Götze, Friedrich; Naumov, Alexey; Spokoiny, Vladimir; Ulyanov, Vladimir

Working Paper

Large ball probabilities, Gaussian comparison and anti-concentration

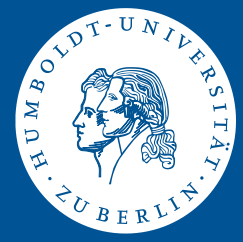
IRTG 1792 Discussion Paper, No. 2018-026

Provided in Cooperation with:
Humboldt University Berlin, International Research Training Group 1792 "High Dimensional Nonstationary Time Series"

Suggested Citation: Götze, Friedrich; Naumov, Alexey; Spokoiny, Vladimir; Ulyanov, Vladimir (2018) : Large ball probabilities, Gaussian comparison and anti-concentration, IRTG 1792 Discussion Paper, No. 2018-026, Humboldt-Universität zu Berlin, International Research Training Group 1792 "High Dimensional Nonstationary Time Series", Berlin

This Version is available at:
https://hdl.handle.net/10419/230737

IRTG 1792 Discussion Paper 2018-026



Large ball probabilities, Gaussian comparison and anti-concentration

Friedrich Götze *
Alexey Naumov *[2]
Vladimir Spokoiny *[2]
Vladimir Ulyanov *[2]

* Bielefeld University, Germany
*[2] National Research University Higher School of Economics, Russia

This research was supported by the Deutsche Forschungsgemeinschaft through the International Research Training Group 1792 "High Dimensional Nonstationary Time Series".

http://irtg1792.hu-berlin.de
ISSN 2568-5619

International Research Training Group 1792

Large ball probabilities, Gaussian comparison and anti-concentration

FRIEDRICH GÖTZE [1] , ALEXEY NAUMOV [2,4,5] , VLADIMIR SPOKOINY [2,4,5,6] and VLADIMIR ULYANOV [2,3].

[1] *Faculty of Mathematics, Bielefeld University, P. O. Box 10 01 31, 33501, Bielefeld, Germany. E-mail:* goetze@math.uni-bielefeld.de

[2] *National Research University Higher School of Economics, 20 Myasnitskaya ulitsa, 101000, Moscow, Russia E-mail:* anaumov@hse.ru, vulyanov@hse.ru

[3] *Moscow State University, Leninskie Gory, 1, Moscow, Russia E-mail:* vulyanov@cs.msu.ru

[4] *Skolkovo Institute of Science and Technology (Skoltech), Skolkovo Innovation Center, Building 3, 143026, Moscow , Russia*

[5] *Institute for Information Transmission Problems RAS, Bolshoy Karetny per. 19, bld.1, 127051, Moscow, Russia.*

[6] *Weierstrass Institute, Mohrenstr. 39, 10117 Berlin, Germany E-mail:* spokoiny@wias-berlin.de

We derive tight non-asymptotic bounds for the Kolmogorov distance between the probabilities of two Gaussian elements to hit a ball in a Hilbert space. The key property of these bounds is that they are dimension-free and depend on the nuclear (Schatten-one) norm of the difference between the covariance operators of the elements and on the norm of the mean shift. The obtained bounds significantly improve the bound based on Pinsker's inequality via the Kullback-Leibler divergence. We also establish an anti-concentration bound for a squared norm of a non-centered Gaussian element in Hilbert space. The paper presents a number of examples motivating our results and applications of the obtained bounds to statistical inference and to high-dimensional CLT.

Keywords: Gaussian comparison, Gaussian anti-concentration inequalities, effective rank, dimension free bounds, Schatten norm, high-dimensional inference.

1. Introduction

In many statistical and probabilistic applications one faces the problem to evaluate how the probability of a ball under a Gaussian measure is affected, if the mean and the covariance operators of this Gaussian measure are slightly changed. Below we present particular examples motivating our results when such "large ball probability" problem naturally arises, including bootstrap validation, Bayesian inference, high-dimensional CLT. This paper presents sharp bounds for the Kolmogorov distance between the probabilities of two Gaussian elements to hit a ball in a Hilbert space. The key property of these bounds

is that they are dimension-free and depend on the nuclear (Schatten-one) norm of the difference between the covariance operators of the elements. We also state a tight dimension free anti-concentration bound for a squared norm of a Gaussian element in Hilbert space which refines the well known results on the density of a chi-squared distribution; see Theorem 2.7.

Section 1.1 presents some application examples where the "large ball probability" issue naturally arises and explains how the new bounds of this paper can be used to improve the existing results. The key observation behind the improvement is that in all mentioned examples we only need to know the properties of Gaussian measures on a class of balls. It means, in particular, that we would like to compare two Gaussian measures on the class of balls instead on the class of all measurable sets. The latter can be upperbounded by general Pinsker's inequality via the Kullback–Leibler divergence. In case of Gaussian measures this divergence can be expressed explicitly in terms of parameters of the underlying measures, see e.g. Spokoiny and Zhilova (2015). However, the obtained bound involves the inverse of the covariance operators of the considered Gaussian measures. In particularly, small eigenvalues have the largest impact which is contra-intuitive if a probability of a ball is considered. Our bounds only involve the operator and Frobenius norms of the related covariance operators and apply even in Hilbert space setup.

The proofs of the present optimal results are based in particular on Theorem 2.6 below. This theorem gives sharp upper bounds for a probability density function $p_{\boldsymbol{\xi}}(x, \boldsymbol{a})$ of $\|\boldsymbol{\xi} - \boldsymbol{a}\|^2$, where $\boldsymbol{\xi}$ is a Gaussian element with zero mean in a Hilbert space $\mathbb{H}$ with norm $\|\cdot\|$ and $\boldsymbol{a} \in \mathbb{H}$. It is well known that $p_{\boldsymbol{\xi}}(x, \boldsymbol{a})$ can be considered as a density function of a weighted sum of non-central χ^2 distributions. An explicit but cumbersome representation for $p_{\boldsymbol{\xi}}(x, \boldsymbol{a})$ in finite dimensional space $\mathbb{H}$ is available (see e.g. Section 18 in Johnson et al. (1994)). However, it involves some special characteristics of the related Gaussian measure which makes it hard to use in specific situations. Our results from Theorem 2.6 and by Lemma B.1 are much more transparent and provide sharp uniform and non-uniform upper bounds on the underlying density respectively.

One can even get two-sided bounds for $p_{\boldsymbol{\xi}}(x, \boldsymbol{a})$ but under additional conditions, see e.g. Christoph et al. (1996). Asymptotic properties of $p_{\boldsymbol{\xi}}(x, \boldsymbol{a})$, small balls probabilities $I\!P\big(\|\boldsymbol{\xi} - a\| \le \varepsilon\big)$, or large deviation bounds $I\!P\big(\|\boldsymbol{\xi}\| \ge 1/\varepsilon\big)$ for small ε can be found e.g. in Bogachev (1998), Ledoux and Talagrand (2002), Li and Shao (2001), Lifshits (2012) and Yurinsky (1995).

The paper is organized as follows: a list of examples motivating our results and possible applications are given in Section 1.1. Section 2 collects the main results. The proofs are given in Section 3. Some technical results and non-uniform upper bounds for $p_{\boldsymbol{\xi}}(x, \boldsymbol{a})$ are presented in the appendix.

1.1. Application examples

This section collects some examples where the developed results seem to be very useful.

1.1.1. Bootstrap validity for the MLE

Consider an independent sample $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$ with a joint distribution $I\!P = \prod_{i=1,\ldots,n} P_i$. The parametric maximum likelihood approach assumes that $I\!P$ belongs to a given parametric family $\bigl(I\!P_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta \subseteq I\!R^p\bigr)$ dominated by a measure $\boldsymbol{\mu}$, that is, $I\!P = I\!P_{\boldsymbol{\theta}^*}$ for $\boldsymbol{\theta}^* \in \Theta$. The corresponding log-likelihood function can be written as a sum of marginal log-likelihoods $\ell_i(Y_i, \boldsymbol{\theta})$:

$$L(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \log \frac{dI\!P_{\boldsymbol{\theta}}}{d\boldsymbol{\mu}}(\boldsymbol{Y}) = \sum_{i=1}^{n} \ell_i(Y_i, \boldsymbol{\theta}), \qquad \ell_i(Y_i, \boldsymbol{\theta}) = \log \frac{dP_{i,\boldsymbol{\theta}}}{d\mu_i}(Y_i).$$

The MLE $\widetilde{\boldsymbol{\theta}}$ of the true parameter $\boldsymbol{\theta}^*$ is defined as the point of maximum of $L(\boldsymbol{\theta})$:

$$\widetilde{\boldsymbol{\theta}} \stackrel{\text{def}}{=} \operatorname*{argmax}_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta}), \qquad L(\widetilde{\boldsymbol{\theta}}) \stackrel{\text{def}}{=} \max_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta}).$$

If the parametric assumption is misspecified, the target $\boldsymbol{\theta}^*$ is defined as the best parametric fit:

$$\boldsymbol{\theta}^* \stackrel{\text{def}}{=} \operatorname*{argmax}_{\boldsymbol{\theta} \in \Theta} I\!E L(\boldsymbol{\theta}).$$

The likelihood based confidence set $\mathcal{E}(\mathfrak{z})$ for the target parameter $\boldsymbol{\theta}^*$ is given by

$$\mathcal{E}(\mathfrak{z}) \stackrel{\text{def}}{=} \bigl\{\boldsymbol{\theta}\colon L(\widetilde{\boldsymbol{\theta}}) - L(\boldsymbol{\theta}) \leq \mathfrak{z}\bigr\}.$$

The value $\mathfrak{z}$ should be selected to ensure the prescribed coverage probability $1 - \alpha$:

$$I\!P\bigl(\boldsymbol{\theta}^* \notin \mathcal{E}(\mathfrak{z})\bigr) \leq \alpha. \tag{1.1}$$

However, it depends on the unknown measure $I\!P$. The bootstrap approach is a resampling technique based on the conditional distribution of the reweighted log-likelihood $L^\flat(\boldsymbol{\theta})$

$$L^\flat(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(Y_i, \boldsymbol{\theta}) w_i^\flat$$

with i.i.d. random weights $w_i^\flat$ given the data $\boldsymbol{Y}$. Below we assume that $w_i^\flat \sim \mathcal{N}(1, 1)$. The bootstrap confidence set is defined as

$$\mathcal{E}^\flat(\mathfrak{z}) \stackrel{\text{def}}{=} \bigl\{\boldsymbol{\theta}\colon \sup_{\boldsymbol{\theta}' \in \Theta} L^\flat(\boldsymbol{\theta}') - L^\flat(\boldsymbol{\theta}) \leq \mathfrak{z}\bigr\}.$$

The bootstrap distribution is perfectly known and the bootstrap quantile $\mathfrak{z}^\flat$ is defined by the condition

$$I\!P^\flat\big(\widetilde{\boldsymbol{\theta}} \notin \mathcal{E}^\flat(\mathfrak{z}^\flat)\big) \;=\; I\!P^\flat\Big(\sup_{\boldsymbol{\theta}\in\Theta} L^\flat(\boldsymbol{\theta}) - L^\flat(\widetilde{\boldsymbol{\theta}}) > \mathfrak{z}^\flat\Big) = \alpha.$$

The bootstrap approach suggests to use $\mathfrak{z}^\flat$ in place of $\mathfrak{z}$ to ensure (1.1) in an asymptotic sense. Bootstrap consistency means that for n large

$$I\!P\big(\boldsymbol{\theta}^* \notin \mathcal{E}(\mathfrak{z}^\flat)\big) \;=\; I\!P\big(L(\widetilde{\boldsymbol{\theta}}) - L(\boldsymbol{\theta}^*) > \mathfrak{z}^\flat\big) \approx \alpha;$$

see e.g. Spokoiny and Zhilova (2015). A proof of this result is quite involved. The key steps are the following two approximations:

$$\begin{aligned} \sup_{\boldsymbol{\theta}\in\Theta} L(\boldsymbol{\theta}) - L(\boldsymbol{\theta}^*) &\approx \frac{1}{2}\big\|\boldsymbol{\xi} + \boldsymbol{a}\big\|^2, \\ \sup_{\boldsymbol{\theta}\in\Theta} L^\flat(\boldsymbol{\theta}) - L^\flat(\widetilde{\boldsymbol{\theta}}) &\approx \frac{1}{2}\big\|\boldsymbol{\xi}^\flat\big\|^2, \end{aligned} \tag{1.2}$$

where $\boldsymbol{\xi}$ is a Gaussian vector with the variance Σ given by

$$\Sigma \stackrel{\text{def}}{=} D^{-1}\,\mathrm{Var}\big[\nabla L(\boldsymbol{\theta}^*)\big]D^{-1}, \qquad D^2 = -\nabla^2 I\!E L(\boldsymbol{\theta}^*),$$

while $\boldsymbol{\xi}^\flat$ is conditionally (given $\boldsymbol{Y}$) Gaussian w.r.t. the bootstrap measure $I\!P^\flat$ with the covariance $\Sigma^\flat$ given by

$$\Sigma^\flat \stackrel{\text{def}}{=} D^{-1}\left(\sum_{i=1}^n \nabla\ell_i(Y_i,\boldsymbol{\theta})\big\{\nabla\ell_i(Y_i,\boldsymbol{\theta})\big\}^\top\right)D^{-1}.$$

