| \right)$$

$$=: \frac{1}{\ell!} S_1 \left(\prod_{i=2}^{k-1} S_i \right) S_k \left(\prod_{i=k+1}^{\ell} S_i \right),$$

say. The same argument as above then shows that the S factors are mutually independent and that $S_i^2 \sim \chi_{d+1-i}^2$ for any $i \in \{1, \ldots, \ell\} \setminus \{k\}$. We may thus focus on the distribution of S_k. To do so, let O be a $d \times d$ orthogonal matrix whose first $k-1$ columns almost surely form an orthonormal basis of the vector space $\mathcal{V}_{k-1}$ spanned by $X_1, \ldots, X_{k-1}$ (the term "almost surely" is associated with the fact that this vector space has dimension $k-1$ with probability one). Spherical symmetry of the X_i's entails that O follows the Haar distribution on the collection of $d \times d$ orthogonal matrices. Letting Λ be the $d \times d$ diagonal matrix with the first $k-1$ diagonal entries equal to one and the remaining ones equal to zero, $O \Lambda O'$ is the matrix of the orthogonal projection onto $\mathcal{V}_{k-1}$. Denoting as $U = (U_1, \ldots, U_d)'$ the first column of O', spherical symmetry of X_k thus yields

$$S_k^2 = \|O(I_d - \Lambda)O'(X_k + \sqrt{\ell}\|y\|e_{1d})\|^2$$

$$= \|(I_d - \Lambda)(O'X_k + \sqrt{\ell}\|y\|U)\|^2 =_{\mathcal{D}} \sum_{r=k}^{d} \left\{ (X_k + \sqrt{\ell}\|y\|U)_r \right\}^2.$$

Now, $X_k + \sqrt{\ell}\|y\|U$, conditional on U, is d-variate normal with mean vector $\sqrt{\ell}\|y\|U$ and covariance matrix I_d, so that, still conditional on U, we have that $S_k^2 \sim \chi_{d-k+1}^2(\delta)$, with $\delta = \ell\|y\|^2 \sum_{r=k}^{d} U_r^2$. Since U is uniformly distributed over the unit sphere $\mathcal{S}^{d-1} = \{x \in \mathbb{R}^d : \|x\| = 1\}$ of $\mathbb{R}^d$,

$$\sum_{r=k}^{d} U_r^2 =_{\mathcal{D}} B_k, \quad \text{with } B_k \sim \text{Beta}\left(\frac{d-k+1}{2}, \frac{k-1}{2}\right);$$

see, e.g., Theorem 1.5.7(ii) in Muirhead [32]. Therefore, $S_k^2 \sim \chi_{d-k+1}^2(\ell\|y\|^2 B_k)$, by which we mean that for any $b \in (0, 1)$, the random variable S_k^2, conditional on $B_k = b$, is $\chi_{d-k+1}^2(\ell\|y\|^2 b)$. Summing up, we proved that

$$W_{\ell,d}(y) =_{\mathcal{D}} \frac{1}{\ell!} Q_d Q_{d-1} \ldots Q_{d-k+2} Q_{d-k+1}(\ell\|y\|^2 B_k) Q_{d-k} Q_{d-k-1} \ldots Q_{d-\ell+1}, \qquad (3.2)$$

where all chi-square factors are mutually independent. Interestingly, this does not seem to agree with the result in Theorem 1.1, that rather states that

$$W_{\ell,d}(y) =_{\mathcal{D}} \frac{1}{\ell!} Q_d(\ell\|y\|^2) Q_{d-1} \ldots Q_{d-\ell+1}. \tag{3.3}$$

This discrepancy therefore establishes a collection of unsuspected distributional identities (indexed by $k \in \{2, \ldots, \ell\}$) involving products of central and non-central chi-square distributions with Beta-distributed non-centrality parameters. We state these identities in Theorem 4.1 in the next section and prove them independently by using the concept of Mellin transform.

4. Some intriguing distributional identities

In this section, we establish the distributional identities that show that the result in (3.2) indeed agrees with the one in (3.3). In this process, we also state auxiliary results we will need when studying multi-variate medians in Section 5. The distributional identities are as follows.

Theorem 4.1. *Fix an integer $d \geq 2$, $m \in \{1, \ldots, d-1\}$ and $\lambda > 0$. Let $B \sim \text{Beta}\left(\frac{m}{2}, \frac{d-m}{2}\right)$ and let $Q_m^2(\lambda B)$ be a random variable that, conditional on $B = b$, is $\chi_m^2(\lambda b)$ (for any $b \in (0, 1)$). Let $Q_d^2 \sim \chi_d^2$ be independent of $Q_m^2(\lambda B)$. Finally, let $Q_d^2(\lambda) \sim \chi_d^2(\lambda)$ and $Q_m^2 \sim \chi_m^2$ be mutually independent. Then, $Q_d^2 Q_m^2(\lambda B) =_{\mathcal{D}} Q_d^2(\lambda) Q_m^2$.*

The Mellin transform of a non-negative random variable Z admitting moments of any order is the mapping $s \mapsto M_Z(s) = \mathbb{E}[Z^{s-1}]$, defined for $s \in [1, \infty)$. For such random variables, the Mellin transform, parallel to the characteristic function, characterizes the distribution; we refer to Epstein [12] for what is probably the first application of the Mellin transform in probability and statistics. It should be noted that this transform was used in Mathai [25] to derive the distribution of some random simplices. What makes the Mellin transform a suitable tool to prove Theorem 4.1 is that, unlike the Fourier transform leading to characteristic functions, the Mellin transform interacts well with products of random variables: obviously, the Mellin transform of a product of independent random variables is the product of their Mellin transforms. To prove Theorem 4.1 by showing that $Q_d^2 Q_m^2(\lambda B)$ and $Q_d^2(\lambda) Q_m^2$ have the same Mellin transform, we will need to compute the Mellin transform of the various factors. This is the role of the following lemma (see the appendix for a proof).

Lemma 4.1. *Let R be a random variable that is almost surely non-negative and has finite moments of any order. Let Z be a random variable that, conditional on $R = r$, is non-central chi-square with d degrees of freedom and non-centrality parameter r. Then, denoting as Γ the Euler Gamma function, the Mellin transform $s \mapsto M_Z(s)$ of Z, for $s \in [1, \infty)$, is such that*

$$M_Z(s) = 2^{s-1} \frac{\Gamma(\frac{d}{2} + s - 1)}{\Gamma(1 - s)} \sum_{k=0}^{\infty} \frac{\Gamma(1 - s + k)}{(-2)^k k! \Gamma(\frac{d}{2} + k)} \mathbb{E}[R^k]$$

if s is not an integer and

$$M_Z(s) = 2^{s-1} \Gamma(\tfrac{d}{2} + s - 1) \sum_{k=0}^{s-1} \frac{\binom{s-1}{k}}{2^k \Gamma(\frac{d}{2} + k)} \mathbb{E}[R^k]$$

if s is an integer.

Taking for R a random variable that is equal to $\lambda (\geq 0)$ almost surely, this result shows that the Mellin transform of the $\chi_d^2(\lambda)$ distribution is such that

$$M(s) = \left(2^{s-1} \frac{\Gamma(\frac{d}{2} + s - 1)}{\Gamma(1-s)} \sum_{k=0}^{\infty} \frac{\Gamma(1 - s + k)}{(-2)^k k! \Gamma(k + \frac{d}{2})} \lambda^k \right) \mathbb{I}[s \notin \mathbb{N}] \tag{4.1}$$

$$+ \left(2^{s-1} \Gamma(\tfrac{d}{2} + s - 1) \sum_{k=0}^{s-1} \frac{\binom{s-1}{k}}{2^k \Gamma(\frac{d}{2} + k)} \lambda^k \right) \mathbb{I}[s \in \mathbb{N}];$$

throughout, $\mathbb{I}[A]$ will denote the indicator function associated with the condition (or event) A. In particular, taking $\lambda = 0$ shows that the Mellin transform of the χ_d^2 distribution is given by

$$M(s) = 2^{s-1} \frac{\Gamma(\frac{d}{2} + s - 1)}{\Gamma(\frac{d}{2})},$$

which is a well-known result; see, e.g., Table 1 in Epstein [12]. We can now prove Theorem 4.1.

Proof of Theorem 4.1. If $R/\lambda \sim \text{Beta}(\alpha, \beta)$, then Lemma 4.1 shows that the Mellin transform of $Q_m^2(R)$ is such that

$$M_{Q_d^2(R)}(s) = 2^{s-1} \frac{\Gamma(\frac{d}{2} + s - 1)\Gamma(\alpha + \beta)}{\Gamma(1-s)\Gamma(\alpha)} \sum_{k=0}^{\infty} \frac{\Gamma(1 - s + k)\Gamma(k + \alpha)}{(-2)^k k! \Gamma(k + \frac{d}{2})\Gamma(k + \alpha + \beta)} \lambda^k$$

if s is not an integer and

$$M_{Q_d^2(R)}(s) = 2^{s-1} \frac{\Gamma(\frac{d}{2} + s - 1)\Gamma(\alpha + \beta)}{\Gamma(\alpha)} \sum_{k=0}^{s-1} \frac{\binom{s-1}{k}\Gamma(k + \alpha)}{2^k \Gamma(\frac{d}{2} + k)\Gamma(k + \alpha + \beta)} \lambda^k$$

if s is an integer. Therefore, at non-integer values of s, the Mellin transform of $Q_d Q_m(\lambda B)$ takes the value

$$\left(2^{s-1} \frac{\Gamma(\frac{d}{2} + s - 1)}{\Gamma(\frac{d}{2})} \right) \left(2^{s-1} \frac{\Gamma(\frac{m}{2} + s - 1)\Gamma(\frac{d}{2})}{\Gamma(1-s)\Gamma(\frac{m}{2})} \sum_{k=0}^{\infty} \frac{\Gamma(1 - s + k)\Gamma(k + \frac{m}{2})}{(-2)^k k! \Gamma(k + \frac{m}{2})\Gamma(k + \frac{d}{2})} \lambda^k \right)$$

$$= 4^{s-1} \frac{\Gamma(\frac{d}{2} + s - 1)\Gamma(\frac{m}{2} + s - 1)}{\Gamma(1-s)\Gamma(\frac{m}{2})} \sum_{k=0}^{\infty} \frac{\Gamma(1 - s + k)}{(-2)^k k! \Gamma(k + \frac{d}{2})} \lambda^k,$$

whereas, at the same value of s, the Mellin transform of $Q_m Q_d(\lambda)$ is equal to

$$\left(2^{s-1} \frac{\Gamma(\frac{m}{2} + s - 1)}{\Gamma(\frac{m}{2})} \right) \left(2^{s-1} \frac{\Gamma(\frac{d}{2} + s - 1)}{\Gamma(1-s)} \sum_{k=0}^{\infty} \frac{\Gamma(1 - s + k)}{(-2)^k k! \Gamma(k + \frac{d}{2})} \lambda^k \right)$$

$$= 4^{s-1} \frac{\Gamma(\frac{d}{2} + s - 1)\Gamma(\frac{m}{2} + s - 1)}{\Gamma(1-s)\Gamma(\frac{m}{2})} \sum_{k=0}^{\infty} \frac{\Gamma(1 - s + k)}{(-2)^k k! \Gamma(k + \frac{d}{2})} \lambda^k.$$
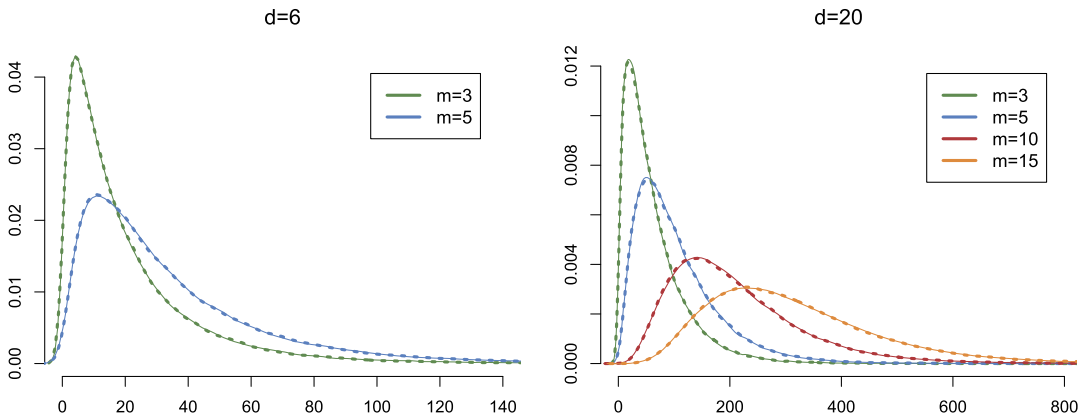
Figure 3. (Left:) Kernel density estimates, for each $m \in \{3, 5\}$ and for $\lambda = 1$, of the density of $Q_d^2 Q_m^2(\lambda B)$ (solid lines) and of the density of $Q_d^2(\lambda) Q_m^2$ (dashed lines) in Theorem 4.1, in each case based on 100 000 mutually independent realizations of these random variables (kernel density estimates are obtained by using the R command `density` with default parameter values). (Right:) The corresponding results for $d = 20$ and $m \in \{3, 5, 10, 15\}$.

