CrossMark

From Low- to High-Dimensional Moments Without Magic

Bernhard G. Bodmann[1] · **Martin Ehler[2]** ·
Manuel Gräf[3]

Abstract We aim to compute the first few moments of a high-dimensional random vector from the first few moments of a number of its low-dimensional projections. To this end, we identify algebraic conditions on the set of low-dimensional projectors that yield explicit reconstruction formulas. We also provide a computational framework, with which suitable projectors can be derived by solving an optimization problem. Finally, we show that randomized projections permit approximate recovery.

Keywords High-dimensional moments · Random matrices · Orthogonal projections · Zonal polynomials

Mathematics Subject Classification (2010) 44A60 · 42C15 · 42C05

✉ Martin Ehler
martin.ehler@univie.ac.at

Bernhard G. Bodmann
bgb@math.uh.edu

Manuel Gräf
mgraef@kfs.oeaw.ac.at

[1] Department of Mathematics, University of Houston, Houston, TX 77204-3008, USA

[2] Department of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria

[3] Acoustics Research Institute, Austrian Academy of Sciences, Wohllebengasse 12-14, 1040 Vienna, Austria

🙋 Springer

1 Introduction

A central problem in dimension reduction, distributed sensing, and many statistical applications is the identification of properties of a high-dimensional random vector from knowledge of marginal distributions, i.e., the distributions of one or more lower-dimensional projections of the random vector. A simple example in statistics is the problem of computing its lowest moments. However, knowledge of some marginal distributions may not be sufficient to identify the first few high-dimensional moments. Here, we shall address the problem of designing low-dimensional projections of the random vector, so that its high-dimensional moments can be computed from the lower-dimensional ones by an explicit formula.

We consider a random vector X in $\mathbb{R}^d$, distributed according to some Borel probability measure. In practice, X could be a random signal that is observed by distributed sensors, each measuring a certain piece of information. Inspired by [8,20], each sensor is modeled as a matrix $Q_j \in \mathbb{R}^{k \times d}$ with full rank $k < d$. Computing with $P_j := Q_j^*(Q_j Q_j^*)^{-1} Q_j$ instead of Q_j, we can effectively turn our measurement matrices into orthogonal projectors $\{P_j\}_{j=1}^n \subset \mathcal{G}_{k,d}$, where $\mathcal{G}_{k,d}$ denotes the set of orthogonal projectors on $\mathbb{R}^d$ with rank k, i.e., P_j is the orthogonal projector onto the row-space of Q_j. A variant of the Cramér–Wold Theorem says that two random vectors $X, Y \in \mathbb{R}^d$ are identically distributed if and only if, for *all* $P \in \mathcal{G}_{k,d}$, the two random vectors PX, PY are identically distributed, cf. [26]. For further related results on projected distributions, we refer to [4,10,14,17]. Here, we do not wish to identify the distribution of X, but restrict us to recover its first few moments. On the other hand, we want to achieve this by observing the moments of a number of low-dimensional projections and combining the information in a process we call moment fusion.

1.1 Moment Fusion

Suppose X is a random vector in $\mathbb{R}^d$ distributed according to some unknown Borel probability measure on $\mathbb{R}^d$. For a fixed integer $p > 0$, our goal is to determine the low-order moments

$$\mathbb{E}X^s, \quad s \in \mathbb{N}^d, \ |s| \le p, \tag{1}$$

from low-order moments of lower-dimensional projections. We use here multi-index notation $X^s = X_1^{s_1} \cdots X_d^{s_d}$ and $|s| = \sum_{j=1}^d s_j$. More specifically, we suppose that we have only access to the first p moments of low-dimensional linear measurements, i.e., for certain matrices $\{Q_j\}_{j=1}^n \subset \mathbb{R}^{k \times d}$ with fixed rank $k < d$, we suppose that we know

$$\mathbb{E}(Q_j X)^s, \quad s \in \mathbb{N}^k, \ |s| \le p. \tag{2}$$

From knowledge of $\{Q_j\}_{j=1}^n$ and the first p moments of the dimension reduced random vectors $Q_j X$, $j = 1, \ldots, n$, in (2), we aim to reconstruct the high-dimensional moments of X in (1).

1.2 Special Examples

Suppose that $x \in \mathbb{R}^d$ is a vector of unknowns. If $\{Q_j\}_{j=1}^n$ are chosen such that

$$\{(Q_j x)^s : j = 1, \ldots, n, \ s \in \mathbb{N}^k, \ |s| \leq p\}$$

spans the space of polynomials in x of degree at most p, then (1) can be computed from (2) by expressing each expected value $\mathbb{E} X^s$, $s \in \mathbb{N}^d$, $|s| \leq p$, in a suitable linear combination of the expected values of $\{\mathbb{E}(Q_j X)^{s'} : j = 1, \ldots, n, \ s' \in \mathbb{N}^k, \ |s'| \leq p\}_{j=1}^n$. We provide an example for $k = 1$:

Example 1.1 (1) If $p = 1$, then we can simply choose $Q_j := e_j^*, j = 1, \ldots, d$, where $\{e_j\}_{j=1}^d$ is the standard orthonormal basis for $\mathbb{R}^d$. So, reconstruction is possible with d projectors.

(2) For $p = 2$, the $\{Q_j\}_{j=1}^d$ together with $Q_{i,j}^+ := e_i^* + e_j^*, i < j$ allow to reconstruct (1) from (2) with $\binom{d+1}{2}$ many low-dimensional measurements.

(3) If $p = 3$, one can check that $\{Q_j\}_{j=1}^n$ with $Q_{i,j}^+, Q_{i,j}^- := e_i^* - e_j^*, i < j$, and $Q_{i,j,k} := e_i^* + e_j^* + e_k^*, i < j < k$, allow reconstruction, so that we use $\binom{d+2}{3}$ many linear measurements.

(4) For $p = 4$, we can choose $\{Q_j\}_{j=1}^n$ with $Q_{i,j}^+, Q_{i,j}^-, \tilde{Q}_{i,j}^+ := e_i^* + 2e_j^*, i < j$, and $Q_{i,j,k}, Q_{i,j,k}^- := e_i^* - e_j^* + e_k^*, \tilde{Q}_{i,j,k}^- := e_i^* + e_j^* - e_k^*, i < j < k$, and $Q_{i,j,k,\ell} := e_i^* + e_j^* + e_k^* + e_\ell^*, i < j < k < \ell$, allow reconstruction, so that we use $\binom{d+3}{4}$ many linear measurements.

Note that the number of linear measurements in Example 1.1 is exactly the dimension of the homogeneous polynomials of degree p in d variables. Similar examples can be derived for more general situations, and the following example deals with $k = 2$:

Example 1.2 (1) If $p = 1$, then the choice $\binom{e_1^*}{e_2^*}, \binom{e_3^*}{e_4^*}, \ldots$, up to $\binom{e_{d-1}^*}{e_d^*}$, for d even or up to $\binom{e_d^*}{e_1^*}$, for d odd, enables us to reconstruct the high-dimensional mean from the lower-dimensional means.

(2) For $p = 2$, moment reconstruction works with the $\binom{d}{2}$ projectors $\binom{e_i^*}{e_j^*}$, for $i < j$.

1.3 Outline and Contribution of this Paper

The present paper is dedicated to go beyond the explicit Examples 1.1 and 1.2, and instead, provide a general strategy for moment reconstruction. Our main contribution is the identification of conditions on the projectors that yield explicit reconstruction formulas. Moreover, such conditions are compatible with numerical schemes, meaning that suitable projectors can be constructed explicitly by minimizing a certain potential function as discussed in Sects. 4 and 5. We also discuss randomized constructions. Our approach stems from applied harmonic analysis and relates to the concept of so-called Grassmannian cubatures, see, for instance, [2,3].

The remainder of the paper is organized as follows: The condition on the projections and the associated reconstruction formula are formulated in Sect. 2 for random vectors X on the unit sphere $\mathbb{S}^{d-1}$. In Sect. 3 we deal with $X \in \mathbb{R}^d$ and either limit us to up to third moments or use rank one projections. Sections 4 and 5 are dedicated to the construction of suitable projectors based on numerical optimization and on a randomization strategy. The results in Sect. 6 imply that suitably randomized constructions can provide approximate moment recovery, with an error bound that holds with overwhelming probability.

2 Main Reconstruction Results for $X \in \mathbb{S}^{d-1}$

In this section, we focus on random vectors X with values in the unit sphere $\mathbb{S}^{d-1}$ of $\mathbb{R}^d$. We will rely on some results on cubatures for polynomial spaces on the Grassmannian manifold, see [1–3,12,13].

2.1 General Moment Reconstruction

We shall make use of the trace inner product $\langle M_1, M_2 \rangle := \operatorname{tr}(M_1 M_2)$, for $M_1, M_2 \in \mathscr{H}_d := \{M \in \mathbb{R}^{d \times d} : M^\top = M\}$. The Grassmann space

$$\mathcal{G}_{k,d} := \{P \in \mathscr{H}_d : P^2 = P, \ \operatorname{tr}(P) = k\}$$

is the set of rank-k orthogonal projections on $\mathbb{R}^d$. Note that $\{Q_j\}_{j=1}^n \subset \mathbb{R}^{k \times d}$ with all matrices having full rank $k < d$, and $P_j := Q_j^*(Q_j Q_j^*)^{-1} Q_j$, for $j = 1, \ldots, n$, yield $\{P_j\}_{j=1}^n \subset \mathcal{G}_{k,d}$. In place of $\{Q_j\}_{j=1}^n$ we shall find conditions on $\{P_j\}_{j=1}^n$ that enable moment reconstruction, i.e., conditions on the respective row-spaces of $\{Q_j\}_{j=1}^n$.

The orthogonal group $\mathcal{O}(d)$ acts transitively on $\mathcal{G}_{k,d}$ by conjugation $P \mapsto UPU^*$, for $P \in \mathcal{G}_{k,d}$ and $U \in \mathcal{O}(d)$. Thus, there is an orthogonally invariant probability measure $\sigma_{k,d}$ on $\mathcal{G}_{k,d}$, which is induced by the Haar measure on $\mathcal{O}(d)$. This measure leads to the *trace moments*

$$\mu_{k,d}(M_1, \ldots, M_t) := \int_{\mathcal{G}_{k,d}} \langle P, M_1 \rangle \cdots \langle P, M_t \rangle d\sigma_{k,d}(P), \qquad \{M_i\}_{i=1}^t \subset \mathscr{H}_d,$$

which were introduced in [12,13]. In the present section, we can restrict ourselves to

$$\mu_{k,d}^t(M) := \mu_{k,d}(M, \ldots, M),$$

where M occurs t times, and in Sect. 3 we shall make use of the more general case. In the following result, we use the notation $E_{x,y} := \frac{1}{2}(xy^* + yx^*)$, for $x, y \in \mathbb{R}^d$.

Theorem 2.1 *For $\alpha \in \mathbb{N}^d$ with $|\alpha| = t$, there are $y_{s,i}^{\alpha} \in \mathbb{S}^{d-1}$ and coefficients $f_{s,i}^{\alpha} \in \mathbb{R}$, such that*

$$x^{\alpha} = \sum_{s=1}^{t} \sum_{i=1}^{m} f_{s,i}^{\alpha} \mu_{k,d}^{s}(E_{x,y_{s,i}^{\alpha}}), \quad \text{for all } x \in \mathbb{S}^{d-1}.$$

Proof Note that [13, Lemma 7.1] yields

$$x_{i_1} \dots x_{i_t} = \frac{1}{t!} \sum_{\emptyset \neq J \subset \{1,\dots,t\}} (-1)^{t+\#J} \left(\sum_{j \in J} x_{i_j} \right)^t, \tag{3}$$

and $\sum_{j \in J} x_{i_j} = \sum_{j \in J} \langle x, e_{i_j} \rangle$ leads to

$$x_{i_1} \dots x_{i_t} = \frac{1}{t!} \sum_{\emptyset \neq J \subset \{1,\dots,t\}} (-1)^{t+\#J} \left(\left\langle x, \sum_{j \in J} e_{i_j} \right\rangle \right)^t. \tag{4}$$

Thus, it is sufficient to check that each $\langle x, y \rangle^t$, for $x, y \in \mathbb{S}^{d-1}$, can be written as a linear combination of terms $\mu_{k,d}^{s}(E_{x,y})$, $s = 1, \dots, t$.

