

On conditional moments of high-dimensional random vectors given lower-dimensional projections

LUKAS STEINBERGER* and HANNES LEEB**

Department of Statistics, University of Vienna, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria.
E-mail: *lukas.steinberger@univie.ac.at; **hannes.leebe@univie.ac.at

One of the most widely used properties of the multivariate Gaussian distribution, besides its tail behavior, is the fact that conditional means are linear and that conditional variances are constant. We here show that this property is also shared, in an approximate sense, by a large class of non-Gaussian distributions. We allow for several conditioning variables and we provide explicit non-asymptotic results, whereby we extend earlier findings of Hall and Li (*Ann. Statist.* **21** (1993) 867–889) and Leeb (*Ann. Statist.* **41** (2013) 464–483).

Keywords: conditional moments; constant conditional variance; high dimensional distribution; linear conditional mean

1. Introduction

1.1. Informal summary

The property of the multivariate Gaussian law, that conditional means are linear and that conditional variances are constant, is used by several fundamental statistical methods, even if these methods per se do not require Gaussianity: the generic linear model is built on the assumption that the conditional mean of the response is linear in the (conditioning) explanatory variables; and the generic homoskedastic linear model rests on the additional assumption that the conditional variance is constant. Linear conditional means and/or constant conditional variances are also assumed, for example, by methods for sufficient dimension reduction such as SIR [11] or SAVE [4], or by certain imputation techniques [14]. Elliptically contoured distributions are characterized by linear conditional means [5]. And methods for spatial statistics such as Kriging rely on Gaussianity mainly through the property that conditional means are linear and that conditional variances are constant.¹ But even though these properties are widely used, in a sense the only distribution that has both linear conditional means and constant conditional variances is the Gaussian (see also Section 1.2).

In this paper, we show that conditional means are approximately linear and that conditional variances are approximately constant, for a large class of multivariate distributions, when the

¹Distributions with linear conditional means and/or constant conditional variances are also studied, for example, in [1,3,8,13,18,19].

conditioning is on lower-dimensional projections. To illustrate our results, consider a random d -vector Z that has a Lebesgue density, and a $d \times p$ matrix B . Conditional on $B'Z$, we show that the mean of Z is linear in $B'Z$, and that the variance/covariance matrix of Z is constant, in an approximate sense. Typically, our approximation error bounds are small if d is sufficiently large relative to p . Our results extend recent findings of [9], where the case $p = 1$ is considered (which, from a modeling perspective, covers only models with one explanatory variable). More precisely, we extend and refine the results of [9] in three directions: First, we allow for the case where $p > 1$, thereby also proving a result that is outlined in [6], Section 5. Second, we derive non-asymptotic and explicit error bounds that hold for fixed d , whereas [6] and [9] only give asymptotic results that hold as $d \rightarrow \infty$; cf. Theorem 2.1. And third, we also give asymptotic results where p is allowed to increase with d ; see Corollary 2.4. In many cases, our error bounds go to zero if $p/\log d \rightarrow 0$.

The rest of the paper is organized as follows: We continue this section with a more detailed description of the results that we derive. Our main results are then stated in Section 2. In Section 3, we provide a number of examples where the assumptions of our main theorem are satisfied and we discuss further extensions of our work. Finally, Section 4 gives a high-level description of the proof. The more technical low-level parts of the proof as well as proofs of results from Section 3 are collected in the online supplementary material [16].

1.2. Outline of results

Consider a random d -vector Z that has a Lebesgue density, and that is centered and standardized so that $\mathbb{E}Z = 0$ and $\mathbb{E}ZZ' = I_d$. And take a $d \times p$ matrix B with orthonormal columns. [While we do rule out degenerate distributions, the requirement that Z is centered and standardized, and the requirement that the columns of B are orthonormal, are inconsequential; cf. Remark 1.1 as well as Section 3.3.] Our objective is to show that the conditional mean and the conditional variance of Z given $B'Z$ are close to what they would be if Z were Gaussian. In the following, we use the notation $\|\cdot\|$ to denote the Euclidean norm of vectors and the spectral norm of matrices; the meaning of $\|\cdot\|$ will always be clear from the context.

Instead of the conditional mean and variance, it will be convenient to focus on the first two conditional moments, that is, on $\mathbb{E}[Z\|B'Z]$ and on $\mathbb{E}[ZZ'\|B'Z]$. If both the expressions

$$\|\mathbb{E}[Z\|B'Z] - BB'Z\| \quad \text{and} \quad \|\mathbb{E}[ZZ'\|B'Z] - (I_d - BB' + BB'ZZ'BB')\|$$

are equal to zero, then the conditional mean of Z given $B'Z$ is linear in $B'Z$, and the corresponding conditional variance is constant in $B'Z$. But the only distribution, which satisfies this for all B , is the Gaussian law; cf. the discussion in [9], page 466. We will show that a weaker form of this requirement, namely that the expressions in the preceding display are close to zero in probability for most B , is satisfied by a much larger class of distributions, provided mainly that d is sufficiently large relative to p .

For the case where $p = 1$, it was shown in [9], for each $t > 0$, that

$$\sup_{B \in \mathbb{G}} \mathbb{P}(\|\mathbb{E}[Z\|B'Z] - BB'Z\| > t) \quad \text{and} \quad (1.1)$$

$$\sup_{B \in \mathbb{G}} \mathbb{P}(\|\mathbb{E}[ZZ' \| B'Z] - (I_d - BB' + BB'ZZ'BB')\| > t) \tag{1.2}$$

converge to zero as $d \rightarrow \infty$, under some conditions, where the sets \mathbb{G} are collections of $d \times p$ matrices with orthonormal columns that become large as $d \rightarrow \infty$. More precisely, for $\nu_{d,p}(\cdot)$ denoting the uniform distribution on the set of all such matrices (i.e., the Haar measure on the Stiefel manifold $\mathcal{V}_{d,p}$), the sets \mathbb{G} satisfy $\nu_{d,p}(\mathbb{G}) \rightarrow 1$ as $d \rightarrow \infty$. [Obviously, \mathbb{G} depends on d and also on p , although this dependence is not shown explicitly in our notation.] In the case where $p = 1$ covered in [9], the sets \mathbb{G} are collections of unit-vectors, and $\nu_{d,1}(\cdot)$ is the uniform distribution on the unit sphere, in \mathbb{R}^d . We derive a non-asymptotic version of this result, that is, explicit upper bounds on (1.1) and (1.2), and also on $1 - \nu_{d,p}(\mathbb{G})$, that hold for fixed d and p , where we allow for $p > 1$; see Theorem 2.1. Moreover, we also provide an asymptotic result where our upper bounds go to zero as $d \rightarrow \infty$, where p may increase with d ; cf. Corollary 2.4. In many cases, our upper bounds are small provided that $p/\log d$ is small. Both our non-asymptotic and our asymptotic result, i.e., both Theorem 2.1 and Corollary 2.4, hold uniformly over classes of distributions for Z , as outlined in Remark 2.2.

Of course, our results rely on further conditions on the distribution of Z (in addition to the existence of a Lebesgue density and the requirements that $\mathbb{E}Z = 0$ and $\mathbb{E}ZZ' = I_d$). In particular, we require that the mean of certain functions of Z , and of i.i.d. copies of Z , is bounded; see the bounds (b1)(a) and (b2), as well as the attending discussion in Section 2. And we require that certain moments of Z are close to what they would be in the Gaussian case; see (b1)(b)–(c). From a statistical perspective, we stress that our results rely on bounds that can be estimated from appropriate data, as outlined in the discussion leading up to Theorem 2.1. One particularly simple example, where these bounds hold, and where the error bounds in Theorem 2.1 get small as d gets large, is the case where the components of Z are independent, with bounded marginal densities and bounded marginal moments of sufficiently large order; see Example 3.1. Finally, we emphasize that (b1) and (b2) do not require that the components of Z are independent.

The results in this paper demonstrate that the requirement of linear conditional means and constant conditional variances (which is quite restrictive as discussed in the second paragraph of this subsection) is actually satisfied, in an approximate sense, by a rather large class of distributions. Some implications, namely to sparse linear modeling, and to sufficient dimension reduction methods like SIR or SAVE, are discussed in [9], Section 1.4. And while the discussion in [9] is hampered by the fact that only situations with $p = 1$, that is, *only models with one explanatory variable*, are covered in that paper, our results show that these considerations extend also to the case where $p > 1$, that is, to more complex models with several explanatory variables.

Remark 1.1. (i) Our requirements, that the random d -vector Z is centered and standardized, and that the matrix B has orthonormal columns, are inconsequential in the following sense: Consider a random d -vector Y such that $\mathbb{E}[Y] = \mu$ and $\text{Var}[Y] = \Sigma$ are both well-defined and finite, and such that Y has a Lebesgue density (which also entails that Σ is invertible). Moreover, consider a $d \times p$ matrix A with linearly independent columns. If Y were Gaussian, we would have $\mathbb{E}[Y \| A'Y] = \mu + \Sigma A(A'\Sigma A)^{-1} A'(Y - \mu)$. In general, one easily verifies that

$$\|\mathbb{E}[Y \| A'Y] - (\mu + \Sigma A(A'\Sigma A)^{-1} A'(Y - \mu))\| \leq \|\Sigma\|^{1/2} \|\mathbb{E}[Z \| B'Z] - BB'Z\|$$

holds for $Z = \Sigma^{-1/2}(Y - \mu)$ and $B = \Sigma^{1/2}A(A'\Sigma A)^{-1/2}$. Note that Z has a Lebesgue density; that Z is centered and standardized so that $\mathbb{E}[Z] = 0$ and $\mathbb{E}[ZZ'] = I_d$; and that the columns of B are orthonormal. In particular, we see that the conditional mean of Y given $A'Y$ is approximately linear if the same is true for the conditional mean of Z given $B'Z$, provided only that the largest eigenvalue of Σ is not too large. A similar consideration applies, mutatis mutandis, to the conditional variance of Y given $A'Y$ and that of Z given $B'Z$. For further details, in particular about the role of Σ , see Section 3.3.

(ii) Conditioning on $B'Z$ is equivalent to conditioning on $BB'Z$, which is the orthogonal projection of Z onto the column space of B . Therefore, we could formulate Theorem 2.1 for collections of p -dimensional subspaces S of \mathbb{R}^d (elements of the Grassmann manifold $\mathcal{G}_{d,p}$) instead of matrices B (from the Stiefel manifold $\mathcal{V}_{d,p}$), and thus replace (1.1) by

$$\sup_{S \in \mathbb{H}} \mathbb{P}(\|\mathbb{E}[Z \| P_S Z] - P_S Z\| > t),$$

with P_S denoting the orthogonal projection matrix for the subspace S . Here, \mathbb{H} denotes the image of the set $\mathbb{G} \subseteq \mathcal{V}_{d,p}$ from (1.1) under the mapping that maps a matrix B into its column space S . Note that the image of the Haar measure on $\mathcal{V}_{d,p}$ under this mapping is the Haar measure on $\mathcal{G}_{d,p}$; see also [2], Theorem 2.2.2(iii). In a similar manner, one can also write (1.2) in terms of the Grassmann manifold, namely as

$$\sup_{S \in \mathbb{H}} \mathbb{P}(\|\mathbb{E}[ZZ' \| P_S Z] - (I_d - P_S + P_S Z(P_S Z)')\| > t).$$

2. Results

We first present our main non-asymptotic result, that is, Theorem 2.1, and the bounds (b1) and (b2) that it relies on. These bounds depend on a constant k that will be chosen as needed later. In Corollary 2.4 and the attending discussion, we then present asymptotic scenarios in which the constants in (b1) and (b2) can be controlled, such that the error bounds in Theorem 2.1 become small. Throughout the following, consider a random d -vector Z that has a Lebesgue density and that satisfies $\mathbb{E}Z = 0$ as well as $\mathbb{E}ZZ' = I_d$. For $k \in \mathbb{N}$, write Z_1, \dots, Z_k for i.i.d. copies of Z , and write S_k for the $k \times k$ Gram-matrix $S_k = (Z_i' Z_j / d)_{i,j=1}^k$. For $g \geq 0$, a monomial of degree g in the elements of $S_k - I_k$ is an expression of the form $G = \prod_{\ell=1}^g (S_k - I_k)_{i_\ell, j_\ell}$ for $(i_\ell, j_\ell) \in \{1, \dots, k\}^2, i_\ell \leq j_\ell, 1 \leq \ell \leq g$ (with the convention that $G = 1$ in case $g = 0$). We say that G has a linear (resp. quadratic) factor if one of the pairs, say (i_1, j_1) , occurs exactly once (resp. twice).

(b1) Fix $k \in \mathbb{N}$.