The vector $\boldsymbol{a}$ in (1.2) is the so called modeling bias and it vanishes if the parametric assumption $I\!P = I\!P_{\boldsymbol{\theta}^*}$ is precisely fulfilled. The matrix Bernstein inequality ensures that $\Sigma^\flat$ is close to Σ in the operator norm for n large; see e.g. Tropp (2012). This yields bootstrap validity under the true parametric assumption in a weak sense. However, for quantifying the quality of the bootstrap approximation one has to measure the distance between two high dimensional Gaussian distributions $\mathcal{N}(\boldsymbol{a},\Sigma)$ and $\mathcal{N}(0,\Sigma^\flat)$. The recent paper Spokoiny and Zhilova (2015) used the approach based on the Pinsker inequality which gives a bound in the total variation distance $\|\cdot\|_{\mathrm{TV}}$ via the Kullback-Leibler divergence between these two measures. A related bound involves the Frobenius norm $\|\cdot\|_{\mathrm{Fr}}$ of the matrix $\Sigma^{-1/2}\Sigma^\flat\Sigma^{-1/2} - I_p$ and the norm of the vector $\boldsymbol{\beta} \stackrel{\text{def}}{=} \Sigma^{-1/2}\boldsymbol{a}$:

$$\big\|\mathcal{N}(\boldsymbol{a},\Sigma) - \mathcal{N}(0,\Sigma^\flat)\big\|_{\mathrm{TV}} \le \frac{1}{2}\Big(\big\|\Sigma^{-1/2}\Sigma^\flat\Sigma^{-1/2} - I_p\big\|_{\mathrm{Fr}} + \big\|\Sigma^{-1/2}\boldsymbol{a}\big\|\Big); \tag{1.3}$$

see e.g. Spokoiny and Zhilova (2015). However, if we limit ourselves to the centered balls then these bounds can be significantly improved. Namely, by the main result of Theorem 2.1 and Corollary 2.2 below, we get under some technical conditions

$$\left| I\!P\Big(\big\| \boldsymbol{\xi} + \boldsymbol{a} \big\|^2 > 2\mathfrak{z}^\flat \Big) - \alpha \right| \leq \frac{\mathtt{C}}{\|\Sigma\|_{\mathrm{Fr}}} \Big(\|\Sigma - \Sigma^\flat\|_1 + \|\boldsymbol{a}\|^2 \Big). \tag{1.4}$$

The "small modeling bias" condition on $\boldsymbol{a}$ from Spokoiny and Zhilova (2015) means that the value $\|\Sigma^{-1/2}\boldsymbol{a}\|$ is small and it ensures that a possible model misspecification does not destroy the validity of the bootstrap. Comparison of (1.4) with (1.3) reveals a number of benefits of (1.4). First, the "shift" term is proportional to the squared norm of the vector $\boldsymbol{a}$, while the bound (1.3) depends on the norm of $\Sigma^{-1/2}\boldsymbol{a}$, i.e. on the whole spectrum of Σ. Normalization by $\Sigma^{-1/2}$ can significantly inflate the vector $\boldsymbol{a}$ in directions where the eigenvalues of Σ are small. In the contrary, the bound (1.4) only involves the squared norm $\|\boldsymbol{a}\|^2$ and the Frobenius norm of Σ, and the improvement from $\big\|\Sigma^{-1/2}\boldsymbol{a}\big\|$ to $\|\boldsymbol{a}\|^2/\|\Sigma\|_{\mathrm{Fr}}$ can be enormous if some eigenvalues of Σ nearly vanish. Further, the Frobenius norm $\big\|\Sigma^{-1/2}\Sigma^\flat\Sigma^{-1/2} - \boldsymbol{I}_p\big\|_{\mathrm{Fr}}$ can be much larger than the ratio $\big\|\Sigma - \Sigma^\flat\big\|_1/\|\Sigma\|_{\mathrm{Fr}}$ by the same reasons.

1.1.2. Prior impact in linear Gaussian modeling

Consider a linear regression model

$$Y_i = \Psi_i^\top \boldsymbol{\theta} + \varepsilon_i$$

The assumption of homogeneous Gaussian errors $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ yields the log-likelihood

$$L(\boldsymbol{\theta}) = -\frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \Psi_i^\top \boldsymbol{\theta})^2 + R = -\frac{1}{2\sigma^2} \big\| \boldsymbol{Y} - \boldsymbol{\Psi}^\top \boldsymbol{\theta} \big\|^2 + R,$$

where the term R does not depend on $\boldsymbol{\theta}$. A Gaussian prior $\mathit{\Pi} = \mathit{\Pi}_G = \mathcal{N}\big(0, G^{-2}\big)$ results in the posterior

$$\boldsymbol{\vartheta}_G \,\big|\, \boldsymbol{Y} \propto \exp\left(L(\boldsymbol{\theta}) - \frac{1}{2}\|G\boldsymbol{\theta}\|^2 \right) \propto \exp\left(-\frac{1}{2\sigma^2} \big\| \boldsymbol{Y} - \boldsymbol{\Psi}^\top \boldsymbol{\theta} \big\|^2 - \frac{1}{2}\|G\boldsymbol{\theta}\|^2 \right).$$

We shall represent the quantity $L_G(\boldsymbol{\theta}) \stackrel{\mathrm{def}}{=} L(\boldsymbol{\theta}) - \frac{1}{2}\|G\boldsymbol{\theta}\|^2$ in the form

$$L_G(\boldsymbol{\theta}) = L_G(\breve{\boldsymbol{\theta}}_G) - \frac{1}{2} \big\| D_G(\boldsymbol{\theta} - \breve{\boldsymbol{\theta}}_G) \big\|^2,$$

where

$$\breve{\boldsymbol{\theta}}_G \stackrel{\mathrm{def}}{=} \big(\boldsymbol{\Psi}\boldsymbol{\Psi}^\top + \sigma^2 G^2 \big)^{-1} \boldsymbol{\Psi}\boldsymbol{Y},$$

$$D_G^2 \stackrel{\mathrm{def}}{=} \sigma^{-2}\boldsymbol{\Psi}\boldsymbol{\Psi}^\top + G^2.$$

In particular, it implies that the posterior distribution $I\!P(\boldsymbol{\vartheta}_G \,|\, \boldsymbol{Y})$ of $\boldsymbol{\vartheta}_G$ given $\boldsymbol{Y}$ is $\mathcal{N}(\breve{\boldsymbol{\theta}}_G, D_G^{-2})$. A contraction property is a kind of concentration of the posterior on the elliptic set

$$E_G(\mathtt{r}) = \bigl\{\boldsymbol{\theta}\colon \|W(\boldsymbol{\theta} - \breve{\boldsymbol{\theta}}_G)\| \le \mathtt{r}\bigr\},$$

where W is a given linear mapping from $I\!R^p$. The desirable credibility property manifests the prescribed conditional probability of $\boldsymbol{\vartheta}_G \in E(\mathtt{r}_G)$ given $\boldsymbol{Y}$ with $\mathtt{r}_G$ defined for a given α by

$$I\!P\Bigl(\bigl\|W\bigl(\boldsymbol{\vartheta}_G - \breve{\boldsymbol{\theta}}_G\bigr)\bigr\| \ge \mathtt{r}_G \,\bigl|\, \boldsymbol{Y}\Bigr) = \alpha. \tag{1.5}$$

Under the posterior measure $\boldsymbol{\vartheta}_G \sim \mathcal{N}(\breve{\boldsymbol{\theta}}_G, D_G^{-2})$, this bound reads as

$$I\!P\bigl(\|\boldsymbol{\xi}_G\| \ge \mathtt{r}_G\bigr) = \alpha \tag{1.6}$$

with a zero mean normal vector $\boldsymbol{\xi}_G \sim \mathcal{N}(0, \Sigma_G)$ for $\Sigma_G = W D_G^{-2} W^\top$. The question of a prior impact can be stated as follows: whether the obtained credible set significantly depends on the prior covariance G. Consider another prior $\Pi_1 = \mathcal{N}(0, G_1^{-2})$ with the covariance matrix G_1^{-2}. The corresponding posterior $\boldsymbol{\vartheta}_{G_1}$ is again normal but now with parameters $\breve{\boldsymbol{\theta}}_{G_1} = \bigl(\boldsymbol{\Psi}\boldsymbol{\Psi}^\top + \sigma^2 G_1^2\bigr)^{-1}\boldsymbol{\Psi}\boldsymbol{Y}$ and $D_{G_1}^2 = \sigma^{-2}\boldsymbol{\Psi}\boldsymbol{\Psi}^\top + G_1^2$. We aim at checking the posterior probability of the credible set $E_G(\mathtt{r}_G)$:

$$I\!P\Bigl(\bigl\|W\bigl(\boldsymbol{\vartheta}_{G_1} - \breve{\boldsymbol{\theta}}_G\bigr)\bigr\| \ge \mathtt{r}_G \,\bigl|\, \boldsymbol{Y}\Bigr).$$

Clearly this probability can be written as

$$I\!P\Bigl(\bigl\|\boldsymbol{\xi}_{G_1} + \boldsymbol{a}\bigr\| \ge \mathtt{r}_G\Bigr)$$

with $\boldsymbol{\xi}_{G_1} \sim \mathcal{N}(0, \Sigma_{G_1})$ for $\Sigma_{G_1} = W D_{G_1}^{-2} W^\top$ and

$$\boldsymbol{a} \stackrel{\mathrm{def}}{=} W\bigl(\breve{\boldsymbol{\theta}}_{G_1} - \breve{\boldsymbol{\theta}}_G\bigr).$$

Therefore,

$$\Bigl|I\!P\Bigl(\bigl\|W\bigl(\boldsymbol{\vartheta}_{G_1} - \breve{\boldsymbol{\theta}}_G\bigr)\bigr\| \ge \mathtt{r}_G \,\bigl|\, \boldsymbol{Y}\Bigr) - \alpha\Bigr| \le \sup_{\mathtt{r}>0}\Bigl|I\!P\Bigl(\bigl\|\boldsymbol{\xi}_{G_1} - \boldsymbol{a}\bigr\| \ge \mathtt{r}\Bigr) - I\!P\Bigl(\bigl\|\boldsymbol{\xi}_G\bigr\| \ge \mathtt{r}\Bigr)\Bigr|.$$

Again, the Pinsker inequality allows to upperbound the total variation distance between the Gaussian measures $\mathcal{N}(0, \Sigma_G)$ and $\mathcal{N}(\boldsymbol{a}, \Sigma_{G_1})$, however the answer is given via the Kullback-Leibler distance between these two measures:

$$\bigl\|\mathcal{N}(0, \Sigma_G) - \mathcal{N}(\boldsymbol{a}, \Sigma_{G_1})\bigr\|_{\mathrm{TV}} \le \mathtt{C}\Bigl(\bigl\|\Sigma_G^{-1/2}\Sigma_{G_1}\Sigma_G^{-1/2} - \boldsymbol{I}_p\bigr\|_{\mathrm{Fr}} + \bigl\|\Sigma_{G_1}^{-1/2}\boldsymbol{a}\bigr\|\Bigr); \tag{1.7}$$

see e.g. Panov and Spokoiny (2015). Results of this paper allow to significantly improve this bound. In particular, only the nuclear norm $\|\Sigma_G - \Sigma_{G_1}\|_1$, the norm of the vector $\boldsymbol{a}$ and the Frobenius norm of Σ_G are involved. If $G^2 \geq G_1^2$, then $\Sigma_G \leq \Sigma_{G_1}$ and

$$\|\Sigma_G - \Sigma_{G_1}\|_1 = \operatorname{tr} \Sigma_{G_1} - \operatorname{tr} \Sigma_G$$

and thus, by the main result of Theorem 2.1 and Corollary 2.2 below, it holds under some technical conditions

$$\left| I\!P\Big(\|W(\vartheta_{G_1} - \breve{\boldsymbol{\theta}}_G)\| \geq \mathtt{r}_G \,\big|\, \boldsymbol{Y} \Big) - \alpha \right| \leq \frac{\mathtt{C}\big(\operatorname{tr} \Sigma_{G_1} - \operatorname{tr} \Sigma_G + \|\boldsymbol{a}\|^2\big)}{\|\Sigma_G\|_{\mathrm{Fr}}}.$$

This new bound significantly outperforms (1.7); see the discussion at the end of Section 1.1.1.

1.1.3. Nonparametric Bayes approach

One of the central question in the *nonparametric Bayes* approach is whether one can use the corresponding credible set as a *frequentist confidence set* for the true underlying mean $I\!E\boldsymbol{Y} = \boldsymbol{f}^* = \boldsymbol{\Psi}^\top \boldsymbol{\theta}^*$. Here we consider the model $\boldsymbol{Y} = \boldsymbol{f}^* + \boldsymbol{\varepsilon} = \boldsymbol{\Psi}^\top \boldsymbol{\theta} + \boldsymbol{\varepsilon}$ in $I\!R^n$ with a homogeneous Gaussian noise $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2 I_n)$ and a Gaussian prior $\mathcal{N}(0, G^{-2})$ on $\boldsymbol{\theta}$. The credible set $E_G(\mathtt{r})$ for ϑ_G yields the credible set $\mathcal{E}_G(\mathtt{r})$ for the corresponding response $\boldsymbol{f} = \boldsymbol{\Psi}^\top \boldsymbol{\theta}$:

$$\mathcal{E}(\mathtt{r}) = \big\{ \boldsymbol{f} = \boldsymbol{\Psi}^\top \boldsymbol{\theta} \colon \|\boldsymbol{A}\boldsymbol{\Psi}^\top(\boldsymbol{\theta} - \breve{\boldsymbol{\theta}}_G)\| \leq \mathtt{r} \big\},$$

with some linear mapping $\boldsymbol{A}$. The radius $\mathtt{r} = \mathtt{r}_G$ is fixed to ensure the prescribed credibility $1 - \alpha$ for the corresponding set $\mathcal{E}(\mathtt{r}_\alpha)$ due to (1.5) or (1.6) with $W = \boldsymbol{A}\boldsymbol{\Psi}^\top$ and $\Sigma_G = \boldsymbol{A}\boldsymbol{\Psi}^\top D_G^{-2} \boldsymbol{\Psi}\boldsymbol{A}^\top = \sigma^2 \boldsymbol{A} \Pi_G \boldsymbol{A}^\top$, with $\Pi_G = \boldsymbol{\Psi}^\top \big(\boldsymbol{\Psi}\boldsymbol{\Psi}^\top + \sigma^2 G^2\big)^{-1} \boldsymbol{\Psi}$. The frequentist coverage probability of the true response $\boldsymbol{f}^*$ is given by

$$I\!P\big(\boldsymbol{f}^* \in \mathcal{E}_G(\mathtt{r})\big) = I\!P\big(\|\boldsymbol{A}(\boldsymbol{f}^* - \boldsymbol{\Psi}^\top \breve{\boldsymbol{\theta}}_G)\| \leq \mathtt{r}\big) = I\!P\big(\|\boldsymbol{A}\boldsymbol{\Psi}^\top(\boldsymbol{\theta}^* - \breve{\boldsymbol{\theta}}_G)\| \leq \mathtt{r}\big).$$