We shall prove the statement by an induction over t. The case $t = 1$ is covered by

$$\langle x, y \rangle = \frac{d}{k} \mu_{k,d}^{1}(E_{x,y}),$$

see, for instance [3].

To consider general t, we need some preparations. A partition of t is an integer vector $\pi = (\pi_1, \dots, \pi_t)$ whose entries are ordered by $\pi_1 \geq \dots \geq \pi_t \geq 0$ and sum up to $t = \sum_{i=1}^{t} \pi_i$. We denote the number of nonzero entries by $l(\pi)$, and the set of partitions π of t with $l(\pi) \leq d$ is denoted by $\mathscr{P}_{t,d}$.

According to invariant theory, cf. [25, Theorem 7.1], the expansion

$$\mu_{k,d}^{t}(M) = \frac{1}{q_{t,d}} \sum_{\pi \in \mathscr{P}_{t,d}} \alpha_{\pi} \operatorname{tr}(M^{\pi_1}) \cdots \operatorname{tr}(M^{\pi_{l(\pi)}}) \tag{5}$$

holds with suitable real-valued coefficients $q_{t,d}$ and α_{π}. For $x, y \in \mathbb{S}^{d-1}$, we observe that $\operatorname{tr}(E_{x,y}^{s})$ is a polynomial of degree s in $\langle x, y \rangle$ with leading coefficient $(\frac{1}{2})^{s-1}$. The latter yields, for $x, y \in \mathbb{S}^{d-1}$, that $\mu_{k,d}^{t}(E_{x,y})$ is a polynomial in $\langle x, y \rangle$ of degree t, i.e.,

$$\mu_{k,d}^{t}(E_{x,y}) = \frac{1}{q_{t,d}} \sum_{\pi \in \mathscr{P}_t} \alpha_{\pi} \left((\frac{1}{2})^{t-l(\pi)} \langle x, y \rangle^t + \sum_{s=1}^{t-1} c_{\pi,s} \langle x, y \rangle^s \right)$$

$$= \frac{\langle x, y \rangle^t}{2^t q_{t,d}} \left(\sum_{\pi \in \mathscr{P}_t} \alpha_{\pi} 2^{l(\pi)} \right) + \sum_{\pi \in \mathscr{P}_t} \sum_{s=1}^{t-1} c_{\pi,s} \langle x, y \rangle^s.$$

One can check that its leading coefficient does not vanish. Indeed, denoting by D_2 a diagonal rank-2 projection matrix, we observe

$$\frac{1}{2^t q_{t,d}} \left(\sum_{\pi \in \mathscr{P}_t} \alpha_\pi 2^{l(\pi)} \right) = \mu_{k,d}^t(D_2) = \int_{\mathcal{G}_{k,d}} \langle P, D_2 \rangle^t d\sigma_{k,d}(P)$$

and the right-hand side is positive since the function $\langle \cdot, D_2 \rangle^t \geq 0$ does not vanish entirely on $\mathcal{G}_{k,d}$. Therefore, we can isolate $\langle x, y \rangle^t$ and write it as a linear combination of $\mu_{k,d}^t(E_{x,y})$ and terms $\langle x, y \rangle^s$, for $s = 1, \ldots, t-1$. By induction, each term $\langle x, y \rangle^s$, $s = 1, \ldots, t-1$ can be written as a linear combination of terms $\mu_{k,d}^s(E_{x,y})$, $s = 1, \ldots, t-1$, which concludes the proof. □

Theorem 2.1 represents monomials by linear combinations of $\mu_{k,d}^s(E_{x,y_{s,i}^\alpha})$, for some $y_{s,i}^\alpha \in \mathbb{S}^{d-1}$. Next, we shall aim to replace the latter with projected monomials. Let us define a function space on $\mathcal{G}_{k,d}$ by

$$\text{Pol}_t^\ell(\mathcal{G}_{k,d}) := \text{span} \left\{ \langle M, \cdot \rangle^s \big|_{\mathcal{G}_{k,d}} : M \in \mathscr{H}_d, \ \text{rank}(M) \leq \ell, \ s \leq t \right\} \quad (6)$$

and introduce the concept of cubatures on $\mathcal{G}_{k,d}$.

Definition 2.2 For $\{P_j\}_{j=1}^n \subset \mathcal{G}_{k,d}$ and $\{\omega_j\}_{j=1}^n \subset \mathbb{R}$, we say that $\{(P_j, \omega_j)\}_{j=1}^n$ is a *cubature for* $\text{Pol}_t^\ell(\mathcal{G}_{k,d})$ if

$$\sum_{j=1}^n \omega_j f(P_j) = \int_{\mathcal{G}_{k,d}} f(P) d\sigma_{k,d}(P), \quad \text{for all } f \in \text{Pol}_t^\ell(\mathcal{G}_{k,d}).$$

Note that the construction of cubatures for $\text{Pol}_t^\ell(\mathcal{G}_{k,d})$ and the properties of the function space $\text{Pol}_t^\ell(\mathcal{G}_{k,d})$ are discussed in more detail in the Sects. 4 and 5, respectively. We can now formulate our first result on moment reconstruction, which is a direct consequence of Theorem 2.1.

Corollary 2.3 *If* $\{(P_j, \omega_j)\}_{j=1}^n$ *is a cubature for* $\text{Pol}_t^2(\mathcal{G}_{k,d})$, *then, for* $\alpha \in \mathbb{N}^d$ *with* $|\alpha| = t$, *there are coefficients* $a_\beta^\alpha \in \mathbb{R}$, *such that, for any random vector* $X \in \mathbb{S}^{d-1}$,

$$\mathbb{E}X^\alpha = \sum_{|\beta| \leq t} a_\beta^\alpha \sum_{j=1}^n \omega_j \mathbb{E}(P_j X)^\beta. \quad (7)$$

Proof We observe that $\langle P_j x, y \rangle = \langle P_j, E_{x,y} \rangle$ holds, for all $x, y \in \mathbb{R}^d$. According to Theorem 2.1 and since $\langle \cdot, E_{x,y} \rangle^t |_{\mathcal{G}_{k,d}} \in \text{Pol}_t^2(\mathcal{G}_{k,d})$, the cubature property yields

$$\mathbb{E}X^\alpha = \sum_{s=1}^t \sum_{i=1}^m f_{s,i}^\alpha \sum_{j=1}^n \omega_j \mathbb{E} \langle P_j X, y_{s,i}^\alpha \rangle^s$$

and the assertion follows by observing that the terms $\mathbb{E}\langle P_j X, y_{s,i}^\alpha \rangle^s$ are linear combinations of moments of order s of $P_j X$. □

Since we are originally given the moments of $Q_j X$, we must still express $\mathbb{E}(P_j X)^\beta$, where $P_j = Q_j^*(Q_j Q_j^*)^{-1} Q_j$ and $\beta \in \mathbb{N}^d$, $|\beta| \leq t$, by means of moments of $Q_j X$. If we suppress the index j, we obtain the multilinear relation

$$\mathbb{E}(PX)_{i_1} \cdots (PX)_{i_t} = \sum_{j_1,\ldots,j_t=1}^{k} \mathbb{E}\left((QX)_{j_1} \cdots (QX)_{j_t}\right) z_{j_1,i_1} \cdots z_{j_t,i_t},$$

where $z_{i,k} = (Q^*(QQ^*)^{-1})_{i,k}$. Thus, the moments of $Q_j X$ enable us to apply (7).

2.2 Explicit Moment Reconstruction

This section is dedicated to explicitly compute the expansion in Corollary 2.3 for $t = 1, 2, 3$ by using a very particular class of polynomial functions. Indeed, zonal polynomials, cf. [9,16,18,19,24], are special multivariate homogeneous polynomials on $\mathcal{H}_d$. These polynomials C_π are indexed by all partitions π of $\mathbb{N}$ and are invariant under orthogonal conjugation. According to [18], see also [12,13], we obtain

$$\mu_{k,d}^t(M) = \sum_{\pi \in \mathscr{P}_{t,d}} \frac{C_\pi(M) C_\pi(D_k)}{C_\pi(D_k)}, \quad \text{for all } M \in \mathcal{H}_d, \tag{8}$$

where D_k is a diagonal matrix in $\mathbb{R}^{d \times d}$ with k ones and zeros elsewhere. Knowledge of the zonal polynomials enabled us in [13] to compute the trace moments for arbitrarily large t and with explicit expressions for $t = 1, 2, 3$:

Theorem 2.4 ([13]) *For all $d \geq 3$ and $k < d$, we have*

$$\mu_{k,d}^1(M) = \frac{1}{q_{1,d}} \alpha_{(1)} \operatorname{tr}(M),$$

$$\mu_{k,d}^2(M) = \frac{1}{q_{2,d}} \left(\alpha_{(1,1)} \operatorname{tr}^2(M) + \alpha_{(2)} \operatorname{tr}(M^2) \right),$$

$$\mu_{k,d}^3(M) = \frac{1}{q_{3,d}} \left(\alpha_{(1,1,1)} \operatorname{tr}^3(M) + \alpha_{(2,1)} \operatorname{tr}(M) \operatorname{tr}(M^2) + \alpha_{(3)} \operatorname{tr}(M^3) \right),$$

holds for all $M \in \mathcal{H}_d$, where $q_{1,d} = d$, $\alpha_{(1)} = k$, and

$$q_{2,d} = (d-1)d(d+2),$$
$$\alpha_{(1,1)} = (d+1)k^2 - 2k$$
$$\alpha_{(2)} = 2k(d-k),$$
$$q_{3,d} = (d-2)(d-1)d(d+2)(d+4),$$

$$\alpha_{(1,1,1)} = (d^2 + 3d - 2)k^3 - 6(d+2)k^2 + 16k,$$
$$\alpha_{(2,1)} = -6(d+2)k^3 + 6(d^2 + 2d + 4)k^2 - 24dk,$$
$$\alpha_{(3)} = 16k^3 - 24dk^2 + 8d^2k.$$

For $d = 2$ and $k = 1$ in Theorem 2.4, the constant $q_{3,d}$ would be zero, but so are $\alpha_{(1,1,1)}$, $\alpha_{(2,1)}$, and $\alpha_{(3)}$. The identity for $\mu_{1,2}^3(M)$ still holds with the modified coefficients

$$q_{3,2} = 48, \quad \alpha_{(1,1,1)} = 1, \quad \alpha_{(2,1)} = 6, \quad \alpha_{(3)} = 8.$$

Theorem 2.4 and the proof of Theorem 2.1 lead to the following explicit moment recovery formulas.