- (a) There are constants $\varepsilon \in [0, 1/2]$ and $\alpha \geq 1$ so that $\mathbb{E}\|\sqrt{d}(S_k - I_k)\|^{2k+1+\varepsilon} \leq \alpha$.
- (b) There are constants $\beta > 0$ and $\xi \in (0, 1/2]$ that satisfy the following: For any monomial $G = G(S_k - I_k)$ in the elements of $S_k - I_k$, whose degree g satisfies $g \leq 2k$, we have $|d^{g/2}\mathbb{E}G - 1| \leq \beta/d^\xi$ if G consists only of quadratic factors in elements above the diagonal, and $|d^{g/2}\mathbb{E}G| \leq \beta/d^\xi$ if G contains a linear factor.

- (c) The constants β and ξ in (b) also satisfy the following: Consider two monomials $G = G(S_k - I_k)$ and $H = H(S_k - I_k)$ of degree g and h , respectively, in the elements of $S_k - I_k$. If G is given by $Z'_1 Z_2 Z'_2 Z_3 \cdots Z'_{g-1} Z_g Z'_g Z_1 / d^g$, if $H = \prod_{\ell=1}^h (S_k - I_k)_{i_\ell, j_\ell}$ with $\{1, \dots, g\} \subseteq \{i_1, j_1, \dots, i_h, j_h\}$, and if $2 \leq h < g \leq k$, then $|d^g \mathbb{E}GH| \leq \beta / d^\xi$.

(b2) For fixed $k \in \mathbb{N}$, there is a constant $D \geq 1$, such that the following holds true: If R is an orthogonal $d \times d$ matrix, then a marginal density of the first $d - k + 1$ components of RZ is bounded by $\binom{d}{k-1}^{1/2} D^{d-k+1}$.

The bounds in (b1) and (b2) essentially guarantee that moments of certain functions of the Gram matrix S_k are either bounded (in (b1)(a) and (b2)) or not too different from what they would be if Z were Gaussian (in (b1)(b)–(c)). We will impose (b1) and (b2) with $k = 2$ when considering conditional means, and with $k = 4$ when considering conditional variances. Clearly, (b1) becomes stronger as k increases. The specific requirements in (b1) are minimal for our current method of proof and the bound in (b2) is chosen in such a way that certain constants γ_1 and γ_2 appearing in Theorem 2.1 do not depend on the dimension d . Other methods of proof, if such can be found, may rely on different conditions. For further discussion and specific examples where our conditions apply, see Section 3.1.

The bounds in (b1) are non-asymptotic versions of condition (t1) in [9], and the bound in (b2) coincides with condition (t2) in that reference. The bounds in (b1)(b)–(c) are written as β/d^ξ , because in Corollary 2.4 we will consider situations where these bounds hold for constants β and ξ that either are both independent of d , or that are such that β is independent of d while ξ depends on d so that $1/d^\xi \rightarrow 0$. In (b2), note that the upper bound on the marginal densities can increase in d . The bound in (b2) appears to be qualitatively different from (b1) in that it does not directly impose restrictions on moments involving the standardized Gram matrix $S_k - I_k$. However, (b2) is used only to bound the p th moment of $\det S_l^{-4(k+1)}$ for $l = 1, \dots, k$; cf. Lemma E.5 and the proof of Proposition 4.4 in Appendix E of the supplement. Just like the bound in (b1), the requirement of a uniform bound on $\max_{l \leq k} \mathbb{E} \det S_l^{-4p(k+1)}$ becomes more restrictive if k increases. From a statistical perspective, we note that the moment-bounds discussed here can be estimated from a sample of independent copies of Z . Indeed, population means like $\mathbb{E} \|S_k - I_k\|^{2k+1+\varepsilon}$, $\mathbb{E}G$, $\mathbb{E}GH$, or $\mathbb{E} \det S_l^{-4p(k+1)}$ as above are readily estimated by appropriate sample means. In this sense, we rely on bounds that can be estimated from data.

Theorem 2.1. *For fixed d , consider a random d -vector Z that has a Lebesgue density f_Z and that is standardized such that $\mathbb{E}Z = 0$ and $\mathbb{E}ZZ' = I_d$.*

- (i) *Suppose that (b1)(a)–(b) and (b2) hold with $k = 2$. Then, for each $p < d$ and for each $\tau \in (0, 1)$, there is a Borel set $\mathbb{G} \subseteq \mathcal{V}_{d,p}$ such that (1.1) is bounded by*

$$\frac{1}{t} d^{-\tau \xi_1} + \frac{\gamma_1}{1 - \tau} \frac{p}{3 \xi_1 \log d} \tag{2.1}$$

for each $t > 0$, and such that

$$\nu_{d,p}(\mathbb{G}^c) \leq \kappa_1 d^{-\tau \xi_1 (1 - \frac{\gamma_1}{\tau} \frac{p}{\xi_1 \log d})}, \tag{2.2}$$

where ξ_1 is given by $\xi_1 = \min\{\xi, \varepsilon/2 + 1/4, 1/2\}/3$ and $\gamma_1 = \max\{g_1, 6 + 2 \log(2D\sqrt{\pi e})\}$. Here, the constant κ_1 depends only on α and β , and g_1 is a global constant.

(ii) Suppose that (b1)(a)–(c) and (b2) hold with $k = 4$. Then, for each $p < d$ and for each $\tau \in (0, 1)$, there is a Borel set $\mathbb{G} \subseteq \mathcal{V}_{d,p}$ so that both (1.1) and (1.2) are bounded by

$$\frac{1}{t} d^{-\tau \xi_2} + \frac{\gamma_2}{1 - \tau} \frac{p}{5 \xi_2 \log d} \tag{2.3}$$

for each $t > 0$, and such that

$$\nu_{d,p}(\mathbb{G}^c) \leq 2\kappa_2 d^{-\tau \xi_2 (1 - \frac{\gamma_2}{\tau} \frac{p}{\xi_2 \log d})}. \tag{2.4}$$

Here, ξ_2 is given by $\xi_2 = \min\{\xi, \varepsilon/2 + 1/4, 1/2\}/5$ and $\gamma_2 = \max\{g_2, 10 + 4 \log(2D\sqrt{\pi e})\}$. The constant κ_2 depends only on α and β , and g_2 is a global constant.

(iii) The set \mathbb{G} in both parts (i) and (ii) can be chosen to have the following additional properties: \mathbb{G} is right-invariant under the action of the orthogonal group of order p and it is left orthogonally equivariant, that is, $\mathbb{G} = \mathbb{G}(f_Z)$ depends on the distribution of Z in such a way that $\mathbb{G}(f_{RZ}) = R\mathbb{G}(f_Z)$, for every $d \times d$ orthogonal matrix R .

The constants g_1, g_2, κ_1 and κ_2 in part (i) and (ii) can be obtained explicitly upon detailed inspection of the proof.

With Theorem 2.1, we aimed to obtain the best possible upper bounds for (1.1), (1.2) and $\nu_{d,p}(\mathbb{G}^c)$ that our current technique of proof delivers. It is likely that better bounds can be obtained under stronger assumptions (like in the case where the components of Z are independent) together with an alternative method of proof. In particular, when bounding (1.1) in Theorem 2.1(i), the term $\frac{\gamma_1}{1-\tau} \frac{p}{3\xi_1 \log d}$ is obtained by bounding $\mathbb{P}(\|B'Z\|^2 > (1 - \tau)3\xi_1 \log(d)/\gamma_1)$ using Chebyshev’s inequality; cf. the proof of Lemma B.2 in the supplement. Under appropriate additional assumptions on the tails of $\|B'Z\|$, this bound can be dramatically improved. The bound on both (1.1) and (1.2) in Theorem 2.1(ii) can be improved in a similar fashion (cf. Section 3.2). When proving Theorem 2.1, we derive upper bounds for (1.1) and (1.2), on the one hand, and for $\nu_{d,p}(\mathbb{G}^c)$, on the other hand, that are antagonistic in the sense that one can be reduced at the expense of the other (namely in the proof of Lemma B.2). For Theorem 2.1, we have balanced these bounds so that both are of the same leading order in d , that is, $d^{-\tau \xi_1}$ in part (i) and $d^{-\tau \xi_2}$ in part (ii).

Remark 2.2. Because the error bounds in Theorem 2.1 depend on Z only through the constants that occur in (b1) and (b2), the theorem a fortiori holds uniformly over the class of all distributions for Z that satisfy (b1) and (b2). For example: Fix constants $\varepsilon, \alpha, \beta$ and ξ as in (b1), fix D as in (b2), and write \mathcal{Z} for the class of all random d -vectors Z that satisfy the bounds (b1)(a)–(c) and (b2) for $k = 4$, that have a Lebesgue density, and that are centered and standardized. Then, for each $Z \in \mathcal{Z}$ and for each $p < d$, there exists a Borel set $\mathbb{G} \subseteq \mathcal{V}_{d,p}$ (that depends on Z), so that (1.1), (1.2) and also $\nu_{d,p}(\mathbb{G}^c)$ are bounded as in Theorem 2.1(ii). Similar considerations also apply, mutatis mutandis, to Theorem 2.1(i) and to Corollary 2.4, which follows.

Remark 2.3. Our results provide conditions under which conditional means are approximately linear and conditional variances are approximately constant, provided that $p/\log d$ is small. Theorem 2.1 provides such a statement for a fixed distribution of Z and for many B . By a slight change of perspective, this also leads to a similar statement that holds for fixed B and many distributions of Z , cf. [17]. We cannot deal with a fixed matrix B and a fixed distribution of Z with our methods. Whether, say, the conditional variance is approximately constant for given B and Z depends on the particulars of B and Z , irrespective of p and d . A few trivial examples, however, are well known. For instance, if the distribution of Z is spherically symmetric, then the conditional expectation of Z given $B'Z$ is exactly linear for every matrix B . Moreover, the conditional expectation is linear and the conditional variance is constant if the components of Z are independent and $B = (e_{j_1}, \dots, e_{j_p})$, where e_j is the j th element of the standard basis in \mathbb{R}^d . See also Section 2.3.4. in Chapter 2 of [15] for a non-trivial example with $p = 1$ and $d = 2$.

Corollary 2.4. For each d , consider a random d -vector $Z^{(d)}$ that has a Lebesgue density and that satisfies $\mathbb{E}Z^{(d)} = 0$ and $\mathbb{E}Z^{(d)}Z^{(d)'} = I_d$. And for each d , suppose that (b1)(a)–(c) and (b2) hold with $Z^{(d)}$ replacing Z and with $k = 4$, such that the constants $\varepsilon, \alpha, \beta$, and D in these bounds do not depend on d , while the constant $\xi = \xi_d$ in (b1) may depend on d as long as $d^{-\xi_d} \rightarrow 0$ as $d \rightarrow \infty$. Moreover, consider a sequence of integers $p_d < d$ such that $p_d/(\xi_d \log d) \rightarrow 0$. Then Theorem 2.1(ii) applies for each d , with $Z^{(d)}$ and p_d replacing Z and p , respectively, and the error bounds provided in the theorem go to zero as $d \rightarrow \infty$.

Corollary 2.4 provides an asymptotic version of Theorem 2.1(ii). Similarly, an asymptotic version of Theorem 2.1(i) can also be obtained, mutatis mutandis. This provides a direct extension of Theorem 2.1 of [9] from the case $p = 1$ covered in that reference to the case where $p > 1$, also allowing for p to grow with d . [Indeed, it is elementary to verify that conditions (t1) and (t2) with $k = 4$ in [9] imply that the conditions of the corollary are satisfied with $\varepsilon = 0$ and for some sequence ξ_d such that $d^{-\xi_d} \rightarrow 0$ as $d \rightarrow \infty$. And if (t1) and (t2) hold with $k = 2$, one obtains conditions that imply an asymptotic version of Theorem 2.1(i).]

If Corollary 2.4 applies with constants ξ_d satisfying $\xi_d \rightarrow \xi_\infty > 0$ (e.g., in the case where the ξ_d do not depend on d , which is also discussed in Example 3.1), the corollary’s requirement on p_d reduces to $p_d = o(\log d)$. In that case, the bounds on $v_{d,p}(\mathbb{G}^c)$ in Theorem 2.1 are of polynomial order in d . But if Corollary 2.4 applies with $\xi_d \rightarrow 0$, then the stronger requirement $p_d = o(\xi_d \log d)$ is needed, and the bounds on $v_{d,p}(\mathbb{G}^c)$ in Theorem 2.1 can be of slower order in d . Note that $d^{-\xi_d} \rightarrow 0$ entails that $\xi_d \log d \rightarrow \infty$, so that the constant sequence $p_d = p$ always satisfies the growth condition in Corollary 2.4.