The aim is to show that the the latter is close to $1 - \alpha$. For the posterior mean $\breve{\boldsymbol{\theta}}_G = \big(\boldsymbol{\Psi}\boldsymbol{\Psi}^\top + \sigma^2 G^2\big)^{-1} \boldsymbol{\Psi}\boldsymbol{Y}$, it holds

$$I\!E\big[\boldsymbol{A}\big(\boldsymbol{f}^* - \boldsymbol{\Psi}^\top \breve{\boldsymbol{\theta}}_G\big)\big] = \boldsymbol{A}\big(I - \Pi_G\big)\boldsymbol{f}^* \stackrel{\mathrm{def}}{=} \boldsymbol{a}.$$

Further,

$$\Sigma \stackrel{\mathrm{def}}{=} \operatorname{Var}\big\{\boldsymbol{A}\big(\boldsymbol{f}^* - \boldsymbol{\Psi}^\top \breve{\boldsymbol{\theta}}_G\big)\big\} = \operatorname{Var}\big\{\boldsymbol{A}\Pi_G\,\boldsymbol{\varepsilon}\big\} = \sigma^2 \boldsymbol{A}\Pi_G^2 \boldsymbol{A}^\top$$

and hence, the vector $\boldsymbol{A}\bigl(\boldsymbol{f}^* - \boldsymbol{\Psi}^\top \breve{\boldsymbol{\theta}}_G\bigr)$ is under $I\!P$ normal with mean $\boldsymbol{a} = \boldsymbol{A}\bigl(I - \Pi_G\bigr)\boldsymbol{f}^*$ and variance $\Sigma = \sigma^2 \boldsymbol{A}\Pi_G^2\boldsymbol{A}^\top$. Therefore,

$$I\!P\bigl(\boldsymbol{f}^* \in \mathcal{E}_G(\mathtt{r})\bigr) \;=\; I\!P\bigl(\bigl\|\boldsymbol{a} + \boldsymbol{\xi}\bigr\| \le \mathtt{r}\bigr).$$

Here $\boldsymbol{\xi} \sim \mathcal{N}(0, \Sigma)$. So, it suffices to compare two probabilities

$$I\!P\bigl(\bigl\|\boldsymbol{a} + \boldsymbol{\xi}\bigr\| \le \mathtt{r}\bigr) \quad \text{vs} \quad I\!P\bigl(\bigl\|\boldsymbol{\xi}_G\bigr\| \le \mathtt{r}\bigr)$$

for all $\mathtt{r} \ge 0$. Existing results cover only very special cases; see e.g. Johnstone (2010); Bontemps (2011); Panov and Spokoiny (2015); Castillo (2012); Castillo and Nickl (2013); Belitser (2017) and references therein. Most of the mentioned results are of asymptotic nature and do not quantify the accuracy of the coverage probability. The results of this paper enable to study this accuracy in a straightforward way. Note first that the covariance operators $\Sigma = \sigma^2 \boldsymbol{A}\Pi_G^2\boldsymbol{A}^\top$ and $\Sigma_G = \sigma^2 \boldsymbol{A}\Pi_G\boldsymbol{A}^\top$ satisfy $\Sigma \le \Sigma_G$. This yields that

$$\bigl\|\Sigma_G - \Sigma\bigr\|_1 \;=\; \operatorname{tr}\Sigma_G - \operatorname{tr}\Sigma\,.$$

Theorem 2.1 and Corollary 2.2 allow to evaluate under some technical conditions the coverage probability of the credibility set

$$\bigl|I\!P\bigl(\boldsymbol{f}^* \notin \mathcal{E}_G(\mathtt{r}_G)\bigr) - \alpha\bigr| \;\le\; \frac{\mathtt{C}\bigl(\operatorname{tr}\Sigma_G - \operatorname{tr}\Sigma + \|\boldsymbol{a}\|^2\bigr)}{\|\Sigma\|_{\mathrm{Fr}}}\,.$$

The right hand-side of this bound can be easily evaluated. The value $\|\boldsymbol{a}\| = \boldsymbol{A}\bigl(I - \Pi_G\bigr)\boldsymbol{f}^*$ is small under usual smoothness assumptions on $\boldsymbol{f}^*$. The difference

$$\operatorname{tr}\Sigma_G - \operatorname{tr}\Sigma \;=\; \sigma^2 \operatorname{tr}\bigl\{\boldsymbol{A}(\Pi_G - \Pi_G^2)\boldsymbol{A}^\top\bigr\}$$

is small under standard condition on the design $\boldsymbol{\Psi}$ and on the spectrum of G^2; see e.g. Spokoiny (2017).

1.1.4. Central Limit Theorem in finite- and infinite-dimensional spaces

Another motivation for the current paper comes from the limit theorem in high-dimensional spaces for convex sets, in particular, for non-centred balls. Applications of smoothing inequalities require to evaluate the probability of hitting the vicinity of a convex set, see e.g. Bentkus (2003), Bentkus (2005). This question is closely related to the anti-concentration inequalities considered below in Theorem 2.7. Recently, significant interest was shown in understanding of the anti-concentration phenomenon for weighted sums of random variables, particularly, in random matrix and number theory. We refer the interested reader to Rudelson and Vershynin (2008), Götze and Zaitsev (2016).

Let $Y_1, \dots, Y_n$ be i.i.d. random vectors in $\mathbb{R}^p$. Assume that all these vectors have zero mean and the covariance operator Σ. Let X be a Gaussian random vector in $\mathbb{R}^p$ with zero mean and the same covariance operator Σ. We are interested to bound

$$\delta(\mathbf{C}) = \sup_{A \in \mathbf{C}} \left| I\!P\left(\frac{Y_1 + \dots + Y_n}{\sqrt{n}} \in A \right) - I\!P(X \in A) \right| \tag{1.8}$$

for some class $\mathbf{C}$ of Borel sets. It is worth emphasizing that the probabilities of hitting the vicinities of a set $A \in \mathbf{C}$, play the crucial role in the form of the bound for $\delta(\mathbf{C})$. Assume the class $\mathbf{C}$ satisfies the following two conditions:

(i) Class $\mathbf{C}$ is invariant under affine symmetric transformations, that is, $\boldsymbol{D}A + \boldsymbol{a} \in \mathbf{C}$ if $\boldsymbol{a} \in \mathbb{R}^p$ and $\boldsymbol{D} : \mathbb{R}^p \to \mathbb{R}^p$ is a linear symmetric invertible operator.

(ii) Class $\mathbf{C}$ is invariant under taking ε-neighborhoods for all $\varepsilon > 0$. More precisely, $A^{\varepsilon}, A^{-\varepsilon} \in \mathbf{C}$ if $A \in \mathbf{C}$, where

$$A^{\varepsilon} = \{x \in \mathbb{R}^p : \rho_A(x) \le \varepsilon\} \text{ and } A^{-\varepsilon} = \{x \in A : B_{\varepsilon}(x) \subset A\},$$

with $\rho_A(x) = \inf_{y \in A} |x - y|$ as the distance between $A \subset \mathbb{R}^p$ and $x \in \mathbb{R}^p$, and $B_{\varepsilon}(x) = \{y \in \mathbb{R}^p : |x - y| \le \varepsilon\}$.

Let X_0 be a Gaussian random vector in $\mathbb{R}^p$ with zero mean and the identity covariance operator I. Assume that the class $\mathbf{C}$ in (1.8) is such that for all $A \in \mathbf{C}$ and $\varepsilon > 0$

$$I\!P(X_0 \in A^{\varepsilon} \backslash A) \le a_p\, \varepsilon, \quad I\!P(X_0 \in A \backslash A^{-\varepsilon}) \le a_p\, \varepsilon, \tag{1.9}$$

where $a_p = a_p(\mathbf{C})$ is the so called isoperimetric constant of $\mathbf{C}$, e.g. taking $\mathbf{C}$ as the class of all convex sets in $\mathbb{R}^p$ we get $a_p \le 4\, p^{1/4}$; see Ball (1993).

It is known (see Bentkus (2005)[Theorem 1.2]) that if $\mathbf{C}$ satisfies conditions (i), (ii) and (1.9) then for some absolute constant C one has

$$\delta(\mathbf{C}) \le C\,(1 + a_p)\, I\!E|Y_1|^3/\sqrt{n}. \tag{1.10}$$

Therefore, the inequalities (1.9), i.e. knowledge of a_p, play the crucial role in the form of the bound (1.10).

We have a similar situation in infinite-dimensional spaces. Though contrary to the finite dimensional case even if $\mathbf{C}$ is a rather small class of "good" subsets, e.g. the class of all balls, the convergence of $I\!P\Big((Y_1 + \dots + Y_n)/\sqrt{n} \in A\Big)$ to $I\!P\big(X \in A\big)$ for each $A \in \mathbf{C}$, implied by the central limit theorem, can not be uniform in $A \in \mathbf{C}$; see e.g. Sazonov (1981)[pp. 69–70]. However, the convergence becomes uniform for a class of all balls with center at some fixed point, say $\boldsymbol{a}$. Such classes naturally appear in various statistical problems; see e.g. Prokhorov and Ulyanov (2013) or our previous application examples. Thus, similar to the inequalities (1.9) we need to get sharp bounds for the

probability $I\!P(x < \|X - \boldsymbol{a}\|^2 < x + \varepsilon)$ for the Gaussian element X in a Hilbert space $\mathbb{H}$. Due to our Theorem 2.7 below, it holds under some technical conditions that

$$I\!P\big(x < \|X - \boldsymbol{a}\|^2 < x + \varepsilon\big) \le \frac{\mathtt{C}\,\varepsilon}{\|\Sigma\|_{\mathrm{Fr}}}$$

for an absolute constant $\mathtt{C}$.

2. Main results

Throughout the paper the following notation are used. We write $a \lesssim b$ ($a \gtrsim b$) if there exists some absolute constant C such that $a \le Cb$ ($a \ge Cb$ resp.). Similarly, $a \asymp b$ means that there exist c, C such that $c\,a \le b \le C\,a$. $I\!R$ (resp. $\mathbb{C}$) denotes the set of all real (resp. complex) numbers. We assume that all random variables are defined on common probability space $(\Omega, \mathfrak{F}, I\!P)$ and take values in a real separable Hilbert space $\mathbb{H}$ with a scalar product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$. If dimension of $\mathbb{H}$ is finite and equals p, we shall write $I\!R^p$ instead of $\mathbb{H}$. Let $I\!E$ be the mathematical expectation with respect to $I\!P$. We also denote by $\mathfrak{B}(\mathbb{H})$ the Borel σ-algebra.

For a self-adjoint operator $\mathbf{A}$ with eigenvalues $\lambda_k(\mathbf{A}), k \ge 1$, let us denote by $\|\mathbf{A}\|$ and $\|\mathbf{A}\|_1$ the operator and nuclear (Schatten-one) norm by $\|\mathbf{A}\| \stackrel{\mathrm{def}}{=} \sup_{\|x\|=1} \|\mathbf{A}x\|$ and

$$\|\mathbf{A}\|_1 \stackrel{\mathrm{def}}{=} \operatorname{tr} |\mathbf{A}| = \sum_{k=1}^{\infty} |\lambda_k(\mathbf{A})|.$$

We suppose below that $\mathbf{A}$ is a nuclear and $\|\mathbf{A}\|_1 < \infty$.

Let $\Sigma_{\boldsymbol{\xi}}$ be a covariance operator of an arbitrary Gaussian random element in $\mathbb{H}$. By $\{\lambda_{k\boldsymbol{\xi}}\}_{k\ge1}$ we denote the set of its eigenvalues arranged in the non-increasing order, i.e. $\lambda_{1\boldsymbol{\xi}} \ge \lambda_{1\boldsymbol{\xi}} \ge \ldots$, and let $\boldsymbol{\lambda}_{\boldsymbol{\xi}} \stackrel{\mathrm{def}}{=} \operatorname{diag}(\lambda_{j\boldsymbol{\xi}})_{j=1}^{\infty}$. Note that $\sum_{j=1}^{\infty} \lambda_{j\boldsymbol{\xi}} < \infty$. Introduce the following quantities

$$\Lambda_{k\boldsymbol{\xi}}^2 \stackrel{\mathrm{def}}{=} \sum_{j=k}^{\infty} \lambda_{j\boldsymbol{\xi}}^2, \quad k = 1, 2,$$

and

$$\varkappa(\Sigma_{\boldsymbol{\xi}}) = \begin{cases} \Lambda_{1\boldsymbol{\xi}}^{-1}, & \text{if } 3\lambda_{1,\boldsymbol{\xi}}^2 \le \Lambda_{1\boldsymbol{\xi}}^2, \\ (\lambda_{1\boldsymbol{\xi}}\Lambda_{2\boldsymbol{\xi}})^{-1/2}, & \text{if } 3\lambda_{1\boldsymbol{\xi}}^2 > \Lambda_{1\boldsymbol{\xi}}^2,\ 3\lambda_{2\boldsymbol{\xi}}^2 \le \Lambda_{2\boldsymbol{\xi}}^2, \\ (\lambda_{1\boldsymbol{\xi}}\lambda_{2\boldsymbol{\xi}})^{-1/2}, & \text{if } 3\lambda_{1\boldsymbol{\xi}}^2 > \Lambda_{1\boldsymbol{\xi}}^2,\ 3\lambda_{2\boldsymbol{\xi}}^2 > \Lambda_{2\boldsymbol{\xi}}^2. \end{cases} \tag{2.1}$$

It is easy to see that $\|\Sigma_{\boldsymbol{\xi}}\|_{\mathrm{Fr}} = \Lambda_{1\boldsymbol{\xi}}$. Moreover, it is straightforward to check that

$$\frac{0.9}{(\Lambda_{1\boldsymbol{\xi}}\Lambda_{2\boldsymbol{\xi}})^{1/2}} \le \varkappa(\Sigma_{\boldsymbol{\xi}}) \le \frac{1.8}{(\Lambda_{1\boldsymbol{\xi}}\Lambda_{2\boldsymbol{\xi}})^{1/2}}.$$

Hence, $\varkappa(\Sigma_{\boldsymbol{\xi}}) \asymp (\Lambda_{1\boldsymbol{\xi}}\Lambda_{2\boldsymbol{\xi}})^{-1/2}$ and therefore equivalent results can be formulated in terms of any of the quantities introduced. The following theorem is our main result.