These two transforms thus take the same value at any $s \in [1, \infty) \setminus \mathbb{N}$. Proceeding similarly, it is readily checked that, at any $s \in [1, \infty) \cap \mathbb{N}$, both Mellin transforms are equal to

$$4^{s-1} \frac{\Gamma(\frac{d}{2} + s - 1)\Gamma(\frac{m}{2} + s - 1)}{\Gamma(\frac{m}{2})} \sum_{k=0}^{s-1} \frac{\binom{s-1}{k}}{2^k \Gamma(\frac{d}{2} + k)} \lambda^k.$$

Consequently, both transforms do coincide, which establishes the result. $\qquad\square$

We conclude this section with a short numerical exercise illustrating the distributional identities above. For various values of d and m, we generated 100 000 mutually independent realizations of the random variables $Q_d^2 Q_m^2(\lambda B)$ and $Q_d^2(\lambda) Q_m^2$ from Theorem 4.1 with $\lambda = 1$. Figure 3, that plots the corresponding kernel density estimates, most clearly indicates that these random variables indeed share the same distribution.

5. Application to multivariate medians

Let P be a probability measure over $\mathbb{R}^d$ and consider the objective function

$$O_{\ell, P}(\mu) := \mathrm{E}_P[W_{\ell, d}(\mu)] = \mathrm{E}_P[m_\ell(\mathrm{Simpl}(X_1, \ldots, X_\ell, \mu))]; \qquad (5.1)$$

all expectations $\mathrm{E}_P[\cdot]$ in the sequel assume that the random variables involved form a random sample from P. Provided that P admits finite first moments (in the sense that $\int_{\mathbb{R}^d} |x_r| \, dP(x)$ exists and is finite for any $r = 1, \ldots, d$), $O_{\ell, P}(\mu)$ is well-defined for any $\mu \in \mathbb{R}^d$ (Lemma A.4), and we then define an *ℓ-median of P* as an arbitrary minimizer of $\mu \mapsto O_{\ell, P}(\mu)$:

$$\mu_{\ell, P} := \arg\min_{\mu \in \mathbb{R}^d} O_{\ell, P}(\mu).$$

For $\ell = d$, which corresponds to the Oja median (see below), existence of such a minimizer was studied in León and Massé [21]. The following result shows existence for a general value of ℓ (see the appendix for a proof).

Theorem 5.1. *Fix an integer $d \geq 1$ and $\ell \in \{1, \ldots, d\}$. Let P be a probability measure over $\mathbb{R}^d$ that admits finite first-order moments. Then, $\mu \mapsto O_{\ell,P}(\mu)$ admits a minimizer over $\mathbb{R}^d$.*

Uniqueness, however, is not guaranteed in general. To identify a unique representative of the set $M_{\ell,P}$ of minimizers, we define *the ℓ-median of P* as the barycenter of $M_{\ell,P}$; since the objective function $\mu \mapsto O_{\ell,P}(\mu)$ is convex (see again Lemma A.4), the barycenter of $M_{\ell,P}$ is itself a minimizer, which justifies the construction (this of course requires that $M_{\ell,P}$ is bounded, which is the case as soon as no $(\ell-1)$-dimensional hyperplane of $\mathbb{R}^d$ has P-probability one; see Dürre and Paindaveine [11]). The ℓ-median provides the *spatial median* for $\ell = 1$ (Brown [8], Chaudhuri [9], Möttönen, Nordhausen and Oja [30]) and the *Oja median* for $\ell = d$ (Oja [35], Hettmansperger, Möttönen and Oja [16], Oja [36], Ollila, Oja and Hettmansperger [39]). For $d = 1$, both medians reduce to the usual univariate median, which explains the terminology. To the best of our knowledge, the multivariate ℓ-medians associated with $\ell \in \{2, \ldots, d-1\}$ have not been considered in the literature.

Now, if a random sample $X_1, \ldots, X_n$ from P is available, then the natural estimator of $\mu_{\ell,P}$ is the sample ℓ-median, which we define as

$$\mu_\ell^{(n)} := \underset{\mu \in \mathbb{R}^d}{\arg\min}\, O_\ell^{(n)}(\mu),$$

where we let

$$O_\ell^{(n)}(\mu) := \frac{1}{\binom{n}{\ell}} \sum_{1 \leq i_1 < \cdots < i_\ell \leq n} m_\ell(\mathrm{Simpl}(X_{i_1}, \ldots, X_{i_\ell}, \mu))$$

(again, the barycenter of the set of minimizers is used if there are more than one minimizer). Our first objective in this section is to study the asymptotic behaviour of the sample ℓ-median; see Brown [8], Arcones [3] or Zhou and Serfling [51] for the spatial median, and Oja and Niinimaa [38], Hettmansperger, Nyblom and Oja [17] or Shen [47] for the Oja median. We do so under three very mild assumptions. First, we will assume that there exists a neighbourhood of $\mu_{\ell,P}$ for which P is $(\ell-1)$-smooth at any μ in the neighbourhood, where we say that P is k-smooth at μ if any k-dimensional hyperplane containing μ has P-probability zero. Of course, this holds if P admits a density with respect to the Lebesgue measure but this allows for more general probability measures, such as, e.g., uniform measures on spheres. Second, we will assume that no ℓ-dimensional hyperplane containing $\mu_{\ell,P}$ has P-probability one (would there be one, then the problem would not be a genuine d-dimensional one; note that we indeed impose $\ell \leq d - 1$ in Theorem 5.2 below). Third, denoting as $\Gamma_{x_1,\ldots,x_\ell}$ the matrix of the orthogonal projection onto the orthogonal complement of the vector space spanned by $x_1 - x_\ell, \ldots, x_{\ell-1} - x_\ell$, we will assume that

$$\mathrm{E}_P\left[\frac{m_{\ell-1}(\mathrm{Simpl}(X_1, \ldots, X_\ell))}{\|\Gamma_{X_1,\ldots,X_\ell}(X_\ell - \mu_{\ell,P})\|} \mathbb{I}[\Gamma_{X_1,\ldots,X_\ell}(X_\ell - \mu_{\ell,P}) \neq 0] \right] < \infty \qquad (5.2)$$

(for $\ell = 1$, we define $\Gamma_x = I_d$ and $m_0(\mathrm{Simpl}(x)) := 1$ for any $x \in \mathbb{R}^d$). The second and third assumptions, which, for $\ell = 1$, are standard when studying the asymptotic behaviour of the sample spatial median (Arcones [3], Möttönen, Nordhausen and Oja [30]) ensure that the Hessian matrix in the following result is well-defined and invertible (see the appendix for a proof of this result).

Theorem 5.2. *Fix an integer $d \geq 2$ and $\ell \in \{1, \ldots, d-1\}$. Let P be a probability measure over $\mathbb{R}^d$ that admits finite second-order moments (in the sense that $\int_{\mathbb{R}^d} x_r^2 \, dP(x)$ exists and is finite for any $r = 1, \ldots, d$), that is $(\ell-1)$-smooth at any μ in a neighbourhood of $\mu_{\ell,P}$, and such that (5.2) holds. Assume further that no ℓ-dimensional hyperplane containing $\mu_{\ell,P}$ has P-probability one. Then,*

$$\sqrt{n}(\mu_\ell^{(n)} - \mu_{\ell,P}) \;=\; H_P^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n T_P(X_i) + o_{\mathrm{P}}(1) \tag{5.3}$$

$$\rightarrow_{\mathcal{D}} \; \mathcal{N}_d\big(0, H_P^{-1} \mathrm{E}_P[T_P(X_\ell) T_P'(X_\ell)] H_P^{-1}\big) \tag{5.4}$$

as n diverges to infinity, where

$$T_P(x)$$
$$:= \mathrm{E}_P\bigg[m_{\ell-1}(\mathrm{Simpl}(X_1, \ldots, X_{\ell-1}, x)) \frac{\Gamma_{X_1,\ldots,X_{\ell-1},x}(x - \mu_{\ell,P})}{\|\Gamma_{X_1,\ldots,X_{\ell-1},x}(x - \mu_{\ell,P})\|} \mathbb{I}[\Gamma_{X_1,\ldots,X_{\ell-1},x}(x - \mu_{\ell,P}) \neq 0] \bigg]$$

is such that $\mathrm{E}_P[T_P(X_\ell) T_P'(X_\ell)]$ is well-defined and where the Hessian matrix H_P of $\mu \mapsto O_{\ell,P}(\mu)$ at $\mu_{\ell,P}$ is well-defined and invertible.

The strategy of the proof we provide for this result requires to show that the objective function $O_{\ell,P}(\mu)$ is twice differentiable under the expectation sign at $\mu_{\ell,P}$. It should be noted that the assumptions of Theorem 5.2 are never satisfied for $\ell = d$ (the only hyperplane with dimension $\ell = d$ containing $\mu_{\ell,P}$, namely $\mathbb{R}^d$ itself, has always probability one). But there is no way to weaken the assumptions to include the case $\ell = d$ in this strategy of proof because, for $\ell = d$, the objective function actually is *not* twice differentiable under the expectation sign at $\mu_{\ell,P}$: inspection of the proof of Lemma A.5 below indeed reveals that the Hessian matrix of $\mu \mapsto m_d(\mathrm{Simpl}(x_1, \ldots, x_d, \mu))$ at $\mu_{d,P}$ is zero P-almost everywhere, so that twice differentiability under the expectation sign at $\mu_{d,P}$ would imply that $H_P = 0$ (which is not the case, e.g, at the d-variate standard normal distribution; see (5.5) below). This is why Theorem 5.2 excludes the case $\ell = d$, which, as mentioned earlier, has already been treated; see, e.g., Hettmansperger, Nyblom and Oja [17] and Shen [47].

We aim at applying Theorem 5.2, and in particular at computing the asymptotic variance in (5.4), at the d-variate standard normal probability measure. As we will show, this will provide a natural application of the results derived in the previous sections. In the rest of this section, thus, P will denote the d-variate standard normal probability measure. First note that since P is centro-symmetric with respect to the origin of $\mathbb{R}^d$ (in the sense that any Borel set B has the same P-probability as its reflection with respect to $0 \in \mathbb{R}^d$), $O_{\ell,P}(-\mu) = O_{\ell,P}(\mu)$ for any μ, which entails that $\mu_{\ell,P} = 0$. Using centro-symmetry of P again, Lemma A.8 then yields

$$T_P(x) = \mathrm{E}_P\bigg[m_{\ell-1}(\mathrm{Simpl}(X_1, \ldots, X_{\ell-1}, 0)) \frac{\Psi x}{\|\Psi x\|} \bigg],$$

where Ψ denotes the matrix of the orthogonal projection onto the orthogonal complement of the vector space spanned by $X_1, \ldots, X_{\ell-1}$. Since P is spherically symmetric about the origin of $\mathbb{R}^d$ and since $m_{\ell-1}(\mathrm{Simpl}(X_1, \ldots, X_{\ell-1}, 0))$ and Ψ are mutually independent, Lemma A.9 provides

$$T_P(x) = \frac{\Gamma(\frac{d-\ell+2}{2})\Gamma(\frac{d}{2})}{\Gamma(\frac{d-\ell+1}{2})\Gamma(\frac{d+1}{2})} \mathrm{E}_P\big[m_{\ell-1}(\mathrm{Simpl}(X_1, \ldots, X_{\ell-1}, 0))\big] \frac{x}{\|x\|},$$

where Γ is the Euler Gamma function. Corollary 2.1(ii) then yields (we use the same notation as in that corollary)

$$T_P(x) = \frac{\Gamma(\frac{d-\ell+2}{2})\Gamma(\frac{d}{2})}{(\ell-1)!\Gamma(\frac{d-\ell+1}{2})\Gamma(\frac{d+1}{2})}\mathrm{E}[Q_d]\mathrm{E}[Q_{d-1}]\ldots\mathrm{E}[Q_{d-\ell+2}]\frac{x}{\|x\|} = \frac{2^{(\ell-1)/2}\Gamma(\frac{d}{2})}{(\ell-1)!\Gamma(\frac{d-\ell+1}{2})}\frac{x}{\|x\|}.$$

To apply Theorem 5.2 at the d-variate standard normal distribution, it remains to evaluate the Hessian matrix H_P. Using Theorem 1.1 and the Mellin transform in (4.1), we obtain

$$\mathrm{E}_P[m_\ell(\mathrm{Simpl}(X_1,\ldots,X_{\ell-1},X_\ell,\mu))] = \frac{1}{\ell!}\mathrm{E}[Q_d(\ell\|\mu\|^2)]\mathrm{E}[Q_{d-1}]\mathrm{E}[Q_{d-2}]\ldots\mathrm{E}[Q_{d-\ell+1}]$$

$$= \frac{2^{\ell/2}\Gamma(\frac{d}{2})\Gamma(\frac{d+1}{2})}{\ell!\Gamma(-\frac{1}{2})\Gamma(\frac{d-\ell+1}{2})}\sum_{k=0}^{\infty}\frac{\Gamma(k-\frac{1}{2})}{(-2)^k k!\Gamma(k+\frac{d}{2})}\ell^k\|\mu\|^{2k},$$

which, after a direct computation, provides

$$H_P = \nabla_\mu^2 \mathrm{E}[m_\ell(\mathrm{Simpl}(X_1,\ldots,X_{\ell-1},X_\ell,\mu))]|_{\mu=0} = \frac{2^{\ell/2}\Gamma(\frac{d+1}{2})}{(\ell-1)!d\Gamma(\frac{d-\ell+1}{2})}I_d. \tag{5.5}$$

Applying Theorem 5.2 and using the equivariance of the sample ℓ-median under translations and under homothetic transformations then establishes the following result.