3. Examples and extensions

3.1. Examples

In this section, we discuss a few simple examples of multivariate distributions for which our assumptions (b1) and (b2) are satisfied and explicit values for the quantities ε and ξ can be given. First, we consider the case of a product distribution on \mathbb{R}^d with moments of sufficiently

high order and bounded component densities. For the proof, we refer the reader to Example A.1 in [10].

Example 3.1 [9]. Suppose that the random d -vector $Z = (z_1, \dots, z_d)'$ has independent components and satisfies $\mathbb{E}Z = 0$, $\mathbb{E}ZZ' = I_d$, and fix $k \in \mathbb{N}$.

(i) If $\mathbb{E}|z_i|^{4k+4} \leq \mu_{4k+4}$, for some universal constant $\mu_{4k+4} > 0$ and for all $i = 1, \dots, d$, then the bounds in (b1)(a)–(b) hold with k as chosen here, with $\varepsilon = \xi = 1/2$ and the constants α and β depend only on k and μ_{4k+4} .

(ii) If $\mathbb{E}|z_i|^{2k+1} \leq \mu_{2k+1}$, for some universal constant $\mu_{2k+1} > 0$ and for all $i = 1, \dots, d$, then the bounds in (b1)(c) hold with k as chosen here, with $\xi = 1/2$ and the constant β depends only on k and μ_{2k+1} .

(iii) If all the marginal Lebesgue densities of the components of Z exist and are bounded by a constant $D \geq 1$ then the bound in (b2) applies for the same constant D and every value of $k \in \{1, \dots, d\}$.

From Example 3.1 we see, in particular, that if the random vector Z has independent components with bounded densities and bounded 12th marginal moments, then the bounds of Theorem 2.1(i) hold, with $\xi_1 = 1/6$ (note that k has to be chosen as $k = 2$ here). If the components of Z even have 20 marginal moments bounded by a universal constant, then also the bounds of Theorem 2.1(ii) hold, with $\xi_2 = 1/10$ (in this case $k = 4$).

The assumptions of Theorem 2.1, however, are not limited to product distributions, as the following examples show.

Example 3.2. Suppose that the random d -vector Z satisfies $\mathbb{E}Z = 0$ and $\mathbb{E}ZZ' = I_d$.

(i) If R is a fixed $d \times d$ orthogonal matrix and Z satisfies any of the bounds (b1)(a)–(c) or (b2) for some values of $k, \alpha, \beta, \varepsilon$ and ξ , then the random vector $Z^* = RZ$ satisfies the same bound with the same constants.

(ii) If r is a scalar random variable taking values in $\{-1, 1\}$ that is independent of Z , and Z satisfies any of the bounds (b1)(a)–(c) or (b2) for some values of $k, \alpha, \beta, \varepsilon$ and ξ , then the random vector $Z^* = rZ$ satisfies the same bound with the same constants.

Examples 3.1 and 3.2 can be combined to produce many multivariate distributions with dependent components that still satisfy the assumptions of Theorem 2.1. For instance, if Z has independent non-Gaussian components with moment and density bounds as in Example 3.1 and R is orthogonal with no zero entry, then, by the Darmois–Skitovich theorem, $Z^* = RZ$ cannot have independent components. Alternatively, if $Z = (z_1, \dots, z_d)'$ is as in Example 3.1 and such that, say, the first two components have non-symmetric distributions, then the first two components of $Z^* = rZ = (z_1^*, \dots, z_d^*)'$, for some non-degenerate random variable r with values in $\{-1, 1\}$, may be dependent. Indeed, for example, take $z_1 \sim z_2 \sim \text{Exp}(1) - 1$ and note that $\mathbb{P}(z_2^* < -1 | z_1^* > 1) = 0 \neq \mathbb{P}(z_2^* < -1)$.

As our last example we discuss a specific case of a spherical distribution. Recall that every spherically symmetric distribution with independent components must be Gaussian. So every spherical non-Gaussian distribution constitutes an example of a multivariate distribution with

dependent components. Also, if Z is spherical, then $\mathbb{E}[Z\|B'Z] = BB'Z$, almost surely, for every $B \in \mathcal{V}_{d,p}$. Hence, the following example is only of interest in connection with Theorem 2.1(ii) on the conditional second moment of Z , since the conclusion of Theorem 2.1(i) is trivially true in this case.

Example 3.3. If Z is uniformly distributed on the d -ball of radius $\sqrt{d+2}$, then $\mathbb{E}Z = 0$ and $\mathbb{E}ZZ' = I_d$. Moreover, for $k \in \{2, 4\}$, at least for all sufficiently large d , Z satisfies (b1) and (b2) with constants $\varepsilon = \xi = 1/2$, $D = 1$, and constants α and β that depend only on k .

Finally, it is worth mentioning that in the case of spherically symmetric Z the structure of the set \mathbb{G} from Theorem 2.1 simplifies dramatically. Indeed, from Theorem 2.1(iii) we see that if Z is spherical, then \mathbb{G} is both left and right-invariant under the action of the appropriate orthogonal groups and thus is either empty or equal to the whole Stiefel manifold $\mathcal{V}_{d,p}$.

3.2. Improved bounds

At the current state of research, we can not say if the bounds provided by Theorem 2.1 are tight, or at least if they are of the optimal rate in d , in the sense that this rate is achieved for some multivariate distribution satisfying conditions (b1) and (b2). However, there are certain distributions for which the bounds of the theorem can be improved substantially.²

First, consider the bounds (2.1) and (2.3), which are only of logarithmic order in d . As mentioned in Section 2, they can be improved considerably if one imposes an appropriate condition on Z . Here, we only consider (2.1) as an example. This bound is derived in the proof of Lemma B.2(i) by the following simple argument involving the cut-off point $M_d = \sqrt{3\xi_1(1-\tau)(\log d)/\gamma_1}$. For $t > 0$,

$$\begin{aligned} & \mathbb{P}(\|\mathbb{E}[Z\|B'Z] - BB'Z\| > t) \\ & \leq \mathbb{P}(\|\mathbb{E}[Z\|B'Z] - BB'Z\| > t, \|B'Z\| \leq M_d) + \mathbb{P}(\|B'Z\| > M_d) \\ & \leq \frac{1}{t} \int_{\|x\| \leq M_d} \|\mathbb{E}[Z\|B'Z = x] - Bx\| d\mathbb{P}_{B'Z}(x) + p/M_d^2. \end{aligned} \tag{3.1}$$

In the proof of Theorem 2.1(i) we choose $\mathbb{G} \subseteq \mathcal{V}_{d,p}$ such that for $B \in \mathbb{G}$ the bound in (3.1) turns into (2.1), while, at the same time, $\nu_{d,p}(\mathbb{G}^c)$ is bounded as in (2.2). Of course, using Markov's inequality to bound $\mathbb{P}(\|B'Z\| > M_d)$ in the preceding display is far from optimal if we have more information on the tails of Z .

Suppose now that the random vector Z , in addition to the assumptions of Theorem 2.1(i), also satisfies the sub-Gaussian tail condition

$$\mathbb{E} \exp(\alpha'Z) \leq \exp(\|\alpha\|^2 \sigma^2/2), \tag{3.2}$$

²Moreover, for each specific distribution, there are typically subsets of the set \mathbb{G} from the theorem, for which the probabilities in (1.1) and (1.2) are substantially smaller than their respective upper bounds (2.1) and (2.3). For instance, if Z has independent components and the columns of B are elements of the standard basis in \mathbb{R}^d , then both probabilities in (1.1) and (1.2) are equal to zero, for all $t > 0$.

for every $\alpha \in \mathbb{R}^d$ and some constant $\sigma > 0$.³ Under this condition, the tail inequality for quadratic forms by [7] applies and yields

$$\mathbb{P}(\|B'Z\|^2 > \sigma^2(p + 2\sqrt{ps} + 2s^2)) \leq e^{-s^2},$$

for all $s > 0$. Suppose that $p < M_d^2/(8\sigma^2)$. Since this restriction also entails that $p < M_d^2/\sigma^2$, the equation $\sigma^2(p + 2\sqrt{ps} + 2s^2) = M_d^2$ has a real positive solution $s_0 = -\sqrt{p}/2 + \sqrt{M_d^2/(2\sigma^2) - p/4}$. Thus, after expanding the square and rearranging terms, we obtain

$$s_0^2 = \frac{M_d^2}{2\sigma^2} - \sqrt{p\left(\frac{M_d^2}{2\sigma^2} - \frac{p}{4}\right)} \geq \frac{M_d^2}{4\sigma^2} = (\log d) \frac{3}{4} \frac{\xi_1(1-\tau)}{\sigma^2\gamma_1},$$

where we have used our restriction on p again. Hence,

$$\mathbb{P}(\|B'Z\|^2 > M_d^2) \leq e^{-s_0^2} \leq d^{-\frac{3}{4} \frac{\xi_1(1-\tau)}{\sigma^2\gamma_1}},$$

and we have managed to replace the term in (2.1) that is only of logarithmic order in d by something that is decreasing polynomially in d . However, since the squared cut-off point M_d^2 is only of logarithmic order in d , the condition that $p < M_d^2/(8\sigma^2)$ still requires $p/\log d$ to be small. At the moment, we do not see a way how to increase the cut-off point to polynomial order in d without simultaneously ruining the bound in (2.2).

Concerning the bounds (2.2) and (2.4), we believe that polynomial rates in d of arbitrarily high order can be achieved under more restrictive assumptions than those maintained here and upon using a more elaborate method of proof. First results in that direction, regarding only the conditional expectation, are in preparation, cf. [12].

3.3. The case of a general covariance matrix

The proof of Theorem 2.1 crucially relies on the assumptions that $\mathbb{E}Z = 0$ and $\mathbb{E}ZZ' = I_d$. However, Theorem 2.1, as it stands, can already be used to generalize substantially beyond the mean zero and unit covariance case. In particular, we can provide a large class of positive definite covariance matrices such that for each element Σ of that class the conclusions of Theorem 2.1 remain valid, provided, of course, that all the relevant quantities are modified to reflect the general covariance structure Σ . The key to this extension is the following observation.

If Y is Gaussian with mean $\mu \in \mathbb{R}^d$ and positive definite covariance matrix Σ , and $A \in \mathcal{V}_{d,p}$, then $\mathbb{E}[Y\|A'Y] = \mu + \Sigma^{1/2}P_{\Sigma^{1/2}A}\Sigma^{-1/2}(Y - \mu)$ and $\mathbb{E}[(Y - \mu)(Y - \mu)'\|A'Y] = \Sigma^{1/2}[I_d - P_{\Sigma^{1/2}A} + P_{\Sigma^{1/2}A}\Sigma^{-1/2}(Y - \mu)(Y - \mu)'\Sigma^{-1/2}P_{\Sigma^{1/2}A}]\Sigma^{1/2}$, where P_{\dots} is the projection matrix corresponding to the column span of the matrix in the subscript. These are our target quantities. Now assume that Y is not necessarily Gaussian but satisfies $Y = \mu + \Sigma^{1/2}Z$, with Z as in

³This is satisfied, for instance, if Z has independent components which all satisfy the one dimensional analogue of (3.2) with the same value of σ .

Theorem 2.1. One easily verifies that

$$\mathbb{E}[Y\|A'Y] - (\mu + \Sigma^{1/2}P_{\Sigma^{1/2}A}\Sigma^{-1/2}(Y - \mu)) = \Sigma^{1/2}(\mathbb{E}[Z\|B'Z] - BB'Z),$$

and

$$\begin{aligned} &\mathbb{E}[(Y - \mu)(Y - \mu)' \| A'Y] \\ &\quad - \Sigma^{1/2}(I_d - P_{\Sigma^{1/2}A} + P_{\Sigma^{1/2}A}\Sigma^{-1/2}(Y - \mu)(Y - \mu)' \Sigma^{-1/2}P_{\Sigma^{1/2}A})\Sigma^{1/2} \\ &= \Sigma^{1/2}(\mathbb{E}[ZZ'\|B'Z] - (I_d - BB' + BB'ZZ'BB'))\Sigma^{1/2}, \end{aligned}$$

where $B = \Sigma^{1/2}A(A'\Sigma A)^{-1/2} \in \mathcal{V}_{d,p}$. Ideally, the norm of these quantities should become small if d is large. Ignoring the additional scaling by the matrix $\Sigma^{1/2}$ of these error terms,⁴ there remains the question of whether the theorem also applies to

$$\mathbb{P}(\|\mathbb{E}[Z\|B'Z] - BB'Z\| > t) \tag{3.3}$$

and

$$\mathbb{P}(\|\mathbb{E}[ZZ'\|B'Z] - (I_d - BB' + BB'ZZ'BB')\| > t), \tag{3.4}$$

instead of (1.1) and (1.2), i.e., if $B = B(\Sigma, A) \in \mathbb{G}$. This raises two questions: For given Σ , how large is the collection of “good” matrices A , that is, how large is the set of A for which $B(\Sigma, A) \in \mathbb{G}$? And: How large is the family of matrices Σ for which the collection of “good” matrices A is large? The latter question is answered by the next result.

