Bernoulli **25**(4A), 2019, 2538–2563
https://doi.org/10.3150/18-BEJ1062

Large ball probabilities, Gaussian comparison and anti-concentration

FRIEDRICH GÖTZE[1], ALEXEY NAUMOV[2,5,*], 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: dimension free bounds; Gaussian anti-concentration inequalities; Gaussian comparison; high-dimensional CLT; high-dimensional inference; Schatten norm

1. Introduction

In many statistical and probabilistic applications, one faces the problem of Gaussian comparison, that is, one has 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, for example, Spokoiny and Zhilova [28]. However, the obtained bound involves the inverse of the covariance operators of the considered Gaussian measures. In particular, 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. Upper bounds for the closeness of two centered Gaussian measures in the class of centered balls in a separable Hilbert space see in Naumov, Spokoiny, Tavyrikov and Ulyanov [21]

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, Kotz and Balakrishnan [15]). 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, for example, Christoph, Prokhorov and Ulyanov [12]. Asymptotic properties of $p_{\boldsymbol{\xi}}(x, \boldsymbol{a})$, small balls probabilities $\mathbb{P}(\|\boldsymbol{\xi} - a\| \le \varepsilon)$, or large deviation bounds $\mathbb{P}(\|\boldsymbol{\xi}\| \ge 1/\varepsilon)$ for small ε can be found for example, in Bogachev [8], Ledoux and Talagrand [17], Li and Shao [18], Lifshits [19] and Yurinsky [31].

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 Appendices A and B respectively.

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

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

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:

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

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

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

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

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

However, it depends on the unknown measure $\mathbb{P}$. 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

$$\mathcal{E}^{\flat}(\mathfrak{z}) \stackrel{\text{def}}{=} \Bigl\{\boldsymbol{\theta} : \sup_{\boldsymbol{\theta}' \in \Theta} L^{\flat}(\boldsymbol{\theta}') - L^{\flat}(\boldsymbol{\theta}) \le \mathfrak{z}\Bigr\}.$$

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

$$\mathbb{P}^{\flat}\bigl(\widetilde{\boldsymbol{\theta}} \notin \mathcal{E}^{\flat}(\mathfrak{z}^{\flat})\bigr) = \mathbb{P}^{\flat}\Bigl(\sup_{\boldsymbol{\theta} \in \Theta} L^{\flat}(\boldsymbol{\theta}) - L^{\flat}(\widetilde{\boldsymbol{\theta}}) > \mathfrak{z}^{\flat}\Bigr) = \alpha.$$

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

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

see, for example, Spokoiny and Zhilova [28]. A proof of this result is quite involved. The key steps are the following two approximations:

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

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

$$
\Sigma \stackrel{\text{def}}{=} D^{-1}\operatorname{Var}\bigl[\nabla L(\boldsymbol{\theta}^*)\bigr]D^{-1}, \qquad D^2 = -\nabla^2 \mathbb{E} L(\boldsymbol{\theta}^*),
$$

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

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

The vector $\boldsymbol{a}$ in (1.2) is the so called modeling bias and it vanishes if the parametric assumption $\mathbb{P} = \mathbb{P}_{\boldsymbol{\theta}^*}$ is precisely fulfilled. The matrix Bernstein inequality ensures that $\Sigma^{\flat}$ is close to Σ in the operator norm for n large; see, for example, Tropp [29]. 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 [28] 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 $\Sigma^{-1/2}\Sigma^{\flat}\Sigma^{-1/2} - \mathbb{I}_p$ and the norm of the vector $\boldsymbol{\beta} \stackrel{\text{def}}{=} \Sigma^{-1/2}\boldsymbol{a}$:

$$
\bigl\|\mathcal{N}(\boldsymbol{a}, \Sigma) - \mathcal{N}\bigl(0, \Sigma^{\flat}\bigr)\bigr\|_{\mathrm{TV}} \le \frac{1}{2}\bigl(\bigl\|\Sigma^{-1/2}\Sigma^{\flat}\Sigma^{-1/2} - \mathbb{I}_p\bigr\|_{\mathrm{Fr}} + \bigl\|\Sigma^{-1/2}\boldsymbol{a}\bigr\|\bigr); \tag{1.3}
$$

see, for example, Spokoiny and Zhilova [28]. 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

$$
\bigl|\mathbb{P}\bigl(\|\boldsymbol{\xi}+\boldsymbol{a}\|^2 > 2\mathfrak{z}^{\flat}\bigr) - \alpha\bigr| \le \frac{\mathtt{C}}{\|\Sigma\|_{\mathrm{Fr}}}\bigl(\bigl\|\Sigma - \Sigma^{\flat}\bigr\|_1 + \|\boldsymbol{a}\|^2\bigr). \tag{1.4}
$$