Corollary 2.5 *Let $X \in \mathbb{S}^{d-1}$ be a random vector with $d \geq 3$.*

(i) *If $\{(P_j, \omega_j)\}_{j=1}^n$ is a cubature for $\mathrm{Pol}_1^2(\mathcal{G}_{k,d})$, then, for $i = 1, \ldots, d$,*

$$\mathbb{E}X_i = A_1 \sum_{j=1}^n \omega_j \mathbb{E}(P_j X)_i, \quad \text{where } A_1 = \frac{d}{k}. \tag{9}$$

(ii) *If $\{(P_j, \omega_j)\}_{j=1}^n$ is a cubature for $\mathrm{Pol}_2^2(\mathcal{G}_{k,d})$, then (9) holds and, for $i, \ell = 1, \ldots, d$,*

$$\mathbb{E}X_i X_\ell = A_2 \sum_{j=1}^n \omega_j \mathbb{E}(P_j X)_i (P_j X)_\ell - B_2 \frac{k}{d} \delta_{i,\ell}, \tag{10}$$

where
$$A_2 = \frac{(d-1)d(d+2)}{k(dk+d-2)}, \quad B_2 = \frac{(d-k)d}{kd(k+1) - 2k}.$$

(iii) *If $\{(P_j, \omega_j)\}_{j=1}^n$ is a cubature for $\mathrm{Pol}_3^2(\mathcal{G}_{k,d})$, then (9) and (10) hold and, for $i, \ell, m = 1, \ldots, d$,*

$$\mathbb{E}X_i X_\ell X_m = A_3 \sum_{j=1}^n \omega_j \mathbb{E}(P_j X)_i (P_j X)_\ell (P_j X)_m$$

$$- \frac{B_3}{3} \sum_{j=1}^n \omega_j \left(\mathbb{E}(P_j X)_i \delta_{\ell,m} + \mathbb{E}(P_j X)_\ell \delta_{i,m} + \mathbb{E}(P_j X)_m \delta_{i,\ell} \right) \frac{k}{d},$$

where

$$A_3 = \frac{(d-2)(d-1)d(d+2)(d+4)}{k(d^2k^2 + 3d^2k + 2d^2 - 6dk - 12d - 4k^2)},$$
$$B_3 = \frac{3d^2(d^2k + 2d^2 - 5dk - 4d + 4k^2 + 2k)}{k^2(d^2k^2 + 3d^2k + 2d^2 + 3dk^2 - 9dk - 12d + 2k^2 - 6k + 16)}.$$

Remark 2.6 Note that (9) is proved by monomial identities, so that it also holds when the expectation is eliminated on both sides.

Proof of Corollary 2.5 For $x, y \in \mathbb{R}^d$, we obtain

$$\operatorname{tr}(E_{x,y}) = \langle x, y \rangle,$$

$$\operatorname{tr}(E_{x,y}^2) = \frac{1}{2}(\langle x, y \rangle^2 + \|x\|^2 \|y\|^2),$$

$$\operatorname{tr}(E_{x,y}^3) = \frac{1}{4}(\langle x, y \rangle^3 + 3\langle x, y \rangle \|x\|^2 \|y\|^2),$$

so that Theorem 2.4 implies, for all $d \geq 3$ and $x, y \in \mathbb{R}^d$,

$$\mu_{k,d}^1(E_{x,y}) = \frac{\alpha_{(1)}}{q_{1,d}} \langle x, y \rangle, \tag{11}$$

$$\mu_{k,d}^2(E_{x,y}) = \frac{2\alpha_{(1,1)} + \alpha_{(2)}}{2q_{2,d}} \langle x, y \rangle^2 + \frac{\alpha_{(2)}}{2q_{2,d}} \|x\|^2 \|y\|^2, \tag{12}$$

$$\mu_{k,d}^3(E_{x,y}) = \frac{4\alpha_{(1,1,1)} + 2\alpha_{(2,1)} + \alpha_{(3)}}{4q_{3,d}} \langle x, y \rangle^3 + \frac{2\alpha_{(2,1)} + \alpha_{(3)}}{4q_{3,d}} \langle x, y \rangle \|x\|^2 \|y\|^2, \tag{13}$$

where the constants $q_{1,d}, q_{2,d}, q_{3,d}$ and $\alpha_{(1)}, \alpha_{(2)}, \alpha_{(3)}, \alpha_{(1,1,1)}, \alpha_{(2,1)}$ are specified in Theorem 2.4.

Suppose now that $\{(P_j, \omega_j)\}_{j=1}^n$ is a cubature of $\operatorname{Pol}_t^2(\mathcal{G}_{k,d})$, so that we obtain

$$\mu_{k,d}^t(E_{x,y}) = \sum_{j=1}^n \omega_j \langle P_j, E_{x,y} \rangle^t = \sum_{j=1}^n \omega_j \langle P_j x, y \rangle^t,$$

and the left-hand sides in (11), (12), and (13) can be replaced with $\sum_{j=1}^n \omega_j \langle P_j x, y \rangle^t$, for $t = 1, 2, 3$, respectively. In the following, we assume $x \in \mathbb{S}^{d-1}$ and $y \in \mathbb{R}^d$. Rearranging terms leads to the following formulas, respectively, and $A_1, A_2, A_3,$ and B_2, B_3 are as in Corollary 2.5. If $\{(P_j, \omega_j)\}_{j=1}^n$ is a cubature for $\operatorname{Pol}_1^2(\mathcal{G}_{k,d})$, then

$$\langle x, y \rangle = A_1 \sum_{j=1}^n \omega_j \langle P_j x, y \rangle. \tag{14}$$

If $\{(P_j, \omega_j)\}_{j=1}^n$ is a cubature for $\operatorname{Pol}_2^2(\mathcal{G}_{k,d})$, then (14) holds and

$$\langle x, y \rangle^2 = A_2 \sum_{j=1}^n \omega_j \langle P_j x, y \rangle^2 - B_2 \|y\|^2 \frac{k}{d}. \tag{15}$$

If $\{(P_j, \omega_j)\}_{j=1}^n$ is a cubature for $\operatorname{Pol}_3^2(\mathcal{G}_{k,d})$, then (14), (15) hold, so that

$$\langle x, y \rangle^3 = A_3 \sum_{j=1}^n \omega_j \langle P_j x, y \rangle^3 - B_3 \|y\|^2 \sum_{j=1}^n \omega_j \langle P_j x, y \rangle \frac{k}{d}. \tag{16}$$

As at the beginning of the proof of Theorem 2.1, [13, Lemma 7.1] yields

$$x_{i_1} \cdots x_{i_t} = \frac{1}{t!} \sum_{J \subset \{1,\ldots,t\}} (-1)^{t + \#J} \left(\left\langle x, \sum_{j \in J} e_{i_j} \right\rangle \right)^t.$$

In order to compute the term $x_{i_1} \ldots x_{i_t}$, we can repeatedly apply (14), (15), (16), respectively, with $y = \sum_{j \in J} e_{i_j}$. Such rearrangements yield the formulas and constants in Corollary 2.5. □

Remark 2.7 The framework that we present in the present paper also allows the explicit computations of higher-order moments beyond $t = 1, 2, 3$. Indeed, if $\{(P_j, \omega_j)\}_{j=1}^n$ is a cubature for $\mathrm{Pol}_t^2(\mathcal{G}_{k,d})$, then we can compute all moments of order t by using the zonal polynomials. However, we do not have one closed formula incorporating t as a variable, but we need to compute the expressions for each t separately. Note that the formulas in Theorem 2.5 are merely based on identities between the corresponding polynomials in the entries of a unit vector $x \in \mathbb{R}^d$.

3 Moment Fusion for $X \in \mathbb{R}^d$

3.1 The General Case for up to Cubic Moments

A homogeneity argument yields that (9) even holds for random $X \in \mathbb{R}^d$. Analogously, considering (10) as a monomial identity with $B_2 \frac{k}{d} \delta_{i,\ell} = B_2 \frac{k}{d} \|x\|^2 \delta_{i,\ell}$, for $x \in \mathbb{S}^{d-1}$, a homogeneity argument yields that, for random $X \in \mathbb{R}^d$,

$$\mathbb{E} X_i X_\ell = A_2 \sum_{j=1}^n \omega_j \mathbb{E}(P_j X)_i (P_j X)_\ell - B_2 \sum_{r=1}^d \sum_{j=1}^n \omega_j \mathbb{E}(P_j X)_r^2 \delta_{i,\ell},$$

provided that $\{(P_j, \omega_j)\}_{j=1}^n$ is a cubature for $\mathrm{Pol}_2^2(\mathcal{G}_{k,d})$.

In order to deal with $X \in \mathbb{R}^d$ for $t = 3$ as well, we observe that the formulas in (11), (12), and (13) hold in more generality, see [13, Theorem 7.3],

$$\mu_{k,d}(X_1) = \frac{1}{q_{1,d}} \alpha_{(1)} \operatorname{tr}(X_1),$$

$$\mu_{k,d}(X_1, X_2) = \frac{1}{q_{2,d}} \left(\alpha_{(1,1)} \operatorname{tr}(X_1) \operatorname{tr}(X_2) + \alpha_{(2)} \operatorname{tr}(X_1 X_2) \right),$$

$$\mu_{k,d}(X_1, X_2, X_3) = \frac{1}{q_{3,d}} \big(\alpha_{(1,1,1)} \operatorname{tr}(X_1) \operatorname{tr}(X_2) \operatorname{tr}(X_3)$$
$$+ \frac{\alpha_{(2,1)}}{3} (\operatorname{tr}(X_1) \operatorname{tr}(X_2 X_3) + \operatorname{tr}(X_2) \operatorname{tr}(X_1 X_3) + \operatorname{tr}(X_3) \operatorname{tr}(X_1 X_2))$$
$$+ \alpha_{(3)} \operatorname{tr}(X_1 X_2 X_3) \big).$$

For $x \in \mathbb{R}^d$ and $y \in \mathbb{S}^{d-1}$, using the above relation gives

$$\mu_{k,d}(E_{x,y}, xx^*) = \frac{\alpha_{(1,1)} + \alpha_{(2)}}{q_{2,d}} \langle x, y \rangle \|x\|^2 = \frac{k(k+2)}{d(d+2)} \langle x, y \rangle \|x\|^2$$

and combined with identity (13), we obtain

$$\langle x, y \rangle^3 = C_{3,d}^{(1)} \mu_{k,d}^3(E_{x,y}) - C_{3,d}^{(2)} \mu_{k,d}(E_{x,y}, xx^*) \tag{17}$$

with $C_{3,d}^{(1)} = \frac{4q_{3,d}}{4\alpha_{(1,1,1)} + 2\alpha_{(2,1)} + \alpha_{(3)}}$ and $C_{3,d}^{(2)} = \frac{2\alpha_{(2,1)} + \alpha_{(3)}}{(4\alpha_{(1,1,1)} + 2\alpha_{(2,1)} + \alpha_{(3)})\frac{k(k+2)}{d(d+2)}}$. Indeed, one can check that the term $4\alpha_{(1,1,1)} + 2\alpha_{(2,1)} + \alpha_{(3)}$ is nonzero. If $\{(P_j, \omega_j)\}_{j=1}^n$ is a cubature for $\text{Pol}_3^2(\mathcal{G}_{k,d})$, then we can apply

$$\mu_{k,d}^3(E_{x,y}) = \sum_{j=1}^n \omega_j \langle P_j x, y \rangle^3, \tag{18}$$

because the mapping $P \mapsto \langle P, E_{x,y} \rangle^3$ is contained in $\text{Pol}_3^2(\mathcal{G}_{k,d})$. In Proposition 5.1 we shall check that $P \mapsto \langle P, E_{x,y} \rangle \langle P, xx^* \rangle$ is also contained in $\text{Pol}_3^2(\mathcal{G}_{k,d})$, so that also

$$\mu_{k,d}(E_{x,y}, xx^*) = \sum_{j=1}^n \sum_{m=1}^d \omega_j \langle P_j x, y \rangle (P_j x)_m^2 \tag{19}$$

holds. The actual moments of order 3 can now be computed from (4) by observing that $\langle P_j x, y \rangle$ yields a linear combination of the terms $(P_j x)_1, \ldots, (P_j x)_d$.