Proposition 3.4. *If Z satisfies the assumptions of Theorem 2.1(i) (or Theorem 2.1(ii)) and $\mathbb{G} \subseteq \mathcal{V}_{d,p}$ is the corresponding subset of the Stiefel manifold, then, for each diagonal positive definite matrix Λ , there exists a collection $\mathbb{U}(\Lambda) = \mathbb{U}(\mathbb{G}, \Lambda) \subseteq \mathcal{O}_d$ of orthogonal matrices, such that the sets*

$$\mathbb{S} := \mathbb{S}(\mathbb{G}) := \{U \Lambda U' : \Lambda = \text{diag}(\lambda_i) > 0, U \in \mathbb{U}(\mathbb{G}, \Lambda)\}$$

and

$$\mathbb{J}(\Sigma) := \mathbb{J}(\Sigma, \mathbb{G}) := \{A \in \mathcal{V}_{d,p} : \Sigma^{1/2}A(A'\Sigma A)^{-1/2} \in \mathbb{G}\},$$

have the following properties:

$$\sup_{\Lambda: \Lambda = \text{diag}(\lambda_i) > 0} v_{d,d}(\mathbb{U}^c(\Lambda)) \quad \text{and} \quad \sup_{\Sigma \in \mathbb{S}} v_{d,p}(\mathbb{J}^c(\Sigma))$$

are bounded by the square root of the right-hand-side of (2.2) (resp. (2.4)). By definition, if Σ is any positive definite covariance matrix and $A \in \mathbb{J}(\Sigma)$, then (3.3) (resp. (3.4)) is bounded by (2.1) (resp. (2.3)) for every $t > 0$.

⁴Whether the scaling by $\Sigma^{1/2}$ matters depends on the specific context of application for these results. Also, the problem can always be circumvented by imposing a boundedness assumption on $\|\Sigma\|$. However, in the context of [17], for example, no such bound is required.

To understand the message of Proposition 3.4, suppose for now that the assumptions of Theorem 2.1(i) are satisfied. Then the set $\mathbb{J}(\Sigma)$ is constructed such that the following holds: If Σ is any positive definite covariance matrix and A is taken from the collection $\mathbb{J}(\Sigma)$, then, for $B = \Sigma^{1/2}A(A'\Sigma A)^{-1/2}$, the expression in (3.3) is bounded by (2.1). In other words, $\mathbb{J}(\Sigma)$ is a collection of “good” matrices A as discussed just before the proposition. Now Proposition 3.4 shows that $\mathbb{J}(\Sigma)$ is large provided that $\Sigma \in \mathbb{S}$, and also that the set \mathbb{S} itself is large. Similar considerations apply, mutatis mutandis, to the conditional variance under the assumptions of Theorem 2.1(ii). In short, for a large class of d -dimensional distributions Z (cf. conditions (b1) and (b2)), for a large set of covariance matrices Σ (given by \mathbb{S}) and for most matrices A from the Stiefel manifold (those contained in $\mathbb{J}(\Sigma)$), the first two conditional moments of $Y = \mu + \Sigma^{1/2}Z$ given $A'Y$ are close to what they would be in the Gaussian case, all provided that $p/\log d$ is small.

4. Proof of Theorem 2.1

The rest of the paper and the on-line supplementary material comprise the proof of Theorem 2.1. The basic strategy of the proof is non-standard and is described in this section. To implement this strategy, we need to deal with several intricate technical challenges. But those can be handled by standard methods from multivariate analysis and probability theory. To keep the main paper short, such technical details are relegated to the on-line supplementary material. Our arguments have the same basic structure as those used in [9]. To prove Theorem 2.1, however, the arguments from [9] require substantial extensions and modifications, because many of the arguments used in that reference are of an asymptotic nature and do not provide explicit error bounds, and because all of these arguments rely heavily on the assumption that p is fixed and equal to 1.

4.1. Two crucial bounds

Throughout, fix $d \in \mathbb{N}$ and let Z be as in Theorem 2.1, i.e., a random d -vector that has a Lebesgue density and that is standardized so that $\mathbb{E}Z = 0$ and $\mathbb{E}ZZ' = I_d$. [The particular assumptions of part (i) and (ii) of Theorem 2.1 will be imposed as needed later.] We will study the following quantities: For a positive integer $p < d$, for $x \in \mathbb{R}^p$, and for $B \in \mathcal{V}_{d,p}$, set $\mu_{x|B} = \mathbb{E}[Z \| B'Z = x]$, $\Delta_{x|B} = \mathbb{E}[ZZ' \| B'Z = x] - (I_d + B(x x' - I_p)B')$, and $h(x|B) = \mathbb{E}[\frac{f(W^{x|B})}{\phi(W^{x|B})}]$, where $f = f_Z$ is a Lebesgue density of Z , $W^{x|B} = Bx + (I_d - BB')V$, and ϕ denotes a Lebesgue density of $V \sim N(0, I_d)$. Note that these quantities, if considered as functions with domain $\mathbb{R}^p \times \mathcal{V}_{d,p}$, can be chosen to be measurable; cf. Lemma B.1 in the on-line supplement. Finally, write $\mathcal{S}_{M,p}$ for the closed ball of radius M in \mathbb{R}^p , i.e., $\mathcal{S}_{M,p} = \{x \in \mathbb{R}^p : \|x\| \leq M\}$.

We now introduce two bounds which will play an essential role in the proof of Theorem 2.1. In each bound, the quantity of interest, which will be introduced shortly, will be bounded by an expression of the form

$$p^{2k+1+\varepsilon} e^{\varepsilon M^2} (2D\sqrt{\pi}e)^{pk} d^{-\min\{\xi, \varepsilon/2+1/4, 1/2\}} \kappa \tag{4.1}$$

for some even integer k , where the precise value of the constants in the bound will depend on the context, i.e., these constants will be chosen as needed later.

The first crucial bound implies the first part of Theorem 2.1: Under the assumptions of the Theorem 2.1(i), we will show that

$$\sup_{x \in \mathcal{S}_{M,p}} \int [\|\mu_{x|B}\|^2 - \|x\|^2] h(x|B)^2 dv_{d,p}(B) \tag{4.2}$$

is bounded by (4.1) for $k = 2$, for every $M > 1$ and every $p \in \mathbb{N}$ such that $d > \max\{4(k + p + 1)M^4, 2k + p(2k + 2)2^{k+3}, p^2\}$, where $\kappa = \kappa_1 \geq 1$ is a constant that depends only on α and β , and where $g = g_1$ is a global constant. The remaining constants occurring here, i.e., ε , α , β , ξ , and D , are those that appear in the bounds (b1)(a)–(b) and (b2) imposed by Theorem 2.1(i). Once that statement has been derived, the proof of Theorem 2.1(i) is easily completed by standard arguments (that are detailed in Lemma B.2(i)).

The second crucial bound similarly delivers the second part of Theorem 2.1: Under the assumptions of Theorem 2.1(ii), we will show that (4.2) and

$$\sup_{x \in \mathcal{S}_{M,p}} \int \text{trace } \Delta_{x|B}^k h(x|B)^k dv_{d,p}(B) \tag{4.3}$$

are both bounded by (4.1) for $k = 4$, for every $M > 1$ and every $p \in \mathbb{N}$ such that $d > \max\{4(k + p + 1)M^4, 2k + p(2k + 2)2^{k+3}, p^2\}$, where $\kappa = \kappa_2 \geq 1$ depends only on α and β , and where $g = g_2$ is a global constant. Again, the remaining constants ε , α , β , ξ , and D are those that appear in the bounds (b1) and (b2) imposed by Theorem 2.1(ii). From this statement, standard arguments complete the proof of Theorem 2.1(ii); cf. Lemma B.2(ii).

It turns out that in order to derive the upper bounds for (4.2) and (4.3), it will be instrumental to show that

$$\sup_{x \in \mathcal{S}_{M,p}} \int [h(x|B) - 1]^2 dv_{d,p}(B) \tag{4.4}$$

is finite. In particular, we will need to establish finiteness of (4.4) for every M and p as in (4.2) to derive the desired bound on (4.2), and for every M and p as in (4.3) for the bound on (4.3). We will in fact show more than that, namely that (4.4) is also bounded by (4.1), under the assumptions of Theorem 2.1(i) and for constants as in (4.2), and also under the assumptions of Theorem 2.1(ii) and for constants as in (4.3).

We pause here for a moment to discuss a weaker version of Theorem 2.1 which also allows us to better appreciate the importance of (4.4) (the exact role of (4.4) in the main argument will become apparent later, after Proposition 4.1): Assume in this paragraph that (4.2), (4.3), and (4.4) are all bounded by (4.1) with $k = 4$, for each $M > 1$, and for each sufficiently large d . [The other constants in the bound, i.e., p , ε , g , D , ξ and κ , are assumed to be fixed, independent of d here.] Under this assumption, we immediately obtain the following weaker version of Corollary 2.4: For each $x \in \mathbb{R}^p$, we have

$$\|\mathbb{E}[Z \| \mathbf{B}'Z = x] - \mathbf{B}x\|^2 \xrightarrow{p} 0,$$

$$\|\mathbb{E}[ZZ' \| \mathbf{B}'Z = x] - (I_d + \mathbf{B}(xx' - I_p)\mathbf{B}')\| \xrightarrow{p} 0$$

as $d \rightarrow \infty$, if \mathbf{B} is a random matrix that is uniformly distributed on $\mathcal{V}_{d,p}$. After noting that $\|\mathbb{E}[Z \| \mathbf{B}'Z = x] - \mathbf{B}x\|^2$ can also be written as $\|\mathbb{E}[Z \| \mathbf{B}'Z = x]\|^2 - \|x\|^2$, this is an easy consequence of Markov’s inequality and Slutsky’s lemma.⁵ [Choose $M \geq \|x\|$ and observe that the upper bound (4.1) converges to zero as $d \rightarrow \infty$, so that also the three quantities in (4.2), (4.3), and (4.4) converge to zero. Now convergence of (4.4) to zero entails that $h(x|\mathbf{B})$ converges to one in squared mean. Convergence of (4.2) to zero implies that $[\|\mathbb{E}[Z \| \mathbf{B}'Z = x]\|^2 - \|x\|^2]h(x|\mathbf{B})^2$ converges to zero in expectation. Similarly, convergence of (4.3) to zero implies that $\text{trace } \Delta_{x|\mathbf{B}}^4 h(x|\mathbf{B})^4$ converges to zero in expectation. Because the involved random variables are all non-negative, the first relation in the preceding display follows from Markov’s inequality and Slutsky’s lemma. The second relation follows in a similar fashion upon observing that the symmetry of $\Delta_{x|\mathbf{B}}$ entails that $\|\Delta_{x|\mathbf{B}}\|^4$ is bounded from above by $\text{trace } \Delta_{x|\mathbf{B}}^4$.]

In this subsection, we have seen that to prove Theorem 2.1(i), it suffices to show, under the assumptions maintained there, that both (4.2) and (4.4) are bounded by (4.1) for constants as in (4.2). And, similarly, to prove Theorem 2.1(ii), it remains to show, under the assumptions maintained there, that (4.2), (4.3) and (4.4) are all bounded by (4.1) for constants as in (4.3).