The “small modeling bias” condition on $\boldsymbol{a}$ from Spokoiny and Zhilova [28] 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 $\Sigma^{-1/2}\boldsymbol{a}$, that is, 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 $\|\Sigma^{-1/2}\boldsymbol{a}\|$ to $\|\boldsymbol{a}\|^2/\|\Sigma\|_{\mathrm{Fr}}$ can be enormous if some eigenvalues of Σ nearly vanish. Further, the Frobenius norm $\|\Sigma^{-1/2}\Sigma^{\flat}\Sigma^{-1/2}-\mathbb{I}_p\|_{\mathrm{Fr}}$ can be much larger than the ratio $\|\Sigma-\Sigma^{\flat}\|_1/\|\Sigma\|_{\mathrm{Fr}}$ by the same reasons.

Note that the approach based on Theorem 2.1 and Corollary 2.2 below was used in Naumov, Spokoiny and Ulyanov [22] to analyze non-asymptotic properties of bootstrap confidence sets for spectral projectors of covariance matrices.

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

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

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

$$\boldsymbol{\vartheta}_G \mid \boldsymbol{Y} \propto \exp\biggl(L(\boldsymbol{\theta}) - \frac{1}{2}\|G\boldsymbol{\theta}\|^2\biggr) \propto \exp\biggl(-\frac{1}{2\sigma^2}\bigl\|\boldsymbol{Y}-\boldsymbol{\Psi}^{\top}\boldsymbol{\theta}\bigr\|^2 - \frac{1}{2}\|G\boldsymbol{\theta}\|^2\biggr).$$

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

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

where

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

In particular, it implies that the posterior distribution $\mathbb{P}(\boldsymbol{\vartheta}_G \mid \boldsymbol{Y})$ 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

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

where W is a given linear mapping from $\mathbb{R}^p$. 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

$$\mathbb{P}\bigl(\bigl\|W(\boldsymbol{\vartheta}_G - \breve{\boldsymbol{\theta}}_G)\bigr\| \ge \mathtt{r}_G \mid \boldsymbol{Y}\bigr) = \alpha. \tag{1.5}$$

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

$$\mathbb{P}\big(\|\boldsymbol{\xi}_G\| \geq \mathtt{r}_G\big) = \alpha \tag{1.6}$$

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 $\breve{\boldsymbol{\theta}}_{G_1} = (\boldsymbol{\Psi}\boldsymbol{\Psi}^\top + \sigma^2 G_1^2)^{-1}\boldsymbol{\Psi}\boldsymbol{Y}$ 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 $E_G(\mathtt{r}_G)$:

$$\mathbb{P}\big(\big\|W(\boldsymbol{\vartheta}_{G_1} - \breve{\boldsymbol{\theta}}_G)\big\| \geq \mathtt{r}_G \mid \boldsymbol{Y}\big).$$

Clearly this probability can be written as

$$\mathbb{P}\big(\|\boldsymbol{\xi}_{G_1} + \boldsymbol{a}\| \geq \mathtt{r}_G\big)$$

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

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

Therefore,

$$\big|\mathbb{P}\big(\big\|W(\boldsymbol{\vartheta}_{G_1} - \breve{\boldsymbol{\theta}}_G)\big\| \geq \mathtt{r}_G \mid \boldsymbol{Y}\big) - \alpha\big| \leq \sup_{\mathtt{r}>0}\big|\mathbb{P}\big(\|\boldsymbol{\xi}_{G_1} + \boldsymbol{a}\| \geq \mathtt{r}\big) - \mathbb{P}\big(\|\boldsymbol{\xi}_G\| \geq \mathtt{r}\big)\big|.$$

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:

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

see, for example, Panov and Spokoiny [23]. 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

$$\big|\mathbb{P}\big(\big\|W(\boldsymbol{\vartheta}_{G_1} - \breve{\boldsymbol{\theta}}_G)\big\| \geq \mathtt{r}_G \mid \boldsymbol{Y}\big) - \alpha\big| \leq \frac{\mathtt{C}(\operatorname{tr}\Sigma_{G_1} - \operatorname{tr}\Sigma_G + \|\boldsymbol{a}\|^2)}{\|\Sigma_G\|_{\mathrm{Fr}}}.$$

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 $\mathbb{E}\boldsymbol{Y} = \boldsymbol{f}^* = \boldsymbol{\Psi}^\top \boldsymbol{\theta}^*$. Here we consider the model $\boldsymbol{Y} = \boldsymbol{f}^* + \boldsymbol{\varepsilon} = \boldsymbol{\Psi}^\top \boldsymbol{\theta} + \boldsymbol{\varepsilon}$ in $\mathbb{R}^n$ with a homogeneous Gaussian noise $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2 \mathbb{I}_n)$ and a Gaussian prior $\mathcal{N}(0, G^{-2})$ on $\boldsymbol{\theta}$. The credible set $E_G(\mathtt{r})$ for $\boldsymbol{\vartheta}_G$ yields the credible set $\mathcal{E}_G(\mathtt{r})$ for the corresponding response $\boldsymbol{f} = \boldsymbol{\Psi}^\top \boldsymbol{\theta}$:

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

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 $\Pi_G = \boldsymbol{\Psi}^\top (\boldsymbol{\Psi}\boldsymbol{\Psi}^\top + \sigma^2 G^2)^{-1} \boldsymbol{\Psi}$. The frequentist coverage probability of the true response $\boldsymbol{f}^*$ is given by

$$\mathbb{P}\bigl(\boldsymbol{f}^* \in \mathcal{E}_G(\mathtt{r})\bigr) = \mathbb{P}\bigl(\bigl\| \boldsymbol{A}\bigl(\boldsymbol{f}^* - \boldsymbol{\Psi}^\top \breve{\boldsymbol{\theta}}_G\bigr) \bigr\| \le \mathtt{r}\bigr) = \mathbb{P}\bigl(\bigl\| \boldsymbol{A}\boldsymbol{\Psi}^\top \bigl(\boldsymbol{\theta}^* - \breve{\boldsymbol{\theta}}_G\bigr) \bigr\| \le \mathtt{r}\bigr).$$

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

$$\mathbb{E}\bigl[\boldsymbol{A}\bigl(\boldsymbol{f}^* - \boldsymbol{\Psi}^\top \breve{\boldsymbol{\theta}}_G\bigr)\bigr] = \boldsymbol{A}(\mathbb{I} - \Pi_G)\boldsymbol{f}^* \stackrel{\text{def}}{=} \boldsymbol{a}.$$

Further,

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

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

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

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

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

for all $\mathtt{r} \ge 0$. Existing results cover only very special cases; see, for example, Johnstone [16], Bontemps [9], Panov and Spokoiny [23], Castillo [10], Castillo and Nickl [11], Belitser [3] 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

$$\|\Sigma_G - \Sigma\|_1 = \operatorname{tr} \Sigma_G - \operatorname{tr} \Sigma.$$

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

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

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

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

is small under standard condition on the design $\boldsymbol{\Psi}$ and on the spectrum of G^2; see, for example, Spokoiny [27].

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, for example, Bentkus [4,5]. 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 [25], Götze and Zaitsev [14].

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

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

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 $A \in \mathbf{C}$, $\boldsymbol{a} \in \mathbb{R}^p$ 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

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

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 $\mathbb{I}$. Assume that the class $\mathbf{C}$ in (1.8) is such that for all $A \in \mathbf{C}$ and $\varepsilon > 0$

$$\mathbb{P}\bigl(X_0 \in A^\varepsilon \backslash A\bigr) \le a_p \varepsilon, \qquad \mathbb{P}\bigl(X_0 \in A \backslash A^{-\varepsilon}\bigr) \le a_p \varepsilon, \tag{1.9}$$

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 $a_p \le 4p^{1/4}$; see Ball [1].