Corollary 5.1. *Fix an integer $d \geq 2$ and $\ell \in \{1,\ldots,d-1\}$. Let P be the d-variate normal distribution with mean vector μ and non-singular covariance matrix $\Sigma = \sigma^2 I_d$. Then,*

$$\sqrt{n}(\mu_\ell^{(n)} - \mu) = \frac{\sigma(d\,\Gamma(\frac{d}{2})\Gamma(\frac{d+2}{2}))^{1/2}}{\sqrt{n}\,\Gamma(\frac{d+1}{2})}\sum_{i=1}^{n}\frac{X_i - \mu}{\|X_i - \mu\|} + o_P(1) \to_{\mathcal{D}} \mathcal{N}_d\left(0, \frac{\Gamma(\frac{d+2}{2})\Gamma(\frac{d}{2})}{\Gamma^2(\frac{d+1}{2})}\Sigma\right)$$

as n diverges to infinity.

Remarkably, neither the influence function of the ℓ-median (see Hampel *et al.* [15])

$$x \mapsto \mathrm{IF}(x; \mu_{\ell,P}, P) = \lim_{\varepsilon \overset{>}{\to} 0}\frac{\mu_{\ell,P_{x,\varepsilon}} - \mu_{\ell,P}}{\varepsilon} = \frac{\sigma(d\,\Gamma(\frac{d}{2})\Gamma(\frac{d+2}{2}))^{1/2}}{\Gamma(\frac{d+1}{2})}\frac{x - \mu}{\|x - \mu\|}$$

(with $P_{x,\varepsilon} = (1-\varepsilon)P + \varepsilon\delta_x$, where δ_x is the Dirac probability measure at x) nor the asymptotic variance of the sample ℓ-median do depend on ℓ. Note that, while Corollary 5.1 in principle does not cover the Oja median (since the case $\ell = d$ was excluded in Theorem 5.2), the influence function of the Oja median (Ollila, Oja and Hettmansperger [39]) and its asymptotic variance (Oja and Niinimaa [38]) at the d-variate standard normal probability measure imply that the result actually holds for $\ell = d$, too. For $\ell = 1$, Corollary 5.1 agrees with the results obtained earlier for the spatial median; we refer, e.g., to Arcones [3] and Zhou and Serfling [51].

We end this section with a numerical exercise. For each sample size $n \in \{30, 100\}$, we generated a collection of 1 000 mutually independent random samples $X_1,\ldots,X_n$ of size n from the ($d = 3$)-variate standard normal distribution. In each replication, we evaluated the sample mean $\hat{\mu}_M = \frac{1}{n}\sum_{i=1}^{n}X_i$ and the sample ℓ-medians $\hat{\mu}_\ell^{(n)}$, $\ell = 1, 2, 3$. The 1-median (the spatial median) was computed by using the function `spatial.median` from the R package `ICSNP`, whereas the 3-median
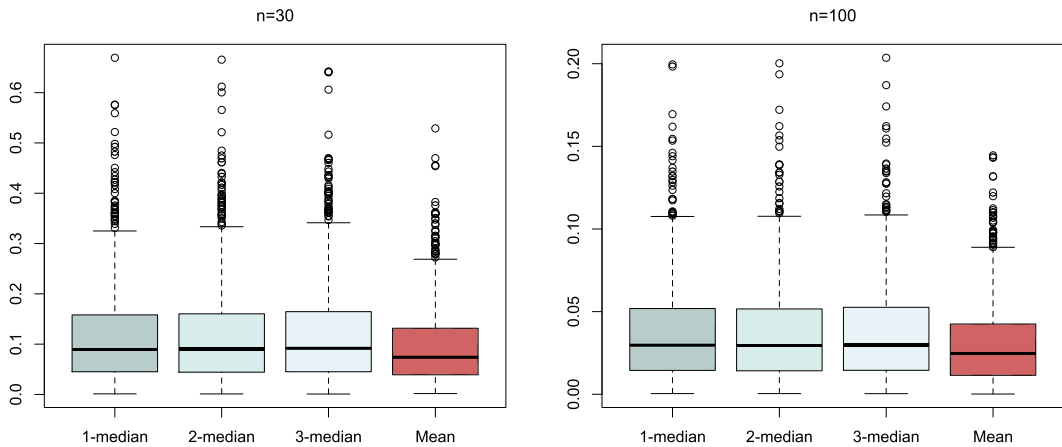
Figure 4. (Left:) Boxplots of the squared Euclidean distances $\|\hat{\mu}^{(n)} - \mu\|^2$, obtained in 1 000 mutually independent random samples $X_1, \ldots, X_n$ of size $n = 30$ from the $(d = 3)$-variate standard normal distribution, between four estimators $\hat{\mu}^{(n)}$ and the true location $\mu = 0$. The estimators considered are the sample ℓ-medians, for $\ell \in \{1, 2, 3\}$, and the sample mean. (Right:) The corresponding results for $n = 100$.

(the Oja median) was evaluated by using the "bounded exact" algorithm in the R package OjaNP; see Fischer *et al.* [13]. The 2-median was computed through convex optimization, more precisely by implementing a gradient descent based on backtracking line search; see Sections 9.2–9.3 from Boyd and Vandenberghe [7]. For each estimator $\hat{\mu}^{(n)}$ and each sample size $n \in \{30, 100\}$, Figure 4 provides the boxplot of the resulting 1 000 squared Euclidean distances $\|\hat{\mu}^{(n)} - \mu\|^2$ between the estimate $\hat{\mu}^{(n)}$ and the true value $\mu = 0$. As expected, the sample mean, which is the efficient estimator of μ under normality, dominates the considered ℓ-medians. Dominance is quite mild, however, which is in line with the fact that in dimension $d = 3$, the common asymptotic covariance matrix of the ℓ-median is only 1.178 times the one of the sample mean. More importantly, the results clearly support the asymptotic equivalence of the three ℓ-medians, that seems to materialize for sample sizes as small as $n = 30$ already. It should be noted that this equivalence quite much relies on the *spherical* nature of the setup considered in this work: if the Gaussian spherical distribution would be made increasingly elliptical, then the efficiency of the Oja median would remain unchanged (since the Oja median is actually equivariant under affine transformations) but the efficiency of the spatial median would deteriorate (Niinimaa and Oja [33]). Under such ellipticity, intermediate ℓ-medians are thus promising objects to achieve a compromise between the nice computability properties of the spatial median and the good efficiency properties of the Oja median.

6. Perspectives for future research

We end the paper by discussing perspectives for future research. First, it is natural to try to extend the results of this paper beyond spherical normal distributions. Getting rid of the normality assumption is complex, though. In particular, it should be noted that the simple proof of Theorem 1.1 we provide in Section 3 focuses on the Gaussian case; indeed, the proof uses the fact that the underlying distribution (i) is spherically symmetric and (ii) has mutually independent marginals, two properties that, according to the Maxwell–Hershell Theorem (see, e.g., Proposition 4.11 in Bilodeau and Brenner [6] or Theorem 1.5.3 in Muirhead [32]) are met at spherical normal distributions only. It is not excluded, however,

that our results can be extended to more general spherical distributions by appropriate conditioning on the corresponding non-Gaussian radii. Getting rid of the sphericity assumption is easy for $\ell = d$, since the volumes of the corresponding simplices behave well under affine transformations (this is the reason why the Oja median is equivariant under affine transformations, which in turn explains that its efficiency is the same under elliptical distributions as under spherical ones; see the end of Section 5). The situation is much more delicate for $\ell < d$. Yet, regarding ℓ-medians, there is some hope that the elliptical results obtained for the spatial median ($\ell = 1$) in Magyar and Tyler [24] can be extended to an arbitrary $\ell \in \{2, \ldots, d-1\}$. An alternative approach would be to modify the definition of ℓ-medians to make them equivariant under affine transformations, as it was done already for the spatial median in Hettmansperger and Randles [18]; see also Serfling [45].

Second, it would be of interest to study the robustness properties of ℓ-medians. While it is expected that, as ℓ increases from 1 to d, the breakdown point of the ℓ-medians will decrease from the maximal 50% breakdown point of the spatial median (Lopuhaä and Rousseeuw [22]) to the minimal 0% breakdown point of the Oja median (Niinimaa, Oja and Tableman [34]), knowing the exact dependence on ℓ would be useful when trying to achieve a balance between robustness and efficiency. In a different perspective, a modern trend in robust statistics has been defining accurate estimators under heavy tails through a "median-of-means" construction, that reduces variability by considering the median of mutually independent means. Implementing this in a multivariate framework of course requires a multivariate median concept, and both the spatial median and componentwise median have been used in this context (Minsker [28], Hsu and Sabato [19]). As hinted in the recent review paper Lugosi and Mendelson [23], the Oja median could also be used in this framework, and it would be interesting to see, more generally, what could be the possible advantages of ℓ-medians in this purpose.

Finally, we mainly focused on multivariate medians in this work when applying our results on anchored simplices, and it would be natural to consider further applications. A natural venue for this is in multivariate nonparametric hypothesis testing, when one wants to test the null hypothesis that the unknown center of the distribution (e.g., the center of symmetry under a centro-symmetric distributional assumption) coincides with a given location in $\mathbb{R}^d$. For this problem, sign tests that are companion procedures to the spatial median and Oja median—hence are based on ($\ell = 1$)-simplices and ($\ell = d$)-simplices, respectively—were proposed and studied in Möttönen and Oja [31] and Hettmansperger, Nyblom and Oja [17], respectively (see also Oja [37]). It would be natural to define a class of ℓ-sign tests (based on ℓ-simplices) that would contain the aforementioned sign tests as particular cases. In this framework, our results on anchored simplices would allow one to obtain explicit expressions for the asymptotic powers of these ℓ-sign tests under contiguous spherical normal alternatives. Another possible inferential application of our results would consist in using the U-statistics considered in the introduction, after suitable standardization of the data, to test for multinormality of the underlying distribution.

Appendix A: Auxiliary proofs

In this technical appendix, we first prove Lemma 2.1 and Lemma 4.1, then we turn to the proof of the results from Section 5.

Proof of Lemma 2.1. Let $Z_1, \ldots, Z_{2d}$ be mutually independent standard normal variables. Then $Q = \sum_{k=1}^{d} Z_{d+k}^2 \sim \chi_d^2$ and

$$T := \sum_{k=1}^{d} (Z_k + \sqrt{\lambda} Z_{d+k})^2$$

is, conditional on Q, non-central chi-square with d degrees of freedom and non-centrality parameter λQ. Therefore, $T =_{\mathcal{D}} R$. The result then follows from the fact that $(Z_k + \sqrt{\lambda} Z_{d+k})/\sqrt{\lambda + 1}$, $k = 1, \ldots, d$, are mutually independent standard normal variables. $\qquad\square$

The proof of Lemma 4.1 requires the following preliminary result.

Lemma A.1. *For any non-negative integer k and any positive real numbers a, b,*

$$\sum_{m=0}^{k} (-1)^m \binom{k}{m} \frac{\Gamma(m+a+b)}{\Gamma(m+a)} = \frac{\Gamma(a+b)}{\Gamma(k+a)} \left(\frac{\Gamma(k-b)}{\Gamma(-b)} \mathbb{I}[b \notin \mathbb{N}] + (-1)^k k! \binom{b}{k} \mathbb{I}[b \in \{k, k+1, \ldots\}] \right),$$

where Γ is the Euler Gamma function.