4.2. Changing the reference measure

Throughout the following, set

$$W_j = \mathbf{B}x + (I_d - \mathbf{B}\mathbf{B}')V_j, \tag{4.5}$$

for $j = 1, \dots, k$, where $\mathbf{B}, V_1, \dots, V_k$ are independent such that \mathbf{B} is a random $d \times p$ matrix with distribution $\nu_{d,p}$ and such that each of the V_i is distributed as $N(0, I_d)$. We call W_1, \dots, W_k the “rotational clones”, in analogy to the name “rotational twins” that the authors of [6] use for the pair W_1, W_2 in case $p = 1$. With this, we may rewrite the integral in (4.4) as

$$\begin{aligned} & \int \left(\mathbb{E} \left[\frac{f(W^{x|B})}{\phi(W^{x|B})} \right] \right)^2 - 2\mathbb{E} \left[\frac{f(W^{x|B})}{\phi(W^{x|B})} \right] + 1 \, d\nu_{d,p}(B) \\ &= \int \mathbb{E} \left[\frac{f(\mathbf{B}x + (I_d - \mathbf{B}\mathbf{B}')V_1)}{\phi(\mathbf{B}x + (I_d - \mathbf{B}\mathbf{B}')V_1)} \right] \mathbb{E} \left[\frac{f(\mathbf{B}x + (I_d - \mathbf{B}\mathbf{B}')V_2)}{\phi(\mathbf{B}x + (I_d - \mathbf{B}\mathbf{B}')V_2)} \right] d\nu_{d,p}(B) \\ &\quad - 2 \int \mathbb{E} \left[\frac{f(\mathbf{B}x + (I_d - \mathbf{B}\mathbf{B}')V_1)}{\phi(\mathbf{B}x + (I_d - \mathbf{B}\mathbf{B}')V_1)} \right] d\nu_{d,p}(B) + 1 \\ &= \mathbb{E} \left[\frac{f(W_1)}{\phi(W_1)} \frac{f(W_2)}{\phi(W_2)} - 1 \right] - 2\mathbb{E} \left[\frac{f(W_1)}{\phi(W_1)} - 1 \right], \tag{4.6} \end{aligned}$$

provided that the expected values in (4.6) are all finite. And, clearly, if both expected values in (4.6) are bounded by (4.1) in absolute value, then (4.6) is bounded by three times the expression

⁵A proof of the first statement in the preceding display was already sketched in [6] as an immediate generalization of the case where $p = 1$ proved therein. See also [9] for further discussion of that result.

in (4.1). To establish the desired bounds on (4.2) and (4.3) it will be convenient to also express the integrals in (4.2) and (4.3) in terms of the W_i . This can be accomplished by virtue of the following proposition.

Proposition 4.1. Fix $d \geq p \geq 1$, and consider a random d -vector Z with Lebesgue density f . Let $V \sim N(0, I_d)$, and write $\phi(\cdot)$ for a Lebesgue density of V . Moreover, for a fixed $d \times p$ -matrix $B \in \mathcal{V}_{d,p}$ and for any $x \in \mathbb{R}^p$, set $W^{x|B} = Bx + (I_d - BB')V$. Then the function $h(\cdot|B) : \mathbb{R}^p \rightarrow \mathbb{R}$ defined by

$$h(x|B) = \mathbb{E} \left[\frac{f(W^{x|B})}{\phi(W^{x|B})} \right]$$

for $x \in \mathbb{R}^p$ is a density of $B'Z$ with respect to the p -variate standard Gaussian measure (i.e., $h(x|B)\phi_p(x)$ is a Lebesgue density of $B'Z$ if ϕ_p denotes a $N_p(0, I_p)$ -density). Moreover, if $\Psi : \mathbb{R}^d \rightarrow \mathbb{R}$ is such that $\Psi(Z)$ is integrable, then a conditional expectation $\mathbb{E}[\Psi(Z)|B'Z = x]$ of $\Psi(Z)$ given $B'Z = x$ satisfies

$$\mathbb{E}[\Psi(Z)|B'Z = x]h(x|B) = \mathbb{E} \left[\Psi(W^{x|B}) \frac{f(W^{x|B})}{\phi(W^{x|B})} \right]$$

whenever $x \in \mathbb{R}^p$ is such that $h(x|B) < \infty$.

Note that this proposition applies under the assumptions of both parts of Theorem 2.1. Assume therefore that Proposition 4.1 is applicable throughout the rest of this subsection. The integral in (4.2) can then be re-written as

$$\begin{aligned} & \int [\|\mu_{x|B}\|^2 - \|x\|^2] h(x|B)^2 d\nu_{d,p}(B) \\ &= \int \|\mu_{x|B} h(x|B)\|^2 d\nu_{d,p}(B) - \|x\|^2 \int h(x|B)^2 d\nu_{d,p}(B) \\ &= \int \mathbb{E} \left[W^{x|B} \frac{f(W^{x|B})}{\phi(W^{x|B})} \right]' \mathbb{E} \left[W^{x|B} \frac{f(W^{x|B})}{\phi(W^{x|B})} \right] d\nu_{d,p}(B) \\ &\quad - \|x\|^2 \mathbb{E} \left[\frac{f(W_1)}{\phi(W_1)} \frac{f(W_2)}{\phi(W_2)} \right] \\ &= \mathbb{E} \left[(W_1' W_2 - \|x\|^2) \frac{f(W_1)}{\phi(W_1)} \frac{f(W_2)}{\phi(W_2)} \right], \end{aligned} \tag{4.7}$$

provided that the expected values in (4.6) and (4.7) are all finite. [Indeed, finiteness of the expected values in (4.6) entails that $\int [h(x|B) - 1]^2 d\nu_{d,p}(B)$ is finite, so that $\nu_{d,p}\{B : h(x|B) = \infty\} = 0$, whence Proposition 4.1 can be used to obtain the second equality in the preceding display. The first and the third equality follow from finiteness of the expected values in (4.6) and (4.7).]

To express the integral in (4.3) in a similar way, define $\Delta_{x|B}(z) : \mathbb{R}^p \times \mathcal{V}_{d,p} \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ by $\Delta_{x|B}(z) = zz' - (I_d + B(xx' - I_p)B')$, and use Proposition 4.1 component-wise with $\Psi_{i,j}(Z) =$

$[\Delta_{x|B}(Z)]_{i,j}$ for all $i, j = 1, \dots, d$ to obtain

$$\begin{aligned} & \int \text{trace } \Delta_{x|B}^k h(x|B)^k d\nu_{d,p}(B) \\ &= \int \text{trace} \left\{ \left(\mathbb{E} [\Delta_{x|B}(Z) \| B'Z = x] h(x|B) \right)^k \right\} d\nu_{d,p}(B) \\ &= \int \text{trace} \left\{ \left(\mathbb{E} \left[\Delta_{x|B}(W^{x|B}) \frac{f(W^{x|B})}{\phi(W^{x|B})} \right] \right)^k \right\} d\nu_{d,p}(B) \\ &= \mathbb{E} \left[\text{trace } \Delta_{x|B}(W_1) \cdots \Delta_{x|B}(W_k) \frac{f(W_1)}{\phi(W_1)} \cdots \frac{f(W_k)}{\phi(W_k)} \right], \end{aligned} \tag{4.8}$$

provided that the expected values in (4.6) and (4.8) are all finite. Lemma C.1 describes how the expression in (4.8) can be written as a weighted sum of expressions that, similarly to (4.7), involve only inner products of the W_i and a product of density ratios. In particular, we find that (4.8) can be written as a weighted sum of terms of the form

$$\begin{aligned} & \sum_{j=1}^k (-1)^{k-j} \binom{k}{j} \mathbb{E} \left[(W'_1 W_2 \cdots W_j W'_j W_1 - d + p - \|x\|^{2j}) \frac{f(W_1)}{\phi(W_1)} \cdots \frac{f(W_k)}{\phi(W_k)} \right], \quad \text{and} \\ & \mathbb{E} \left[\left(\prod_{i=1}^m W'_{j_{i-1}+1} W_{j_{i-1}+2} W'_{j_{i-1}+2} \cdots W'_{j_{i-1}} W_{j_i} - \|x\|^{2(j_m-m)} \right) \frac{f(W_1)}{\phi(W_1)} \cdots \frac{f(W_k)}{\phi(W_k)} \right] \end{aligned} \tag{4.9}$$

for $m \geq 1$ and indices j_0, \dots, j_m satisfying $j_0 = 0$, $j_m < k$, and $j_{i-1} + 1 < j_i$ whenever $1 \leq i \leq m$, provided that the expected values in (4.9) are all finite. [Note that these requirements entail that $m \leq k/2$, and that there are no more than $\binom{k}{m}$ choices for the indices j_0, \dots, j_m in the second expected value in (4.9).] Lemma C.1 also shows that the weights in this expansion of (4.8) only depend on k and on x , and are polynomials in $\|x\|^2$. Note that hence all the weights are bounded, in absolute value and uniformly in $x \in \mathcal{S}_{M,p}$, by $e^{c(k)M^2}$ for some constant $c(k)$ that depends only on k . In particular, if the expected values in (4.9) are all bounded by (4.1) in absolute value, then the same is true for (4.8) or, equivalently, (4.3) upon replacing the constants g and κ in (4.1) by, say, $g + c(k)$ and $(1 + (k/2)\binom{k}{k/2})\kappa$, respectively.

In this subsection, we have seen how the integrals in (4.2), (4.3) and (4.4) can be re-written as weighted sums of expected values involving the rotational clones, provided these expected values are all finite. To prove Theorem 2.1(i), it thus remains to show, under the maintained assumptions, that the expected values in (4.6) and (4.7) are all bounded by (4.1) in absolute value, uniformly in $x \in \mathcal{S}_{M,p}$, and for constants as in (4.2). And Theorem 2.1(ii) follows if, under the assumptions of that theorem, the expected values in (4.6) and (4.7) as well as all the expressions in (4.9) are bounded by (4.1) in absolute value, uniformly in $x \in \mathcal{S}_{M,p}$, and for constants as in (4.3).

4.3. The joint density of the “rotational clones”

Proposition 4.2. *For integers $1 \leq p < d$ and $1 \leq k \leq d - p$, let $x \in \mathbb{R}^p$, and let W_1, \dots, W_k be as in (4.5). Then W_1, \dots, W_k have a joint density $\varphi_x(w_1, \dots, w_k)$ with respect to Lebesgue measure which satisfies*

$$\begin{aligned} & \frac{\varphi_x(w_1, \dots, w_k)}{\phi(w_1) \cdots \phi(w_k)} \\ &= \left(\frac{d}{2}\right)^{-\frac{pk}{2}} \prod_{i=1}^k \frac{\Gamma((d-i+1)/2)}{\Gamma((d-p-i+1)/2)} \det(S_k)^{-\frac{p}{2}} \left(1 - \frac{\|x\|^2}{d} \iota' S_k^{-1} \iota\right)^{\frac{d-p-k-1}{2}} e^{\frac{k}{2}\|x\|^2} \end{aligned}$$

if S_k is invertible with $\|x\|^2 \iota' S_k^{-1} \iota < d$, and $\varphi_x(w_1, \dots, w_k) = 0$ otherwise, where $S_k = (w_i' w_j / d)_{i,j=1}^k$ denotes the $k \times k$ matrix of scaled inner products of the w_i , and $\iota = (1, \dots, 1)'$ denotes an appropriate vector of ones.

If, in addition, $k < d - p - 1$, then the normalizing constant in the preceding display, that is, the quantity $\eta(d, p, k) = (d/2)^{-kp/2} \prod_{i=1}^k \frac{\Gamma((d-i+1)/2)}{\Gamma((d-p-i+1)/2)}$ satisfies

$$0 < \eta(d, p, k) \leq \exp\left[\frac{p^2}{d} \left(1 - \frac{p+k-1}{d}\right)^{-1} \frac{k^2}{2}\right].$$

Note that Proposition 4.2 applies whenever p, d , and k are as in (4.2) or (4.3). For p, d , and k as in Proposition 4.2, we can re-express the expression in (4.6) as

$$\begin{aligned} & \mathbb{E}\left[\frac{f(W_1)}{\phi(W_1)} \frac{f(W_2)}{\phi(W_2)} - 1\right] - 2\mathbb{E}\left[\frac{f(W_1)}{\phi(W_1)} - 1\right] \\ &= \int_{\mathbb{R}^d \times \mathbb{R}^d} \left(\frac{f(w_1)}{\phi(w_1)} \frac{f(w_2)}{\phi(w_2)} - 1\right) \varphi_x(w_1, w_2) dw_1 dw_2 \\ & \quad - 2 \int_{\mathbb{R}^d} \left(\frac{f(w_1)}{\phi(w_1)} - 1\right) \varphi_x(w_1) dw_1 \\ &= \mathbb{E}\left[\frac{\varphi_x(Z_1, Z_2)}{\phi(Z_1)\phi(Z_2)} - 1\right] - 2\mathbb{E}\left[\frac{\varphi_x(Z_1)}{\phi(Z_1)} - 1\right], \end{aligned} \tag{4.10}$$

we can rewrite (4.7) as

$$\mathbb{E}\left[(Z_1' Z_2 - \|x\|^2) \frac{\varphi_x(Z_1, Z_2)}{\phi(Z_1)\phi(Z_2)}\right], \tag{4.11}$$

and the expressions in (4.9) can be written as

$$\begin{aligned} & \sum_{j=1}^k (-1)^{k-j} \binom{k}{j} \mathbb{E}\left[(Z_1' Z_2 \cdots Z_j Z_j' Z_1 - d + p - \|x\|^2)^j \frac{\varphi_x(Z_1, \dots, Z_k)}{\phi(Z_1) \cdots \phi(Z_k)}\right], \\ & \mathbb{E}\left[\left(\prod_{i=1}^m Z_{j_{i-1}+1}' Z_{j_{i-1}+2} \cdots Z_{j_{i-1}}' Z_{j_i} - \|x\|^{2(j_m-m)}\right) \frac{\varphi_x(Z_1, \dots, Z_k)}{\phi(Z_1) \cdots \phi(Z_k)}\right], \end{aligned} \tag{4.12}$$

for $m \geq 1$ and indices j_0, \dots, j_m satisfying $j_0 = 0$, $j_m < k$, and $j_{i-1} + 1 < j_i$ whenever $1 \leq i \leq m$.