It is known (see Bentkus [5], Theorem 1.2) that if $\mathbf{C}$ satisfies conditions (i), (ii) and (1.9) then for some absolute constant C one has

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

Therefore, the inequalities (1.9), that is, 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, for example, the class of all balls, the convergence of $\mathbb{P}((Y_1 + \cdots + Y_n)/\sqrt{n} \in A)$ to $\mathbb{P}(X \in A)$ for each $A \in \mathbf{C}$, implied by the central limit theorem, can not be uniform in $A \in \mathbf{C}$; see, for example, Sazonov [26], 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, for example, Prokhorov and Ulyanov [24] or our previous application examples. Thus, similar to the inequalities (1.9) we need to get sharp bounds for the probability $\mathbb{P}(x < \|X - \boldsymbol{a}\|^2 < x + \varepsilon)$ 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

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

for an absolute constant 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 $ca \le b \le Ca$. $\mathbb{R}$ (resp. $\mathbb{C}$) denotes the set of all real (resp. complex) numbers. We assume that all random variables are defined on common probability space $(\Omega, \mathfrak{F}, \mathbb{P})$ 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 $\mathbb{R}^p$ instead of $\mathbb{H}$. Let $\mathbb{E}$ be the mathematical expectation with respect to $\mathbb{P}$. 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

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

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

Let Σ_{ξ} be a covariance operator of an arbitrary Gaussian random element in $\mathbb{H}$. By $\{\lambda_{k\xi}\}_{k\ge1}$ we denote the set of its eigenvalues arranged in the non-increasing order, that is, $\lambda_{1\xi} \ge \lambda_{2\xi} \ge \cdots$, and let $\boldsymbol{\lambda}_{\xi} \stackrel{\mathrm{def}}{=} \operatorname{diag}(\lambda_{j\xi})_{j=1}^{\infty}$. Note that $\sum_{j=1}^{\infty} \lambda_{j\xi} < \infty$. Introduce the following quantities

$$\Lambda_{k\xi}^2 \stackrel{\mathrm{def}}{=} \sum_{j=k}^{\infty} \lambda_{j\xi}^2, \qquad k = 1, 2, \tag{2.1}$$

and

$$\varkappa(\Sigma_\xi) = \begin{cases} \Lambda_{1\xi}^{-1}, & \text{if } 3\lambda_{1,\xi}^2 \le \Lambda_{1\xi}^2, \\ (\lambda_{1\xi}\Lambda_{2\xi})^{-1/2}, & \text{if } 3\lambda_{1\xi}^2 > \Lambda_{1\xi}^2, 3\lambda_{2\xi}^2 \le \Lambda_{2\xi}^2, \\ (\lambda_{1\xi}\lambda_{2\xi})^{-1/2}, & \text{if } 3\lambda_{1\xi}^2 > \Lambda_{1\xi}^2, 3\lambda_{2\xi}^2 > \Lambda_{2\xi}^2. \end{cases} \tag{2.2}$$

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

$$\frac{0.9}{(\Lambda_{1\xi}\Lambda_{2\xi})^{1/2}} \le \varkappa(\Sigma_\xi) \le \frac{1.8}{(\Lambda_{1\xi}\Lambda_{2\xi})^{1/2}}. \tag{2.3}$$

Hence, $\varkappa(\Sigma_\xi) \asymp (\Lambda_{1\xi}\Lambda_{2\xi})^{-1/2}$ 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 Σ_ξ and Σ_η, respectively. For any $\boldsymbol{a} \in \mathbb{H}$*

$$\begin{aligned} &\sup_{x>0}\big|\mathbb{P}\big(\|\boldsymbol{\xi} - \boldsymbol{a}\| \le x\big) - \mathbb{P}\big(\|\boldsymbol{\eta}\| \le x\big)\big| \\ &\quad\lesssim \big\{\varkappa(\Sigma_\xi) + \varkappa(\Sigma_\eta)\big\}\big(\|\boldsymbol{\lambda}_\xi - \boldsymbol{\lambda}_\eta\|_1 + \|\boldsymbol{a}\|^2\big). \end{aligned} \tag{2.4}$$

The proof of Theorem 2.1 is given in Section 3.