We now collect the resulting expressions for all third moments:

Corollary 3.1 *Let $X \in \mathbb{R}^d$ be a random vector with $d \geq 3$, let the constants A_1, A_2 and B_2 be as above, and let $i, h, l \in \{1, 2, \ldots, d\}$ with $i \neq h \neq l \neq i$.*

(i) *If $\{(P_j, \omega_j)\}_{j=1}^n$ is a cubature for $\text{Pol}_1^2(\mathcal{G}_{k,d})$, then*

$$\mathbb{E}X_i = A_1 \sum_{j=1}^n \omega_j \mathbb{E}(P_j X)_i. \tag{20}$$

(ii) *If $\{(P_j, \omega_j)\}_{j=1}^n$ is a cubature for $\text{Pol}_2^2(\mathcal{G}_{k,d})$, then (20) holds and*

$$\mathbb{E}X_i^2 = A_2 \sum_{j=1}^n \omega_j \mathbb{E}(P_j X)_i^2 - B_2 \sum_{j=1}^n \sum_{m=1}^d \omega_j \mathbb{E}(P_j X)_m^2, \tag{21}$$

$$\mathbb{E}X_i X_\ell = A_2 \sum_{j=1}^n \omega_j \mathbb{E}(P_j X)_i (P_j X)_\ell. \tag{22}$$

(iii) *If $\{(P_j, \omega_j)\}_{j=1}^n$ is a cubature for $\text{Pol}_3^2(\mathcal{G}_{k,d})$, then (20), (21), (22) hold and*

$$\mathbb{E}X_i^3 = C_{3,d}^{(1)} \sum_{j=1}^{n} \omega_j \mathbb{E}(P_j X)_i^3 - C_{3,d}^{(2)} \sum_{j=1}^{n} \sum_{m=1}^{d} \omega_j \mathbb{E}(P_j X)_i (P_j X)_m^2, \quad (23)$$

$$\mathbb{E}X_i^2 X_h = C_{3,d}^{(1)} \sum_{j=1}^{n} \omega_j \mathbb{E}(P_j X)_i^2 (P_j X_h) - \frac{1}{3} C_{3,d}^{(2)} \sum_{j=1}^{n} \sum_{m=1}^{d} \omega_j \mathbb{E}(P_j X)_h (P_j X)_m^2, \tag{24}$$

$$\mathbb{E}X_i X_h X_\ell = C_{3,d}^{(1)} \sum_{j=1}^{n} \omega_j \mathbb{E}(P_j X)_i (P_j X)_h (P_j X)_\ell. \tag{25}$$

Proof The first and second moments have already been discussed prior to the statement of the corollary. For the third moments, the expression for $\mathbb{E}X_i^3$ results from choosing $y = e_i$ in (17), (18), and (19), which yields (23).

Next, we address (24). The choices $y_+ = \frac{1}{\sqrt{2}}(e_i + e_h)$ and $y_- = \frac{1}{\sqrt{2}}(e_i + e_h)$ yield

$$x_i^2 x_h = \frac{\sqrt{2}}{3}[\langle x, y_+ \rangle^3 - \langle x, y_- \rangle^3] - \frac{1}{3}\langle x, e_h \rangle^3.$$

Applying (17), (18), and (19) leads to

$$\mathbb{E}X_i^2 X_h = C_{3,d}^{(1)} \sum_{j=1}^{n} \omega_j ((P_j X)_i^2 (P_j X)_h + \frac{1}{3}(P_j X)_h^3)$$

$$- \frac{2}{3} C_{3,d}^{(2)} \sum_{j=1}^{n} \sum_{m=1}^{d} \omega_j (P_j X)_h (P_j X)_m^2 - \frac{1}{3}\mathbb{E}X_h^3$$

and inserting the expression for $\mathbb{E}X_h^3$ then reduces this to (24).

Finally, for $\mathbb{E}X_i X_h X_\ell$, we observe

$$x_i x_h x_\ell = \frac{1}{24}((x_i + x_h + x_\ell)^3 + (x_i - x_h - x_\ell)^3 - (x_i + x_h - x_\ell)^3 - (x_i - x_h + x_\ell)^3).$$

By using $y_{+++} = \frac{1}{\sqrt{3}}(e_i + e_h + e_\ell)$, $y_{+--} = \frac{1}{\sqrt{3}}(e_i - e_h - e_\ell)$, $y_{--+} = -\frac{1}{\sqrt{3}}(e_i + e_j - e_\ell)$, and $y_{-+-} = -\frac{1}{\sqrt{3}}(e_i - e_h + e_\ell)$, we obtain

$$\mathbb{E}X_i X_h X_\ell = \frac{\sqrt{3}}{8}\mathbb{E}[\langle X, y_{+++} \rangle^3 + \langle X, y_{+--} \rangle^3 + \langle X, y_{--+} \rangle^3 + \langle X, y_{-+-} \rangle^3],$$

and a calculation using (17), (18), and (19) leads to (25). $\qquad \square$

3.2 All Moments from Projections onto One-Dimensional Subspaces

In the previous section, we have outlined the recovery of the moments for $t = 1, 2, 3$ and general k. To address all moments $t > 3$, we now restrict us to $k = 1$.

Let us denote the permutation group of $\{1, \ldots, t\}$ by S_t. We say a permutation $s \in S_t$ is associated to a partition π and denote this by $s \sim \pi$ if there is a set of cycles $\{c_i\}_{i=1}^{m}$ such that $s = (c_1) \cdots (c_m)$ and the cardinality of c_i equals π_i, for $i = 1, \ldots, m$. Note that we also use the standard notation $c_i \in s$ for a cycle c_i occurring in s. For $\{M_i\}_{i=1}^{t} \subset \mathcal{H}_d$, we use a cycle index $M_{c_i} := M_{c_{i,1}} \cdots M_{c_{i,\ell_i}}$, where $c_i = (c_{i,1} \ldots c_{i,\ell_i})$.

To clarify notation, we provide a simple example.

Example 3.2 For $t = 4$, let the permutation s be given by $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 2 & 4 \end{pmatrix}$, and suppose we have a set of matrices $\{M_i\}_{i=1}^{4}$, then s has the cyclic representation $(c_1)(c_2) = (132)(4)$ and is associated to the partition $(3, 1, 0, 0)$. This implies $M_{c_1} = M_1 M_3 M_2$ and $M_{c_2} = M_4$.

Due to the orthogonal invariance of the Haar measure, the Grassmannian trace moments are invariant under the orthogonal group $\mathcal{O}_d$, i.e.,

$$\mu_{k,d}(U M_1, \ldots, U M_t) = \mu_{k,d}(M_1, \ldots, M_t), \quad \text{for all } U \in \mathcal{O}_d.$$

A general result in invariant theory, cf. [25], and the invariance of $\mu_{k,d}$ under permutations yield

$$\mu_{k,d}(M_1, \ldots, M_t) = \sum_{\pi \in \mathcal{P}_t} \alpha_\pi \sum_{\substack{s \in S_t \\ s \sim \pi}} \prod_{c \in s} \text{tr}(M_c), \tag{26}$$

where $\alpha_\pi \in \mathbb{R}$, see also (5) for $M_1 = \ldots = M_t$.

Proposition 3.3 *For $d \geq t$, $t \in \mathbb{N}_0$, and provided that $k = 1$, the expansion (26) of the trace moments possesses only positive coefficients α_π, $\pi \in \mathcal{P}_t$.*

Proof For any fixed permutation $\sigma : \{1, \ldots, t\} \to \{1, \ldots, t\}$, we consider the matrices

$$M_i = \begin{cases} e_i e_{\sigma(i)}^* + e_{\sigma(i)} e_i^*, & i \neq \sigma(i) \\ e_i e_i^*, & i = \sigma(i), \end{cases} \quad i = 1, \ldots, t,$$

where $\{e_i\}_{i=1}^{d} \in \mathbb{R}^d$ is the standard basis.

Now, let $s \in S_t$ be another arbitrary permutation with some cycle $c \in s$. We denote the cardinality of c by l. From

$$\text{tr}(M_c) = \text{tr}(M_{c_1} \cdots M_{c_l}) = \sum_{k_1, \ldots, k_l = 1}^{d} (M_{c_1})_{k_1, k_2} (M_{c_2})_{k_2, k_3} \cdots (M_{c_{l-1}})_{k_{l-1}, k_l} (M_{c_l})_{k_l, k_1}$$

we conclude by the definition of M_i that the indices k_i contribute to the sum if and only if $k_i, k_{i+1 \bmod l} \in \{c_i, \sigma(c_i)\}$ for all $i = 1, \ldots, l$. Equivalently, it must hold that $k_i \in \{c_i, \sigma(c_i)\} \cap \{c_{i-1 \bmod l}, \sigma(c_{i-1 \bmod l})\}$, $i = 1, \ldots, l$. Since $c_i \neq c_j$ and $\sigma(c_i) \neq \sigma(c_j)$ for $i \neq j$, this can only happen if $c_i = \sigma(c_{i-1 \bmod l})$ for all $i = 1, \ldots, l$ or $c_{i-1 \bmod l} = \sigma(c_i)$ for all $i = 1, \ldots, l$. Hence, the trace of M_c vanishes if and only if neither the cycle c nor its inverse c^{-1} are contained in σ. More precisely,

$$\text{tr}(M_c) = \begin{cases} 1, & c \in \sigma \text{ or } c^{-1} \in \sigma, \\ 0, & \text{else.} \end{cases}$$

Using these observations we obtain

$$\mu_{1,d}(M_1, \ldots, M_t) = \sum_{\pi \in \mathscr{P}_t} \alpha_\pi \sum_{\substack{s \in S_t \\ s \sim \pi}} \prod_{c \in s} \text{tr}(M_c)$$

$$= \alpha_{\pi_\sigma} \#\{s \in S_t : c \text{ or } c^{-1} \in \sigma, \forall c \in s\},$$

where π_σ is the partition associated to σ. Hence, π_σ is a fraction of the trace moment $\mu_{1,d}(M_1, \ldots, M_t)$. It remains to verify that the latter is positive.

Together with the definition of the trace moments $\mu_{1,d}(M_1, \ldots, M_t)$ and those of M_i we arrive at

$$\mu_{1,d}(M_1, \ldots, M_t) = \int_{\mathcal{O}_d} \prod_{i=1}^t \langle O D_1 O^*, M_i \rangle dO$$

$$= \int_{\mathcal{O}_d} \prod_{i=1}^t \langle O e_1 (O e_1)^*, M_i \rangle dO$$

$$= \int_{\mathcal{O}_d} \prod_{i=1}^t 2^{\#\{i,\sigma(i)\}} O_{1,i} O_{1,\sigma(i)} dO$$

$$= \int_{\mathcal{O}_d} \Big(\prod_{i=1}^t 2^{\#\{i,\sigma(i)\}} O_{1,i} \Big) \Big(\prod_{i=1}^t O_{1,\sigma(i)} \Big) dO.$$

Since σ is a permutation, we obtain

$$\mu_{1,d}(M_1, \ldots, M_t) = \int_{\mathcal{O}_d} \Big(\prod_{i=1}^t 2^{\#\{i,\sigma(i)\}} O_{1,i} \Big) \Big(\prod_{i=1}^t O_{1,i} \Big) dO$$

$$= \int_{\mathcal{O}_d} \prod_{i=1}^t 2^{\#\{i,\sigma(i)\}} (O_{1,i})^2 dO > 0$$

and the assertion follows. $\qquad\square$

Proposition 3.4 *For fixed* $m, \ell \in \mathbb{N}_0$ *and* $d \in \mathbb{N}$ *with* $d \geq m + \ell$ *there are coefficients* $\{a_i\}_{i=0}^{\lfloor m/2 \rfloor} \in \mathbb{R}$ *such that, for all* $x \in \mathbb{R}^d$, $y \in \mathbb{S}^{d-1}$, *it holds*

$$\langle x, y \rangle^m \|x\|^{2\ell} = \sum_{i=0}^{\lfloor m/2 \rfloor} a_i \mu_{1,d}^{(m-2i, \ell+i)}(E_{x,y}, xx^*). \tag{27}$$

Proof Let us first note that by the identity (26) the trace moments $\mu_{1,d}^{(m,\ell)}(E_{x,y}, xx^*)$, $m, \ell \in \mathbb{N}_0$, can be written as polynomials in $\langle x, y \rangle$, $\|x\|^2$ and $\|y\|^2$. Hence, together with the homogeneity in x and y we infer the representation

$$\mu_{1,d}^{(m,\ell)}(E_{x,y}, xx^*) = \sum_{i=0}^{\lfloor m/2 \rfloor} b_i^{(m,\ell)} \langle x, y \rangle^{m-2i} \|x\|^{2(i+\ell)}, \qquad x \in \mathbb{R}^d, y \in \mathbb{S}^{d-1}, \quad (28)$$

for some coefficients $b_i^{(m,\ell)} \in \mathbb{R}$. Moreover, we have

$$b_0^{(m,\ell)} > 0, \qquad d \geq m + \ell, \tag{29}$$

which follows from Proposition 3.3 and the fact that the coefficients of $\langle x, y \rangle^m \|x\|^{2\ell}$ in any term of the form

$$\prod_{i=1}^{r} \mathrm{tr} \Big(\prod_{j=1}^{s_i} (E_{x,y})^{m_{i,j}} (xx^*)^{\ell_{i,j}} \Big), \quad \text{with} \quad \sum_{i=1}^{r}\sum_{j=1}^{s_i} m_{i,j} = m, \quad \sum_{i=1}^{r}\sum_{j=1}^{s_i} \ell_{i,j} = \ell$$

are positive.