In this subsection, we have seen how the integrals in (4.2), (4.3) and (4.4) can be re-written as weighted sums of expected values involving i.i.d. copies of Z and the density of the rotational clones, provided these expected values are all finite. Theorem 2.1(i) now follows if we can show that the expected values in (4.10) and (4.11) are all bounded by (4.1) in absolute value, uniformly in $x \in \mathcal{S}_{M,p}$, and for constants as in (4.2), under the assumptions of that theorem. Similarly, Theorem 2.1(ii) follows, provided the expected values in (4.10) and (4.11) as well as all the expressions in (4.12) are bounded by (4.1) in absolute value, uniformly in $x \in \mathcal{S}_{M,p}$, and for constants as in (4.3), under the assumptions of that theorem.

4.4. Two sufficient conditions

For an even integer k , consider the quantities

$$\mathbb{E} \left[\left(\prod_{i=1}^m Z'_{j_{i-1}+1} Z_{j_{i-1}+2} \cdots Z'_{j_i-1} Z_{j_i} \right) \frac{\varphi_x(Z_1, \dots, Z_l)}{\phi(Z_1) \cdots \phi(Z_l)} \right] - \|x\|^{2(j_m-m)} \tag{4.13}$$

for $l = 1, \dots, k$, for each $m \geq 0$, and for each set of indices j_0, \dots, j_m that satisfies $j_0 = 0$, $j_m \leq l$ and $j_{i-1} + 1 < j_i$ whenever $1 \leq i \leq m$. And, again for even k , consider

$$\sum_{j=1}^k (-1)^{k-j} \binom{k}{j} \mathbb{E} \left[(Z'_1 Z_2 \cdots Z_j Z'_j Z_1 - d + p - 1) \right. \\ \left. \times \frac{\varphi_x(Z_1, \dots, Z_k)}{\phi(Z_1) \cdots \phi(Z_k)} \right] - (1 - \|x\|^2)^k. \tag{4.14}$$

If the expressions of the form (4.13) are all bounded by (4.1), in absolute value and with constants as in (4.2), then both the expected values in (4.10) and (4.11) are also bounded by (4.1), again in absolute value and for constants as in (4.2). Indeed, the two expected values in (4.10) are special cases of (4.13), namely with $m = 0$ and $l = 1$, and with $m = 0$ and $l = 2$, respectively. Similarly, one sees that (4.11) equals

$$\mathbb{E} \left[(Z'_1 Z_2) \frac{\varphi_x(Z_1, Z_2)}{\phi(Z_1)\phi(Z_2)} - \|x\|^2 \right] - \|x\|^2 \mathbb{E} \left[\frac{\varphi_x(Z_1, Z_2)}{\phi(Z_1)\phi(Z_2)} - 1 \right].$$

Note that the two expected values in the preceding display are special cases of (4.13), namely with $m = 1, l = 2$ and with $m = 0, l = 2$. If these special cases of (4.13) are both bounded by (4.1) in absolute value, uniformly in $x \in \mathcal{S}_{M,p}$, and for constants as in (4.2), then the expression in the preceding display is similarly bounded by the product of (4.1) and $1 + M^2$. It is now easy to see that the resulting upper bound on the expression in the preceding display, and hence also on (4.11), is itself upper bounded by an expression of the form (4.1) for constants as in (4.2).

Similarly, if the expressions of the form (4.13) and also (4.14) are bounded by (4.1), in absolute value and with constants as in (4.3), then the expected values in (4.10) and (4.11) as well as all

the expressions in (4.12) are also bounded by (4.1), again in absolute value and for constants as in (4.3). Indeed, (4.10) can be bounded as claimed by arguing as in the preceding paragraph. For (4.12), we rewrite the first expression in that display as

$$\begin{aligned} & \sum_{j=1}^k (-1)^{k-j} \binom{k}{j} \mathbb{E} \left[(Z'_1 Z_2 \cdots Z_j Z'_j Z_1 - d + p - \|x\|^{2j}) \frac{\varphi_x(Z_1, \dots, Z_k)}{\phi(Z_1) \cdots \phi(Z_k)} \right] \\ &= \sum_{j=1}^k (-1)^{k-j} \binom{k}{j} \mathbb{E} \left[(Z'_1 Z_2 \cdots Z_j Z'_j Z_1 - d + p - 1) \frac{\varphi_x(Z_1, \dots, Z_k)}{\phi(Z_1) \cdots \phi(Z_k)} \right] \\ & \quad + \mathbb{E} \left[\frac{\varphi_x(Z_1, \dots, Z_k)}{\phi(Z_1) \cdots \phi(Z_k)} \right] \sum_{j=1}^k (-1)^{k-j} \binom{k}{j} (1 - \|x\|^{2j}) \\ &= (4.14) + (1 - \|x\|^2)^k - \mathbb{E} \left[\frac{\varphi_x(Z_1, \dots, Z_k)}{\phi(Z_1) \cdots \phi(Z_k)} \right] (1 - \|x\|^2)^k \\ &= (4.14) - (1 - \|x\|^2)^k \mathbb{E} \left[\frac{\varphi_x(Z_1, \dots, Z_k)}{\phi(Z_1) \cdots \phi(Z_k)} - 1 \right], \end{aligned}$$

where the second equality is obtained from the binomial formula upon recalling that k is even. From this, the first expression in (4.12) can be bounded by an expression of the form (4.1), namely by first bounding both (4.13) and (4.14) with $m = 0$ and $l = k$ by (4.1), by using the fact that $(1 - \|x\|^2)^k \leq 2^k M^{2k}$ for $x \in \mathcal{S}_{M,p}$, and by adjusting the constants κ and g in (4.1) accordingly. The second expression in (4.12) can be bounded in a similar fashion upon using appropriate bounds on (4.13).

In this subsection, we have seen how bounds on the expressions of the form (4.13) and on (4.14) can be used to prove both parts of Theorem 2.1. In particular, Theorem 2.1(i) follows if, under the assumptions of that theorem, the expressions of the form (4.13) are all bounded by (4.1), in absolute value, uniformly in $x \in \mathcal{S}_{M,p}$, and for constants as in (4.2). And Theorem 2.1(ii) follows if the expressions of the form (4.13) and also (4.14) are all bounded by (4.1), in absolute value, uniformly in $x \in \mathcal{S}_{M,p}$, and for constants as in (4.3), under the assumptions of that theorem.

4.5. Approximating the density ratio

Proposition 4.3. Fix $M > 1$, positive integers k, d , and p , such that $d > p^2$ and $d > 4(k + p + 1)M^4$ and let $x \in \mathcal{S}_{M,p}$. For a collection of d -vectors w_1, \dots, w_k , define the $k \times k$ -matrix $S_k = (w'_i w_j / d)_{i,j=1}^k$. Then the density ratio $\frac{\varphi_x(w_1, \dots, w_k)}{\phi(w_1) \cdots \phi(w_k)}$ can be expanded as

$$\frac{\varphi_x(w_1, \dots, w_k)}{\phi(w_1) \cdots \phi(w_k)} = \psi_x(S_k - I_k) + \Delta,$$

where the quantities on the right hand side have the following properties:

ψ_x is a polynomial of degree k in the elements of $S_k - I_k$ whose coefficients are bounded by $p^k M^{2(k+2)} C_\psi(k)$, where $C_\psi(k)$ depends only on k . In particular, we may write

$$\psi_x(S_k - I_k) = \sum_{H \in \mathcal{M}_k} \mathcal{C}(H) H(S_k - I_k),$$

where \mathcal{M}_k is the set of all monomials in the entries of a symmetric $k \times k$ -matrix (i.e., in $k(k+1)/2$ variables) up to degree k and $\mathcal{C}(H) \in \mathbb{R}$ is the coefficient in ψ_x corresponding to the monomial H , which satisfies $|\mathcal{C}(H)| \leq p^k M^{2(k+2)} C_\psi(k)$. In addition, the coefficients $\mathcal{C}(H)$ are invariant under permutations in the following sense: Define the function g by $g(w_1, \dots, w_k) = S_k - I_k$. If $H, G \in \mathcal{M}_k$ are such that $H \circ g(w_1, \dots, w_k) = G \circ g(w_{\pi(1)}, \dots, w_{\pi(k)})$, for some permutation π of k elements and every choice of $w_1, \dots, w_k \in \mathbb{R}^d$, then $\mathcal{C}(H) = \mathcal{C}(G)$.

Moreover, there exists a constant $\xi(k) > 2k$ that depends only on the value of k , such that whenever $\|S_k - I_k\| < 1/(p\xi(k))$, the remainder term Δ satisfies $|\Delta| \leq p^{k+1} M^{2(k+2)} e^{\frac{k}{2} M^2} \|S_k - I_k\|^{k+1} C_\Delta(k)$, where $C_\Delta(k)$ is a constant that depends only on k .

Note that Proposition 4.3 applies whenever M, k, d , and p are either as in (4.2) or as in (4.3). The proposition suggests to replace the density ratio $\frac{\varphi_x(w_1, \dots, w_k)}{\phi(w_1) \cdots \phi(w_k)}$ by the polynomial $\psi_x(S_k - I_k)$. For a fixed even integer k , we therefore consider, as approximations to the expressions in (4.13), the quantities

$$\mathbb{E} \left[\left(\prod_{i=1}^m Z'_{j_{i-1}+1} Z_{j_{i-1}+2} \cdots Z'_{j_{i-1}} Z_{j_i} \right) \psi_x(S_l - I_l) \right] - \|x\|^{2(j_m-m)} \tag{4.15}$$

for $l = 1, \dots, k$, for each $m \geq 0$, and for each set of indices j_0, \dots, j_m that satisfies $j_0 = 0, j_m \leq l$ and $j_{i-1} + 1 < j_i$ whenever $1 \leq i \leq m$. And as approximation to (4.14), we consider

$$\sum_{j=1}^k (-1)^{k-j} \binom{k}{j} \mathbb{E}[(Z'_1 Z_2 \cdots Z_j Z'_j Z_1 - d + p - 1) \psi_x(S_k - I_k)] - (1 - \|x\|^2)^k. \tag{4.16}$$

In order for these approximations to be useful we have to make sure that the difference between (4.14) and (4.16) as well as the difference between (4.13) and (4.15) can be controlled. The following proposition provides us with the appropriate tool.

Proposition 4.4. Fix positive integers d and k . Moreover, let $M > 1$ and $p \in \mathbb{N}$ be such that $d > \max\{4(k + p + 1)M^4, 2k + p(2k + 2)2^{k+3}, p^2\}$. Let Z be a random d -vector such that $\mathbb{E}Z = 0$ and $\mathbb{E}ZZ' = I_d$, and such that bounds (b1)(a) and (b2) obtain with k as chosen here. Write Z_1, \dots, Z_k for i.i.d. copies of Z . Finally, fix $l \in \{1, \dots, k\}$, let $S_l = (Z'_i Z_j / d)_{i,j=1}^l$, and let $H(S_l - I_l)$ be a (fixed) monomial in the elements of $S_l - I_l$ whose degree, denoted by $\deg(H)$, satisfies $0 \leq \deg(H) \leq l$. Then

$$\sup_{x \in \mathcal{S}_{M,p}} \mathbb{E} \left[d^{\frac{l+\deg(H)}{2}} |H(S_l - I_l)| \left| \frac{\varphi_x(Z_1, \dots, Z_l)}{\phi(Z_1) \cdots \phi(Z_l)} - \psi_x(S_l - I_l) \right| \right]$$

is bounded by

$$p^{2k+1+\varepsilon} M^{2(k+2)} e^{\frac{k}{2} M^2} (2D\sqrt{\pi\varepsilon})^{pk} \alpha d^{-\varepsilon/2-1/4}$$

times a constant that depends only on k . Here $\varepsilon \in [0, 1/2]$ and $\alpha \geq 1$ are the constants from (b1)(a) and D is the constant from (b2).