Theorem 2.1. *Let $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ be Gaussian elements in $\mathbb{H}$ with zero mean and covariance operators $\Sigma_{\boldsymbol{\xi}}$ and $\Sigma_{\boldsymbol{\eta}}$ respectively. For any $\boldsymbol{a} \in \mathbb{H}$*

$$\sup_{x>0} |I\!P(\|\boldsymbol{\xi} - \boldsymbol{a}\| \le x) - I\!P(\|\boldsymbol{\eta}\| \le x)| \\ \lesssim \Big\{\varkappa(\Sigma_{\boldsymbol{\xi}}) + \varkappa(\Sigma_{\boldsymbol{\eta}})\Big\}\Big(\|\boldsymbol{\lambda}_{\boldsymbol{\xi}} - \boldsymbol{\lambda}_{\boldsymbol{\eta}}\|_1 + \|\boldsymbol{a}\|^2\Big). \tag{2.2}$$

The proof of Theorem 2.1 is given in Section 3.

We can see that the obtained bounds can be expressed in terms of the specific characteristics of the matrices $\Sigma_{\boldsymbol{\xi}}$ and $\Sigma_{\boldsymbol{\eta}}$ such as their operator and the Frobenius norms rather than the dimension p. Another nice feature of the obtained bounds is that they do not involve the inverse of $\Sigma_{\boldsymbol{\xi}}$ or $\Sigma_{\boldsymbol{\eta}}$. In other words, small or vanishing eigenvalues of $\Sigma_{\boldsymbol{\xi}}$ or $\Sigma_{\boldsymbol{\eta}}$ do not affect the obtained bounds in the contrary to the Pinsker bound. Similarly, only the squared norm $\|\boldsymbol{a}\|^2$ of the shift $\boldsymbol{a}$ shows up in the results, while the Pinsker bound involves $\|\Sigma_{\boldsymbol{\xi}}^{-1/2}\boldsymbol{a}\|$ which can be very large or infinite if $\Sigma_{\boldsymbol{\xi}}$ is not well conditioned.

The representation (2.1) mimics well the three typical situations: in the "large-dimensional case" with three or more significant eigenvalues $\lambda_{j\boldsymbol{\xi}}$, one can take $\varkappa(\Sigma_{\boldsymbol{\xi}}) = \|\Sigma_{\boldsymbol{\xi}}\|_{\mathrm{Fr}}^{-1} = \lambda_{1\boldsymbol{\xi}}^{-1}$. In the "two dimensional" case, when the sum $\Lambda_{k\boldsymbol{\xi}}^2$ is of the order $\lambda_{k\boldsymbol{\xi}}^2$ for $k = 1, 2$, the bound only depends on the product $(\lambda_{1\boldsymbol{\xi}}\lambda_{2\boldsymbol{\xi}})^{-1/2}$. The intermediate case of a spike model with one large eigenvalue $\lambda_{1\boldsymbol{\xi}}$ and many small eigenvalues $\lambda_{j\boldsymbol{\xi}}, j \ge 2$, the bound depends on $(\lambda_{1\boldsymbol{\xi}}\Lambda_{2\boldsymbol{\xi}})^{-1/2}$.

As it was mentioned earlier, the result of Theorem 2.2 may be equivalently formulated in a "unified" way in terms of $(\Lambda_{1\boldsymbol{\xi}}\Lambda_{2\boldsymbol{\xi}})^{-1/2}$. Moreover, we specify the bound (2.2) in the "high-dimensional" case which means at least three significantly positive eigenvalues of the matrices $\Sigma_{\boldsymbol{\xi}}$ and $\Sigma_{\boldsymbol{\eta}}$.

***Corollary* 2.2.** Let $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ be Gaussian elements in $\mathbb{H}$ with zero mean and covariance operators $\Sigma_{\boldsymbol{\xi}}$ and $\Sigma_{\boldsymbol{\eta}}$ respectively. Then for any $\boldsymbol{a} \in \mathbb{H}$

$$\sup_{x>0} |I\!P(\|\boldsymbol{\xi} - \boldsymbol{a}\| \le x) - I\!P(\|\boldsymbol{\eta}\| \le x)| \\ \lesssim \left(\frac{1}{(\Lambda_{1\boldsymbol{\xi}}\Lambda_{2\boldsymbol{\xi}})^{1/2}} + \frac{1}{(\Lambda_{1\boldsymbol{\eta}}\Lambda_{2\boldsymbol{\eta}})^{1/2}}\right)\Big(\|\boldsymbol{\lambda}_{\boldsymbol{\xi}} - \boldsymbol{\lambda}_{\boldsymbol{\eta}}\|_1 + \|\boldsymbol{a}\|^2\Big).$$

Moreover, assume that

$$3\|\Sigma_{\boldsymbol{\xi}}\|^2 \le \|\Sigma_{\boldsymbol{\xi}}\|_{\mathrm{Fr}}^2 \quad \text{and} \quad 3\|\Sigma_{\boldsymbol{\eta}}\|^2 \le \|\Sigma_{\boldsymbol{\eta}}\|_{\mathrm{Fr}}^2.$$

Then for any $\boldsymbol{a} \in \mathbb{H}$

$$\sup_{x>0} |I\!P(\|\boldsymbol{\xi} - \boldsymbol{a}\| \le x) - I\!P(\|\boldsymbol{\eta}\| \le x)| \\ \lesssim \left(\frac{1}{\|\Sigma_{\boldsymbol{\xi}}\|_{\mathrm{Fr}}} + \frac{1}{\|\Sigma_{\boldsymbol{\eta}}\|_{\mathrm{Fr}}}\right)\Big(\|\boldsymbol{\lambda}_{\boldsymbol{\xi}} - \boldsymbol{\lambda}_{\boldsymbol{\eta}}\|_1 + \|\boldsymbol{a}\|^2\Big).$$

We complement the result of Theorem 2.1 and Corollary 2.2 with several additional remarks. The first remark is that by the Weilandt–Hoffman inequality, $\|\boldsymbol{\lambda}_{\boldsymbol{\xi}} - \boldsymbol{\lambda}_{\boldsymbol{\eta}}\|_1 \le \|\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}\|_1$, see e.g. Markus (1964). This yields the bound in terms of the nuclear norm of the difference $\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}$, which may be more useful in a number of applications.

***Corollary* 2.3.** Under conditions of Theorem 2.1 we have

$$\sup_{x>0} \big|I\!P\big(\|\boldsymbol{\xi} - \boldsymbol{a}\| \le x\big) - I\!P(\|\boldsymbol{\eta}\| \le x)\big| \lesssim \Big\{\varkappa(\Sigma_{\boldsymbol{\xi}}) + \varkappa(\Sigma_{\boldsymbol{\eta}})\Big\}\Big(\|\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}\|_1 + \|\boldsymbol{a}\|^2\Big).$$

The right-hand-side of (2.2) does not change if we exchange $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ in Theorem 2.1 and its Corollaries hold for the balls with the same shift $\boldsymbol{a}$. In particular, the following corollary is true.

***Corollary* 2.4.** Under conditions of Theorem 2.1 we have

$$\sup_{x>0}\Big|I\!P(\|\boldsymbol{\xi} - \boldsymbol{a}\| \le x) - I\!P(\|\boldsymbol{\eta} - \boldsymbol{a}\| \le x)\Big| \lesssim \Big\{\varkappa(\Sigma_{\boldsymbol{\xi}}) + \varkappa(\Sigma_{\boldsymbol{\eta}})\Big\}\Big(\|\boldsymbol{\lambda}_{\boldsymbol{\xi}} - \boldsymbol{\lambda}_{\boldsymbol{\eta}}\|_1 + \|\boldsymbol{a}\|^2\Big).$$

The result of Theorem 2.1 may be also rewritten in terms of the operator norm

$$\|\Sigma_{\boldsymbol{\xi}}^{-1/2}\Sigma_{\boldsymbol{\eta}}\Sigma_{\boldsymbol{\xi}}^{-1/2} - I\|.$$

Indeed, using the inequality $\|\mathbf{AB}\|_1 \le \|\mathbf{A}\|_1\|\mathbf{B}\|$ we immediately obtain the following corollary.

***Corollary* 2.5.** Under conditions of Theorem 2.1 we have

$$\sup_{x>0} |I\!P(\|\boldsymbol{\xi} - \boldsymbol{a}\| \le x) - I\!P(\|\boldsymbol{\eta}\| \le x)| \\ \lesssim \Big\{\varkappa(\Sigma_{\boldsymbol{\xi}}) + \varkappa(\Sigma_{\boldsymbol{\eta}})\Big\}\Big(\operatorname{tr}\big(\Sigma_{\boldsymbol{\xi}}\big)\,\|\Sigma_{\boldsymbol{\xi}}^{-1/2}\Sigma_{\boldsymbol{\eta}}\Sigma_{\boldsymbol{\xi}}^{-1/2} - I\| + \|\boldsymbol{a}\|^2\Big).$$

We now discuss the origin of the value $\varkappa(\Sigma_{\boldsymbol{\xi}})$ which appears in the main theorem and its corollaries. Analysing the proof of Theorem 2.1 one may find out that it is necessary to get an upper bound for a probability density function (p.d.f.) $p_{\boldsymbol{\xi}}(x)$ (resp. $p_{\boldsymbol{\eta}}(x)$) of $\|\boldsymbol{\xi}\|^2$ (resp. $\|\boldsymbol{\eta}\|^2$) and the more general p.d.f. $p_{\boldsymbol{\xi}}(x, \boldsymbol{a})$ of $\|\boldsymbol{\xi} - \boldsymbol{a}\|^2$ for all $\boldsymbol{a} \in \mathbb{H}$. The same arguments remain true for $p_{\boldsymbol{\eta}}(x)$. The following theorem provides uniform bounds.

Theorem 2.6. *Let $\boldsymbol{\xi}$ be a Gaussian element in $\mathbb{H}$ with zero mean and covariance operator $\Sigma_{\boldsymbol{\xi}}$. Then it holds for any $\boldsymbol{a}$ that*

$$\sup_{x\geq 0} p_{\boldsymbol{\xi}}(x,\boldsymbol{a}) \lesssim \varkappa(\Sigma_{\boldsymbol{\xi}}) \tag{2.3}$$

with $\varkappa(\Sigma_{\boldsymbol{\xi}})$ from (2.1). In particular, $\varkappa(\Sigma_{\boldsymbol{\xi}}) \lesssim (\Lambda_{1\boldsymbol{\xi}}\Lambda_{2\boldsymbol{\xi}})^{-1/2}$.

The proof of this theorem will be given in Section 3.

Since $\boldsymbol{\xi} \stackrel{\mathrm{d}}{=} \sum_{j=1}^{\infty}\sqrt{\lambda_{j\boldsymbol{\xi}}}Z_j\mathbf{e}_{j\boldsymbol{\xi}}$, we obtain that $\|\boldsymbol{\xi}\|^2 \stackrel{\mathrm{d}}{=} \sum_{j=1}^{\infty}\lambda_{j\boldsymbol{\xi}}Z_j^2$. Here and in what follows $\{\mathbf{e}_{j\boldsymbol{\xi}}\}_{j=1}^{\infty}$ is the orthonormal basis formed by the eigenvectors of $\Sigma_{\boldsymbol{\xi}}$ corresponding to $\{\lambda_{1\boldsymbol{\xi}}\}_{j=1}^{\infty}$. In the case $\mathbb{H}=\mathbb{R}^p$, $\boldsymbol{a}=0, \Sigma_{\boldsymbol{\xi}} \asymp I$ one has that the distribution of $\|\boldsymbol{\xi}\|^2$ is close to standard χ^2 with p degrees of freedom and

$$\sup_{x\geq 0} p_{\boldsymbol{\xi}}(x,0) \asymp p^{-1/2}.$$

Hence, the bound (2.3) gives the right dependence on p because $\varkappa(\Sigma_{\boldsymbol{\xi}}) \asymp p^{-1/2}$. However, a lower bound for $\sup_{x\geq 0} p_{\boldsymbol{\xi}}(x,\boldsymbol{a})$ in the general case is still an open question. Another possible extension is a non-uniform upper bound for the p.d.f. of $\|\boldsymbol{\xi}-\boldsymbol{a}\|^2$. In this direction for any $\lambda > \lambda_{1\boldsymbol{\xi}}$ we can prove that

$$p_{\boldsymbol{\xi}}(x,\boldsymbol{a}) \leq \frac{\exp\bigl(-(x^{1/2}-\|\boldsymbol{a}\|)^2/(2\lambda)\bigr)}{\sqrt{2\lambda_{1\boldsymbol{\xi}}\lambda_{2\boldsymbol{\xi}}}} \prod_{j=3}^{\infty}(1-\lambda_{j\boldsymbol{\xi}}/\lambda)^{-1/2};$$

see Lemma B.1 and remark after it in the Appendix. It is still an open question whether it is possible to replace the $\lambda_{k\boldsymbol{\xi}}$'s in the denominator by $\Lambda_{k\boldsymbol{\xi}}$, $k=1,2$.