We see that the obtained bounds are expressed in terms of the specific characteristics of the matrices Σ_ξ and Σ_η 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 Σ_ξ or Σ_η. In other words, small or vanishing eigenvalues of Σ_ξ or Σ_η 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 $\|\Sigma_\xi^{-1/2}\boldsymbol{a}\|$ which can be very large or infinite if Σ_ξ is not well conditioned.

Let us consider $\varkappa(\Sigma_\xi)$ in the first factor on the r.h.s of (2.4): $\varkappa(\Sigma_\xi) + \varkappa(\Sigma_\eta)$. The representation (2.2) mimics well the three typical situations: in the "large-dimensional case" with three or more significant eigenvalues $\lambda_{j\xi}$, one can take $\varkappa(\Sigma_\xi) = \|\Sigma_\xi\|_{\mathrm{Fr}}^{-1} = \lambda_{1\xi}^{-1}$. In the "two dimensional" case, when the sum $\Lambda_{k\xi}^2$ is of the order $\lambda_{k\xi}^2$ for $k = 1, 2$, we have that $\varkappa(\Sigma_\xi)$ behaves as the product $(\lambda_{1\xi}\lambda_{2\xi})^{-1/2}$. In the intermediate case of a spike model with one large eigenvalue $\lambda_{1\xi}$ and many small eigenvalues $\lambda_{j\xi}$, $j \ge 2$, we have that $\varkappa(\Sigma_\xi)$ behaves as $(\lambda_{1\xi}\Lambda_{2\xi})^{-1/2}$.

As it was mentioned earlier (see (2.3)), the result of Theorem 2.1 may be equivalently formulated in a "unified" way in terms of $(\Lambda_{1\xi}\Lambda_{2\xi})^{-1/2}$ and $(\Lambda_{1\eta}\Lambda_{2\eta})^{-1/2}$. Moreover, we specify the bound (2.4) in the "high-dimensional" case, $3\|\Sigma_\xi\|^2 \le \|\Sigma_\xi\|_{\mathrm{Fr}}^2$, $3\|\Sigma_\eta\|^2 \le \|\Sigma_\eta\|_{\mathrm{Fr}}^2$, which means at least three significantly positive eigenvalues of the matrices Σ_ξ and Σ_η. In this case $\Lambda_{2\xi}^2 \ge 2\Lambda_{1\xi}^2/3$, $\Lambda_{2\eta}^2 \ge 2\Lambda_{1\eta}^2/3$ and we get the following corollary.

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 $\boldsymbol{a} \in \mathbb{H}$*

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

Moreover, assume that

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

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

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

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, for example, Markus [20]. 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*

$$\sup_{x>0}\bigl|\mathbb{P}\bigl(\|\boldsymbol{\xi}-\boldsymbol{a}\| \le x\bigr) - \mathbb{P}\bigl(\|\boldsymbol{\eta}\| \le x\bigr)\bigr| \lesssim \bigl\{\varkappa(\Sigma_{\boldsymbol{\xi}}) + \varkappa(\Sigma_{\boldsymbol{\eta}})\bigr\}\bigl(\|\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}\|_1 + \|\boldsymbol{a}\|^2\bigr).$$

Since the right-hand side of (2.4) does not change if we exchange $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$, Theorem 2.1 and its corollaries hold for the balls with the same shift $\boldsymbol{a}$. In particular, the following corollary is true, that improves Theorem 4 in Barsov and Ulyanov [2]

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

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

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

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

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*

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

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*

$$\sup_{x\ge 0} p_{\boldsymbol{\xi}}(x,\boldsymbol{a})\lesssim\varkappa(\Sigma_{\boldsymbol{\xi}}) \tag{2.5}$$

with $\varkappa(\Sigma_{\boldsymbol{\xi}})$ from (2.2). 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 $\boldsymbol{\xi}\overset{\mathsf{d}}{=}\sum_{j=1}^{\infty}\sqrt{\lambda_{j\boldsymbol{\xi}}}Z_j\mathbf{e}_{j\boldsymbol{\xi}}$, we obtain that $\|\boldsymbol{\xi}\|^2\overset{\mathsf{d}}{=}\sum_{j=1}^{\infty}\lambda_{j\boldsymbol{\xi}}Z_j^2$. 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 $\{\lambda_{j\boldsymbol{\xi}}\}_{j=1}^{\infty}$. In the case $\mathbb{H}=\mathbb{R}^p$, $\boldsymbol{a}=0$, $\Sigma_{\boldsymbol{\xi}}\asymp\mathbb{I}$ one has that the distribution of $\|\boldsymbol{\xi}\|^2$ is close to standard χ^2 with p degrees of freedom and