Now, the statement (3.4) will follow by induction over m. Therefore, let $m \geq 2$, $\ell \in \mathbb{N}_0$ with $d \geq m + \ell$ be given and assume that the statement (3.4) holds for all $m' = m - 2i$, $\ell' = \ell + i$, $i = 1, \ldots, \lfloor m/2 \rfloor$. Using equation (28) we obtain

$$\mu_{1,d}^{(m,\ell)}(E_{x,y}, xx^*) = b_0^{(m,\ell)} \langle x, y \rangle^m \|x\|^{2\ell} + \sum_{i=1}^{\lfloor m/2 \rfloor} b_i^{(m,\ell)} \langle x, y \rangle^{m-2i} \|x\|^{2(\ell+i)}.$$

Since $d \geq m + \ell > m' + \ell' = m + \ell - i$, $i = 1, \ldots, \lfloor m/2 \rfloor$, we can expand the sum on the right-hand side by the induction hypothesis into trace moments $\mu_{1,d}^{(m-2i,\ell+i)}(E_{x,y}, xx^*)$, $i = 1, \ldots, \lfloor m/2 \rfloor$. Hence, using $b_0^{(m,\ell)} \neq 0$, see (29), we can rearrange terms and arrive at the statement (27). It remains to show the induction base with the cases $m \in \{0, 1\}$, $\ell \in \mathbb{N}_0$, $d \geq m + \ell$.

For $m = 0$, $\ell \in \mathbb{N}_0$ and $d \in \mathbb{N}$ we observe by orthogonal invariance

$$\mu_{1,d}^{(0,\ell)}(E_{x,y}, xx^*) = \mu_{1,d}^{\ell}(xx^*) = \mu_{1,d}^{\ell}(D_1) \|x\|^{2\ell}, \qquad x \in \mathbb{R}^d.$$

The term $\mu_{1,d}^{\ell}(D_1)$ is positive and has been explicitly computed in [3]:

$$\mu_{1,d}^{\ell}(D_1) = \frac{(1/2)_\ell}{(d/2)_\ell}, \qquad (a)_\ell := a(a+\ell)\cdots(a+\ell-1).$$

Hence, the assertion follows for $m = 0$, $\ell \in \mathbb{N}_0$.

For $m = 1$, $\ell \in \mathbb{N}_0$ and $d \in \mathbb{N}$ we find by (26) that

$$\mu_{1,d}^{(1,\ell)}(E_{x,y}, xx^*) = \langle x, y \rangle \|x\|^{2\ell} \sum_{\pi \in \mathscr{P}_{\ell+1}} \alpha_\pi \sum_{\substack{s \in S_{\ell+1} \\ s \sim \pi}} 1, \qquad x \in \mathbb{R}^d, \quad y \in \mathbb{S}^{d-1}.$$

Moreover, we can check that the coefficient of $\langle x, y \rangle \|x\|^{2\ell}$ is nonzero by observing

$$\sum_{\pi \in \mathscr{P}_t} \alpha_\pi \sum_{\substack{s \in S_{\ell+1} \\ s \sim \pi}} 1 = \mu_{1,d}^{\ell+1}(I_1) > 0,$$

which concludes the proof. $\qquad\qquad\square$

Corollary 3.5 *If $\{(P_j, \omega_j)\}_{j=1}^n$ is a cubature for $\mathrm{Pol}_t^3(\mathcal{G}_{1,d})$, then, for $\alpha \in \mathbb{N}^d$ with $|\alpha| = t \leq d$, there are coefficients $a_\beta^\alpha \in \mathbb{R}$, such that, for any random vector $X \in \mathbb{R}^d$,*

$$\mathbb{E}X^\alpha = \sum_{|\beta|=t} a_\beta^\alpha \sum_{j=1}^n \omega_j \mathbb{E}(P_j X)^\beta. \tag{30}$$

Proof For any $i = 0, \ldots, \lfloor t/2 \rfloor$, the function $F : \mathcal{G}_{1,d} \to \mathbb{R}$ defined by $P \mapsto \langle P, E_{x,y} \rangle^{t-2i} \langle P, xx^* \rangle^{2i}$ is contained in $\mathrm{Pol}_t^3(\mathcal{G}_{1,d})$, see part two of Proposition 4.1 in the subsequent Sect. 4 for details. Thus, the cubature property yields

$$\mu_{1,d}^{(t-2i,i)}(E_{x,y_j}, xx^*) = \sum_{j=1}^n \omega_j \langle P_j, E_{x,y} \rangle^{t-2i} \langle P_j, xx^* \rangle^i$$

$$= \sum_{j=1}^n \omega_j \langle P_j x, y \rangle^{t-2i} \|P_j x\|^{2i}.$$

Note that $\langle P_j x, y \rangle^{t-2i}$ and $\|P_j x\|^{2i}$ are linear combinations of monomials in $P_j x$ of degree $t - 2i$ and $2i$, respectively. Hence, their product is a linear combination of monomials in $P_j x$ of degree t. Applying (4) and invoking Proposition 3.4 for $\ell = 0$ concludes the proof. $\qquad\square$

4 Constructing Cubatures for $\mathrm{Pol}_t^2(\mathcal{G}_{k,d})$

In this section, we shall derive a general framework for the construction of cubatures for $\mathrm{Pol}_t^2(\mathcal{G}_{k,d})$ that are needed to apply our results in Theorem 2.5. For general existence results of cubatures, we refer to [11], and explicit group theoretical constructions are provided in [7]. In the following, we shall discuss random constructions as well as deterministic constructions based on the solution of an optimization problem.

4.1 Random Construction

For $n, m \geq \dim(\mathrm{Pol}_t^\ell(\mathcal{G}_{k,d}))$, it follows from classical arguments that there are $\{M_i\}_{i=1}^m \subset \mathcal{H}_d^\ell$ and $\{P_j\}_{j=1}^n \subset \mathcal{G}_{k,d}$ such that the matrix

$$(\langle M_i, P_j \rangle^t)_{\substack{i=1\ldots,m \\ j=1,\ldots,n}} \tag{31}$$

has rank $\dim(\mathrm{Pol}_t^\ell(\mathcal{G}_{k,d}))$. We can now compute weights $\omega := \{\omega\}_{j=1}^n$ by solving the linear system of equations

$$\sum_{j=1}^n \langle M_i, P_j \rangle^t \omega_j = \mu_{k,d}^t(M_i), \quad i = 1, \ldots, m,$$

which yields a cubature $\{(P_j, \omega_j)\}_{j=1}^n$. Note that the weights $\{\omega_j\}_{j=1}^n$ are not necessarily nonnegative.

We claim that M_i and P_j can be chosen in a random fashion. Indeed, we observe that both spaces, $\mathcal{G}_{k,d}$ and $\mathcal{H}_d^\ell$, can be parametrized analytically, so that there is $D > 0$ and a surjective analytic mapping $F : \mathbb{R}^D \to (\mathcal{H}_d^\ell)^m \times (\mathcal{G}_{k,d})^n$. Let us assume $n = m = \dim(\mathrm{Pol}_t^\ell(\mathcal{G}_{k,d}))$ for simplicity. Otherwise, we can extract a submatrix. We now define

$$G : (\mathcal{H}_d^\ell)^n \times (\mathcal{G}_{k,d})^n \to \mathbb{R}$$
$$\left((P_i)_{i=1}^n, (M_j)_{j=1}^n\right) \mapsto \det\left((\langle P_i, M_j \rangle^t)_{i,j}\right).$$

Since F is surjective, the mapping $G \circ F : \mathbb{R}^D \to \mathbb{R}$ is not identically zero. Moreover, $G \circ F$ is analytic, so that $(G \circ F)^{-1}(\{0\}) \subset \mathbb{R}^D$ has Lebesgue measure zero and, hence, is a zero set with respect to any continuous probability measure ν on $\mathbb{R}^D$. Thus, the parametrization F enables a random choice in $(\mathcal{H}_d^\ell)^m \times (\mathcal{G}_{k,d})^n$, so that the matrix (31) has rank $\dim(\mathrm{Pol}_t^\ell(\mathcal{G}_{k,d}))$ with probability one with respect to ν. In other words, $G^{-1}(\{0\})$ is a zero set with respect to the induced probability measure ν_F on $(\mathcal{H}_d^\ell)^m \times (\mathcal{G}_{k,d})^n$. Thus, (31) has rank $\dim(\mathrm{Pol}_t^\ell(\mathcal{G}_{k,d}))$ with probability one and weights $\{\omega_j\}_{j=1}^n$ can be computed.

Let us also verify that (31) having rank $\dim(\mathrm{Pol}_t^\ell(\mathcal{G}_{k,d}))$ is a generic property. Indeed, both spaces, $\mathcal{G}_{k,d}$ and $\mathcal{H}_d^\ell$, are real algebraic varieties that are irreducible, cf. [5,21], so that also $(\mathcal{H}_d^\ell)^m \times (\mathcal{G}_{k,d})^n$ is irreducible. Without loss of generality, we can restrict us to $n = m = \dim(\mathrm{Pol}_t^\ell(\mathcal{G}_{k,d}))$ again. Note that G is a polynomial map and, hence, is Zariski continuous. Therefore, the set $U := \{u \in (\mathcal{H}_d^\ell)^m \times (\mathcal{G}_{k,d})^n : G(u) \neq 0\}$ is Zariski open. Classical arguments yield that U cannot be empty, so that irreducibility yields that U is Zariski dense. Thus, we have verified that, for $n, m \geq \dim(\mathrm{Pol}_t^\ell(\mathcal{G}_{k,d}))$, there is a nonempty Zariski open and dense subset U in $(\mathcal{H}_d^\ell)^m \times (\mathcal{G}_{k,d})^n$ such that the matrix

$$(\langle M_i, P_j \rangle^t)_{\substack{i=1,\dots,m \\ j=1,\dots,n}}, \quad \text{where} \quad \left((M_i)_{i=1}^m, (P_j)_{j=1}^n\right) \in U,$$

has rank $\dim(\mathrm{Pol}_t^\ell(\mathcal{G}_{k,d}))$.