Proof of Lemma A.1. (i) Consider first the case where b is not an integer. Denoting the rising factorial as $z^{(m)} = \Gamma(z+m)/\Gamma(z)$, we then have

$$\frac{\Gamma(a)}{\Gamma(a+b)} \sum_{m=0}^{k} (-1)^m \binom{k}{m} \frac{\Gamma(m+a+b)}{\Gamma(m+a)} = \sum_{m=0}^{k} (-1)^m \binom{k}{m} \frac{(a+b)^{(m)}}{a^{(m)}}$$

$$= \sum_{m=0}^{\infty} (-1)^m k(k-1)\ldots(k-m+1) \frac{(a+b)^{(m)}}{m! a^{(m)}} = \sum_{m=0}^{\infty} \frac{(-k)^{(m)}(a+b)^{(m)}}{m! a^{(m)}} = {}_2F_1(-k, a+b; a; 1),$$

where ${}_2F_1(\alpha, \beta; \gamma; z)$ is the Gauss hypergeometric function. The linear transformation rule

$$_2F_1(\alpha, \beta; \gamma; z) = \frac{\Gamma(\gamma)\Gamma(\gamma - \alpha - \beta)}{\Gamma(\gamma - \alpha)\Gamma(\gamma - \beta)} {}_2F_1(\alpha, \beta; \alpha + \beta + 1 - \gamma; 1 - z)$$

$$+ \frac{\Gamma(\gamma)\Gamma(\alpha + \beta - \gamma)}{\Gamma(\alpha)\Gamma(\beta)} (1 - z)^{\gamma - \alpha - \beta} {}_2F_1(\gamma - \alpha, \gamma - \beta; 1 + \gamma - \alpha - \beta; 1 - z)$$

(that holds whenever $\gamma - \alpha - \beta \notin \mathbb{Z}$; see (15.3.6) in Abramowitz and Stegun [1]) then entails

$$\frac{\Gamma(a)}{\Gamma(a+b)} \sum_{m=0}^{k} (-1)^m \binom{k}{m} \frac{\Gamma(m+a+b)}{\Gamma(m+a)} = \frac{\Gamma(a+b)\Gamma(k-b)}{\Gamma(k+a)\Gamma(-b)},$$

as was to be showed. (ii) Let us thus turn to the case where b is a positive integer. From continuity, we have

$$\sum_{m=0}^{k} (-1)^m \binom{k}{m} \frac{\Gamma(m+a+b)}{\Gamma(m+a)} = \lim_{z \to b} \frac{\Gamma(a+z)\Gamma(k-z)}{\Gamma(k+a)\Gamma(-z)} = \frac{\Gamma(a+b)}{\Gamma(k+a)} \lim_{z \to b} \frac{\Gamma(k-z)}{\Gamma(-z)}.$$

If $b < k$, then this limit is zero (since $\Gamma(k-z) \to \Gamma(k-b) \in \mathbb{R}$ and $|\Gamma(-z)| \to \infty$), whereas if $b \geq k$, then Euler's reflection formula $\Gamma(z)\Gamma(1-z) = \pi/\sin(\pi z)$ (for $z \notin \mathbb{Z}$) and L'Hôpital's rule provide

$$\sum_{m=0}^{k} (-1)^m \binom{k}{m} \frac{\Gamma(m+a+b)}{\Gamma(m+a)} = \frac{\Gamma(a+b)}{\Gamma(k+a)} \lim_{z \to b} \frac{\Gamma(z+1)\sin(\pi(z+1))}{\Gamma(z-k+1)\sin(\pi(z-k+1))}$$

$$= (-1)^k \frac{\Gamma(a+b)\Gamma(b+1)}{\Gamma(k+a)\Gamma(b-k+1)} = (-1)^k \frac{\Gamma(a+b)b!}{\Gamma(k+a)(b-k)!} = (-1)^k k! \frac{\Gamma(a+b)}{\Gamma(k+a)} \binom{b}{k},$$

which establishes the result. □

Proof of Lemma 4.1. Working conditionally on R, Equation (7) from Wells, Anderson and Cell [50] provides

$$E[Z_L^{s-1}|R] = e^{-R/2} \sum_{m=0}^{\infty} \frac{2^{s-m-1}\Gamma(m+\frac{d}{2}+s-1)}{m!\Gamma(m+\frac{d}{2})} R^m$$

$$= 2^{s-1} \sum_{m,n=0}^{\infty} \frac{(-1)^n 2^{-(m+n)}\Gamma(m+\frac{d}{2}+s-1)}{m!n!\Gamma(m+\frac{d}{2})} R^{m+n}.$$

Therefore, taking expectations (treating separately odd and even values of n, the monotone convergence theorem allows us to take expectation termwise in the righthand side) yields

$$M_Z(s) = 2^{s-1} \sum_{m,n=0}^{\infty} \frac{(-1)^n 2^{-(m+n)}\Gamma(m+\frac{d}{2}+s-1)}{m!n!\Gamma(m+\frac{d}{2})} E[R^{m+n}]$$

$$= 2^{s-1} \sum_{m,n=0}^{\infty} \binom{m+n}{m} \frac{(-1)^n 2^{-(m+n)}\Gamma(m+\frac{d}{2}+s-1)}{(m+n)!\Gamma(m+\frac{d}{2})} E[R^{m+n}]$$

$$= 2^{s-1} \sum_{k=0}^{\infty} \sum_{m=0}^{k} \binom{k}{m} \frac{(-1)^{k+m} 2^{-k}\Gamma(m+\frac{d}{2}+s-1)}{k!\Gamma(m+\frac{d}{2})} E[R^k]$$

$$= 2^{s-1} \sum_{k=0}^{\infty} \frac{1}{(-2)^k k!} \left(\sum_{m=0}^{k} (-1)^m \binom{k}{m} \frac{\Gamma(m+\frac{d}{2}+s-1)}{\Gamma(m+\frac{d}{2})} \right) E[R^k].$$

The result then follows from Lemma A.1. □

We turn to the proof of the results from Section 5. Let us recall some key notation introduced in that section. For $x_1, \ldots, x_\ell \in \mathbb{R}^d$, we denote as $\Gamma_{x_1,\ldots,x_\ell}$ the matrix of the orthogonal projection onto the orthogonal complement of the vector space spanned by $x_1 - x_\ell, \ldots, x_{\ell-1} - x_\ell$; for $\ell = 1$, we let $\Gamma_x := I_d$ for any $x \in \mathbb{R}^d$. Recall also that we put $m_0(\text{Simpl}(x)) := 1$ for any $x \in \mathbb{R}^d$. The following result will play an important role in the subsequent proofs.

Lemma A.2. *Fix a positive integer d and $\ell \in \{1, \ldots, d\}$. Then, for any $x_1, \ldots, x_{\ell+1} \in \mathbb{R}^d$,*

$$m_\ell(\text{Simpl}(x_1, \ldots, x_\ell, x_{\ell+1})) = \frac{1}{\ell} m_{\ell-1}(\text{Simpl}(x_1, \ldots, x_\ell)) \|\Gamma(x_{\ell+1} - x_\ell)\|,$$

where we wrote $\Gamma = \Gamma_{x_1,\ldots,x_\ell}$.

Proof of Lemma A.2. Since the result is trivial for $\ell = 1$, we restrict to $\ell > 1$. We then have

$$m_\ell(\text{Simpl}(x_1, \ldots, x_\ell, x_{\ell+1}))$$

$$= \frac{1}{\ell!} m_\ell(\text{Parall}(x_1 - x_\ell, \ldots, x_{\ell-1} - x_\ell, x_{\ell+1} - x_\ell, 0))$$

$$= \frac{1}{\ell!} \|x_1 - x_\ell\| \left(\prod_{i=2}^{\ell-1} \|\pi_\perp(x_i - x_\ell; x_1 - x_\ell, \ldots, x_{i-1} - x_\ell)\| \right) \|\Gamma(x_{\ell+1} - x_\ell)\|,$$

where $\pi_\perp(z; z_1, \ldots, z_k)$ still denotes the orthogonal projection of z onto the orthogonal complement of the vector space spanned by $z_1, \ldots, z_k$. It follows that

$$m_\ell(\text{Simpl}(x_1, \ldots, x_\ell, x_{\ell+1}))$$

$$= \frac{1}{\ell!} m_{\ell-1}(\text{Parall}(x_1 - x_\ell, \ldots, x_{\ell-1} - x_\ell, 0)) \|\Gamma(x_{\ell+1} - x_\ell)\|$$

$$= \frac{1}{\ell} m_{\ell-1}(\text{Simpl}(x_1 - x_\ell, \ldots, x_{\ell-1} - x_\ell, 0)) \|\Gamma(x_{\ell+1} - x_\ell)\|$$

$$= \frac{1}{\ell} m_{\ell-1}(\text{Simpl}(x_1, \ldots, x_\ell)) \|\Gamma(x_{\ell+1} - x_\ell)\|,$$

as was to be proved. □

The proof of Theorem 5.1 crucially relies on Lemma A.4 below, which itself requires the following preliminary result.

Lemma A.3. *Fix an integer $d \geq 2$, $\ell \in \{2, \ldots, d\}$ and $u \in \mathbb{R}^d \setminus \{0\}$. Let P be a probability measure over $\mathbb{R}^d$ such that no $(\ell - 1)$-dimensional hyperplane has P-probability one. Let $X_1, \ldots, X_\ell$ form a random sample from P. Then, the probability that $u, X_1 - X_\ell, \ldots, X_{\ell-1} - X_\ell$ span a vector space of dimension ℓ is positive.*

Proof of Lemma A.3. Pick $z_1, \ldots, z_{\ell+1} \in \mathbb{R}^d$ not contained in an $(\ell - 1)$-dimensional hyperplane of $\mathbb{R}^d$ and such that $P[\{z_i\}_\varepsilon] > 0$ for any $\varepsilon > 0$ and any $i = 1, \ldots, \ell + 1$; throughout this proof, for any subset A of $\mathbb{R}^d$, we write $A_\varepsilon := \{y \in \mathbb{R}^d : \|y - x\| < \varepsilon \text{ for some } x \in A\}$.

We first prove that such z_i's exist. Let us start by showing that there exists $z_1 \in \mathbb{R}^d$ such that $P[\{z_1\}_\varepsilon] > 0$ for any $\varepsilon > 0$. Ad absurdum, assume that for any $z \in \mathbb{R}^d$, there exists $\varepsilon(z) > 0$ such that $P[\{z\}_{\varepsilon(z)}] = 0$. Take then $R > 0$ so that $B(0, R) = \{x \in \mathbb{R}^d : \|x\| \leq R\}$ has a positive P-probability. From compactness, the cover $\{\{x\}_{\varepsilon(x)} : x \in B(0, R)\}$ of $B(0, R)$ admits a finite subcover $\{\{x_i\}_{\varepsilon(x_i)} : i = 1, \ldots, N\}$, which yields

$$P[B(0, R)] \leq \sum_{i=1}^{N} P[\{x_i\}_{\varepsilon(x_i)}] = 0, \tag{A.1}$$

a contradiction. Assume now that, for some $r \in \{1, \ldots, \ell\}$, we could pick $z_1, \ldots, z_r$ not contained in an $(r - 2)$-dimensional hyperplane of $\mathbb{R}^d$ and such that $P[\{z_i\}_\varepsilon] > 0$ for any $\varepsilon > 0$ and any $i = 1, \ldots, r$.

Let us show that we can then find z_{r+1} such that $z_1, \ldots, z_r, z_{r+1}$ are not contained in an $(r-1)$-dimensional hyperplane of $\mathbb{R}^d$ and such that $P[\{z_i\}_\varepsilon] > 0$ for any $\varepsilon > 0$ and any $i = 1, \ldots, r+1$. Denote as H the $(r-1)$-dimensional hyperplane containing $z_1, \ldots, z_r$, and let us assume, ad absurdum, that for any $z \in \mathbb{R}^d \setminus H$, there exists $\varepsilon(z) > 0$ such that $P[\{z\}_{\varepsilon(z)}] = 0$. By assumption, $P[H] < 1$. Take then $\delta > 0$ small enough to have $P[H_\delta] < 1$ and R large enough to have $P[B(0, R) \setminus H_\delta] > 0$. Proceeding as above, we can find a finite cover of this compact set $B(0, R) \setminus H_\delta$ that is of the form $\{\{x_i\}_{\varepsilon(x_i)} : i = 1, \ldots, N\}$, which, as in (A.1), implies that $P[B(0, R) \setminus H_\delta] = 0$, a contradiction. This argument can be repeated until obtaining d-vectors $z_1, \ldots, z_{\ell+1}$ meeting the properties stated in the previous paragraph.