A direct corollary of Theorem 2.6 is the following theorem which states for a rather general situation a dimension-free anti-concentration inequality for the squared norm of a Gaussian element $\boldsymbol{\xi}$. In the "high dimensional situation", this anti-concentration bound only involves the Frobenius norm of $\Sigma_{\boldsymbol{\xi}}$.

Theorem 2.7 (ε-band of the squared norm of a Gaussian element)**.** *Let $\boldsymbol{\xi}$ be a Gaussian element in $\mathbb{H}$ with zero mean and a covariance operator $\Sigma_{\boldsymbol{\xi}}$. Then for arbitrary $\varepsilon > 0$, one has*

$$\sup_{x>0} \mathbb{P}(x < \|\boldsymbol{\xi}-\boldsymbol{a}\|^2 < x+\varepsilon) \lesssim \varkappa(\Sigma_{\boldsymbol{\xi}})\,\varepsilon \tag{2.4}$$

with $\varkappa(\Sigma_{\boldsymbol{\xi}})$ from (2.1). In particular, $\varkappa(\Sigma_{\boldsymbol{\xi}})$ can be replaced by $(\Lambda_{1\boldsymbol{\xi}}\,\Lambda_{2\boldsymbol{\xi}})^{-1/2}$.

We finish this section showing that the structure of estimates in Theorem 2.1 and Theorem 2.7 is the right one.

For simplicity, we consider the case of centred ball, i.e. $\boldsymbol{a} = 0$ and denote $\varkappa(\Sigma_{\boldsymbol{\xi}}, \Sigma_{\boldsymbol{\eta}}) \stackrel{\text{def}}{=} \max\{\varkappa(\Sigma_{\boldsymbol{\xi}}), \varkappa(\Sigma_{\boldsymbol{\eta}})\}$. We show in the special case that

$$\limsup \left(\sup_{x>0} \frac{|I\!P(\|\boldsymbol{\xi}\| \le x) - I\!P(\|\boldsymbol{\eta}\| \le x)|}{\varkappa(\Sigma_{\boldsymbol{\xi}}, \Sigma_{\boldsymbol{\eta}}) \|\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}\|_1} \right) \ge \mathtt{C}_1, \tag{2.5}$$

where $\mathtt{C}_1$ is some absolute positive constant and $\limsup$ is taken w.r.t. $\max(\lambda_{2\boldsymbol{\xi}}, \lambda_{2\boldsymbol{\eta}}) \downarrow 0$. Hence, in general it is impossible to obtain the upper bound in Theorem 2.1, such that it doesn't tend to infinity when $\lambda_{2\boldsymbol{\xi}}$ (or $\lambda_{2\boldsymbol{\eta}}$) tends to zero. To show (2.5) we construct the following example. Let $\boldsymbol{\xi}$ be a Gaussian vector in $I\!R^3$ with zero mean and covariance matrix $\Sigma_{\boldsymbol{\xi}} = \operatorname{diag}(\lambda_{1\boldsymbol{\xi}}, \lambda_{2\boldsymbol{\xi}}, \lambda_{3\boldsymbol{\xi}})$. Similarly, let $\boldsymbol{\eta}$ be a Gaussian vector with zero mean and covariance matrix $\Sigma_{\boldsymbol{\eta}} = \operatorname{diag}(\lambda_{1\boldsymbol{\eta}}, \lambda_{2\boldsymbol{\eta}}, \lambda_{3\boldsymbol{\eta}})$. Then

$$\sup_{x>0} |I\!P(\|\boldsymbol{\xi}\| \le x) - I\!P(\|\boldsymbol{\eta}\| \le x)| \ge \Big| I\!P(\|\boldsymbol{\xi}\| \le \sqrt{R}) - I\!P(\|\boldsymbol{\eta}\| \le \sqrt{R}) \Big|,$$

for some R which will be chosen later. Put

$$\mathcal{E}_1 \stackrel{\text{def}}{=} \Big\{ (x_1, x_2, x_3) \in I\!R^3 : \sum_{j=1}^{3} \lambda_{j\boldsymbol{\xi}} x_j^2 \le R \Big\}, \quad \mathcal{E}_2 \stackrel{\text{def}}{=} \Big\{ (x_1, x_2, x_3) \in I\!R^3 : \sum_{j=1}^{3} \lambda_{j\boldsymbol{\eta}} x_j^2 \le R \Big\}.$$

Let us take $\lambda_{1\boldsymbol{\xi}} = \lambda_{1\boldsymbol{\eta}}, \lambda_{2\boldsymbol{\xi}} = \lambda_{2\boldsymbol{\eta}}, \lambda_{3\boldsymbol{\eta}} = \lambda_{3\boldsymbol{\xi}}(1+\varepsilon)$ for some $0 < \varepsilon < 1$. This choice gives $\|\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}\|_1 = \varepsilon\lambda_3$ and $\varkappa(\Sigma_{\boldsymbol{\xi}}, \Sigma_{\boldsymbol{\eta}}) \asymp (\lambda_{1\boldsymbol{\xi}}\lambda_{2\boldsymbol{\xi}})^{-1/2}$. It is straightforward to check that

$$\begin{aligned} |I\!P(\|\boldsymbol{\xi}\| \le R) - I\!P(\|\boldsymbol{\eta}\| \le R)| &= \frac{1}{(2\pi)^{3/2}} \int_{\mathcal{E}_1 \setminus \mathcal{E}_2} \exp\Big(-\frac{x_1^2 + x_2^2 + x_3^2}{2}\Big)\, dx_1\, dx_2\, dx_3 \\ &\ge \frac{1}{(2\pi)^{3/2}} (|\mathcal{E}_1| - |\mathcal{E}_2|) \exp\Big[-\frac{R}{2}\Big(\frac{1}{\lambda_{1\boldsymbol{\xi}}} + \frac{1}{\lambda_{2\boldsymbol{\xi}}} + \frac{1}{\lambda_{3\boldsymbol{\xi}}}\Big)\Big], \end{aligned}$$

where $|\mathcal{E}_i|$ is a volume of the ellipsoid $|\mathcal{E}_i|$, $i = 1, 2$. Applying formula for the volume of an ellipsoid we obtain

$$|\mathcal{E}_1| - |\mathcal{E}_2| = \frac{4\pi R^{3/2} \|\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}\|_1}{3\sqrt{\lambda_{1\boldsymbol{\xi}}\lambda_{2\boldsymbol{\xi}}}\lambda_{3\boldsymbol{\xi}}^{3/2}\sqrt{1+\varepsilon}(1+\sqrt{1+\varepsilon})} > \frac{\pi \|\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}\|_1}{\sqrt{\lambda_{1\boldsymbol{\xi}}\lambda_{2\boldsymbol{\xi}}}} \left(\frac{R}{2\lambda_{3\boldsymbol{\xi}}}\right)^{3/2}.$$

We take $R = 2\lambda_{3\boldsymbol{\xi}}$. Then

$$\left(\frac{R}{2\lambda_{3\boldsymbol{\xi}}}\right)^{3/2} \exp\Big(-\frac{R}{2\lambda_{3\boldsymbol{\xi}}}\Big) \ge e^{-1} \ge \frac{1}{3}.$$

Hence,

$$\Big| I\!P(\|\boldsymbol{\xi}\| \le \sqrt{R}) - I\!P(\|\boldsymbol{\eta}\| \le \sqrt{R}) \Big| \ge \frac{\|\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}\|_1}{16\sqrt{\lambda_{1\boldsymbol{\xi}}\lambda_{2\boldsymbol{\xi}}}} \exp\Big[-\Big(\frac{\lambda_{3\boldsymbol{\xi}}}{\lambda_{1\boldsymbol{\xi}}} + \frac{\lambda_{3\boldsymbol{\xi}}}{\lambda_{2\boldsymbol{\xi}}}\Big)\Big] \ge \mathtt{C}_1 \frac{\|\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}\|_1}{\sqrt{\lambda_{1\boldsymbol{\xi}}\lambda_{2\boldsymbol{\xi}}}},$$

where $\mathtt{C}_1 \stackrel{\text{def}}{=} \exp(-2)/16$. From the last inequality we may conclude (2.5).

We now turn to the case $\mathbb{H} = I\!R^1$. Here, one may get a two-sided inequality. First, we derive an upper bound. Let $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ be normal variables with zero mean and variances $\lambda_{\boldsymbol{\xi}}$ and $\lambda_{\boldsymbol{\eta}}$ resp. Without loss of generality we may assume that $\lambda_{\boldsymbol{\xi}} < \lambda_{\boldsymbol{\eta}}$. Then

$$\sup_{x>0} |I\!P(\|\boldsymbol{\xi}\| \le x) - I\!P(\|\boldsymbol{\eta}\| \le x)| = \frac{2}{\sqrt{2\pi}} \sup_{x>0} \int_{x/\sqrt{\lambda_{\boldsymbol{\eta}}}}^{x/\sqrt{\lambda_{\boldsymbol{\xi}}}} e^{-y^2/2}\, dy$$
$$\le \frac{\|\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}\|_1}{\sqrt{\lambda_{\boldsymbol{\eta}}\lambda_{\boldsymbol{\xi}}}(\sqrt{\lambda}_{\boldsymbol{\xi}} + \sqrt{\lambda}_{\boldsymbol{\eta}})} \sup_{x>0} \bigl(x \exp\bigl(-x^2/(2\lambda_{\boldsymbol{\eta}})\bigr)\bigr) \lesssim \frac{\|\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}\|_1}{\lambda_{\boldsymbol{\xi}}}.$$

We also have the following lower bound:

$$\sup_{x>0} |I\!P(\|\boldsymbol{\xi}\| \le x) - I\!P(\|\boldsymbol{\eta}\| \le x)| = \frac{2}{\sqrt{2\pi}} \sup_{x>0} \int_{x/\sqrt{\lambda_{\boldsymbol{\eta}}}}^{x/\sqrt{\lambda_{\boldsymbol{\xi}}}} e^{-y^2/2}\, dy$$
$$\ge \frac{2\,\|\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}\|_1\, x_0 \exp\bigl(-x_0^2/(2\lambda_{\boldsymbol{\xi}})\bigr)}{\sqrt{2\pi}\sqrt{\lambda_{\boldsymbol{\eta}}\lambda_{\boldsymbol{\xi}}}(\sqrt{\lambda}_{\boldsymbol{\xi}} + \sqrt{\lambda}_{\boldsymbol{\eta}})} \gtrsim \frac{\|\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}\|_1}{\lambda_{\boldsymbol{\eta}}},$$

where $x_0 \stackrel{\text{def}}{=} \sqrt{\lambda_{\boldsymbol{\xi}}}$.

Similar arguments can be applied in the case of Theorem 2.7. The right-hand side of (2.4) essentially depends on the first two eigenvalues of $\Sigma_{\boldsymbol{\xi}}$. In general, it is impossible to get similar bounds of order $O(\varepsilon)$ with dependence on $\lambda_{1\boldsymbol{\xi}}$ only. In fact, let $\mathbb{H} = I\!R^2$ and $\lambda_{1\boldsymbol{\xi}} = 1$ and $\lambda_{2\boldsymbol{\xi}} = 0$ (i.e. $\boldsymbol{\xi}$ has the generate Gaussian distribution). Then for all positive $\varepsilon \le \log 2$ one has

$$\sup_{x>0} I\!P(x < \|\boldsymbol{\xi}\|^2 < x + \varepsilon) \ge \varepsilon^{1/2}/(2\sqrt{\pi}).$$

3. Proofs of the main results

This section collects the proofs of the main results.

Proof of Theorem 2.6. Let $\{\mathbf{e}_j\}_{j=1}^{\infty}$ be an orthonormal basis in $\mathbb{H}$ formed by the eigenvectors of $\Sigma_{\boldsymbol{\xi}}$ corresponding to eigenvalues $\{\lambda_{1\boldsymbol{\xi}}\}_{j=1}^{\infty}$. In what follows we omit the index $\boldsymbol{\xi}$ from the notation. Put $a_j \stackrel{\text{def}}{=} \langle a, \mathbf{e}_j \rangle$ and $\xi_j \stackrel{\text{def}}{=} \langle \boldsymbol{\xi}, \mathbf{e}_j \rangle$. Then ξ_j, $j \ge 1$, are independent $\mathcal{N}(0, \lambda_j)$ r.v. Let $g_j(x)$, $j \ge 1$, (resp. $f_j(t)$) be the p.d.f (resp. c.f.) of $(\xi_j - a_j)^2$. Moreover, let $g(m, x), m \ge 1$ (resp. $\overline{g}(m, x), m \ge 1$) be the p.d.f. of $\sum_{j=1}^{m}(\xi_j - a_j)^2$ (resp. $\sum_{j=m+1}^{\infty}(\xi_j - a_j)^2$). We also introduce the c.f. $f(m, t)$ of $g(m, x)$. As

$$p(x, \boldsymbol{a}) \le \int_{-\infty}^{\infty} g(m, y)\, \overline{g}(m, x - y)\, dy \le \sup_{x \ge 0} g(m, x), \tag{3.1}$$

we may restrict ourselves to the finite dimensional case only, e.d., $\mathbb{H} = I\!R^m$, where m is some large integer. Hence, in what follows we will assume that $\boldsymbol{\xi}$ is a m dimensional vector.

We separately consider three cases correspondingto the definition (2.1) of $\varkappa(\Sigma_{\boldsymbol{\xi}})$:

1. $3\lambda_1^2 \le \Lambda_1^2$;
2. $3\lambda_1^2 \ge \Lambda_1^2$, $3\lambda_2^2 \ge \Lambda_2^2$;
3. $3\lambda_1^2 \ge \Lambda_1^2$, $3\lambda_2^2 \le \Lambda_2^2$.