4.2 Deterministic Construction

Here, we present the design of cubatures as the solution of an optimization problem. As in [12], we shall apply the theory of reproducing kernel Hilbert spaces. We first define a measure $\nu_{\ell,d}$ on $\mathscr{H}_d^\ell := \{M \in \mathscr{H}_d : \mathrm{rank}(M) \le \ell\}$ by

$$\nu_{\ell,d}(\mathcal{A}) := \int_{\mathbb{S}^{\ell-1}} \int_{\mathcal{O}_d} 1_{\mathcal{A}}(O^* \mathrm{diag}(\lambda_1, \dots, \lambda_\ell, 0, \dots, 0)O) d\lambda dO,$$

where $1_{\mathcal{A}}$ is the indicator function of the set $\mathcal{A}$. It is not hard to see that the mapping

$$K_t^\ell : \mathcal{G}_{k,d} \times \mathcal{G}_{k,d} \to \mathbb{R}$$
$$(P_1, P_2) \mapsto \int_{\mathscr{H}_d^\ell} \langle P_1, M \rangle^t \langle M, P_2 \rangle^t d\nu_{\ell,d}(M)$$

is a positive definite kernel on $\mathcal{G}_{k,d}$. Next, we check that the function spaces under consideration are spanned by the shifts of K_t^ℓ.

Proposition 4.1 *If ℓ and t are nonnegative integers, then*

$$\mathrm{Pol}_t^\ell(\mathcal{G}_{k,d}) = \mathrm{span}\{\langle M, \cdot \rangle^t \big|_{\mathcal{G}_{k,d}} : M \in \mathscr{H}_d, \ \mathrm{rank}(M) = \ell\}$$
$$= \mathrm{span}\{K_t^\ell(P, \cdot)|_{\mathcal{G}_{k,d}} : P \in \mathcal{G}_{k,d}\}.$$

If ℓ_1, ℓ_2 and t_1, t_2 are nonnegative integers, then

$$\mathrm{Pol}_{t_1}^{\ell_1}(\mathcal{G}_{k,d}) \cdot \mathrm{Pol}_{t_2}^{\ell_2}(\mathcal{G}_{k,d}) \subset \mathrm{Pol}_{t_1+t_2}^{\ell_1+\ell_2}(\mathcal{G}_{k,d}).$$

Proof To verify the first equality, we must check that the left-hand side is contained in the right-hand side. We first define

$$\widetilde{\mathrm{Pol}}_t^\ell(\mathcal{G}_{k,d}) := \mathrm{span}\{\langle M, \cdot \rangle^t \big|_{\mathcal{G}_{k,d}} : M \in \mathscr{H}_d^\ell\}.$$

Since the rank ℓ matrices are dense in $\mathscr{H}_d^\ell$, we obtain

$$\widetilde{\mathrm{Pol}}_t^\ell(\mathcal{G}_{k,d}) = \mathrm{span}\{\langle M, \cdot \rangle^t \big|_{\mathcal{G}_{k,d}} : M \in \mathscr{H}_d, \ \mathrm{rank}(M) = \ell\}.$$

Thus, the first equality holds if we can verify that the spaces $\widetilde{\mathrm{Pol}}_t^\ell(\mathcal{G}_{k,d})$ are an ascending sequence in t, i.e., $\widetilde{\mathrm{Pol}}_t^\ell(\mathcal{G}_{k,d}) \subset \widetilde{\mathrm{Pol}}_{t+1}^\ell(\mathcal{G}_{k,d})$. To do so, we first aim to verify

$$\widetilde{\mathrm{Pol}}_t^\ell(\mathcal{G}_{k,d}) = \bigoplus_{t_1+\ldots+t_\ell=t} \widetilde{\mathrm{Pol}}_{t_1}^1(\mathcal{G}_{k,d}) \cdots \widetilde{\mathrm{Pol}}_{t_\ell}^1(\mathcal{G}_{k,d}). \tag{32}$$

The spectral decomposition yields that the left-hand side is contained in the right-hand side. To verify the reverse set inclusion, we must check that

$$P \mapsto \langle P, x_1 x_1^* \rangle^{t_1} \cdots \langle P, x_\ell x_\ell^* \rangle^{t_\ell} \in \widetilde{\mathrm{Pol}}_t^\ell(\mathcal{G}_{k,d}), \quad \text{for all } t_1 + \ldots + t_\ell = t. \tag{33}$$

We now observe that [13, Lemma 7.1] as already used in (3) yields (33). Thus, (32) is satisfied. It was checked in [3] that $\widetilde{\mathrm{Pol}}_t^1(\mathcal{G}_{k,d}) \subset \widetilde{\mathrm{Pol}}_{t+1}^1(\mathcal{G}_{k,d})$ holds, so that (32) implies $\widetilde{\mathrm{Pol}}_t^\ell(\mathcal{G}_{k,d}) \subset \widetilde{\mathrm{Pol}}_{t+1}^\ell(\mathcal{G}_{k,d})$, which yields the first equality.

The second part of the proposition follows from the first equality and (32).

We now take care of the second equality. Since $\mathcal{M} := \mathrm{span}\{\langle \cdot, P_1 \rangle^t \langle \cdot, P_2 \rangle^t |_{\mathscr{H}_d^\ell} : P_1, P_2 \in \mathcal{G}_{k,d}\}$ is finite-dimensional, classical arguments let us infer that there are $\{M_j\}_{j=1}^m \subset \mathscr{H}_d^\ell$ and numbers $\{\omega_j\}_{j=1}^m \in \mathbb{R}$ such that, for all $P_1, P_2 \in \mathcal{G}_{k,d}$,

$$\sum_{j=1}^m \omega_j \langle M_j, P_1 \rangle^t \langle M_j, P_2 \rangle^t = \int_{\mathscr{H}_d^\ell} \langle M, P_1 \rangle^t \langle M, P_2 \rangle^t dv_{\ell,d}(M), \tag{34}$$

cf. [15, Theorem 6.1]. By applying (34), we derive

$$\begin{aligned} K_t^\ell(P, \cdot) &= \int_{\mathscr{H}_d^\ell} \langle M, P \rangle^t \langle M, \cdot \rangle^t dv_{\ell,d}(M) \\ &= \sum_{j=1}^m \omega_j \langle M_j, P \rangle^t \langle M_j, \cdot \rangle^t \in \mathrm{Pol}_t^\ell(\mathcal{G}_{k,d}). \end{aligned}$$

Thus, we have verified that $\mathrm{span}\{K_t^\ell(P, \cdot)|_{\mathcal{G}_{k,d}} : P \in \mathcal{G}_{k,d}\} \subset \mathrm{Pol}_t^\ell(\mathcal{G}_{k,d})$. To verify the reverse inclusion, we shall check that $\dim(\mathrm{span}\{K_t^\ell(P, \cdot)|_{\mathcal{G}_{k,d}} : P \in \mathcal{G}_{k,d}\}) \geq \dim(\mathrm{Pol}_t^\ell(\mathcal{G}_{k,d}))$.

We first observe that

$$\dim(\mathrm{span}\{\langle \cdot, M \rangle^t |_{\mathcal{G}_{k,d}} : M \in \mathscr{H}_d^\ell\}) = \dim(\mathrm{span}\{\langle P, \cdot \rangle^t |_{\mathscr{H}_d^\ell} : P \in \mathcal{G}_{k,d}\}), \tag{35}$$

which is a general principle that holds in much more generality, see [13, Proof of Lemma 5.5] for details. Note that the left-hand side of (35) is $\dim(\mathrm{Pol}_t^\ell(\mathcal{G}_{k,d}))$, and we shall denote this number by r here. Then according to (35), there are $\{P_j\}_{j=1}^r \subset \mathcal{G}_{k,d}$ such that $\{\langle P_j, \cdot \rangle^t |_{\mathscr{H}_d^\ell}\}_{j=1}^r$ is a basis for $\mathrm{span}\{\langle P, \cdot \rangle^t |_{\mathscr{H}_d^\ell} : P \in \mathcal{G}_{k,d}\}$. If we can verify that the matrix $K := \left(K_t^\ell(P_i, P_j)\right)_{i,j=1}^r$ is nonsingular, then $\{K_t^\ell(P_j, \cdot)\}_{j=1}^r$ is linearly

independent, which concludes the proof. Indeed, suppose that $\alpha^* K \alpha = 0$, then we obtain

$$
\begin{aligned}
0 &= \sum_{i,j} \alpha_i \alpha_j K_t^\ell(P_i, P_j) \\
&= \int_{\mathscr{H}_d^\ell} \sum_{i,j} \alpha_i \alpha_j \langle P_i, M \rangle^t \langle M, P_j \rangle^t d\nu_{\ell,d}(M) \\
&= \int_{\mathscr{H}_d^\ell} \Big(\sum_i \alpha_i \langle P_i, M \rangle^t \Big)^2 d\nu_{\ell,d}(M).
\end{aligned}
$$

This implies $\sum_i \alpha_i \langle P_i, M \rangle^t = 0$, for all $M \in \mathrm{supp}(\nu_{\ell,d})$. Since $\langle P_j, \cdot \rangle^t |_{\mathscr{H}_d^\ell}$ are homogeneous polynomials, the latter also holds for all $M \in \mathscr{H}_d^\ell$. The linear independence of $\{\langle P_j, \cdot \rangle^t |_{\mathscr{H}_d^\ell}\}_{j=1}^r$ implies that we must have $\alpha_1, \ldots, \alpha_r = 0$. Thus, K is indeed nonsingular, and this concludes the proof. $\qquad\square$

Remark 4.2 The end of the above proof shows that the special form of $\nu_{\ell,d}$ is not important, and any measure with sufficiently large support would work.

The kernel K_t^ℓ induces an inner product (and hence also a norm $\|\cdot\|_{K_t^\ell}$) on $\mathrm{Pol}_t^\ell(\mathcal{G}_{k,d})$ by

$$
\langle f, g \rangle_{\mathrm{Pol}_t^\ell} := \sum_{i,j} \alpha_i \beta_j K_t^\ell(P_i, \tilde{P}_j), \tag{36}
$$

where $f = \sum_i \alpha_i K_t^\ell(P_i, \cdot)$ and $g = \sum_j \beta_j K_t^\ell(\tilde{P}_i, \cdot)$. Note that the expression (36) does not depend on the special choice of P_i and $\tilde{P}_j$. The induced norm enables us to introduce approximate cubatures:

Definition 4.3 We say that $\{(P_j, \omega_j)\}_{j=1}^n$ is an ϵ-*approximate cubature for* $\mathrm{Pol}_t^\ell(\mathcal{G}_{k,d})$ with respect to K_t^ℓ if

$$
\sup_{f \in \mathrm{Pol}_t^\ell(\mathcal{G}_{k,d}),\ \|f\|_{K_t^\ell}=1} \Big| \sum_{j=1}^n \omega_j f(P_j) - \int_{\mathcal{G}_{k,d}} f(P) d\sigma_{k,d}(P) \Big| \leq \epsilon. \tag{37}
$$

Apparently, an ϵ-approximate cubature for $\mathrm{Pol}_t^\ell(\mathcal{G}_{k,d})$ yields

$$
\Big| \sum_{j=1}^n \omega_j f(P_j) - \int_{\mathcal{G}_{k,d}} f(P) d\sigma_{k,d}(P) \Big| \leq \epsilon \|f\|_{K_t^\ell}, \quad \text{for all } f \in \mathrm{Pol}_t^\ell(\mathcal{G}_{k,d}).
$$

In order to numerically find ϵ-approximate cubatures, we consider the modified fusion frame potential

$$
\sum_{i,j} \omega_i \omega_j K_t^\ell(P_i, P_j). \tag{38}
$$

By following the lines in [12] for the standard fusion frame potential, see also [3], we derive that

$$c_t^\ell := \int_{\mathcal{G}_{k,d}} \int_{\mathcal{G}_{k,d}} K_t^\ell(P, Q) d\sigma_{k,d}(P) d\sigma_{k,d}(Q) \tag{39}$$

is a lower bound on (38), and the gap

$$\sum_{i,j} \omega_i \omega_j K_t^\ell(P_i, P_j) - c_t^\ell \geq 0$$

is exactly the squared cubature error, i.e.,

$$\sum_{i,j} \omega_i \omega_j K_t^\ell(P_i, P_j) - c_t^\ell = \sup_{f \in \mathrm{Pol}_t^\ell(\mathcal{G}_{k,d}), \, \|f\|_{K_t^\ell} = 1} \left| \sum_{j=1}^n \omega_j f(P_j) - \int_{\mathcal{G}_{k,d}} f(P) d\sigma_{k,d}(P) \right|^2.$$

Indeed, if (38) can be minimized numerically, then a proper cubature or at least an ϵ-approximate cubature can be obtained, where ϵ relates to machine precision provided there exists a corresponding cubature for this choice of n. However, numerical evaluation of the kernel K_t^ℓ may be difficult in practice. In the subsequent section, we shall circumvent such difficulties by considering cubatures for larger spaces that enable us to work with a simpler kernel.