Assume now that u belongs to the span of $z_1 - z_{\ell+1}, \ldots, z_{k-1} - z_{\ell+1}, z_{k+1} - z_{\ell+1}, \ldots, z_\ell - z_{\ell+1}$ for any $k = 1, \ldots, \ell$ (it is obvious what we mean with this for $k \in \{1, \ell\}$). This implies that there exists an $\ell \times \ell$ matrix A, whose diagonal entries are equal to zero, such that $(z_1 - z_{\ell+1} \ldots z_\ell - z_{\ell+1})A = (u \ldots u)$. Since the rank of A is larger than or equal to two and the rank of $(u \ldots u)$ is equal to one, the matrix $(z_1 - z_{\ell+1} \ldots z_\ell - z_{\ell+1})$ cannot be of maximal rank ℓ, which implies that $z_1, \ldots, z_{\ell+1}$ belong to an $(\ell-1)$-dimensional hyperplane, a contradiction. Without loss of generality, we may thus assume that u does not belong to the span of $z_1 - z_{\ell+1}, \ldots, z_{\ell-1} - z_{\ell+1}$. This entails that $u, z_1 - z_{\ell+1}, \ldots, z_{\ell-1} - z_{\ell+1}$ span a vector space of dimension ℓ, hence that there exists $\varepsilon > 0$ such that, irrespective of $x_1 \in \{z_1\}_\varepsilon, \ldots, x_\ell \in \{z_\ell\}_\varepsilon, x_{\ell+1} \in \{z_{\ell+1}\}_\varepsilon$, we have that $u, x_1 - x_\ell, \ldots, x_{\ell-1} - x_\ell$ span a vector space of dimension ℓ. It follows that

$$P[u, X_1 - X_\ell, \ldots, X_{\ell-1} - X_\ell \text{ span a vector space of dimension } \ell]$$

$$\geq P[X_1 \in \{z_1\}_\varepsilon] \ldots P[X_\ell \in \{z_\ell\}_\varepsilon]P[X_{\ell+1} \in \{z_{\ell+1}\}_\varepsilon] > 0,$$

as was to be proved. $\qquad\square$

Lemma A.4. *Fix a positive integer d and $\ell \in \{1, \ldots, d\}$. Let P be a probability measure over $\mathbb{R}^d$ that admits finite first moments. Then, (i) $O_{\ell,P}(\mu)$ is well-defined for any $\mu \in \mathbb{R}^d$; (ii) the mapping $\mu \mapsto O_{\ell,P}(\mu)$ is convex over $\mathbb{R}^d$; (iii) if no $(\ell-1)$-dimensional hyperplane of $\mathbb{R}^d$ has P-probability one, then $\mu \mapsto O_{\ell,P}(\mu)$ is coercive, in the sense that $(O_{\ell,P}(\mu_k)) \to \infty$ for any sequence (μ_k) such that $(\|\mu_k\|) \to \infty$.*

Proof of Lemma A.4. (i) For any $\mu \in \mathbb{R}^d$, one has

$$\mathrm{E}_P[m_\ell(\mathrm{Simpl}(X_1, \ldots, X_\ell, \mu))] = \frac{1}{\ell!}\mathrm{E}_P[m_\ell(\mathrm{Parall}(X_1 - \mu, \ldots, X_\ell - \mu, 0))]$$

$$\leq \frac{1}{\ell!}\prod_{i=1}^\ell \mathrm{E}_P[\|X_i - \mu\|] = \frac{1}{\ell!}(\mathrm{E}_P[\|X_1 - \mu\|])^\ell,$$

which is finite since P admits finite first moments. (ii) Lemma A.2 readily entails that, for any $x_1, \ldots, x_\ell \in \mathbb{R}^d$, the mapping $\mu \mapsto m_\ell(\mathrm{Simpl}(x_1, \ldots, x_\ell, \mu))$ is convex over $\mathbb{R}^d$. It directly follows that

$$\mu \mapsto O_{\ell,P}(\mu) = \int_{\mathbb{R}^{\ell d}} m_\ell(\mathrm{Simpl}(x_1, \ldots, x_\ell, \mu)) \, dP(x_1) \ldots dP(x_\ell)$$

is convex over $\mathbb{R}^d$. (iii) Ad absurdum, let (μ_k) be a sequence such that $(\|\mu_k\|) \to \infty$ and for which $(O_{\ell,P}(\mu_k))$ does not diverge to infinity. Since the unit sphere $\mathcal{S}^{d-1}$ of $\mathbb{R}^d$ is compact, we may

assume, upon extraction of a subsequence, that $(\mu_k/\|\mu_k\|)$ converges to $u \in \mathcal{S}^{d-1}$ (note that $\|\mu_k\| > 0$ for k large enough). Still with $\Gamma = \Gamma_{x_1,\ldots,x_\ell}$, Lemma A.2 then provides

$$\frac{O_{\ell,P}(\mu_k)}{\|\mu_k\|} = \frac{1}{\ell} \int_{\mathbb{R}^{\ell d}} m_{\ell-1}(\mathrm{Simpl}(x_1,\ldots,x_\ell)) \frac{\|\Gamma(\mu_k - x_\ell)\|}{\|\mu_k\|} \, dP(x_1) \ldots dP(x_\ell),$$

so that Fatou's lemma yields

$$\liminf_{k\to\infty} \frac{O_{\ell,P}(\mu_k)}{\|\mu_k\|} \geq \frac{1}{\ell} \int_{\mathbb{R}^{\ell d}} m_{\ell-1}(\mathrm{Simpl}(x_1,\ldots,x_\ell)) \left(\liminf_{k\to\infty} \frac{\|\Gamma(\mu_k - x_\ell)\|}{\|\mu_k\|} \right) dP(x_1) \ldots dP(x_\ell)$$

$$= \frac{1}{\ell} \int_{\mathbb{R}^{\ell d}} m_{\ell-1}(\mathrm{Simpl}(x_1,\ldots,x_\ell)) \|\Gamma u\| \, dP(x_1) \ldots dP(x_\ell)$$

$$=: \mathcal{I},$$

say. If $\ell = 1$, then $\mathcal{I} = \|u\| > 0$, which implies that $(O_{\ell,P}(\mu_k))$ diverges to infinity, a contradiction. Assume then that $\ell \geq 2$. We trivially have

$$\mathcal{I} \geq \frac{1}{\ell} \int_{B_u} m_{\ell-1}(\mathrm{Simpl}(x_1 - x_\ell, \ldots, x_{\ell-1} - x_\ell, 0)) \|\Gamma u\| \, dP(x_1) \ldots dP(x_\ell),$$

where B_u collects the values of $x = (x_1, x_2, \ldots, x_\ell)$ for which $u, x_1 - x_\ell, x_2 - x_\ell, \ldots, x_{\ell-1} - x_\ell$ span a vector space of dimension ℓ. Clearly, the integrand in the last integral is everywhere positive and, from Lemma A.3, B_u has a positive P-probability. Thus, $\mathcal{I} > 0$, which implies that $(O_{\ell,P}(\mu_k))$ diverges to infinity, a contradiction. This establishes the result. □

We can now prove Theorem 5.1.

Proof of Theorem 5.1. We consider two cases. (i) If there exists an $(\ell - 1)$-dimensional hyperplane H such that $P[H] = 1$, then any $\mu \in H$ satisfies $O_{\ell,P}(\mu) = \mathrm{E}_P[m_\ell(\mathrm{Simpl}(X_1,\ldots,X_\ell,\mu))] = 0$, hence is a minimizer of $\mu \mapsto O_{\ell,P}(\mu)$. (ii) If there is no such hyperplane H, then $\mu \mapsto O_{\ell,P}(\mu)$ is continuous (since it is convex) and coercive; see Lemma A.4. Coercivity implies that there exists $R > 0$ such that $O_{\ell,P}(\mu) \geq O_{\ell,P}(0)$ for any $\mu \notin B(0, R) = \{x \in \mathbb{R}^d : \|x\| \leq R\}$. Therefore, any minimum of $\mu \mapsto O_{\ell,P}(\mu)$ in $B(0, R)$ (existence follows from continuity of $\mu \mapsto O_{\ell,P}(\mu)$ and compactness of $B(0, R)$) is a minimum of $\mu \mapsto O_{\ell,P}(\mu)$ over $\mathbb{R}^d$. □

Lemmas A.5 and A.7 below are needed to prove Theorem 5.2.

Lemma A.5. *Fix an integer $d \geq 2$ and $\ell \in \{1, \ldots, d - 1\}$. Let P be a probability measure over $\mathbb{R}^d$ that admits finite first moments, that is $(\ell - 1)$-smooth at any μ in an open neighbourhood of μ_0, and such that*

$$\mathrm{E}_P\left[\frac{m_{\ell-1}(\mathrm{Simpl}(X_1,\ldots,X_\ell))}{\|\Gamma_{X_1,\ldots,X_\ell}(X_\ell - \mu_0)\|} \mathbb{I}[\Gamma_{X_1,\ldots,X_\ell}(X_\ell - \mu_0) \neq 0] \right] < \infty.$$

Then, $\mu \mapsto O_{\ell,P}(\mu)$ is twice differentiable at μ_0, with gradient

$$\mathrm{E}[\nabla_\mu m_\ell(\mathrm{Simpl}(X_1,\ldots,X_\ell,\mu))|_{\mu=\mu_0} \mathbb{I}[(X_1,\ldots,X_\ell) \notin C_{\mu_0}]] \tag{A.2}$$

and Hessian matrix

$$E[\nabla^2_\mu m_\ell(\mathrm{Simpl}(X_1,\ldots,X_\ell,\mu))|_{\mu=\mu_0}\mathbb{I}[(X_1,\ldots,X_\ell)\notin C_{\mu_0}]], \tag{A.3}$$

with $C_\mu := \{x = (x_1,\ldots,x_\ell) \in \mathbb{R}^{\ell d} : \Gamma_{x_1,\ldots,x_\ell}(x_\ell - \mu) = 0\}$.

Proof of Lemma A.5. Let V be an open neighbourhood of μ_0 such that P is $(\ell-1)$-smooth at any $\mu \in V$. Fix $\mu_1 \in V$, $v \in \mathbb{R}^d \setminus \{0\}$, and $x = (x_1,\ldots,x_\ell) \notin C_{\mu_1}$. Writing $\Gamma = \Gamma_{x_1,\ldots,x_\ell}$, $m_\ell(x,\mu) := m_\ell(\mathrm{Simpl}(x_1,\ldots,x_\ell,\mu))$ and $m_{\ell-1}(x) := m_{\ell-1}(\mathrm{Simpl}(x_1,\ldots,x_\ell))$, it follows from Lemma A.2 that

$$\frac{1}{t}(m_\ell(x,\mu_1+tv) - m_\ell(x,\mu_1)) = \frac{1}{t\ell}m_{\ell-1}(x)(\|\Gamma(\mu_1+tv-x_\ell)\| - \|\Gamma(\mu_1-x_\ell)\|)$$

$$= \frac{1}{t\ell}m_{\ell-1}(x)\frac{t^2 v'\Gamma v + 2tv'\Gamma(\mu_1-x_\ell)}{\|\Gamma(\mu_1+tv-x_\ell)\| + \|\Gamma(\mu_1-x_\ell)\|}$$

$$\to \frac{1}{\ell}m_{\ell-1}(x)\frac{v'\Gamma(\mu_1-x_\ell)}{\|\Gamma(\mu_1-x_\ell)\|}$$

as $t \to 0$. Thus, for any $x \notin C_{\mu_1}$, the mapping $\mu \mapsto m_\ell(x,\mu)$ admits a directional derivative in direction v at μ_1, given by

$$\frac{\partial m_\ell(x,\mu_1)}{\partial v} = \frac{1}{\ell}m_{\ell-1}(x)\frac{v'\Gamma(\mu_1-x_\ell)}{\|\Gamma(\mu_1-x_\ell)\|}, \tag{A.4}$$

which in particular makes the integrand in (A.2) well-defined (for $x \in C_{\mu_1}$, the gradient in (A.2) may simply be defined as the zero vector, say). Now, note that $x = (x_1,\ldots,x_\ell) \in C_{\mu_1}$ if and only if x_ℓ belongs to the smallest hyperplane, $H(x_1,\ldots,x_{\ell-1},\mu_1)$ say, containing $x_1,\ldots,x_{\ell-1},\mu$, so that $(\ell-1)$-smoothness of P at μ_1 implies that

$$P[X = (X_1,\ldots,X_\ell) \in C_{\mu_1}] = P[X_\ell \in H(X_1,\ldots,X_{\ell-1},\mu_1)] \tag{A.5}$$

$$= \int_{\mathbb{R}^{(\ell-1)d}} P[X_\ell \in H(x_1,\ldots,x_{\ell-1},\mu_1)]\,dP(x_1)\ldots dP(x_{\ell-1}) = 0.$$

Therefore, writing $dP(x)$ instead of $dP(x_1)\ldots dP(x_\ell)$,

$$\int_{\mathbb{R}^{\ell d}} \left|\frac{1}{t}(m_\ell(x,\mu_1+tv) - m_\ell(x,\mu_1)) - \frac{\partial m_\ell(x,\mu_1)}{\partial v}\mathbb{I}[x \notin C_{\mu_1}]\right|dP(x)$$

$$= \frac{1}{\ell}\int_{\mathbb{R}^{\ell d}\setminus C_{\mu_1}} m_{\ell-1}(x)\left|\frac{\|\Gamma(\mu_1+tv-x_\ell)\| - \|\Gamma(\mu_1-x_\ell)\|}{t} - \frac{v'\Gamma(\mu_1-x_\ell)}{\|\Gamma(\mu_1-x_\ell)\|}\right|dP(x) \to 0,$$

in view of the Lebesgue's dominated convergence theorem; indeed, the inequality $|\|y\| - \|z\|| \le \|y - z\|$ and the Cauchy–Schwarz inequality show that the integrand in the last integral is upper-bounded by the function $x \mapsto 2m_{\ell-1}(x)\|v\|$ that is P-integrable (see Dürre and Paindaveine [11]) and does not

depend on t. Thus,

$$\frac{O_{\ell,P}(\mu_1 + tv) - O_{\ell,P}(\mu_1)}{t} - \mathrm{E}_P\left[\frac{\partial m_\ell(X, \mu_1)}{\partial v}\mathbb{I}[X \notin C_{\mu_1}]\right]$$

$$= \int_{\mathbb{R}^{\ell d}}\left\{\frac{1}{t}(m_\ell(x, \mu_1 + tv) - m_\ell(x, \mu_1)) - \frac{\partial m_\ell(x, \mu_1)}{\partial v}\mathbb{I}[x \notin C_{\mu_1}]\right\}dP(x) \to 0.$$