This proposition applies under the assumptions of Theorem 2.1(i) and for constants as in (4.2), and also under the assumptions of Theorem 2.1(ii) and for constants as in (4.3). Under the assumptions of Proposition 4.4, consider first the difference of (4.14) and (4.16). This difference is a sum of k terms (with k an even integer), where the modulus of the j th term is bounded by

$$\begin{aligned} & \binom{k}{j} \mathbb{E} \left[\left| Z'_1 Z_2 \cdots Z_j Z'_j Z_1 - d + p - 1 \right| \left| \frac{\varphi_x(Z_1, \dots, Z_k)}{\phi(Z_1) \cdots \phi(Z_k)} - \psi_x(S_k - I_k) \right| \right] \\ & \leq \binom{k}{j} \mathbb{E} \left[d^{\frac{k+j}{2}} |H_j(S_k - I_k)| \left| \frac{\varphi_x(Z_1, \dots, Z_k)}{\phi(Z_1) \cdots \phi(Z_k)} - \psi_x(S_k - I_k) \right| \right] \\ & \quad + \binom{k}{j} \mathbb{E} \left[d^{\frac{k+0}{2}} |H_0(S_k - I_k)| \left| \frac{\varphi_x(Z_1, \dots, Z_k)}{\phi(Z_1) \cdots \phi(Z_k)} - \psi_x(S_k - I_k) \right| \right], \end{aligned}$$

with $H_0(S_k - I_k) = 1$, $H_1(S_k - I_k) = Z'_1 Z_1/d - 1$ and, for $j \geq 2$, $H_j(S_k - I_k) = d^{-j} Z'_1 Z_2 \cdots Z_j Z'_j Z_1$, with $\deg(H_j) = j$ for $j = 0, 1, \dots, k$, and where we have used $|p - 1| \leq |-d + p - 1| \leq d^{k/2}$. Thus Proposition 4.4 yields an upper bound for the supremum over $\mathcal{S}_{M,p}$ of the absolute value of this difference. Taken together, the difference of (4.14) and (4.16) can be bounded, in absolute value and uniformly over $x \in \mathcal{S}_{M,p}$, by the upper bound from Proposition 4.4 multiplied by $2 \sum_{j=1}^k \binom{k}{j} = 2(2^k - 1)$ (which is a constant that depends only on k). In the same way, one sees that also the absolute difference between (4.13) and (4.15) can be bounded, uniformly over $x \in \mathcal{S}_{M,p}$, by the quantity given by Proposition 4.4. Note that the bound in Proposition 4.4 is itself upper bounded by (4.1), for example, by choosing $\kappa = \kappa(k, \alpha)$ and $g = k/2 + 2(k + 2)$.

The arguments in this subsection entail that it suffices to bound the approximating quantities (4.15) and (4.16): Theorem 2.1(i) follows if, under the assumptions of that theorem, the expressions of the form (4.15) are all bounded by (4.1), in absolute value, uniformly over $x \in \mathcal{S}_{M,p}$, and for constants as in (4.2). Similarly, Theorem 2.1(ii) follows if, under the assumptions of that theorem, the expressions of the form (4.15) and also (4.16) are all bounded by (4.1), in absolute value, uniformly over $x \in \mathcal{S}_{M,p}$, and for constants as in (4.3).

4.6. Comparing the approximating quantities with Gaussian expressions

We now compare (4.15) and (4.16) with the analogous expressions where the Z_1, \dots, Z_k are replaced by i.i.d. standard normal d -vectors V_1, \dots, V_k (and the Gram matrices $S_l = (Z'_i Z_j/d)_{i,j=1}^l$ for $l = 1, \dots, k$ are replaced by the corresponding Gram matrices of the V_i , i.e., by $S_l^* = (V'_i V_j/d)_{i,j=1}^l$). In particular, we show that (4.15), with the Z_i replaced by the V_i , can be controlled as desired, and we then show that the difference of (4.15) and of (4.15) with the Z_i

replaced by the V_i can also be controlled. This will lead to the desired bound on (4.15). A similar strategy is also employed to bound (4.16).

The first step is to bound (4.15) and (4.16) with the Z_i replaced by the V_i in both displays. More explicitly, for an even integer k , we want to bound the quantities

$$\mathbb{E} \left[\left(\prod_{i=1}^m V'_{j_{i-1}+1} V_{j_{i-1}+2} \cdots V'_{j_{i-1}} V_{j_i} \right) \psi_x(S_l^* - I_l) \right] - \|x\|^{2(j_m-m)}, \tag{4.17}$$

for $l = 1, \dots, k$, for each $m \geq 0$, and for each set of indices j_0, \dots, j_m that satisfies $j_0 = 0, j_m \leq l$ and $j_{i-1} + 1 < j_i$ whenever $1 \leq i \leq m$. And we also want to bound

$$\sum_{j=1}^k (-1)^{k-j} \binom{k}{j} \mathbb{E}[(V'_1 V_2 \cdots V_j V'_j V_1 - d + p - 1) \psi_x(S_k^* - I_k)] - (1 - \|x\|^2)^k. \tag{4.18}$$

Note that these expressions can be viewed as approximations, in the sense of Proposition 4.3, to the expressions in the two preceding displays with $\psi_x(S_l^* - I_l)$ and $\psi_x(S_k^* - I_k)$ replaced by $\varphi_x(V_1, \dots, V_l)/(\phi(V_1) \cdots \phi(V_l))$ and $\varphi_x(V_1, \dots, V_k)/(\phi(V_1) \cdots \phi(V_k))$, respectively. But with that replacement, the resulting expressions are equal to zero, as is easily verified (cf. Lemma F.1 for details). It is now elementary to verify, for a standard Gaussian d -vector V , that the bound (b1) with V replacing Z is satisfied with $\varepsilon = 1/2$ (and therefore for all $\varepsilon \in [0, 1/2]$), with $\xi = 1/2$ (and thus for all $\xi \in (0, 1/2]$), and with α and β replaced by constants α^* and β^* that depend only on k (see also the proof of Example A.1(i) in [10]). Clearly also the bound (b2) holds with V replacing Z (e.g., with $D = 1$). Therefore, all the arguments so far concerning the d -vectors Z_1, \dots, Z_k also apply to the Gaussian d -vectors V_1, \dots, V_k . In particular, we see that (4.17) and (4.18) are bounded in absolute value by the quantity given in Proposition 4.4, and hence also by (4.1), uniformly in $x \in \mathcal{S}_{M,p}$ and with constants that depend only on k .

It remains to bound the difference of (4.15) and (4.17) as well as the difference of (4.16) and (4.18). In particular, to complete the proof of Theorem 2.1(i), we need to show, under the assumptions of that theorem, that the absolute difference of each expression of the form (4.15) and the corresponding expression in (4.17) is bounded by (4.1), uniformly in $x \in \mathcal{S}_{M,p}$ and for constants as in (4.2). And to finish proving Theorem 2.1(ii), we have to show, under the assumptions of that theorem, that the absolute difference of each expression of the form (4.15) and the corresponding expression in (4.17) and also the absolute difference of (4.16) and (4.18) are all bounded by (4.1), uniformly in $x \in \mathcal{S}_{M,p}$ and for constants as in (4.3).

Proposition 4.5. *Suppose the random d -vector Z satisfies the bounds (b1)(a) and (b1)(b) for a fixed positive integer $k \leq 4$. For each $d \geq 1$, let Z_1, \dots, Z_k be i.i.d. copies of Z , set $S_k = (Z'_i Z_j / d)_{i,j=1}^k$, and let G and H be two (fixed) monomials in the elements of $S_k - I_k$ of degree g and h , respectively, where $\max\{g, h\} \leq k$. Finally, define G^* and H^* as G and H , but with the Z_1, \dots, Z_k replaced by i.i.d. standard Gaussian d -vectors, and consider*

$$\mathbb{E}[d^g (G - \mathbb{E}[G])H] - \mathbb{E}[d^g (G^* - \mathbb{E}[G^*])H^*]. \tag{4.19}$$

- (i) We then have $|\mathbb{E}[H] - \mathbb{E}[H^*]| \leq d^{-h/2}(\alpha + \alpha^*)$.
- (ii) Assume that G is given by the monomial

$$\prod_{i=1}^m (S_k - I_k)_{j_{i-1}+1, j_{i-1}+2} (S_k - I_k)_{j_{i-1}+2, j_{i-1}+3} \cdots (S_k - I_k)_{j_{i-1}, j_i}, \quad (4.20)$$

for some $m \geq 0$ and for indices j_0, \dots, j_m that satisfy $j_0 = 0$, $j_m \leq k$ and $j_{i-1} + 1 < j_i$ whenever $1 \leq i \leq m$. Then the expression in (4.19) is bounded in absolute value by $d^{-\min\{\xi, 1/2\}} \max\{\alpha + \alpha^*, \beta + \beta^*\}$.

- (iii) Assume that G is given by the monomial

$$(S_k - I_k)_{1,2} (S_k - I_k)_{2,3} \cdots (S_k - I_k)_{j-1,j} (S_k - I_k)_{j,1} \quad (4.21)$$

and note that the degree of G is j and satisfies $1 \leq j \leq k$. Moreover, assume that also the bound (b1)(c) is satisfied with k as chosen here. Then the expression in (4.19) is bounded in absolute value by $d^{-\min\{\xi, 1/2\}}(\beta(\alpha + 1) + \beta^*(\alpha^* + 1) + \beta^2 + \beta^{*2})$, unless either (a) $H = (S_k - I_k)_{a,a}$ for some a satisfying $1 \leq a \leq j$, (b) $H = (S_k - I_k)_{a,b}$ with $1 \leq a < b \leq j$, or (c) $H = ((S_k - I_k)_{a,b})^2$ with $1 \leq a < b \leq j$. In case (a), the expression in (4.19) is equal to $\text{Var}[Z'_1 Z_1]/d - 2$; in case (b), it is equal to $\mathbb{E}[(Z'_1 Z_2)^3]/d$; and in case (c), it equals $\text{Var}[(Z'_1 Z_2)^2]/d^2 - 2(1 + 3/d)$.

Here, α, β and ξ are the constants from (b1) and α^* and β^* are the analogous quantities to α and β if Z is replaced by a standard Gaussian d -vector. Note that α^* and β^* can be chosen such that they depend only on the value of k .

For later use, we note that Proposition 4.5(i)–(ii) applies under the assumptions of Theorem 2.1(i), and that Proposition 4.5(i)–(iii) applies under the assumptions of Theorem 2.1(ii).

Assume for the moment that Proposition 4.3 and Proposition 4.5(i)–(ii) applies, and consider the difference between (4.15) and (4.17), for some integer $k \leq 4$, and for indices l, m, j_0, \dots, j_m such that $1 \leq l \leq k$, $m \geq 0$, $j_0 = 0$, $j_m \leq l$ and $j_{i-1} + 1 < j_i$ whenever $1 \leq i \leq m$. If $m = 0$, the difference of interest is simply

$$\mathbb{E}[\psi_x(S_l - I_l)] - \mathbb{E}[\psi_x(S_l^* - I_l)]. \quad (4.22)$$

For $x \in \mathcal{S}_{M,p}$, recall that $\psi_x(S_l - I_l)$ is a weighted sum of monomials in $S_k - I_k$ of degree up to k , where the weights are all bounded in absolute value by $p^k M^{2(k+2)} C_\psi(k)$, where $C_\psi(k)$ is a constant that depends only on k (use Proposition 4.3 with l replacing k , and note that $l \leq k$). In that sum, the number of summands depends only on k . Therefore, (4.22) is a weighted sum of terms of the form $\mathbb{E}[H] - \mathbb{E}[H^*]$ as in Proposition 4.5(i) and thus we see that (4.22) is bounded by $p^k M^{2(k+2)} d^{-1/2} \kappa$ in absolute value and uniformly in $x \in \mathcal{S}_{M,p}$, where κ depends only on α and k . [Note that for $\text{deg}(H) = 0$ the difference in question is equal to zero.] This, in turn, is obviously bounded by the quantity in (4.1) uniformly in $x \in \mathcal{S}_{M,p}$, for an appropriate choice of g , for example, $g = 2(k + 2)$, that depends only on k and for $\kappa = \kappa(k, \alpha)$ depending only on k

and α . If $m > 0$, the difference of interest reads

$$\begin{aligned} & \mathbb{E} \left[\left(\prod_{i=1}^m Z'_{j_{i-1}+1} Z_{j_{i-1}+2} \cdots Z'_{j_{i-1}} Z_{j_i} \right) \psi_x(S_l - I_l) \right] \\ & \quad - \mathbb{E} \left[\left(\prod_{i=1}^m V'_{j_{i-1}+1} V_{j_{i-1}+2} \cdots V'_{j_{i-1}} V_{j_i} \right) \psi_x(S_l^* - I_l) \right] \\ & = \mathbb{E}[d^{\deg(G)}(G - \mathbb{E}[G])\psi_x(S_l - I_l)] - \mathbb{E}[d^{\deg(G^*)}(G^* - \mathbb{E}[G^*])\psi_x(S_l^* - I_l)], \end{aligned}$$

where G and G^* are as in Proposition 4.5(ii) and thus have mean zero. Again, if $x \in \mathcal{S}_{M,p}$, then we are dealing with a weighted sum of expressions of the form (4.19) whose weights are all bounded in absolute value by $p^k M^{2(k+2)} C_\psi(k)$ (by Proposition 4.3). Arguing as in the case where $m = 0$ and now using Proposition 4.5(ii), we see that the expression in the preceding display is bounded by (4.1) in absolute value and uniformly in $x \in \mathcal{S}_{M,p}$ for a constant $g = g(k)$ depending only on k and for $\kappa = \kappa(k, \alpha, \beta)$ depending only on k, α , and β .