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

Hence, the bound (2.5) gives the right dependence on p because $\varkappa(\Sigma_{\boldsymbol{\xi}})\asymp p^{-1/2}$. However, a lower bound for $\sup_{x\ge 0}p_{\boldsymbol{\xi}}(x,\boldsymbol{a})$ 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

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

see Lemma B.1 and remark after it in Appendix B. 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*

$$\sup_{x>0} \mathbb{P}\big(x < \|\boldsymbol{\xi} - \boldsymbol{a}\|^2 < x + \varepsilon\big) \lesssim \varkappa(\Sigma_{\boldsymbol{\xi}})\varepsilon \tag{2.6}$$

with $\varkappa(\Sigma_{\boldsymbol{\xi}})$ from (2.2). In particular, $\varkappa(\Sigma_{\boldsymbol{\xi}})$ can be replaced by $(\Lambda_{1\boldsymbol{\xi}}\Lambda_{2\boldsymbol{\xi}})^{-1/2}$.

We finish this section by some lower bounds that justify the structure of estimates in Theorem 2.1 and Theorem 2.7.

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 that in the case $\mathbb{H} = \mathbb{R}^2$ there exist covariance operators $\Sigma_{\boldsymbol{\xi}}$ and $\Sigma_{\boldsymbol{\eta}}$ and some absolute positive constant $\mathtt{C}_1$ such that $\varkappa(\Sigma_{\boldsymbol{\xi}}, \Sigma_{\boldsymbol{\eta}}) \asymp (\lambda_{1\boldsymbol{\xi}}\lambda_{2\boldsymbol{\xi}})^{-1/2}$ and

$$\sup_{x>0}\big|\mathbb{P}\big(\|\boldsymbol{\xi}\| \le x\big) - \mathbb{P}\big(\|\boldsymbol{\eta}\| \le x\big)\big| \ge \mathtt{C}_1(\lambda_{1\boldsymbol{\xi}}\lambda_{2\boldsymbol{\xi}})^{-1/2}\|\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}\|_1, \tag{2.7}$$

that is, in this case the lower bound coincides up to an absolute constant with the upper bound in Theorem 2.1. To show (2.7), we consider the following example. Let $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ be the Gaussian random vectors in $\mathbb{R}^2$ with zero means and covariance matrices $\Sigma_{\boldsymbol{\xi}} = \operatorname{diag}(\lambda_{1\boldsymbol{\xi}}, \lambda_{2\boldsymbol{\xi}})$ and $\Sigma_{\boldsymbol{\eta}} = \operatorname{diag}(\lambda_{1\boldsymbol{\eta}}, \lambda_{2\boldsymbol{\eta}})$ resp. Then

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

for some R which will be chosen later. Put

$$\mathcal{E}_1 \stackrel{\text{def}}{=} \left\{(x_1, x_2) \in \mathbb{R}^2 : \sum_{j=1}^2 \lambda_{j\boldsymbol{\xi}} x_j^2 \le R\right\}, \qquad \mathcal{E}_2 \stackrel{\text{def}}{=} \left\{(x_1, x_2) \in \mathbb{R}^2 : \sum_{j=1}^2 \lambda_{j\boldsymbol{\eta}} x_j^2 \le R\right\}.$$