This proves that the mapping $\mu \mapsto O_{\ell,P}(\mu)$ admits a directional derivative in direction v at μ_1, given by (see (A.4))

$$\mathrm{E}_P\left[\frac{\partial m_\ell(X, \mu_1)}{\partial v}\mathbb{I}[X \notin C_{\mu_1}]\right]$$

$$= \frac{1}{\ell}\mathrm{E}_P\left[m_{\ell-1}(\mathrm{Simpl}(X_1, \ldots, X_\ell))\frac{v'\Gamma_{X_1,\ldots,X_\ell}(\mu_1 - X_\ell)}{\|\Gamma_{X_1,\ldots,X_\ell}(\mu_1 - X_\ell)\|}\mathbb{I}[(X_1, \ldots, X_\ell) \notin C_{\mu_1}]\right].$$

We turn to second-order differentiability at μ_0. Proceeding as above, it is easy to show that, for any $x \notin C_{\mu_0}$,

$$\frac{\partial^2 m_\ell(x, \mu_0)}{\partial w \partial v} = \lim_{t \to 0}\frac{1}{t}\left(\frac{\partial m_\ell(x, \mu_0 + tw)}{\partial v} - \frac{\partial m_\ell(x, \mu_0)}{\partial v}\right)$$

$$= \frac{1}{\ell\|\Gamma(\mu_0 - x_\ell)\|}m_{\ell-1}(x)v'\Gamma\left(I_d - \frac{(\mu_0 - x_\ell)(\mu_0 - x_\ell)'}{\|\Gamma(\mu_0 - x_\ell)\|^2}\right)\Gamma w,$$

so that the integrand in (A.3) is well-defined (parallel to the gradient in (A.2), the Hessian matrix in (A.3) may be defined as the zero matrix for $x \in C_{\mu_0}$). Note that for any $x \notin C_{\mu_0}$, using the Cauchy–Schwarz inequality, the inequality $\|(v/\|v\|) - (w/\|w\|)\| \le 2\|v - w\|/\|w\|$, and the inequality $\|\|y\| - \|z\|\| \le \|y - z\|$, we have

$$\left|\frac{1}{t}\left(\frac{v'\Gamma(\mu_0 + tw - x_\ell)}{\|\Gamma(\mu_0 + tw - x_\ell)\|} - \frac{v'\Gamma(\mu_0 - x_\ell)}{\|\Gamma(\mu_0 - x_\ell)\|}\right) - \frac{1}{\|\Gamma(\mu_0 - x_\ell)\|}v'\Gamma\left(I_d - \frac{(\mu_0 - x_\ell)(\mu_0 - x_\ell)'}{\|\Gamma(\mu_0 - x_\ell)\|^2}\right)\Gamma w\right|$$

$$\le \frac{1}{|t|}\|v\|\left\|\frac{\Gamma(\mu_0 + tw - x_\ell)}{\|\Gamma(\mu_0 + tw - x_\ell)\|} - \frac{\Gamma(\mu_0 - x_\ell)}{\|\Gamma(\mu_0 - x_\ell)\|}\right\| + \frac{2\|\Gamma v\|\|\Gamma w\|}{\|\Gamma(\mu_0 - x_\ell)\|}$$

$$\le \frac{2}{|t|}\|v\|\frac{\|\Gamma(\mu_0 + tw - x_\ell) - \Gamma(\mu_0 - x_\ell)\|}{\|\Gamma(\mu_0 - x_\ell)\|} + \frac{2\|v\|\|w\|}{\|\Gamma(\mu_0 - x_\ell)\|}$$

$$= 2\|v\|\frac{\|\Gamma w\|}{\|\Gamma(\mu_0 - x_\ell)\|} + \frac{2\|v\|\|w\|}{\|\Gamma(\mu_0 - x_\ell)\|}$$

$$\le \frac{4\|v\|\|w\|}{\|\Gamma(\mu_0 - x_\ell)\|},$$

where, by assumption, the last expression defines, after multiplication by $m_{\ell-1}(x)$, a P-integrable function of $x = (x_1, \dots, x_\ell)$ over $\mathbb{R}^{\ell d} \setminus C_{\mu_0}$. Since this function does not depend on t, Lebesgue's dominated convergence theorem entails that

$$\int_{\mathbb{R}^{\ell d}} \left| \frac{1}{t} \left(\frac{\partial m_\ell(x, \mu_0 + tw)}{\partial v} - \frac{\partial m_\ell(x, \mu_0)}{\partial v} \right) - \frac{\partial^2 m_\ell(x, \mu_0)}{\partial v \partial w} \mathbb{I}[x \notin C_{\mu_0}] \right| dP(x)$$

$$= \frac{1}{\ell} \int_{\mathbb{R}^{\ell d} \setminus C_{\mu_0}} m_{\ell-1}(x) \left| \frac{1}{t} \left(\frac{v' \Gamma(\mu_0 + tw - x_\ell)}{\|\Gamma(\mu_0 + tw - x_\ell)\|} - \frac{v' \Gamma(\mu_0 - x_\ell)}{\|\Gamma(\mu_0 - x_\ell)\|} \right) \right.$$

$$\left. - \frac{1}{\|\Gamma(\mu_0 - x_\ell)\|} v' \Gamma \left(I_d - \frac{(\mu_0 - x_\ell)(\mu_0 - x_\ell)'}{\|\Gamma(\mu_0 - x_\ell)\|^2} \right) \Gamma w \right| dP(x) \to 0,$$

as $t \to 0$, where we used $(\ell - 1)$-smoothness of P at μ_0. Therefore,

$$\frac{1}{t} \left(\frac{\partial O_{\ell, P}(\mu_0 + tw)}{\partial v} - \frac{\partial O_{\ell, P}(\mu_0)}{\partial v} \right) - \mathrm{E}_P \left[\frac{\partial^2 m_\ell(X, \mu_0)}{\partial v \partial w} \mathbb{I}[X \notin C_{\mu_0}] \right]$$

$$= \int_{\mathbb{R}^{\ell d}} \left\{ \frac{1}{t} \left(\frac{\partial m_\ell(x, \mu_0 + tw)}{\partial v} - \frac{\partial m_\ell(x, \mu_0)}{\partial v} \right) - \frac{\partial^2 m_\ell(x, \mu_0)}{\partial v \partial w} \mathbb{I}[x \notin C_{\mu_0}] \right\} dP(x) \to 0.$$

Thus, the mapping $\mu \mapsto O_{\ell, P}(\mu)$ is twice differentiable at μ_0, with Hessian matrix $H_P = (H_{P,ij})$, $i, j = 1, \dots, d$, where

$$H_{P,ij} = \mathrm{E}_P \left[\frac{\partial^2 m_\ell(X, \mu_0)}{\partial e_{id} \partial e_{jd}} \mathbb{I}[X \notin C_{\mu_0}] \right] \tag{A.6}$$

$$= \frac{1}{\ell} \mathrm{E}_P \left[\frac{m_{\ell-1}(\mathrm{Simpl}(X_1, \dots, X_\ell))}{\|\Gamma_X(\mu_0 - X_\ell)\|} \right.$$

$$\left. \times e_{id}' \Gamma_X \left(I_d - \frac{(\mu_0 - X_\ell)(\mu_0 - X_\ell)'}{\|\Gamma_X(\mu_0 - X_\ell)\|^2} \right) \Gamma_X e_{jd} \mathbb{I}[(X_1, \dots, X_\ell) \notin C_{\mu_0}] \right]$$

involves $\Gamma_X = \Gamma_{X_1, \dots, X_\ell}$ and where e_{id} denotes the ith vector of the canonical basis of $\mathbb{R}^d$. $\qquad \square$

The proof of Lemma A.7 requires the following preliminary result (its proof, that can be done strictly along the same lines as the proof of Lemma A.3, is omitted).

Lemma A.6. *Fix an integer $d \geq 2$, $\ell \in \{1, \dots, d-1\}$ and $v \in \mathbb{R}^d \setminus \{0\}$. Let P be a probability measure over $\mathbb{R}^d$ such that no ℓ-dimensional hyperplane containing $\mu(\in \mathbb{R}^d)$ has P-probability one. Let $X_1, \dots, X_\ell$ form a random sample from P. Then, the probability that $v, X_1 - \mu, \dots, X_\ell - \mu$ span a vector space of dimension $\ell + 1$ is positive.*

Lemma A.7. *Let the assumptions of Theorem 5.2 hold. Then, the Hessian matrix $H_P = (H_{P,ij})$ (see (A.6)) is positive definite.*

Proof of Lemma A.7. As the Hessian matrix of a convex function, H_P is positive semi-definite. Ad absurdum, assume then that there exists $v \in \mathbb{R}^d \setminus \{0\}$ such that

$$v'H_P v = \frac{1}{\ell} \mathrm{E}_P \left[\frac{m_{\ell-1}(\mathrm{Simpl}(X_1, \ldots, X_\ell))}{\|\Gamma_X(X_\ell - \mu_{\ell,P})\|} \right. \tag{A.7}$$

$$\left. \times v' \Gamma_X \left(I_d - \frac{(X_\ell - \mu_{\ell,P})(X_\ell - \mu_{\ell,P})'}{\|\Gamma_X(X_\ell - \mu_{\ell,P})\|^2} \right) \Gamma_X v \mathbb{I}[\Gamma_X(X_\ell - \mu_{\ell,P}) \neq 0] \right] = 0.$$

Note that $(\ell-1)$-smoothness of P at $\mu_{\ell,P}$ implies that $P[m_\ell(\mathrm{Simpl}(X_1, \ldots, X_\ell, \mu_{\ell,P})) = 0] = 0$ (this is shown by conditioning with respect to $X_1, \ldots, X_{\ell-1}$, as in (A.5)), which, in view of Lemma A.2, implies that $P[m_{\ell-1}(\mathrm{Simpl}(X_1, \ldots, X_\ell)) = 0] = 0$ and $P[\Gamma_X(X_\ell - \mu_{\ell,P}) = 0] = 0$. Consequently, (A.7) yields that

$$\left(I_d - \frac{(X_\ell - \mu_{\ell,P})(X_\ell - \mu_{\ell,P})'}{\|\Gamma_X(X_\ell - \mu_{\ell,P})\|^2} \right) \Gamma_X v \mathbb{I}[\Gamma_X(X_\ell - \mu_{\ell,P}) \neq 0] = 0$$

with P-probability one, which, since Γ_X is symmetric and idempotent, implies that

$$\left(I_d - \frac{\Gamma_X(X_\ell - \mu_{\ell,P})(\Gamma_X(X_\ell - \mu_{\ell,P}))'}{\|\Gamma_X(X_\ell - \mu_{\ell,P})\|^2} \right) \Gamma_X v \mathbb{I}[\Gamma_X(X_\ell - \mu_{\ell,P}) \neq 0] = 0$$

P-almost surely. Since $\Gamma_X(X_\ell - \mu_{\ell,P}) \neq 0$ with P-probability one, $\Gamma_X v$ must be proportional to $\Gamma_X(X_\ell - \mu_{\ell,P})$ with P-probability one. Consequently, there must exist $\lambda_1, \ldots, \lambda_\ell \in \mathbb{R}$ (that may depend on $X_1, \ldots, X_\ell$) such that $v - \lambda_\ell(X_\ell - \mu_{\ell,P}) = \sum_{i=1}^{\ell-1} \lambda_i(X_i - X_\ell)$ P-almost surely, that is, such that

$$v = \sum_{i=1}^{\ell-1} \lambda_i(X_i - \mu_{\ell,P}) + \left(\lambda_\ell - \sum_{i=1}^{\ell-1} \lambda_i \right)(X_\ell - \mu_{\ell,P}),$$

P-almost surely. Thus, with P-probability one, $v, X_1 - \mu_{\ell,P}, \ldots, X_\ell - \mu_{\ell,P}$ span a vector space of dimension strictly less than $\ell + 1$, which, in view of Lemma A.6, contradicts the assumption that no ℓ-dimensional hyperplane containing $\mu_{\ell,P}$ has P-probability one. $\qquad\square$

We can now prove Theorem 5.2.