We start with the case 1. It is straightforward to check that

$$|f_j(t)| \le \frac{1}{(1+4\lambda_j^2 t^2)^{1/4}}, \qquad j = 1, \dots, m. \tag{3.2}$$

By the inverse formula

$$\begin{aligned} p(x, \boldsymbol{a}) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} \prod_{j=1}^{m} f_j(t)\, dt \\ &\le \frac{1}{2\pi} \int_{-\infty}^{+\infty} \prod_{j=1}^{m} |f_j(t)|\, dt \le \frac{1}{2\pi} \int_{-\infty}^{+\infty} \prod_{j=1}^{m} \frac{1}{(1+4\lambda_j^2 t^2)^{1/4}}\, dt. \end{aligned}$$

Now Lemma A.2 implies the desired bound.

The proof in case 2 follows from the Lemma B.1 in Section B. However, as long as a uniform bound is concerned, one can simplify the proof. Indeed, similarly to (3.1) one can show that for $m \ge 2$

$$g(m, x) \le \sup_{x \ge 0} g(2, x).$$

It is straightforward to check that

$$g_j(x) = \frac{1}{2\sqrt{2\pi x \lambda_j}} \Big[\exp\Big(-\frac{(x^{1/2} - a_j)^2}{2\lambda_j} \Big) + \exp\Big(-\frac{(x^{1/2} + a_j)^2}{2\lambda_j} \Big) \Big] \le \frac{1}{\sqrt{2\pi x \lambda_j}}. \tag{3.3}$$

This inequality implies that

$$g(2, x) = \int_0^x g_1(x-y) g_2(y)\, dy \le \frac{1}{2\pi\sqrt{\lambda_1 \lambda_2}} \int_0^x (x-y)^{-1/2} y^{-1/2}\, dy = \frac{1}{2\sqrt{\lambda_1 \lambda_2}}.$$

It remains to use the fact that the r.h.s. of the previous inequality can also be bounded by $\mathtt{C}/\sqrt{\Lambda_1 \Lambda_2}$.

Finally we consider the case 3. Define $w_j \stackrel{\text{def}}{=} \lambda_j^2/\Lambda_2^2$ for $j \ge 2$ and rewrite $\|\boldsymbol{\xi}\|^2$ as follows

$$\|\boldsymbol{\xi}\|^2 \stackrel{\text{d}}{=} (\xi_1 - a_1)^2 + \Lambda_2\, \eta,$$

where $\eta \stackrel{\text{def}}{=} \sum_{j=2}^{m} \sqrt{w_j}\,(Z_j - a'_j)^2$, $a'_j \stackrel{\text{def}}{=} a_j/\sqrt{\lambda_j}$, $Z_j \sim \mathcal{N}(0,1)$. Let p_η be the p.d.f. of random variable η. The bound (3.3) implies

$$g(m,x) \le \frac{1}{\sqrt{2\pi\lambda_1}} \int_0^{x/\Lambda_2} \frac{p_\eta(z)}{\sqrt{x - \Lambda_2\, z}}\, dz \le \frac{\mathtt{C}}{\sqrt{\lambda_1 \Lambda_2}} \sup_{x>0} \int_0^x \frac{p_\eta(z)}{\sqrt{x-z}}\, dz. \tag{3.4}$$

Note that $p_\eta(z)$ is bounded by some absolute constant. Indeed, by the inverse formula

$$p_\eta(z) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itz} \prod_{j=2}^{m} \overline{f}_j(t)\, dt,$$

where $\overline{f}_j(t)$ is the characteristic function of $\sqrt{w_j}\,(Z_j - a'_j)^2$ for $j = 2, \ldots, m$. Similarly to (3.2) we can bound $|\overline{f}_j(t)| \le (1 + 4\, w_j\, t^2)^{-1/4}$ and

$$p_\eta(z) \le \frac{1}{2\pi} \int_{-\infty}^{+\infty} \prod_{j=2}^{m} |\overline{f}_j(t)|\, dt \le \frac{1}{2\pi} \int_{-\infty}^{+\infty} \prod_{j=2}^{m} \frac{1}{(1 + 4\, w_j\, t^2)^{1/4}}\, dt\,.$$

In view of $\sum_{j \ge 2} w_j = 1$, Lemma A.2 implies

$$\sup_z p_\eta(z) \lesssim 1.$$

Combining this bound with (3.3) and (3.4) yields the upper bound of order $\bigl(\lambda_1 \Lambda_2\bigr)^{-1/2} \asymp \bigl(\Lambda_1 \Lambda_2\bigr)^{-1/2}$ in case (3). This completes the proof of the theorem. □

***Remark* 3.1.** We would like to remark that instead of Lemma A.2 one may also apply an alternative approach from Ulyanov (1987)[Lemma 5].

Proof of Theorems 2.1. We split the proof into two parts. In the first part we study the case $\boldsymbol{a} = 0$. The second part is devoted to the case $\Sigma_{\boldsymbol{\xi}} = \Sigma_{\boldsymbol{\eta}}$. The final estimate will follow by combining the two obtained estimates and the triangular inequality.

Case I: $\boldsymbol{a} = 0$.

Without loss of generality we may assume that $\Sigma_{\boldsymbol{\xi}} = \boldsymbol{\lambda}_{\boldsymbol{\xi}}, \Sigma_{\boldsymbol{\eta}} = \boldsymbol{\lambda}_{\boldsymbol{\eta}}$, where $\boldsymbol{\lambda}_{\boldsymbol{\xi}} \stackrel{\text{def}}{=} \operatorname{diag}(\lambda_{1\boldsymbol{\xi}}, \lambda_{2\boldsymbol{\xi}}, \ldots)$, $\boldsymbol{\lambda}_{\boldsymbol{\eta}} \stackrel{\text{def}}{=} \operatorname{diag}(\lambda_{1\boldsymbol{\eta}}, \lambda_{2\boldsymbol{\eta}}, \ldots)$ and $\lambda_{1\boldsymbol{\xi}} \ge \lambda_{1\boldsymbol{\xi}} \ge \ldots$ and similarly in decreasing order for $\lambda_{i\boldsymbol{\eta}}$'s.

Fix any $s : 0 \le s \le 1$. Let $Z(s)$ be a Gaussian random element in $\mathbb{H}$ with zero mean and diagonal covariance operator $\mathbf{V}(s)$:

$$\mathbf{V}(s) \stackrel{\text{def}}{=} s\boldsymbol{\lambda}_{\boldsymbol{\xi}} + (1-s)\boldsymbol{\lambda}_{\boldsymbol{\eta}}.$$

Denote by $f(t,s)$ (resp. $p(x,s)$) the characteristic function (resp. p.d.f.) of $\|Z(s)\|^2$. Let $\lambda_1(s) \geq \lambda_2(s) \geq \dots$ be the eigenvalues of $\mathbf{V}(s)$ and introduce the diagonal resolvent operator $\mathbf{G}(t,s) \stackrel{\text{def}}{=} (I - 2it\,\mathbf{V}(s))^{-1}$. Recall that $\|Z(s)\|^2 \stackrel{\text{d}}{=} \sum_{j=1}^n \lambda_j(s) Z_j^2$, where $Z_j, j \geq 1$, are i.i.d. $\mathcal{N}(0,1)$ r.v. Then it is straightforward to check that a characteristic function $f(t,s)$ of $\|Z(s)\|^2$ can be written as

$$f(t,s) \,=\, I\!E \exp\{it\|Z(s)\|^2\} = \exp\Big\{ -\frac{1}{2} \operatorname{tr} \log\big(I - 2it\mathbf{V}(s)\big)\Big\},$$

where for an operator $\mathbf{A}$ and the identity operator I we use notation

$$\log(I + \mathbf{A}) = \mathbf{A} \int_0^1 (I + y\mathbf{A})^{-1} dy.$$

It is well known, see e.g. Chung (2001)[§6.2, p. 168], that for a continues d.f. $F(x)$ with c.f. $f(t)$ we may write

$$F(x) = \frac{1}{2} + \frac{i}{2\pi} \lim_{T\to\infty} \text{V.P.} \int_{|t|\leq T} e^{-itx} f(t) \frac{dt}{t}.$$

Let us fix an arbitrary $x > 0$. Then

$$I\!P(\|\boldsymbol{\xi}\|^2 < x) - I\!P(\|\boldsymbol{\eta}\|^2 < x) = \frac{i}{2\pi} \lim_{T\to\infty} \text{V.P.} \int_{|t|\leq T} \frac{f(t,1) - f(t,0)}{t} e^{-itx}\, dt.$$

By the Newton-Leibnitz formula

$$f(t,1) - f(t,0) = \int_0^1 \frac{\partial f(t,s)}{\partial s}\, ds.$$

It is straightforward to check that

$$\frac{\partial f(t,s)/\partial s}{t} = i f(t,s) \operatorname{tr}\big\{(\boldsymbol{\lambda}_{\boldsymbol{\xi}} - \boldsymbol{\lambda}_{\boldsymbol{\eta}})\mathbf{G}(t,s)\big\}.$$

Changing the order of integration we get

$$\begin{aligned} I\!P(\|\boldsymbol{\xi}\|^2 < x) \;-\; & I\!P(\|\boldsymbol{\eta}\|^2 < x) \\ &= -\frac{1}{2\pi} \int_0^1 \int_{-\infty}^{\infty} \operatorname{tr}\left\{(\boldsymbol{\lambda}_{\boldsymbol{\xi}} - \boldsymbol{\lambda}_{\boldsymbol{\eta}})\mathbf{G}(t,s)\right\} f(t,s) e^{-itx}\, dt\, ds. \end{aligned} \tag{3.5}$$

Since $\mathbf{G}(t,s)$ is the diagonal operator with $(1 - 2it\lambda_j(s))^{-1}$ on the diagonal, we may fix s and j and consider the following quantity

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} (1 - 2it\lambda_j(s))^{-1} f(t,s) e^{-itx}\, dt.$$

Let $\overline{Z}_j(s), j \geq 1$ be independent exponentially distributed r.v. with parameter $1/(2\lambda_j(s))$ (we write $\mathrm{Exp}(2\lambda_j(s))$), which are also independent of $Z_k, k \geq 1$. Then

$$I\!\!E e^{it\overline{Z}_j(s)} = (1 - 2it\lambda_j(s))^{-1}.$$

Moreover, $(1 - 2it\lambda_j(s))^{-1} f(t,s)$ is the characteristic function of $\overline{Z}_j(s) + \|Z(s)\|^2$. Let $p_j(x,s)$ be the corresponding p.d.f. Then

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} (1 - 2it\lambda_j(s))^{-1} f(t,s) e^{-itx}\, dt = p_j(x,s).$$

Denote by $\mathbf{P}(x,s)$ a diagonal operator with $p_j(x,s)$ on the main diagonal. Then we may conclude that

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{tr}\left\{(\boldsymbol{\lambda}_{\boldsymbol{\xi}} - \boldsymbol{\lambda}_{\boldsymbol{\eta}})\mathbf{G}(t,s)\right\} f(t,s) e^{-itx}\, dt = \mathrm{tr}\left\{(\boldsymbol{\lambda}_{\boldsymbol{\xi}} - \boldsymbol{\lambda}_{\boldsymbol{\eta}})\mathbf{P}(x,s)\right\}.$$

It is clear that the absolute value of the last term is bounded above by

$$\|\boldsymbol{\lambda}_{\boldsymbol{\xi}} - \boldsymbol{\lambda}_{\boldsymbol{\eta}}\|_1 \max_j \sup_{x \geq 0} p_j(x,s)$$

and we need to bound uniformly each $p_j(x,s)$. For any j:

$$p_j(x,s) = \int_{-\infty}^{\infty} p(y,s)\overline{p}_j(x-y,s)\, dy \leq \sup_{x \geq 0} p(x,s),$$

where $\overline{p}_j(x,s)$ is the p.d.f. of $\overline{Z}_j(s)$. Applying Theorem 2.6 we obtain

$$\sup_{x \geq 0} p(x,s) \lesssim \varkappa(\Sigma(s)),$$

where $\varkappa(\Sigma(s))$ is from (2.1). It remains to integrate over s to obtain

$$\sup_{x>0} \Big| I\!\!P(\|\boldsymbol{\xi}\|^2 < x) - I\!\!P(\|\boldsymbol{\eta}\|^2 < x) \Big| \leq \Big\{ \varkappa(\Sigma_{\boldsymbol{\xi}}) + \varkappa(\Sigma_{\boldsymbol{\eta}}) \Big\} \big\| \boldsymbol{\lambda}_{\boldsymbol{\xi}} - \boldsymbol{\lambda}_{\boldsymbol{\eta}} \big\|_1.$$

Case II: $\Sigma_{\boldsymbol{\xi}} = \Sigma_{\boldsymbol{\eta}}$ *and* $\boldsymbol{a} \neq 0$.