5 Construction of Cubatures for $\mathrm{Pol}_t(\mathcal{G}_{k,d})$

5.1 Cubatures from Optimization Procedures

This section is dedicated to derive cubatures from a numerical scheme that is indeed easy to implement. We define polynomials of degree at most t on $\mathcal{G}_{k,d}$ by

$$\mathrm{Pol}_t(\mathcal{G}_{k,d}) := \{\text{polynomials of degree at most } t \text{ on } \mathcal{H}_d \text{ restricted to } \mathcal{G}_{k,d}\}. \tag{40}$$

Note that $\mathrm{Pol}_t(\mathcal{G}_{k,d})$ satisfies the product property that is usually associated with polynomial spaces, i.e.,

$$\mathrm{span}\left(\mathrm{Pol}_{t_1}(\mathcal{G}_{k,d}) \cdot \mathrm{Pol}_{t_2}(\mathcal{G}_{k,d}) \right) = \mathrm{Pol}_{t_1+t_2}(\mathcal{G}_{k,d}),$$

see, for instance, [12]. It is known that these spaces can be rewritten as

$$\mathrm{Pol}_t(\mathcal{G}_{k,d}) = \mathrm{span}\{\langle M, \cdot \rangle^t \big|_{\mathcal{G}_{k,d}} : M \in \mathcal{H}_d\},$$

see, for instance, [3,12]. Obviously, $\mathrm{Pol}_t^\ell(\mathcal{G}_{k,d})$ is contained in $\mathrm{Pol}_t(\mathcal{G}_{k,d})$. In the following proposition, we explore when equality holds:

Proposition 5.1 *For $0 \leq \ell < d$ and $0 \leq t$, we have*

$$\mathrm{Pol}_t^\ell(\mathcal{G}_{k,d}) = \mathrm{Pol}_t(\mathcal{G}_{k,d}), \quad \text{for } \ell \geq \min\{k, t\}.$$

Proof For $\ell \geq k$, the equality is standard cf. [12]. We derive from (32) that

$$\text{Pol}_t^t(\mathcal{G}_{k,d}) = \text{span}\left(\text{Pol}_1^1(\mathcal{G}_{k,d}) \cdots \text{Pol}_1^1(\mathcal{G}_{k,d})\right),$$

where the product has t terms, holds, and the findings in [3] yield that the right-hand side equals $\text{Pol}_t(\mathcal{G}_{k,d})$. Therefore, $\ell \geq t$ also yields $\text{Pol}_t^\ell(\mathcal{G}_{k,d}) = \text{Pol}_t(\mathcal{G}_{k,d})$.

Note that $\{(P_j, \omega_j)\}_{j=1}^n$ being a cubature for $\text{Pol}_3^2(\mathcal{G}_{k,d})$ as used in Corollary 2.5 already implies that it is also a cubature for $\text{Pol}_2(\mathcal{G}_{k,d}) = \text{Pol}_2^2(\mathcal{G}_{k,d})$ and for $\text{Pol}_1(\mathcal{G}_{k,d}) = \text{Pol}_1^2(\mathcal{G}_{k,d})$. It should also be mentioned that the space $\text{Pol}_t^2(\mathcal{G}_{1,d})$ in Corollary 3.5 is the same as $\text{Pol}_t(\mathcal{G}_{1,d})$. Hence, for $k = 1$, we were dealing with the space (40) all along.

A computational approach for cubatures for $\text{Pol}_t(\mathcal{G}_{k,d})$ is discussed in [12]. Since $\text{Pol}_t^2(\mathcal{G}_{k,d})$ is a subset, this approach yields also cubatures for $\text{Pol}_t^2(\mathcal{G}_{k,d})$. By refining some ideas in [12], we shall introduce ϵ-approximate cubatures for the kernel

$$K_t : \mathcal{G}_{k,d} \times \mathcal{G}_{k,d} \to \mathbb{R}$$
$$(P_1, P_2) \mapsto \langle P_1, P_2 \rangle^t.$$

Indeed, K_t is a positive definite kernel on $\mathcal{G}_{k,d}$ and its shifts generate $\text{Pol}_t(\mathcal{G}_{k,d})$, i.e.,

$$\text{Pol}_t(\mathcal{G}_{k,d}) = \text{span}\{K_t(P, \cdot) : P \in \mathcal{G}_{k,d}\}.$$

The kernel K_t induces an inner product on $\text{Pol}_t(\mathcal{G}_{k,d})$ analogously to (36) and, in turn, also a norm $\|\cdot\|_{K_t}$.

Definition 5.2 We say that $\{(P_j, \omega_j)\}_{j=1}^n$ is an ϵ-approximate cubature for $\text{Pol}_t(\mathcal{G}_{k,d})$ with respect to K_t if

$$\sup_{f \in \text{Pol}_t(\mathcal{G}_{k,d}),\ \|f\|_{K_t}=1} \left| \sum_{j=1}^n \omega_j f(P_j) - \int_{\mathcal{G}_{k,d}} f(P) d\sigma_{k,d}(P) \right| \leq \epsilon.$$

In the following, we shall describe that ϵ-approximate cubatures for $\text{Pol}_t(\mathcal{G}_{k,d})$ can be computed by numerical schemes as at the end of Sect. 4.2. Indeed, the potential

$$\sum_{i,j} \omega_i \omega_j K_t(P_i, P_j) \tag{41}$$

can be bounded from below by

$$\lambda_t := \int_{\mathcal{G}_{k,d}} \int_{\mathcal{G}_{k,d}} K_t(P, P') d\sigma_{k,d}(P) d\sigma_{k,d}(P'), \tag{42}$$

so that

$$\sum_{i,j} \omega_i \omega_j K_t(P_i, P_j) - \lambda_t \geq 0.$$

As in the previous section, this gap is exactly the squared cubature error, i.e.,

$$\sum_{i,j} \omega_i \omega_j K_t(P_i, P_j) - \lambda_t = \sup_{f \in \mathrm{Pol}_t(\mathcal{G}_{k,d}),\ \|f\|_{K_t}=1} \left| \sum_{j=1}^{n} \omega_j f(P_j) - \int_{\mathcal{G}_{k,d}} f(P) d\sigma_{k,d}(P) \right|^2.$$

It is remarkable that (42) can be computed exactly by analytical tools, so that the outcome of numerical optimization schemes minimizing (41) can be compared with λ_t, see [12] for further details and examples of successful minimization outcomes. Indeed, the easier structure of the kernel K_t generating $\mathrm{Pol}_t(\mathcal{G}_{k,d})$ makes this approach more amenable to numerical optimization than the setting of $\mathrm{Pol}_t^{\ell}(\mathcal{G}_{k,d})$ presented in the previous section.

5.2 Approximate Cubatures from Randomized Projections

We now examine to what extent a random choice of projections gives an approximate cubature. Let us call a (Borel)-probability measure $\nu_{k,d}$ on $\mathcal{G}_{k,d}$ a *probabilistic cubature* for $\mathrm{Pol}_t(\mathcal{G}_{k,d})$ if

$$\int_{\mathcal{G}_{k,d}} f(P) d\nu_{k,d}(P) = \int_{\mathcal{G}_{k,d}} f(P) d\sigma_{k,d}(P), \quad \text{for all } f \in \mathrm{Pol}_t(\mathcal{G}_{k,d}). \qquad (43)$$

Note that any cubature for $\mathrm{Pol}_t(\mathcal{G}_{k,d})$ can be considered as a finitely supported probabilistic cubature, provided that the weights are nonnegative. Another example, of course, is $\sigma_{k,d}$ itself.

In the remainder of this section, we let each $\omega_j = \frac{1}{n}$ and choose each P_j according to a probabilistic cubature $\nu_{k,d}$. In that case, $\{P_j\}_{j=1}^{n}$ is a collection of random matrices and the expected value of the gap, that is, the squared cubature error, can be computed explicitly. Denoting the expectation with respect to the random choice of $\{P_j\}_{j=1}^{n}$ by $\mathbb{E}_P$, and using that $\mathbb{E}_P K_t(P_i, P_i) = k^t$ and $\mathbb{E}_P K_t(P_i, P_j) = \lambda_t$ if $i \neq j$, we get

$$\mathbb{E}_P \left[\frac{1}{n^2} \sum_{i,j=1}^{n} K_t(P_i, P_j) - \lambda_t \right] = \frac{k^t}{n} + \frac{n(n-1)}{n^2} \lambda_t - \lambda_t = \frac{1}{n}(k^t - \lambda_t).$$

Thus, letting n grow faster than k_t ensures that the expected value of the gap becomes arbitrarily small. In the following theorem, we show that this expected behavior happens with overwhelming probability.

Theorem 5.3 *If $\{P_j\}_{j=1}^{n}$ are chosen independently identically distributed with respect to a probabilistic cubature $\nu_{k,d}$ for $\mathrm{Pol}_t(\mathcal{G}_{k,d})$ and $\tau > 0$, then*

$$\mathbb{P}\left(\frac{1}{n^2} \sum_{i,j=1}^{n} K_t(P_i, P_j) - \lambda_t - \frac{1}{n}(k^t - \lambda_t) \geq \frac{\tau^2 k^t}{n} \right) \leq 4e^{-\Psi_\tau(n)} r_\tau(n),$$

where

$$\Psi_\tau(n) = \frac{\tau^2/2}{(1 - \lambda_t/k^t) + \frac{\tau}{3\sqrt{n}}}, \qquad r_\tau(n) = 1 + \frac{6}{n\tau^2 \ln^2(1 + \frac{\tau}{\sqrt{n}(1-\lambda_t/k^t)})}.$$

Proof First, we note that $\langle P_i, P_j \rangle^t = \langle P_i^{\otimes t}, P_j^{\otimes t} \rangle$, where the Hilbert–Schmidt inner product on the right-hand side is on the Hilbert space $(\mathbb{R}^{d \times d})^{\otimes t} \simeq \mathbb{R}^{d^2 t}$. Thus, $\sum_{i,j} K_t(P_i, P_j) = \| \sum_j P_j^{\otimes t} \|_{HS}^2$ holds.