Proof of Theorem 5.2. All convergences in this proof are as $n \to \infty$. Let $G_n(\mu) := n O_\ell^{(n)}(\mu)$,

$$\eta_n = \frac{\sqrt{n}}{\binom{n}{\ell}} \sum_{1 \leq i_1 < \cdots < i_\ell \leq n} \nabla_\mu m_\ell(\mathrm{Simpl}(X_{i_1}, \ldots, X_{i_\ell}, \mu))|_{\mu = \mu_{\ell,P}} \mathbb{I}[(X_{i_1}, \ldots, X_{i_\ell}) \notin C] \tag{A.8}$$

and

$$V_n = \frac{1}{2\binom{n}{\ell}} \sum_{1 \leq i_1 < \cdots < i_\ell \leq n} \nabla_\mu^2 m_\ell(\mathrm{Simpl}(X_{i_1}, \ldots, X_{i_\ell}, \mu))|_{\mu = \mu_{\ell,P}} \mathbb{I}[(X_{i_1}, \ldots, X_{i_\ell}) \notin C], \tag{A.9}$$

with $C := C_{\mu_{\ell,P}}$ (see Lemma A.5 for the definition of C_μ). To make the notation lighter, we write I in the rest of the proof for a multivariate index $(i_1 < \cdots < i_\ell)$ as in (A.8)–(A.9) and $\mathcal{I}_n$ for the corresponding collection of I's. For any I, we then denote as X_I the collection of observations indexed

by I, that is, $(X_{i_1}, \ldots, X_{i_\ell})$. Finally, X will stand for $(X_1, \ldots, X_\ell)$. With this notation, write, for any d-vector h,

$$R_n := G_n(\mu_{\ell,P} + n^{-1/2}h) - G_n(\mu_{\ell,P}) - h'\eta_n - h'V_n h = R_{n1} + R_{n2},$$

with

$$R_{n1} = \frac{n}{\binom{n}{\ell}} \sum_{I \in \mathcal{I}_n} \Big\{ m_\ell(\mathrm{Simpl}(X_I, \mu_{\ell,P} + n^{1/2}h)) - m_\ell(\mathrm{Simpl}(X_I, \mu_{\ell,P}))$$

$$- \frac{1}{\sqrt{n}} h'\nabla_\mu m_\ell(\mathrm{Simpl}(X_I, \mu))|_{\mu=\mu_{\ell,P}} - \frac{1}{2n} h'\nabla_\mu^2 m_\ell(\mathrm{Simpl}(X_I, \mu))|_{\mu=\mu_{\ell,P}} h \Big\} \mathbb{I}[X_I \notin C]$$

and

$$R_{n2} = \frac{n}{\binom{n}{\ell}} \sum_{I \in \mathcal{I}_n} \Big\{ m_\ell(\mathrm{Simpl}(X_I, \mu_{\ell,P} + n^{1/2}h)) - m_\ell(\mathrm{Simpl}(X_I, \mu_{\ell,P})) \Big\} \mathbb{I}[X_I \in C].$$

Since $(\ell - 1)$-smoothness of P at $\mu_{\ell,P}$ implies that $P[X \in C] = 0$ (see (A.5)), we have that, for any $\varepsilon > 0$,

$$P[|R_{n2}| > \varepsilon] \leq P[R_{n2} \neq 0] \leq P[\cup_{I \in \mathcal{I}_n} [X_I \in C]] \leq \sum_{I \in \mathcal{I}_n} P[X_I \in C] = 0,$$

so that $R_{n2} = o_P(1)$. Let us thus focus on R_{n1}, which we write as

$$R_{n1} = \frac{1}{\binom{n}{\ell}} \sum_{I \in \mathcal{I}_n} (J_n(X_I) - J(X_I)),$$

with

$$J_n(x) := n \big\{ m_\ell(\mathrm{Simpl}(x, \mu_{\ell,P} + n^{-1/2}h)) - m_\ell(\mathrm{Simpl}(x, \mu_{\ell,P}))$$

$$- \frac{1}{\sqrt{n}} h'\nabla_\mu m_\ell(\mathrm{Simpl}(x, \mu))|_{\mu=\mu_{\ell,P}} \big\} \mathbb{I}[x \notin C]$$

and

$$J(x) := \frac{1}{2} h'\nabla_\mu^2 m_\ell(\mathrm{Simpl}(x, \mu))|_{\mu=\mu_{\ell,P}} h \, \mathbb{I}[x \notin C].$$

Lemma A.5 entails that $J_n(x) \to J(x)$ for any x and that

$$\mathrm{E}_P[J_n(X)] - \mathrm{E}_P[J(X)] = n \Big\{ O_{\ell,P}(\mu_{\ell,P} + n^{-1/2}h) - O_{\ell,P}(\mu_{\ell,P}) - \frac{1}{\sqrt{n}} h'\nabla_\mu O_{\ell,P}(\mu))|_{\mu=\mu_{\ell,P}}$$

$$- \frac{1}{2n} h'\nabla_\mu^2 O_{\ell,P}(\mu))|_{\mu=\mu_{\ell,P}} h \Big\} \to 0.$$

Convexity implies that $J_n(x)$ and $J(x)$ are non-negative for any x, so that Scheffé's lemma entails that $\mathrm{E}_P[|J_n(X) - J(X)|] = o(1)$. It follows that $\mathrm{E}_P[|R_{n1}|] \leq \mathrm{E}_P[|J_n(X) - J(X)|] = o(1)$, so that R_{n1}, hence also R_n, are $o_P(1)$.

The consistency result for U-statistics on page 190 from Serfling [46] and Lemma A.5 then yield

$$
\begin{aligned}
V_n &= \frac{1}{2\binom{n}{\ell}} \sum_{I \in \mathcal{I}_n} \nabla_\mu^2 m_\ell(\mathrm{Simpl}(X_I, \mu))|_{\mu=\mu_{\ell,P}} \mathbb{I}[X_I \notin C] \\
&= \frac{1}{2} \mathrm{E}[\nabla_\mu^2 m_\ell(\mathrm{Simpl}(X_1, \ldots, X_\ell, \mu))|_{\mu=\mu_{\ell,P}} \mathbb{I}[(X_1, \ldots, X_\ell) \notin C]] + o_\mathrm{P}(1) \\
&= \frac{1}{2} \nabla_\mu^2 \mathrm{E}[m_\ell(\mathrm{Simpl}(X_1, \ldots, X_\ell, \mu))]|_{\mu=\mu_{\ell,P}} + o_\mathrm{P}(1) \\
&= \frac{1}{2} H_P + o_\mathrm{P}(1),
\end{aligned}
\tag{A.10}
$$

where, in view of Lemma A.7, H_P is invertible. Now, since $\mu_{\ell,P}$ is a minimizer of $\mu \mapsto O_{\ell,P}(\mu)$, Lemma A.5 provides

$$
\mathrm{E}[\nabla_\mu m_\ell(\mathrm{Simpl}(X_1, \ldots, X_\ell, \mu))|_{\mu=\mu_{\ell,P}} \mathbb{I}[(X_1, \ldots, X_\ell) \notin C]] = \nabla_\mu O_{\ell,P}(\mu)|_{\mu=\mu_{\ell,P}} = 0.
$$

The usual Hoeffding's representation result for U-statistics (see, e.g., Theorem 12.3 from van der Vaart [48]) then entails that

$$
\eta_n - \frac{\ell}{\sqrt{n}} \sum_{i=1}^n S_P(X_i) = o_\mathrm{P}(1),
\tag{A.11}
$$

where we let

$$
S_P(x_\ell) := \mathrm{E}_P[\nabla_\mu m_\ell(\mathrm{Simpl}(X_1, \ldots, X_{\ell-1}, x_\ell, \mu))|_{\mu=\mu_{\ell,P}} \mathbb{I}[(X_1, \ldots, X_{\ell-1}, x_\ell) \notin C]].
$$

Writing $\Gamma = \Gamma_{X_1, \ldots, X_{\ell-1}, x_\ell}$, working as in Lemma A.5 then yields

$$
\begin{aligned}
S_P(x_\ell) &= \frac{1}{\ell} \mathrm{E}_P\left[m_{\ell-1}(\mathrm{Simpl}(X_1, \ldots, X_{\ell-1}, x_\ell)) \frac{\Gamma(\mu_{\ell,P} - x_\ell)}{\|\Gamma(\mu_{\ell,P} - x_\ell)\|} \mathbb{I}[(X_1, \ldots, X_{\ell-1}, x_\ell) \notin C] \right] \\
&= \frac{1}{\ell} \mathrm{E}_P\left[m_{\ell-1}(\mathrm{Simpl}(X_1, \ldots, X_{\ell-1}, x_\ell)) \frac{\Gamma(\mu_{\ell,P} - x_\ell)}{\|\Gamma(\mu_{\ell,P} - x_\ell)\|} \mathbb{I}[\Gamma(\mu_{\ell,P} - x_\ell) \neq 0] \right]. \quad \text{(A.12)}
\end{aligned}
$$

Since the $S_P(X_i)$'s are independent and identically distributed with a common distribution that, by proceeding strictly as in the proof of Lemma A.4(i), is shown to admit finite second-order moments, we have that $\eta_n = O_\mathrm{P}(1)$. Therefore, Theorem 3 from Arcones [4] applies and yields

$$
\sqrt{n}(\mu_\ell^{(n)} - \mu_{\ell,P}) = -\frac{1}{2} V_n^{-1} \eta_n + o_\mathrm{P}(1),
$$

which, using (A.10)–(A.12), establishes the Bahadur representation result in (5.3). Of course, (5.4) then readily follows from the multivariate central limit theorem.   □

We conclude by establishing two results we used to derive Corollary 5.1 from Theorem 5.2.

Lemma A.8. *Fix a positive integer d and $\ell \in \{1, \ldots, d\}$. Fix $x_1, \ldots, x_\ell \in \mathbb{R}^d$ such that $H(x_1, \ldots, x_\ell)$, the smallest hyperplane containing $x_1, \ldots, x_\ell$, has dimension $\ell - 1$ and fix $\mu \in \mathbb{R}^d \setminus H(x_1, \ldots, x_\ell)$.*

Let $C = (x_1 - x_\ell \ldots x_{\ell-1} - x_\ell)$ and $D = (x_1 - \mu \ldots x_{\ell-1} - \mu)$. Then,

$$m_{\ell-1}(\mathrm{Simpl}(x_1, \ldots, x_\ell)) \frac{\Gamma(x_\ell - \mu)}{\|\Gamma(x_\ell - \mu)\|}$$

$$= m_{\ell-1}(\mathrm{Simpl}(x_1, \ldots, x_{\ell-1}, \mu)) \frac{\Psi(x_\ell - \mu)}{\|\Psi(x_\ell - \mu)\|} + \ell\, m_\ell(\mathrm{Simpl}(x_1, \ldots, x_\ell, \mu)) \Phi D (D'\Phi D)^{-1} 1_{\ell-1},$$

where $\Gamma = I_d - C(C'C)^{-1}C'$ (resp., $\Psi = I_d - D(D'D)^{-1}D'$) is the matrix of the orthogonal projection onto the orthogonal complement of the vector space spanned by the columns of C (resp., by the columns of D) and where $\Phi = I_d - (x_\ell - \mu)(x_\ell - \mu)'/\|x_\ell - \mu\|^2$ is the matrix of the orthogonal projection onto the orthogonal complement of the vector space spanned by $x_\ell - \mu$.

Proof of Lemma A.8. Define

$$v := \frac{\Gamma(x_\ell - \mu)}{\|\Gamma(x_\ell - \mu)\|^2} \quad \text{and} \quad w := \frac{\Psi(x_\ell - \mu)}{\|\Psi(x_\ell - \mu)\|^2} + \Phi D (D'\Phi D)^{-1} 1_{\ell-1}.$$

Since

$$D'\Gamma(x_\ell - \mu) - 1_{\ell-1}(x_\ell - \mu)'\Gamma(x_\ell - \mu) = (D - (x_\ell - \mu)1'_{\ell-1})'\Gamma(x_\ell - \mu) = C'\Gamma(x_\ell - \mu) = 0,$$

we have $D'\Gamma(x_\ell - \mu)/\|\Gamma(x_\ell - \mu)\|^2 = 1_{\ell-1}$, which shows that $(x_i - \mu)'v = (x_i - \mu)'w$ for any $i = 1, \ldots, \ell - 1$. Since a direct computation yields $(x_\ell - \mu)'v = (x_\ell - \mu)'w$, we have that $y'v = y'w$ for any y in the span $\mathcal{V}$ of $x_1 - \mu, \ldots, x_{\ell-1} - \mu, x_\ell - \mu$. Now, for any y in the orthogonal complement of $\mathcal{V}$,

$$y'v = \frac{y'(x_\ell - \mu)}{\|\Gamma(x_\ell - \mu)\|^2} = 0 = \frac{y'(x_\ell - \mu)}{\|\Psi(x_\ell - \mu)\|^2} + y'D(D'\Phi D)^{-1}1_{\ell-1} = y'w,$$

so that $v = w$. This implies that

$$\ell\, m_\ell(\mathrm{Simpl}(x_1, \ldots, x_\ell, \mu)) \frac{\Gamma(x_\ell - \mu)}{\|\Gamma(x_\ell - \mu)\|^2}$$

$$= \ell\, m_\ell(\mathrm{Simpl}(x_1, \ldots, x_{\ell-1}, \mu, x_\ell)) \frac{\Psi(x_\ell - \mu)}{\|\Psi(x_\ell - \mu)\|^2} + \ell\, m_\ell(\mathrm{Simpl}(x_1, \ldots, x_\ell, \mu)) \Phi D (D'\Phi D)^{-1} 1_{\ell-1}.$$

The result then follows from Lemma A.2. $\qquad\square$

Lemma A.9. *Fix an integer $d \geq 2$ and $k \in \{1, \ldots, d - 1\}$. Let P be a probability measure over $\mathbb{R}^d$ that is spherically symmetric about the origin of $\mathbb{R}^d$ and satisfies $P[\{0\}] = 0$. Let $X_1, \ldots, X_k$ form a random sample from P and let $\mathcal{V}$ be the vector space spanned by $X_1, \ldots, X_k$. Denote as Ψ the matrix of the orthogonal projection onto the orthogonal complement of $\mathcal{V}$. Then,*

$$\mathrm{E}\left[\frac{\Psi x}{\|\Psi x\|}\right] = \frac{\Gamma(\frac{d-k+1}{2})\Gamma(\frac{d}{2})}{\Gamma(\frac{d-k}{2})\Gamma(\frac{d+1}{2})} \frac{x}{\|x\|}$$

for any non-zero d-vector x, where Γ is the Euler Gamma function.