Let us take $\lambda_{1\boldsymbol{\xi}} = \lambda_{1\boldsymbol{\eta}}, \lambda_{2\boldsymbol{\eta}}/2 < \lambda_{2\boldsymbol{\xi}} < \lambda_{2\boldsymbol{\eta}}$. This choice gives $\|\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}\|_1 = \lambda_{2\boldsymbol{\eta}} - \lambda_{2\boldsymbol{\xi}}$ 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

$$\begin{aligned}\big|\mathbb{P}\big(\|\boldsymbol{\xi}\| \le \sqrt{R}\big) - \mathbb{P}\big(\|\boldsymbol{\eta}\| \le \sqrt{R}\big)\big| &= \frac{1}{2\pi}\int_{\mathcal{E}_1\setminus\mathcal{E}_2} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right) dx_1\, dx_2 \\ &\ge \frac{1}{2\pi}\big(|\mathcal{E}_1| - |\mathcal{E}_2|\big)\exp\left[-\frac{R}{2}\left(\frac{1}{\lambda_{1\boldsymbol{\xi}}} + \frac{1}{\lambda_{2\boldsymbol{\xi}}}\right)\right],\end{aligned}$$

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

$$|\mathcal{E}_1| - |\mathcal{E}_2| \ge \frac{\pi R\|\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}\|_1}{4\sqrt{2}\sqrt{\lambda_{1\boldsymbol{\xi}}\lambda_{2\boldsymbol{\xi}}}\lambda_{2\boldsymbol{\xi}}}.$$

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

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

Hence,

$$\left|\mathbb{P}\left(\|\boldsymbol{\xi}\| \le \sqrt{R}\right) - \mathbb{P}\left(\|\boldsymbol{\eta}\| \le \sqrt{R}\right)\right| \ge \frac{\|\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}\|_1}{12\sqrt{2}\sqrt{\lambda_{1\boldsymbol{\xi}}\lambda_{2\boldsymbol{\xi}}}}\exp\left[-\frac{\lambda_{2\boldsymbol{\xi}}}{\lambda_{1\boldsymbol{\xi}}}\right] \ge \mathrm{c}_1\frac{\|\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}\|_1}{\sqrt{\lambda_{1\boldsymbol{\xi}}\lambda_{2\boldsymbol{\xi}}}},$$

where $\mathrm{c}_1 \stackrel{\text{def}}{=} \exp(-1)/(12\sqrt{2})$. From the last inequality, we get (2.7).

However it is still an open question to get a lower bound in Theorem 2.1 even in the case of centered balls. The problem to get a lower bound in Theorem 2.6 is open as well. Partly the last problem was solved in Christoph, Prokhorov and Ulyanov [12], Theorem 1.

We now turn to the case $\mathbb{H} = \mathbb{R}^1$. 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

$$\begin{aligned}
&\sup_{x>0}\left|\mathbb{P}\left(\|\boldsymbol{\xi}\| \le x\right) - \mathbb{P}\left(\|\boldsymbol{\eta}\| \le x\right)\right| \\
&\quad= \frac{2}{\sqrt{2\pi}}\sup_{x>0}\int_{x/\sqrt{\lambda_{\boldsymbol{\eta}}}}^{x/\sqrt{\lambda_{\boldsymbol{\xi}}}} e^{-y^2/2}\,dy \\
&\quad\le \frac{\|\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}\|_1}{\sqrt{\lambda_{\boldsymbol{\eta}}\lambda_{\boldsymbol{\xi}}}(\sqrt{\lambda_{\boldsymbol{\xi}}} + \sqrt{\lambda_{\boldsymbol{\eta}}})}\sup_{x>0}\left(x\exp\left(-x^2/(2\lambda_{\boldsymbol{\eta}})\right)\right) \lesssim \frac{\|\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}\|_1}{\lambda_{\boldsymbol{\xi}}}.
\end{aligned}$$

We also have the following lower bound:

$$\begin{aligned}
&\sup_{x>0}\left|\mathbb{P}\left(\|\boldsymbol{\xi}\| \le x\right) - \mathbb{P}\left(\|\boldsymbol{\eta}\| \le x\right)\right| \\
&\quad= \frac{2}{\sqrt{2\pi}}\sup_{x>0}\int_{x/\sqrt{\lambda_{\boldsymbol{\eta}}}}^{x/\sqrt{\lambda_{\boldsymbol{\xi}}}} e^{-y^2/2}\,dy \\
&\quad\ge \frac{2\|\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}\|_1 x_0\exp(-x_0^2/(2\lambda_{\boldsymbol{\xi}}))}{\sqrt{2\pi}\sqrt{\lambda_{\boldsymbol{\eta}}\lambda_{\boldsymbol{\xi}}}(\sqrt{\lambda_{\boldsymbol{\xi}}} + \sqrt{\lambda_{\boldsymbol{\eta}}})} \gtrsim \frac{\|\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}\|_1}{\lambda_{\boldsymbol{\eta}}},
\end{aligned}$$