We define the averaged tensor power $\Lambda_t = \mathbb{E}_P P_1^{\otimes t}$. For $P_0 \in \mathcal{G}_{k,d}$, we can compute

$$\langle P_0^{\otimes t}, \Lambda_t \rangle = \left\langle P_0^{\otimes t}, \int_{\mathcal{G}_{k,d}} P^{\otimes t} d\nu_{k,d}(P) \right\rangle = \int_{\mathcal{G}_{k,d}} \langle P_0, P \rangle^t d\nu_{k,d}(P).$$

Since $\nu_{k,d}$ is a probabilistic cubature and $\langle P_0, \cdot \rangle^s \in \mathrm{Pol}_t(\mathcal{G}_{k,d})$, we obtain

$$\langle P_0^{\otimes t}, \Lambda_t \rangle = \int_{\mathcal{G}_{k,d}} \langle P_0, P \rangle^t d\sigma_{k,d}(P).$$

Let $U \in \mathcal{O}_d$ be such that $U^* D_k U = P_0$, where D_k denotes the diagonal matrix with k ones and zeros elsewhere. The commutativity of the trace and the orthogonal invariance of $\sigma_{k,d}$ yield

$$\langle P_0^{\otimes t}, \Lambda_t \rangle = \int_{\mathcal{G}_{k,d}} \langle U^* D_k U, P \rangle^t d\sigma_{k,d}(P)$$

$$= \int_{\mathcal{G}_{k,d}} \langle D_k, U P U^* \rangle^t d\sigma_{k,d}(P)$$

$$= \int_{\mathcal{G}_{k,d}} \langle D_k, P \rangle^t d\sigma_{k,d}(P).$$

By applying the probabilistic cubature property once more, we derive

$$\langle P_0^{\otimes t}, \Lambda_t \rangle = \int_{\mathcal{G}_{k,d}} \langle D_k, P \rangle^t d\nu_{k,d}(P) = \langle D_k^{\otimes t}, \Lambda_t \rangle.$$

Thus, the term $\langle P_0^{\otimes t}, \Lambda_t \rangle$ does not depend on the particular choice of $P_0 \in \mathcal{G}_{k,d}$. Averaging over all P_0 with respect to $\sigma_{k,d}$ then implies that for each i,

$$\langle P_i^{\otimes t}, \Lambda_t \rangle = \int_{\mathcal{G}_{k,d}} \langle P_0^{\otimes t}, \Lambda_t \rangle d\sigma_{k,d}(P_0) = \|\Lambda_t\|_{HS}^2. \tag{44}$$

Similarly to the above computations, the probabilistic cubature property also yields

$$\|\Lambda_t\|_{HS}^2 = \lambda_t. \tag{45}$$

By applying (44) and (45), taking $Y_j = (P_j^{\otimes t} - \Lambda_t)/k^{t/2}$ then gives

$$\|Y_j\|_{HS}^2 = 1 - \lambda_t/k^t.$$

Hence, $\|Y_j\|_{HS} \leq 1$ and $\mathbb{E}_P Y_j = 0$, so that Minsker's vector-valued Bernstein inequality [23, Corollary 5.1] provides, for all $\tau > 0$,

$$\mathbb{P}\left(\left\|\frac{1}{n}\sum_{j=1}^{n} Y_j\right\|_{HS}^2 > \tau^2/n\right) \leq 4e^{-\Psi_\tau(n)}r_\tau(n),$$

where Ψ_τ and r_τ are as stated. To finish the proof, we observe that

$$k^t \left\|\frac{1}{n}\sum_{j=1}^{n} Y_j\right\|_{HS}^2 = \frac{1}{n^2}\sum_{i,j=1}^{n} K_t(P_i, P_j) - \lambda_t - \frac{1}{n}(k^t - \lambda_t).$$

$\square$

When n tends to infinity, then $r_\tau(n) \to 1 + \frac{6(1-\lambda_t/k^t)^2}{\tau^4}$ and $\Psi_\tau(n) \to \frac{\tau^2/2}{1-\lambda_t/k^t}$. Thus, for large n, the distribution of the gap concentrates near zero at the same rate as the expected value. Since the gap is the square of the maximal cubature error, we conclude a probabilistic construction of approximate cubatures.

Corollary 5.4 *If $\{P_j\}_{j=1}^{n}$ are chosen independently from a probabilistic cubature for* $\mathrm{Pol}_t(\mathcal{G}_{k,d})$ *and $\tau > 0$, then a* $\sqrt{\frac{(1+\tau^2)k^t - \lambda_t}{n}}$*-approximate cubature for* $\mathrm{Pol}_t(\mathcal{G}_{k,d})$ *with respect to K_t is obtained with probability bounded below by $1 - 4e^{-\Psi_\tau(n)}r_\tau(n)$.*

For related results on random matrices, we refer to [6,22,27,28].

6 Error Propagation for ϵ-Approximate Cubatures

The numerical optimization approach in general can provide cubatures up to machine precision only. Therefore, we are dealing with ϵ-approximate cubatures and this is also what we obtain from the random constructions. In these cases, the moment reconstruction formulas in Corollary 2.5 hold up to some error term:

Theorem 6.1 *Let $X \in \mathbb{S}^{d-1}$ be a random vector and $\{(P_j, \omega_j)\}_{j=1}^{n}$ be an ϵ-approximate cubature for* $\mathrm{Pol}_t(\mathcal{G}_{k,d})$ *with respect to K_t. Then (7) in Corollary 2.3 holds up to a constant c_α times ϵ, i.e., for $\alpha \in \mathbb{N}^d$, $|\alpha| = t$,*

$$\left|\mathbb{E}X^\alpha - \sum_{|\beta| \leq t} a_\beta^\alpha \sum_{j=1}^{n} \omega_j \mathbb{E}(P_j X)^\beta\right| \leq \epsilon c_\alpha. \tag{46}$$

If $k = 1$ and X is random vector in $\mathbb{R}^d$, then (30) in Corollary 3.5 holds up to a constant times $\epsilon\mathbb{E}\|X\|^t$.

The above theorem verifies that the cubature error propagates in a linear fashion when it comes to the moment reconstruction formulas. It should be mentioned though that the constant c_α depends on k and d.

Proof According to Theorem 2.1, we derive, for $x \in \mathbb{S}^{d-1}$

$$x^\alpha = \sum_{s=1}^{t} \sum_{i=1}^{m} f_{s,i} \mu_{k,d}^s (E_{x,y_i})$$

$$= \sum_{s=1}^{t} \sum_{i=1}^{m} f_{s,i} \int_{\mathcal{G}_{k,d}} \langle P, E_{x,y_i} \rangle^s d\sigma_{k,d}(P)$$

Since the function $F_x^\alpha = \sum_{s=1}^{t} \sum_{i=1}^{m} f_{s,i}^\alpha \langle \cdot, E_{x,y_{s,i}^\alpha} \rangle^s$ is an element in $\mathrm{Pol}_t(\mathcal{G}_{k,d})$, the cubature property yields

$$\left| x^\alpha - \sum_{s=1}^{t} \sum_{i=1}^{m} f_{s,i}^\alpha \sum_{j=1}^{n} \omega_j \langle P_j, E_{x,y_{s,i}^\alpha} \rangle^s \right| \leq \epsilon \|F_x\|_{K_t}.$$

The coefficients a_β^α in Corollary 2.3 are used with $c_\alpha := \sup_{x \in \mathbb{S}^{d-1}} \|F_x^\alpha\|_{K_t}$ to derive (46).

The second part of the theorem can be verified in an analogous fashion, so we omit the details. $\square$

7 Concluding Remarks

Our results appear to match reasonable characteristics in distributed sensing. We require a rather large set of sensors (projectors) and we assume that the high-dimensional signal is modeled by means of a probability distribution. The sensors are deterministic and can even be given by the experimental setup as long as we are able to find weights such that projectors and weights altogether form a cubature. Each sensor must reconstruct the first few moments of the projection marginal distribution, which may allow in practice for fewer data samples than for estimating the marginal distribution itself. In the end, the first few moments of the high-dimensional random signal can be computed with low costs by a closed formula.

As far as we know, the present paper is a first attempt to address this type of moment recovery problem with tools from harmonic analysis. Further investigations are necessary to combine those ideas with proper statistical estimation techniques, in which the low-dimensional moments are estimated from acquired data. This is intended to be addressed in forthcoming work.

References

1. Bachoc, C., Bannai, E., Coulangeon, R.: Codes and designs in Grassmannian spaces. Discrete Math. **277**, 15–28 (2004)
2. Bachoc, C., Coulangeon, R., Nebe, G.: Designs in Grassmannian spaces and lattices. J. Algebraic Comb. **16**, 5–19 (2002)
3. Bachoc, C., Ehler, M.: Tight p-fusion frames. Appl. Comput. Harmon. Anal. **35**(1), 1–15 (2013)
4. Bélisle, C., Massé, J.C., Ransford, T.: When is a probability measure determined by infinitely many projections? Ann. Probab. **25**(2), 767–786 (1997)
5. Bochnak, J., Coste, M., Roy, M.: Real Algebraic Geometry. Springer, Berlin (1998)
6. Bodmann, B.G.: Random fusion frames are nearly equiangular and tight. Linear Algebra Appl. **439**, 1401–1414 (2013)
7. Calderbank, A.R., Hardin, R.H., Rains, E.M., Shor, P.W., Sloane, N.J.A.: A group-theoretic framework for the construction of packings in Grassmannian spaces. J. Algebraic Comb. **9**, 129–140 (1997)
8. Casazza, P.G., Kutyniok, G., Li, S.: Fusion frames and distributed processing. Appl. Comput. Harmon. Anal. **25**(1), 114–132 (2008)
9. Chikuse, Y.: Statistics on Special Manifolds. Lecture Notes in Statistics. Springer, New York (2003)
10. Cuesta-Albertos, J., Fraiman, R., Ransford, T.: A sharp form of the Cramér–Wold theorem. J. Theor. Probab. **20**(2), 201–209 (2007)
11. de la Harpe, P., Pache, C.: Cubature formulas, geometrical designs, reproducing kernels, and Markov operators. In: Infinite Groups: Geometric, Combinatorial and Dynamical Aspects, vol. 248, pp. 219–267 . Birkhäuser, Basel (2005)
12. Ehler, M., Gräf, M.: Reproducing kernels for the irreducible components of polynomial spaces on unions of Grassmannians (2017). arXiv:1411.5865v4
13. Ehler, M., Gräf, M., Kiraly, F.: Phase retrieval using cubatures of positive semidefinite matrices. Waves Wavel Fractals Adv. Anal. **1**(1), 32–50 (2015)
14. Gilbert, W.M.: Projections of probability distributions. Acta Math. Acad. Sci. Hung. **6**(1–2), 195–198 (1955)
15. Gräf, M.: Efficient Algorithms for the Computation of Optimal Quadrature Points on Riemannian Manifolds. Universitäts Verlag, Chemnitz (2013)
16. Gross, K., St, D., Richards, P.: Special functions of matrix argument. I: algebraic induction, zonal polynomials and hypergeometric functions. Trans. Am. Math. Soc. **301**, 781–811 (1987)
17. Heppes, A.: On the determination of probability distributions of more dimensions by their projections. Acta Math. Hung. **7**(3–4), 403–410 (1956)
18. James, A.T.: Distributions of matrix variates and latent roots derived from normal samples. Ann. Math. Stat. **35**(2), 475–501 (1964)
19. James, A.T., Constantine, A.G.: Generalized Jacobi polynomials as spherical functions of the Grassmann manifold. Proc. Lond. Math. Soc. **29**(3), 174–192 (1974)
20. Kutyniok, G., Pezeshki, A., Calderbank, R., Liu, T.: Robust dimension reduction, fusion frames, and Grassmannian packings. Appl. Comput. Harmon. Anal. **26**(1), 64–76 (2009)
21. Kutz, E.R.: Cohen–Macaulay rings and ideal theory in rings of invariants of algebraic groups. Trans. Am. Math. Soc. **194**, 115–129 (1974)
22. Levina, E., Vershynin, R.: Partial estimation of covariance matrices. Prob. Theory Rel. Fields **153**, 405–419 (2012)
23. Minsker, S.: On some extensions of Bernstein's inequality for self-adjoint operators (2017). arXiv:1112.5448v2
24. Muirhead, R.J.: Aspects of Multivariate Statistical Theory. Wiley, New York (1982)
25. Procesi, C.: The invariant theory of $n \times n$ matrices. Adv. Math. **19**, 306–381 (1976)
26. Rényi, A.: On projections of probability distributions. Acta Math. Acad. Sci. Hung. **3**(3), 131–142 (1952)
27. Vershynin, R.: Spectral norm of products of random and deterministic matrices. Prob. Theory Relat. Fields **150**, 471–509 (2011)
28. Vershynin, R.: How close is the sample covariance matrix to the actual covariance matrix? J. Theor. Probab. **25**, 655–686 (2012)