Proof of Lemma A.9. Fix an arbitrary non-zero d-vector x. By assumption, $\Psi = O \Lambda O'$ almost surely, where Λ is idempotent with trace $d - k$ and O follows the Haar distribution on the collection of $d \times d$ orthogonal matrices. In particular, $P \Psi P' =_{\mathcal{D}} P$ for any $d \times d$ orthogonal matrix P. For any such P, the quantity $g(x) := \mathrm{E}[\Psi x / \|\Psi x\|]$ satisfies

$$g(Px) = P\mathrm{E}\left[\frac{P'\Psi P x}{\|P'\Psi P x\|}\right] = P\mathrm{E}\left[\frac{\Psi x}{\|\Psi x\|}\right] = Pg(x).$$

Fix then $v \in \mathbb{R}^d$ with $v'x = 0$. Picking P so that $Px = x$ and $Pv = -v$, we obtain

$$v'g(x) = v'g(Px) = v'Pg(x) = -v'g(x),$$

so that $v'g(x) = 0$. Therefore, $g(x) = \lambda x / \|x\|$ for some $\lambda \in \mathbb{R}$. To determine λ, note that, since Ψ is symmetric and idempotent, we have, with $u := x / \|x\|$,

$$\lambda = \frac{1}{\|x\|}x'g(x) = \mathrm{E}\left[\frac{x'\Psi x}{\|x\|\|\Psi x\|}\right] = \mathrm{E}\left[\frac{\|\Psi x\|}{\|x\|}\right] = \mathrm{E}[\|O\Lambda O'u\|] = \mathrm{E}[\|\Lambda U\|],$$

where $U = (U_1, \ldots, U_d)'$ is uniformly distributed over the unit sphere of $\mathbb{R}^d$. Since Theorem 1.5.7(ii) in Muirhead [32] entails that $\|\Lambda U\|^2 = \sum_{r=1}^{d-k} U_r^2$ is Beta$(\frac{d-k}{2}, \frac{k}{2})$, the result then follows. $\qquad\square$

Acknowledgements

Funding

References

[1] Abramowitz, M. and Stegun, I.A. (1970). *Handbook of Mathematical Functions*, 9th ed. New York: Dover Publications. MR0208797

[2] Anderson, W.J. (1986). On certain random simplices in $\mathbb{R}^n$. *J. Multivariate Anal.* **19** 265–272. MR0853057 https://doi.org/10.1016/0047-259X(86)90031-X

[3] Arcones, M.A. (1995). Asymptotic normality of multivariate trimmed means. *Statist. Probab. Lett.* **25** 43–53. MR1364817 https://doi.org/10.1016/0167-7152(94)00204-L

[4] Arcones, M.A. (1998). Asymptotic theory for M-estimators over a convex kernel. *Econometric Theory* **14** 387–422. MR1650029 https://doi.org/10.1017/S0266466698144018

[5] Baryshnikov, Y.M. and Vitale, R.A. (1994). Regular simplices and Gaussian samples. *Discrete Comput. Geom.* **11** 141–147. MR1254086 https://doi.org/10.1007/BF02574000

[6] Bilodeau, M. and Brenner, D. (1999). *Theory of Multivariate Statistics. Springer Texts in Statistics*. New York: Springer. MR1705291

[7] Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge: Cambridge Univ. Press. MR2061575 https://doi.org/10.1017/CBO9780511804441

[8] Brown, B.M. (1983). Statistical uses of the spatial median. *J. Roy. Statist. Soc. Ser. B* **45** 25–30. MR0701072

[9] Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *J. Amer. Statist. Assoc.* **91** 862–872. MR1395753 https://doi.org/10.2307/2291681

[10] Dahel, S. and Giri, N. (1994). Some distributions related to a noncentral Wishart distribution. *Comm. Statist. Theory Methods* **23** 229–237. MR1261743 https://doi.org/10.1080/03610929408831249

[11] Dürre, A. and Paindaveine, D. (2021). Affine-equivariant inference for multivariate location under L_p loss functionals. Submitted.

[12] Epstein, B. (1948). Some applications of the Mellin transform in statistics. *Ann. Math. Stat.* **19** 370–379. MR0029128 https://doi.org/10.1214/aoms/1177730201

[13] Fischer, D., Mosler, K., Möttönen, J., Nordhausen, K., Pokotylo, O. and Vogel, D. (2020). Computing the Oja median in R: The package OjaNP. *J. Stat. Softw.* **92**.

[14] Grote, J., Kabluchko, Z. and Thäle, C. (2019). Limit theorems for random simplices in high dimensions. *ALEA Lat. Am. J. Probab. Math. Stat.* **16** 141–177. MR3903027 https://doi.org/10.30757/alea.v16-06

[15] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986). *Robust Statistics*: *The Approach Based on Influence Functions*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. New York: Wiley. MR0829458

[16] Hettmansperger, T.P., Möttönen, J. and Oja, H. (1997). Affine-invariant multivariate one-sample signed-rank tests. *J. Amer. Statist. Assoc.* **92** 1591–1600. MR1615268 https://doi.org/10.2307/2965430

[17] Hettmansperger, T.P., Nyblom, J. and Oja, H. (1994). Affine invariant multivariate one-sample sign tests. *J. Roy. Statist. Soc. Ser. B* **56** 221–234. MR1257809

[18] Hettmansperger, T.P. and Randles, R.H. (2002). A practical affine equivariant multivariate median. *Biometrika* **89** 851–860. MR1946515 https://doi.org/10.1093/biomet/89.4.851

[19] Hsu, D. and Sabato, S. (2016). Loss minimization and parameter estimation with heavy tails. *J. Mach. Learn. Res.* **17** Paper No. 18, 40. MR3491112

[20] Kabluchko, Z. and Zaporozhets, D. (2020). Absorption probabilities for Gaussian polytopes and regular spherical simplices. *Adv. in Appl. Probab.* **52** 588–616. MR4123647 https://doi.org/10.1017/apr.2020.7

[21] León, C.A. and Massé, J.-C. (1993). La médiane simpliciale d'Oja: Existence, unicité et stabilité. *Canad. J. Statist.* **21** 397–408. MR1254286 https://doi.org/10.2307/3315703

[22] Lopuhaä, H.P. and Rousseeuw, P.J. (1991). Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *Ann. Statist.* **19** 229–248. MR1091847 https://doi.org/10.1214/aos/1176347978

[23] Lugosi, G. and Mendelson, S. (2019). Mean estimation and regression under heavy-tailed distributions: A survey. *Found. Comput. Math.* **19** 1145–1190. MR4017683 https://doi.org/10.1007/s10208-019-09427-x

[24] Magyar, A. and Tyler, D.E. (2011). The asymptotic efficiency of the spatial median for elliptically symmetric distributions. *Sankhya B* **73** 165–192. MR2924190 https://doi.org/10.1007/s13571-011-0032-x

[25] Mathai, A.M. (1982). On a conjecture in geometric probability regarding asymptotic normality of a random simplex. *Ann. Probab.* **10** 247–251. MR0637392

[26] Mathai, A.M. (1999). Random p-content of a p-parallelotope in Euclidean n-space. *Adv. in Appl. Probab.* **31** 343–354. MR1724556 https://doi.org/10.1239/aap/1029955138

[27] Miles, R.E. (1971). Isotropic random simplices. *Adv. in Appl. Probab.* **3** 353–382. MR0309164 https://doi.org/10.2307/1426176

[28] Minsker, S. (2015). Geometric median and robust estimation in Banach spaces. *Bernoulli* **21** 2308–2335. MR3378468 https://doi.org/10.3150/14-BEJ645

[29] Möttönen, J., Hettmansperger, T.P., Oja, H. and Tienari, J. (1998). On the efficiency of affine invariant multivariate rank tests. *J. Multivariate Anal.* **66** 118–132. MR1648529 https://doi.org/10.1006/jmva.1998.1740

[30] Möttönen, J., Nordhausen, K. and Oja, H. (2010). Asymptotic theory of the spatial median. In *Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis*: *A Festschrift in Honor of Professor Jana Jurečková. Inst. Math. Stat.* (*IMS*) *Collect.* **7** 182–193. Beachwood, OH: IMS. MR2808378

[31] Möttönen, J. and Oja, H. (1995). Multivariate spatial sign and rank methods. *J. Nonparametr. Stat.* **5** 201–213. MR1346895 https://doi.org/10.1080/10485259508832643

[32] Muirhead, R.J. (2005). *Aspects of Multivariate Statistical Theory*. Hoboken, NJ: Wiley. MR0652932

[33] Niinimaa, A. and Oja, H. (1995). On the influence functions of certain bivariate medians. *J. Roy. Statist. Soc. Ser. B* **57** 565–574. MR1341324

[34] Niinimaa, A., Oja, H. and Tableman, M. (1990). The finite-sample breakdown point of the Oja bivariate median and of the corresponding half-samples version. *Statist. Probab. Lett.* **10** 325–328. MR1069910 https://doi.org/10.1016/0167-7152(90)90050-H

[35] Oja, H. (1983). Descriptive statistics for multivariate distributions. *Statist. Probab. Lett.* **1** 327–332. MR0721446 https://doi.org/10.1016/0167-7152(83)90054-8

[36] Oja, H. (1999). Affine invariant multivariate sign and rank tests and corresponding estimates: A review. *Scand. J. Stat.* **26** 319–343. MR1712063 https://doi.org/10.1111/1467-9469.00152

[37] Oja, H. (2010). *Multivariate Nonparametric Methods with R: An Approach Based on Spatial Signs and Ranks*. *Lecture Notes in Statistics* **199**. New York: Springer. MR2598854 https://doi.org/10.1007/978-1-4419-0468-3

[38] Oja, H. and Niinimaa, A. (1985). Asymptotic properties of the generalized median in the case of multivariate normality. *J. Roy. Statist. Soc. Ser. B* **47** 372–377. MR0816103

[39] Ollila, E., Oja, H. and Hettmansperger, T.P. (2002). Estimates of regression coefficients based on the sign covariance matrix. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 447–466. MR1924300 https://doi.org/10.1111/1467-9868.00344

[40] Paouris, G., Pivovarov, P. and Valettas, P. (2017). Gaussian convex bodies: A non-asymptotic approach. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov.* (*POMI*) **457** 286–316. MR3723587 https://doi.org/10.1007/s10958-019-04256-3

[41] Paouris, G., Pivovarov, P. and Zinn, J. (2014). A central limit theorem for projections of the cube. *Probab. Theory Related Fields* **159** 701–719. MR3230006 https://doi.org/10.1007/s00440-013-0518-8

[42] Pivovarov, P. (2010). On determinants and the volume of random polytopes in isotropic convex bodies. *Geom. Dedicata* **149** 45–58. MR2737677 https://doi.org/10.1007/s10711-010-9462-2

[43] Ruben, H. (1979). The volume of an isotropic random parallelotope. *J. Appl. Probab.* **16** 84–94. MR0520939 https://doi.org/10.1017/s0021900200046222

[44] Ruben, H. and Miles, R.E. (1980). A canonical decomposition of the probability measure of sets of isotropic random points in $\mathbb{R}^n$. *J. Multivariate Anal.* **10** 1–18. MR0569792 https://doi.org/10.1016/0047-259X(80)90077-9

[45] Serfling, R. (2010). Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardisation. *J. Nonparametr. Stat.* **22** 915–936. MR2738875 https://doi.org/10.1080/10485250903431710

[46] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. *Wiley Series in Probability and Mathematical Statistics*. New York: Wiley. MR0595165

[47] Shen, G. (2008). Asymptotics of Oja median estimate. *Statist. Probab. Lett.* **78** 2137–2141. MR2458023 https://doi.org/10.1016/j.spl.2008.02.004

[48] van der Vaart, A.W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge: Cambridge Univ. Press. MR1652247 https://doi.org/10.1017/CBO9780511802256

[49] Vitale, R.A. (2007). Multivariate medians and measure-symmetrization. In *Complex Datasets and Inverse Problems*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **54** 260–267. Beachwood, OH: IMS. MR2459194 https://doi.org/10.1214/074921707000000201

[50] Wells, W.T., Anderson, R.L. and Cell, J.W. (1962). The distribution of the product of two central or noncentral chi-square variates. *Ann. Math. Stat.* **33** 1016–1020. MR0142170 https://doi.org/10.1214/aoms/1177704469

[51] Zhou, W. and Serfling, R. (2008). Multivariate spatial U-quantiles: A Bahadur-Kiefer representation, a Theil-Sen estimator for multiple regression, and a robust dispersion estimator. *J. Statist. Plann. Inference* **138** 1660–1678. MR2400473 https://doi.org/10.1016/j.jspi.2007.05.043