Now suppose that the assumptions of Theorem 2.1(i) are satisfied. The argument in the preceding paragraph with $k = 2$ shows that the absolute difference between (4.15) and (4.17) is bounded by (4.1) uniformly in $x \in \mathcal{S}_{M,p}$ and for constants as in (4.2) (because k is fixed and equals 2 here). This completes the proof of Theorem 2.1(i).

Finally, assume that the conditions of Theorem 2.1(ii) are met (and note that hence $k = 4$). Arguing as in the second-to-last paragraph, we see that the absolute difference between (4.15) and (4.17) is bounded by (4.1) uniformly in $x \in \mathcal{S}_{M,p}$ and for constants as in (4.3) (again because k is fixed here). It remains to deal with the difference of (4.16) and (4.18). It is not difficult to rewrite this difference as

$$\begin{aligned} & \sum_{j=1}^k (-1)^{k-j} \binom{k}{j} \{ \mathbb{E}[(Z'_1 Z_2 \cdots Z_j Z'_j Z_1 - d + p - 1)\psi_x(S_k - I_k)] \\ & \quad - \mathbb{E}[(V'_1 V_2 \cdots V_j V'_j V_1 - d + p - 1)\psi_x(S_k^* - I_k)] \} \\ & = \sum_{j=1}^k (-1)^{k-j} \binom{k}{j} \{ \mathbb{E}[d^{\deg(G_j)}(G_j - \mathbb{E}[G_j])\psi_x(S_k - I_k)] \\ & \quad - \mathbb{E}[d^{\deg(G_j^*)}(G_j^* - \mathbb{E}[G_j^*])\psi_x(S_k^* - I_k)] \} \\ & \quad + \sum_{j=1}^k (-1)^{k-j} \binom{k}{j} (p - 1) \{ \mathbb{E}[\psi_x(S_k - I_k)] - \mathbb{E}[\psi_x(S_k^* - I_k)] \}, \end{aligned} \tag{4.23}$$

where G_j is shorthand for $Z'_1 Z_1/d - 1$ if $j = 1$ and for $Z'_1 Z_2 \cdots Z_j Z'_j Z_1/d^j$ if $j > 1$, and where G_j^* denotes the corresponding quantity computed from the V_i . Clearly, $\mathbb{E}[G_1] = 0$, while, for $j > 1$, $\mathbb{E}[G_j] = d^{-j+1}$, which verifies the equality in the previous display. Now the expected value in the last line of the preceding display can be treated similarly to (4.22) with $l = k$. Noting that

$(p - 1) \sum_{j=1}^k (-1)^{k-j} \binom{k}{j} = (p - 1)$, we see that the expression in the last line of the preceding display is bounded, in absolute value and uniformly in $x \in \mathcal{S}_{M,p}$, by (4.1) for constants as in (4.3). We are left with the first sum on the right-hand side in (4.23). Recalling that $\psi_x(S_k - I_k)$ and $\psi_x(S_k^* - I_k)$ are both weighted sums of monomials, we can rewrite this sum as

$$\begin{aligned} & \sum_{j=1}^k (-1)^{k-j} \binom{k}{j} \sum_{H \in \mathcal{M}_k} \mathcal{C}(H) \{ \mathbb{E}[d^{\deg(G_j)}(G_j - \mathbb{E}[G_j])H] \\ & - \mathbb{E}[d^{\deg(G_j^*)}(G_j^* - \mathbb{E}[G_j^*])H^*] \}, \end{aligned} \tag{4.24}$$

where \mathcal{M}_k and $\mathcal{C}(H)$, $H \in \mathcal{M}_k$, are as in Proposition 4.3, and where H^* is defined as H but with the Z_i replaced by the V_i . Note that (4.24) is a weighted sum of expressions of the form (4.19), where G_j and G_j^* are given by (4.21) defined with the Z_i and the V_i respectively. The coefficients $\mathcal{C}(H)$ of the terms in the polynomial $\varphi_x(S_k - I_k)$, as defined in Proposition 4.3, can be bounded, in absolute value and uniformly in $x \in \mathcal{S}_{M,p}$, by $p^k M^{2(k+2)} C_\psi(k)$. And using Proposition 4.5(iii), the difference of expectations in (4.24) can be bounded by $d^{-\min\{\xi, 1/2\}}(\beta(\alpha + 1) + \beta^*(\alpha^* + 1) + \beta^2 + \beta^{*2})$, unless the case (a), (b), or (c) in Proposition 4.5(iii) occurs. Taken together, it is now elementary to verify that those terms in (4.24) that do not correspond to the cases (a), (b), or (c) in the proposition are bounded, in absolute value and uniformly in $x \in \mathcal{S}_{M,p}$, by an expression of the form (4.1) with constants as in (4.3).

To deal with the remaining terms in (4.24), consider first those terms where G_j and H are as in case (a) of Proposition 4.5(iii). For each of these terms, the monomial H is of the form $(S_k - I_k)_{a,a}$ for some a with $1 \leq a \leq j$. Because the coefficients $\mathcal{C}(H)$ are invariant under permutations in the sense of Proposition 4.3, we see that the coefficient $\mathcal{C}(H)$ is the same number whenever case (a) occurs, and we denote this number by $\mathcal{C}_{(a)}$ in the following. Moreover, whenever G_j and H are such that the case (a) occurs, then the difference in expectations in (4.24) is equal to the same number, namely $\text{Var}[Z'_1 Z_1]/d - 2$, by Proposition 4.5(iii). Finally, for fixed j (and hence for fixed G_j), we note that the number of monomials H , where G_j and H are as in case (a), is equal to j . Putting the pieces together, we see that the combined contribution of those terms in (4.24), where G_j and H are as in case (a) of Proposition 4.5(iii), is equal to

$$\mathcal{C}_{(a)} (\text{Var}[Z'_1 Z_1]/d - 2) \sum_{j=1}^k \binom{k}{j} (-1)^{k-j} j.$$

But the sum in the preceding display can be written as $k \sum_{j=0}^{k-1} \binom{k-1}{j} (-1)^{k-1-j} = k(1 - 1)^{k-1} = 0$ (recall that we have chosen $k = 4$ here). Hence, the combined contribution of these terms is zero.

Consider now those terms in (4.24) where G_j and H are as in case (b) of Proposition 4.5(iii). For each of these terms, the monomial H is of the form $(S_k - I_k)_{a,b}$ for some a and b with $1 \leq a < b \leq j$. As in the preceding paragraph, we see that the coefficient $\mathcal{C}(H)$ is the same number whenever case (b) occurs, and we denote this number by $\mathcal{C}_{(b)}$; cf. Proposition 4.3. And whenever G_j and H are such that the case (b) occurs, then the difference in expectations in (4.24) is equal to the same number, namely $\mathbb{E}[(Z'_1 Z_1)^3]/d$; cf. Proposition 4.5(ii). Finally, for fixed j

(and hence for fixed G_j), there are $\binom{j}{2}$ monomials H so that G_j and H are as in case (b). The contribution of those terms in (4.24), where G_j and H are as in case (b) of Proposition 4.5(ii), is therefore equal to

$$C_{(b)} \mathbb{E}[(Z_1' Z_1)^3 / d] \sum_{j=1}^k \binom{k}{j} (-1)^{k-j} \binom{j}{2}.$$

The sum in the preceding display can be written as $\binom{k}{2} \sum_{j=0}^{k-2} \binom{k-2}{j} (-1)^{k-2-j} = 0$. A similar argument shows that the combined contribution of the terms where case (c) occurs is also zero.

In summary, we see, under the assumptions of Theorem 2.1(ii), that the difference of (4.15) and (4.17) as well as the difference of (4.16) and (4.18) can be bounded by (4.1) in absolute value, uniformly in $x \in \mathcal{S}_{M,p}$, and for constants as in (4.3). With this, also the proof of Theorem 2.1(ii) is complete.

Acknowledgments

The Austrian Science Fund (FWF) supports first author with project P 28233-N32 and the second author with projects P 28233-N32 and P 26354-N26.

Supplementary Material

Supplement to “On conditional moments of high-dimensional random vectors given lower-dimensional projections” (DOI: [10.3150/16-BEJ888SUPP](https://doi.org/10.3150/16-BEJ888SUPP); .pdf). The online supplement contains the proofs of the results in Section 3, as well as several more technical arguments that are used in Section 4 including, in particular, the proofs of Propositions 4.1 through 4.5.

References

- [1] Bryc, W. and Wesolowski, J. (2005). Conditional moments of q -Meixner processes. *Probab. Theory Related Fields* **131** 415–441. [MR2123251](#)
- [2] Chikuse, Y. (2003). *Statistics on Special Manifolds. Lecture Notes in Statistics* **174**. New York: Springer. [MR1960435](#)
- [3] Cohen, S.N. (2012). Quasi-sure analysis, aggregation and dual representations of sublinear expectations in general spaces. *Electron. J. Probab.* **17** no. 62, 15. [MR2959068](#)
- [4] Cook, R. D. and Weisberg, S. (1991). Discussion of “Sliced inverse regression for dimension reduction”, by K.-C. Li. *J. Amer. Statist. Assoc.* **86** 328–332.
- [5] Eaton, M.L. (1986). A characterization of spherical distributions. *J. Multivariate Anal.* **20** 272–276. [MR0866075](#)
- [6] Hall, P. and Li, K.-C. (1993). On almost linearity of low-dimensional projections from high-dimensional data. *Ann. Statist.* **21** 867–889. [MR1232523](#)
- [7] Hsu, D., Kakade, S.M. and Zhang, T. (2012). A tail inequality for quadratic forms of subgaussian random vectors. *Electron. Commun. Probab.* **17** no. 52, 6. [MR2994877](#)
- [8] Jones, M.C. (1998). Constant local dependence. *J. Multivariate Anal.* **64** 148–155. [MR1621930](#)

- [9] Leeb, H. (2013). On the conditional distributions of low-dimensional projections from high-dimensional data. *Ann. Statist.* **41** 464–483. [MR3099110](#)
- [10] Leeb, H. (2013). Supplement to “On the conditional distributions of low-dimensional projections from high-dimensional data.” DOI:[10.1214/12-AOS1081SUPP](#).
- [11] Li, K.-C. (1991). Sliced inverse regression for dimension reduction. *J. Amer. Statist. Assoc.* **86** 316–342. [MR1137117](#)
- [12] Milović, I. (2016). Conditional means of low-dimensional projections from high-dimensional data. Explicit error bounds. Ph.D. thesis, University of Vienna.
- [13] Plucińska, A. (1983). On a stochastic process determined by the conditional expectation and the conditional variance. *Stochastics* **10** 115–129. [MR0716819](#)
- [14] Qu, A., Lindsay, B.G. and Lu, L. (2010). Highly efficient aggregate unbiased estimating functions approach for correlated data with missing at random. *J. Amer. Statist. Assoc.* **105** 194–204. [MR2656049](#)
- [15] Steinberger, L. (2015). Statistical inference in high-dimensional linear regression based on simple working models. Ph.D. thesis, University of Vienna.
- [16] Steinberger, L. and Leeb, H. (2016). Supplement to “On conditional moments of high-dimensional random vectors given lower-dimensional projections.” DOI:[10.3150/16-BEJ888SUPP](#).
- [17] Steinberger, L. and Leeb, H. (2016). Prediction when fitting simple models to high-dimensional data. Working paper.
- [18] Tarpey, T. and Sanders, R.D. (2004). Linear conditional expectation for discretized distributions. *J. Appl. Stat.* **31** 361–372. [MR2061388](#)
- [19] Wesolowski, J. (1993). Stochastic processes with linear conditional expectation and quadratic conditional variance. *Probab. Math. Statist.* **14** 33–44. [MR1267516](#)

Received June 2014 and revised June 2016