We may rotate $\boldsymbol{\xi}$ such that $\Sigma_{\boldsymbol{\xi}} = \Lambda_{\boldsymbol{\xi}}$. Then we have to replace $\boldsymbol{a}$ by appropriate $\overline{\boldsymbol{a}}$, but $\|\boldsymbol{a}\| = \|\overline{\boldsymbol{a}}\|$. Fix any $s : 0 \leq s \leq 1$. Let $\overline{\boldsymbol{a}}(s) \stackrel{\mathrm{def}}{=} \overline{\boldsymbol{a}}\sqrt{s}$. Introduce the diagonal operator $\mathbf{G}(t) \stackrel{\mathrm{def}}{=} (I - 2it\,\Lambda_{\boldsymbol{\xi}})^{-1}$. It is straightforward to check that a characteristic function $f(t, \overline{\boldsymbol{a}}(s))$ of $\|\boldsymbol{\xi} - \overline{\boldsymbol{a}}(s)\|^2$ can be written as

$$\begin{aligned} f(t, \overline{\boldsymbol{a}}(s)) &= I\!\!E \exp\{it\|\boldsymbol{\xi} - \overline{\boldsymbol{a}}(s)\|^2\} \\ &= \exp\left\{ it \left(s\|\overline{\boldsymbol{a}}\|^2 + s\langle \mathbf{G}(t)\overline{\boldsymbol{a}}, \overline{\boldsymbol{a}} \rangle - \frac{1}{2it} \mathrm{tr}\log\big(I - 2it\Lambda_{\boldsymbol{\xi}}\big) \right) \right\}. \end{aligned}$$

Repeating the arguments from the proof of Theorem 2.1 we obtain (compare with (3.5))

$$\begin{aligned}&I\!P(\|\boldsymbol{\xi}-\boldsymbol{a}\|^2<x)-I\!P(\|\boldsymbol{\xi}\|^2<x)\\&\quad=-\frac{1}{2\pi}\int_0^1\int_{-\infty}^{\infty}\Big[\|\overline{\boldsymbol{a}}\|^2+\langle\mathbf{G}(t)\overline{\boldsymbol{a}},\overline{\boldsymbol{a}}\rangle\Big]f(t,\overline{\boldsymbol{a}}(s))e^{-itx}\,dt\,ds.\end{aligned}$$

Moreover, we may rewrite the last equation as follows

$$\begin{aligned}&I\!P(\|\boldsymbol{\xi}-\boldsymbol{a}\|^2<x)-I\!P(\|\boldsymbol{\xi}\|^2<x)\\&\quad=-\|\boldsymbol{a}\|^2\int_0^1 p(x,\overline{\boldsymbol{a}}(s))\,ds-\sum_{j=1}^{\infty}[\overline{a}_j]^2\int_0^1 p_j(x,\overline{\boldsymbol{a}}(s))\,ds,\end{aligned}$$

where $p(x,\overline{\boldsymbol{a}}(s))$, $p_j(x,\overline{\boldsymbol{a}}(s))$ are p.d.f of $\|\boldsymbol{\xi}-\overline{\boldsymbol{a}}(s)\|^2$ and $\overline{Z}_j+\|\boldsymbol{\xi}-\overline{\boldsymbol{a}}(s)\|^2$ resp. Here $\overline{Z}_j$ is a random variable with exponential distribution $\mathrm{Exp}(2\lambda_{j\boldsymbol{\xi}})$. It remains to apply Theorem 2.6 and integrate over s. □

4. Acknowledgements

We would like to thank the Associate Editor and the Reviewer for helpful comments and suggestions.

F. Götze was supported by the German Research Foundation (DFG) through the Collaborative Research Center 1283: "Taming uncertainty and profiting from randomness and low regularity in analysis, stochastics and their applications". A. Naumov was supported RFBR N 16-31-00005 and President's of Russian Federation Grant for young scientists N 4596.2016.1. V. Spokoiny was supported by the Russian Science Foundation (project no. 14 50 00150). Financial support by the German Research Foundation (DFG) through the Collaborative Research Center 1294 "Data Assimilation – The Seamless Integration of Data and Models" is gratefully acknowledged.

Appendix A: Technical results

Lemma A.1. *It holds*

$$\sup_{0<a\le 1} a\int_0^{\infty}\frac{1}{(1+t^2)^{a+1/2}}dt\le \mathtt{C}, \tag{A.1}$$

and

$$\sup_{a\ge 1} a^{1/2}\int_0^{\infty}\frac{1}{(1+t^2)^{a+1/2}}dt\le \mathtt{C}. \tag{A.2}$$

Proof. Define

$$H(a) \stackrel{\text{def}}{=} \int_0^\infty \frac{1}{(1+t^2)^{a+1/2}}\, dt. \tag{A.3}$$

Obviously, $H(a)$ monotonously decreases in a. Integration by parts implies for $a > 0$

$$\begin{aligned}\int_0^\infty \frac{t^2}{(1+t^2)^{a+3/2}}\, dt &= -\frac{1}{2a+1}\int_0^\infty t\, d\left(\frac{1}{(1+t^2)^{a+1/2}}\right) \\ &= \frac{1}{2a+1}\int_0^\infty \frac{1}{(1+t^2)^{a+1/2}}\, dt = \frac{H(a)}{2a+1}\,.\end{aligned}$$

At the same time, for $a > 0$

$$\int_0^\infty \frac{t^2}{(1+t^2)^{a+3/2}}\, dt = \int_0^\infty \frac{1+t^2}{(1+t^2)^{a+3/2}}\, dt - \int_0^\infty \frac{1}{(1+t^2)^{a+3/2}}\, dt = H(a) - H(a+1)\,.$$

This implies a recurrent relation

$$H(a+1) = \frac{a}{a+1/2} H(a)\,.$$

For $a \in [0,1]$, it implies

$$aH(a) = (a+1/2)H(a+1) \le \frac{3}{2}\, H(1) = \mathtt{C}$$

and (A.1) follows. For $a = a_0 + k$ with $a_0 \in \bigl[1,2\bigr]$ and an integer $k \ge 0$, we use that

$$\begin{aligned}\sqrt{a}\, H(a) &= \sqrt{a}\, \frac{(a-1)(a-2)\ldots a_0}{(a-1/2)(a-3/2)\ldots(a_0+1/2)} H(a_0) \\ &= \frac{\sqrt{a(a-1)}}{a-1/2}\, \frac{\sqrt{(a-1)(a-2)}}{a-3/2} \ldots \frac{\sqrt{(a_0+1)a_0}}{a_0+1/2}\, \sqrt{a_0}\, H(a_0) \le \sqrt{2}\, H(1) = \mathtt{C}.\end{aligned}$$

This proves (A.2). □

Lemma A.2. *Let* $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_p$ *and*

$$3\lambda_1^2 \le \Lambda^2 \stackrel{\text{def}}{=} \sum_{j=1}^p \lambda_j^2\,.$$

Define

$$h_j(t) \stackrel{\text{def}}{=} \frac{1}{(1+\lambda_j^2 t^2)^{1/4}}\,, \qquad j = 1, \ldots, p.$$

Then it holds

$$\int_0^\infty \prod_{j=1}^p h_j(t)\,dt \lesssim \frac{1}{\Lambda}\,.$$

Proof. Let q_j be a set of positive numbers with $q_j \geq 3$ and $\sum_j q_j^{-1} = 1$. A specific choice will be given later. By the Hölder inequality

$$\int_0^\infty \prod_{j=1}^p h_j(t)\,dt \leq \prod_{j=1}^p \Big(\int_0^\infty |h_j(t)|^{q_j}\,dt\Big)^{1/q_j}\,.$$

Further, for each j, by the change of variable $\lambda_j t = u$

$$\int_0^\infty |h_j(t)|^{q_j}\,dt = \int_0^\infty \frac{dt}{(1+\lambda_j^2 t^2)^{q_j/4}} = \lambda_j^{-1}\int_0^\infty \frac{du}{(1+u^2)^{q_j/4}} = \lambda_j^{-1} H(q_j/4-1/2)$$

with $H(\cdot)$ from (A.3). Therefore, by (A.2) of Lemma A.1 in view of $q_j/4 - 1/2 \geq 1/4$

$$\int_0^\infty \prod_{j=1}^p h_j(t)\,dt \leq \prod_{j=1}^p \Big(\frac{1}{\lambda_j H(q_j/4-1/2)}\Big)^{1/q_j} \lesssim \prod_{j=1}^p \Big(\frac{1}{\lambda_j\sqrt{q_j/4-1/2}}\Big)^{1/q_j}\,. \quad \text{(A.4)}$$

Now we fix q_j by the condition

$$\lambda_j^2(q_j/4-1/2) = \tau,$$

where the constant τ is determined by $\sum_{j=1}^p q_j^{-1} = 1$. This yields

$$\frac{1}{q_j} = \frac{\lambda_j^2}{4\tau + 2\lambda_j^2}\,, \qquad \sum_{j=1}^p \frac{\lambda_j^2}{4\tau+2\lambda_j^2} = 1,$$

and obviously $\tau \leq \Lambda^2/4$ and $\tau + \lambda_1^2/2 \geq \Lambda^2/4$. The condition $3\lambda_1^2 \leq \Lambda^2$ implies

$$q_j = \frac{4\tau}{\lambda_j^2} + 2 \geq \frac{\Lambda^2 - 2\lambda_1^2}{\lambda_1^2} + 2 \geq 3, \qquad j \leq p.$$

Also

$$\tau \geq \frac{1}{4}\big(\Lambda^2 - 2\lambda_1^2\big) \geq \frac{1}{4}\Big(\Lambda^2 - \frac{2\Lambda^2}{3}\Big) \gtrsim \Lambda^2\,.$$

Now it follows from (A.4) that

$$\int_0^\infty \prod_{j=1}^p h_j(t)\,dt \lesssim \left(\frac{1}{\sqrt{\tau}}\right)^{q_1^{-1}+\ldots+q_p^{-1}} \lesssim \frac{1}{\Lambda}$$

as required. □

Appendix B: A non-uniform bound for the density of a weighted non-central χ^2 distribution

Lemma B.1. *Let $\boldsymbol{\xi}$ be a Gaussian element in $\mathbb{H}$ with zero mean and covariance operator $\Sigma_{\boldsymbol{\xi}}$. For any $\boldsymbol{a} \in \mathbb{H}$ and all $\lambda > \lambda_{1\boldsymbol{\xi}}$*

$$p_{\boldsymbol{\xi}}(x, \boldsymbol{a}) \le \frac{\exp\bigl(-(x^{1/2} - \|\boldsymbol{a}\|)^2/(2\,\lambda)\bigr)}{\sqrt{2\lambda_{1\boldsymbol{\xi}}\lambda_{2\boldsymbol{\xi}}}} \prod_{j=3}^\infty (1 - \lambda_{j\boldsymbol{\xi}}/\lambda)^{-1/2}. \tag{B.1}$$

***Remark* B.1.** The infinite product in the r.h.s. of (B.1) is convergent. Indeed, taking logarithm and using $\log(1+x) \ge x/(x+1)$ for $x > -1$ we obtain

$$0 < -\frac{1}{2}\log \prod_{j=3}^\infty (1 - \lambda_{j\boldsymbol{\xi}}/\lambda) \le \frac{1}{2(\lambda - \lambda_{1\boldsymbol{\xi}})} \sum_{j=3}^\infty \lambda_{j\boldsymbol{\xi}} < \infty,$$

where we also used the fact that $\Sigma_{\boldsymbol{\xi}}$ is a nuclear and $\|\Sigma_{\boldsymbol{\xi}}\|_1 < \infty$. Taking $\lambda = \|\Sigma_{\boldsymbol{\xi}}\|_1$ we get $\prod_{j=3}^\infty (1 - \lambda_{j\boldsymbol{\xi}}/\lambda)^{-1/2} \le \sqrt{e}$.

Proof. We will use the notation from the proof of Theorem 2.6. We rewrite $g_j(x)$ as follows

$$g_j(x) = \frac{1}{\sqrt{2\pi x \lambda_j}} d_j(x),$$

where

$$d_j(x) \stackrel{\text{def}}{=} d_j(\lambda_j, x) \stackrel{\text{def}}{=} \frac{1}{2}\Bigl[\exp\bigl(-(x^{1/2} - a_j)^2/(2\lambda_j)\bigr) + \exp\bigl(-(x^{1/2} + a_j)^2/(2\lambda_j)\bigr)\Bigr].$$

It is straightforward to check that for $a \ge b \ge 0$

$$((a-b)^{1/2} - c)^2 + (b^{1/2} - d)^2 \ge (a^{1/2} - (c^2 + d^2)^{1/2})^2,$$

and

$$d_j(x) \le \exp\bigl(-(x^{1/2} - |a_j|)^2/(2\lambda_j)\bigr).$$

We have for all $j = 1, 2, \dots$ and any $\lambda > \lambda_1$

$$g_j(x) \le \frac{1}{\sqrt{2\pi x \lambda_j}} \exp\bigl(-(x^{1/2} - |a_j|)^2/(2\lambda)\bigr) d_j(\lambda\lambda_j/(\lambda - \lambda_j), x). \tag{B.2}$$

Moreover,

$$(2\pi x)^{-1/2}(\lambda - \lambda_j)^{1/2}/(\lambda\lambda_j)^{1/2}\, d_j(\lambda\lambda_j/(\lambda - \lambda_j), x) \tag{B.3}$$

is the density function of $\bigl(\sqrt{\lambda/(\lambda - \lambda_j)}\,\xi_j - a_j\bigr)^2$. These inequalities imply

$$\begin{aligned} g(2,x) &= \int_0^x g_1(x-y) g_2(y)\, dy \\ &\le \frac{1}{2\pi\sqrt{\lambda_1\lambda_2}} \exp\bigl(-(x^{1/2} - (a_1^2 + a_2^2)^{1/2})^2/(2\lambda)\bigr) \int_0^x (x-y)^{-1/2} y^{-1/2}\, dy \\ &= \frac{1}{2\sqrt{\lambda_1\lambda_2}} \exp\bigl(-(x^{1/2} - (a_1^2 + a_2^2)^{1/2})^2/(2\lambda)\bigr). \end{aligned}$$

Similarly, applying the last inequality, (B.2) and (B.3) we obtain

$$\begin{aligned} g(3,x) &= \int_0^x g(2, x-y) g_3(y)\, dy \\ &\le \frac{1}{2\sqrt{\lambda_1\lambda_2}\sqrt{2\pi\lambda_3}} \exp\bigl(-(x^{1/2} - (a_1^2 + a_2^2 + a_3^2)^{1/2})^2/(2\lambda)\bigr) \\ &\quad \times \int_0^x \frac{d_j(\lambda\lambda_3/(\lambda - \lambda_3), y)}{y^{1/2}}\, dy \\ &\le \frac{1}{2\sqrt{\lambda_1\lambda_2}} \exp\bigl(-(x^{1/2} - (a_1^2 + a_2^2 + a_3^2)^{1/2})^2/(2\lambda)\bigr) \left(1 - \frac{\lambda_3}{\lambda}\right)^{-1/2}. \end{aligned}$$

By induction we get

$$g(m,x) \le \frac{1}{2\sqrt{\lambda_1\lambda_2}} \exp\Bigl(-\frac{(x^{1/2} - (a_1^2 + \ldots + a_m^2)^{1/2})^2}{2\lambda}\Bigr) \prod_{j=3}^{m} \left(1 - \frac{\lambda_j}{\lambda}\right)^{-1/2}. \tag{B.4}$$

Now take an arbitrary $\varepsilon > 0$ and any integer $m > 0$. Let $0 < \mu < 1/(2\lambda_j)$ for all $j \ge m+1$. Without loss of generality we assume that at least two $\lambda_j, j \ge m+1$, are non-zero. Otherwise the arguments are simpler. By Markov's inequality we obtain

$$I\!P\left(\sum_{j=m+1}^{\infty} \xi_j^2 \ge \varepsilon^2\right) \le e^{-\mu\varepsilon^2} \prod_{j=m+1}^{\infty} I\!E e^{\mu\xi_j^2} = e^{-\mu\varepsilon^2} \prod_{j=m+1}^{\infty} \frac{1}{\sqrt{1 - 2\mu\lambda_j}}.$$

Choosing $\mu \stackrel{\text{def}}{=} 1/(2\sum_{j=m+1}^{\infty}\lambda_j)$ we get

$$I\!P\left(\sum_{j=m+1}^{\infty}\xi_j^2 \geq \varepsilon^2\right) \leq 2\exp\left\{-\varepsilon^2\left(2\sum_{j=m+1}^{\infty}\lambda_j\right)^{-1}\right\}.$$

Hence, there exists $M_1 = M_1(\varepsilon)$ such that for all $m \geq M$

$$I\!P\left(\sum_{j=m+1}^{\infty}\xi_j^2 \geq \varepsilon^2\right) \leq \varepsilon^2.$$

For any $m \geq 1$ we obtain

$$\sum_{j=m+1}^{\infty}(\xi_j - a_j)^2 \leq 2\left(\sum_{j=m+1}^{\infty}\xi_j^2 + \sum_{j=m+1}^{\infty}a_j^2\right).$$

We choose $M_2 = M_2(\varepsilon)$ such that $\sum_{j=m+1}^{\infty}a_j^2 \leq \varepsilon^2$. Hence, for $M = M_1 + M_2$ we obtain the following inequality

$$I\!P(x-\varepsilon \leq \|\boldsymbol{\xi}-\boldsymbol{a}\|^2 \leq x+\varepsilon) \leq I\!P\left(x-\varepsilon-4\varepsilon^2 \leq \sum_{j=1}^{m}(\xi_j-a_j)^2 \leq x+\varepsilon\right) + \varepsilon^2.$$

The last inequality implies

$$I\!P(x-\varepsilon \leq \|\boldsymbol{\xi}-a\|^2 \leq x+\varepsilon) \leq \varepsilon^2 + (2\varepsilon+4\varepsilon^2)\sup_{y\in T(\varepsilon,x)}g(m,y),$$

where $T(\varepsilon,x) \stackrel{\text{def}}{=} \{y\in\mathbb{R}^1 : x-\varepsilon-4\varepsilon^2 \leq y \leq x+\varepsilon\}$. Dividing the right-hand side of the previous inequality by ε we obtain (B.1) from (B.4) as ε tends to 0. □

References

K. Ball. The reverse isoperimetric problem for Gaussian measure. *Discrete Comput. Geom.*, 10(4):411–420, 1993. ISSN 0179-5376. . URL http://dx.doi.org/10.1007/BF02573986.

E. Belitser. On coverage and local radial rates of credible sets. *Ann. Statist.*, 45(3): 1124–1151, 06 2017. . URL http://dx.doi.org/10.1214/16-AOS1477.

V. Bentkus. On the dependence of the Berry-Esseen bound on dimension. *J. Statist. Plann. Inference*, 113(2):385–402, 2003. ISSN 0378-3758. . URL http://dx.doi.org/10.1016/S0378-3758(02)00094-0.

V. Bentkus. A Lyapunov-type bound in R^d. *Probab. Theory and Appl.*, 49(2):311–323, 2005. . URL http://dx.doi.org/10.1137/S0040585X97981123.

V. Bogachev. *Gaussian Measures*, volume 62 of *Mathematical Surveys and Monographs*. American Mathematical Society, USA, 1998.

D. Bontemps. Bernstein-von Mises theorems for Gaussian regression with increasing number of regressors. *Ann. Statist.*, 39(5):2557–2584, 2011. ISSN 0090-5364. . URL http://dx.doi.org/10.1214/11-AOS912.

I. Castillo. A semiparametric Bernstein - von Mises theorem for Gaussian process priors. *Probability Theory and Related Fields*, 152:53–99, 2012. ISSN 0178-8051. URL http://dx.doi.org/10.1007/s00440-010-0316-5. 10.1007/s00440-010-0316-5.

I. Castillo and R. Nickl. Nonparametric Bernstein–von Mises theorems in Gaussian white noise. *Ann. Statist.*, 41(4):1999–2028, 2013. ISSN 0090-5364. . URL http://dx.doi.org/10.1214/13-AOS1133.

G. Christoph, Y. Prokhorov, and V. Ulyanov. On distribution of quadratic forms in gaussian random variables. *Theory of Probability and its Applications*, 40(2):250–260, 1996. ISSN 0040-585X; 1095-7219. . URL https://doi.org/10.1137/1140028.

K. Chung. *A course in probability theory*. Academic Press, Inc., San Diego, CA, third edition, 2001. ISBN 0-12-174151-6.

F. Götze and A. Zaitsev. New applications of Arak's inequalities to the Littlewood-Offord problem, 2016. URL http://arxiv.org/abs/arXiv:1611.00831. arXiv:1611.00831.

N. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions, Vol.1, Second Edition*. John Wiley and Sons, Inc., New York, 1994. ISBN 978-0-471-58495-7.

I. Johnstone. High dimensional Bernstein–von Mises: simple examples. In *Borrowing strength: theory powering applications—a Festschrift for Lawrence D. Brown*, volume 6 of *Inst. Math. Stat. Collect.*, pages 87–98. Inst. Math. Statist., Beachwood, OH, 2010.

M. Ledoux and M. Talagrand. *Probability in Banach Spaces. Isoperimetry and Processes*, volume 23 of *A Series of Modern Surveys in Mathematics*. Springer–Verlag, Berlin, Heidelberg, 2002. ISBN 978-3-642-20211-7.

W. Li and Q. Shao. Gaussian processes: Inequalities, small ball probabilities and applications. In *STOCHASTIC PROCESSES: THEORY AND METHODS*, volume 19 of *Handbook of Statistics*, pages 533–598. Elsevier, New York, 2001.

M. Lifshits. *Lectures on Gaussian Processes*. Springer Publishing Company, Incorporated, 1st edition, 2012. ISBN 978-3-642-24938-9.

A. Markus. The eigen- and singular values of the sum and product of linear operators. *Russian Mathematical Surveys*, 19(4):91–120, 1964. ISSN 0042-1316.

M. Panov and V. Spokoiny. Finite Sample Bernstein - von Mises Theorem for Semiparametric Problems. *Bayesian Analysis*, 10(3):665–710, 09 2015. . URL http://dx.doi.org/10.1214/14-BA926.

Y. Prokhorov and V. Ulyanov. Some approximation problems in statistics and probability. In *Limit theorems in probability, statistics and number theory*, volume 42 of *Springer Proc. Math. Stat.*, pages 235–249. Springer, Heidelberg, 2013. . URL http://dx.doi.org/10.1007/978-3-642-36068-8_11.

M. Rudelson and R. Vershynin. The Littlewood-Offord problem and invertibility of random matrices. *Adv. Math.*, 218(2):600–633, 2008. ISSN 0001-8708. . URL

http://dx.doi.org/10.1016/j.aim.2008.01.010.

V. Sazonov. *Normal approximation—some recent advances*, volume 879 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin-New York, 1981. ISBN 3-540-10863-7.

V. Spokoiny. Penalized maximum likelihood estimation and effective dimension. *AIHP*, 53(1):389–429, 2017. . arXiv:1205.0498.

V. Spokoiny and M. Zhilova. Bootstrap confidence sets under model misspecification. *Ann. Statist.*, 43(6):2653–2675, 2015. ISSN 0090-5364. . URL http://dx.doi.org/10.1214/15-AOS1355. arXiv:1410.0347.

J. Tropp. User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.*, 12(4):389–434, 2012. ISSN 1615-3375. . URL http://dx.doi.org/10.1007/s10208-011-9099-z.

V. Ulyanov. Asymptotic expansions for distributions of sums of independent random variables in H. *Theory of Probability and its Applications*, 31(1):25–39, 1987. ISSN 0040-585X. . URL https://doi.org/10.1137/1131002.

V. Yurinsky. *Sums and Gaussian Vectors*. Springer–Verlag, Berlin, Heidelberg, 1995. ISBN 3-540-60311-5.

IRTG 1792 Discussion Paper Series 2018

For a complete list of Discussion Papers published, please visit irtg1792.hu-berlin.de.

001 "Data Driven Value-at-Risk Forecasting using a SVR-GARCH-KDE Hybrid" by Marius Lux, Wolfgang Karl Härdle and Stefan Lessmann, January 2018.

002 "Nonparametric Variable Selection and Its Application to Additive Models" by Zheng-Hui Feng, Lu Lin, Ruo-Qing Zhu asnd Li-Xing Zhu, January 2018.

003 "Systemic Risk in Global Volatility Spillover Networks: Evidence from Option-implied Volatility Indices " by Zihui Yang and Yinggang Zhou, January 2018.

004 "Pricing Cryptocurrency options: the case of CRIX and Bitcoin" by Cathy YH Chen, Wolfgang Karl Härdle, Ai Jun Hou and Weining Wang, January 2018.

005 "Testing for bubbles in cryptocurrencies with time-varying volatility" by Christian M. Hafner, January 2018.

006 "A Note on Cryptocurrencies and Currency Competition" by Anna Almosova, January 2018.

007 "Knowing me, knowing you: inventor mobility and the formation of technology-oriented alliances" by Stefan Wagner and Martin C. Goossen, February 2018.

008 "A Monetary Model of Blockchain" by Anna Almosova, February 2018.

009 "Deregulated day-ahead electricity markets in Southeast Europe: Price forecasting and comparative structural analysis" by Antanina Hryshchuk, Stefan Lessmann, February 2018.

010 "How Sensitive are Tail-related Risk Measures in a Contamination Neighbourhood?" by Wolfgang Karl Härdle, Chengxiu Ling, February 2018.

011 "How to Measure a Performance of a Collaborative Research Centre" by Alona Zharova, Janine Tellinger-Rice, Wolfgang Karl Härdle, February 2018.

012 "Targeting customers for profit: An ensemble learning framework to support marketing decision making" by Stefan Lessmann, Kristof Coussement, Koen W. De Bock, Johannes Haupt, February 2018.

013 "Improving Crime Count Forecasts Using Twitter and Taxi Data" by Lara Vomfell, Wolfgang Karl Härdle, Stefan Lessmann, February 2018.

014 "Price Discovery on Bitcoin Markets" by Paolo Pagnottoni, Dirk G. Baur, Thomas Dimpfl, March 2018.

015 "Bitcoin is not the New Gold - A Comparison of Volatility, Correlation, and Portfolio Performance" by Tony Klein, Hien Pham Thu, Thomas Walther, March 2018.

016 "Time-varying Limit Order Book Networks" by Wolfgang Karl Härdle, Shi Chen, Chong Liang, Melanie Schienle, April 2018.

017 "Regularization Approach for NetworkModeling of German EnergyMarket" by Shi Chen, Wolfgang Karl Härdle, Brenda López Cabrera, May 2018.

018 "Adaptive Nonparametric Clustering" by Kirill Efimov, Larisa Adamyan, Vladimir Spokoiny, May 2018.

019 "Lasso, knockoff and Gaussian covariates: a comparison" by Laurie Davies, May 2018.

IRTG 1792, Spandauer Straße 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.

IRTG 1792 Discussion Paper Series 2018

For a complete list of Discussion Papers published, please visit irtg1792.hu-berlin.de.

020 "A Regime Shift Model with Nonparametric Switching Mechanism" by Haiqiang Chen, Yingxing Li, Ming Lin and Yanli Zhu, May 2018.

021 "LASSO-Driven Inference in Time and Space" by Victor Chernozhukov, Wolfgang K. Härdle, Chen Huang, Weining Wang, June 2018.

022 " Learning from Errors: The case of monetary and fiscal policy regimes" by Andreas Tryphonides, June 2018.

023 "Textual Sentiment, Option Characteristics, and Stock Return Predictability" by Cathy Yi-Hsuan Chen, Matthias R. Fengler, Wolfgang Karl Härdle, Yanchu Liu, June 2018.

024 "Bootstrap Confidence Sets For Spectral Projectors Of Sample Covariance" by A. Naumov, V. Spokoiny, V. Ulyanov, June 2018.

025 "Construction of Non-asymptotic Confidence Sets in 2 -Wasserstein Space" by Johannes Ebert, Vladimir Spokoiny, Alexandra Suvorikova, June 2018.

025 "Large ball probabilities, Gaussian comparison and anti-concentration" by Friedrich Götze, Alexey Naumov, Vladimir Spokoiny, Vladimir Ulyanov, June 2018.

IRTG 1792, Spandauer Straße 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.