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 (2.6) 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 $\mathbb{H} = \mathbb{R}^2$ and $\lambda_{1\boldsymbol{\xi}} = 1$ and $\lambda_{2\boldsymbol{\xi}} = 0$ (i.e., $\boldsymbol{\xi}$ has the degenerate Gaussian distribution). Then for all positive $\varepsilon \le \log 2$ one has

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

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 $a_j \overset{\text{def}}{=} \langle a, \mathbf{e}_j \rangle$ and $\xi_j \overset{\text{def}}{=} \langle \boldsymbol{\xi}, \mathbf{e}_j \rangle$. Then ξ_j, $j \geq 1$, are independent $\mathcal{N}(0, \lambda_j)$ r.v. Let $g_j(x)$, $j \geq 1$, (resp. $f_j(t)$) be the p.d.f (resp. c.f.) of $(\xi_j - a_j)^2$. Moreover, let $g(m, x), m \geq 1$ (resp. $\overline{g}(m, x), m \geq 1$) 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 $\sum_{j=1}^{m} (\xi_j - a_j)^2$. As

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

we may restrict ourselves to the finite dimensional case only, e.d., $\mathbb{H} = \mathbb{R}^m$, 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 corresponding to the definition (2.2) of $\varkappa(\Sigma_{\boldsymbol{\xi}})$:

1. $3\lambda_1^2 \leq \Lambda_1^2$;
2. $3\lambda_1^2 \geq \Lambda_1^2, 3\lambda_2^2 \geq \Lambda_2^2$;
3. $3\lambda_1^2 \geq \Lambda_1^2, 3\lambda_2^2 \leq \Lambda_2^2$.

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

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

By the inverse formula

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

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 $m \geq 2$

$$g(m, x) \leq \sup_{x \geq 0} g(2, x).$$

It is straightforward to check that

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

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 $\mathsf{C}/\sqrt{\Lambda_1\Lambda_2}$.

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

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

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

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

We will show 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 $j = 2, \dots, m$. Similarly to (3.2), we can bound $|\overline{f}_j(t)| \le (1+4w_jt^2)^{-1/4}$ and

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

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

$$\sup_z p_\eta(z) \lesssim 1. \tag{3.5}$$

We divide integral in r.h.s. of (3.4) into two parts: $x - z < 1, 0 \le z \le x$ and $x - z \ge 1, 0 \le z \le x$. For the first part, we use (3.5), for the second part we estimate integrand by $p_\eta(z)$ and use $\int_0^\infty p_\eta(z)\,dz = 1$. This bound yields the upper bound of order $(\lambda_1\Lambda_2)^{-1/2} \asymp (\Lambda_1\Lambda_2)^{-1/2}$ in case (3). This completes the proof of the theorem. $\square$

Remark 3.1. Notice that instead of Lemma A.2 one may also apply an alternative approach from Ul'yanov [30], Lemma 5, and Bobkov, Chistyakov and Götze [7], Lemma 7.1.

Proof of Theorem 2.1. We split the proof into two parts. In the first part, we study the case $\boldsymbol{a}=0$ and show that

$$\sup_{x>0}\big|\mathbb{P}\big(\|\boldsymbol{\xi}\|\le x\big)-\mathbb{P}\big(\|\boldsymbol{\eta}\|\le x\big)\big|\lesssim\big\{\varkappa(\varSigma_{\boldsymbol{\xi}})+\varkappa(\varSigma_{\boldsymbol{\eta}})\big\}\|\boldsymbol{\lambda}_{\boldsymbol{\xi}}-\boldsymbol{\lambda}_{\boldsymbol{\eta}}\|_1. \tag{3.6}$$

The second part is devoted to the case $\varSigma_{\boldsymbol{\xi}}=\varSigma_{\boldsymbol{\eta}}$. We prove

$$\sup_{x>0}\big|\mathbb{P}\big(\|\boldsymbol{\xi}-\boldsymbol{a}\|\le x\big)-\mathbb{P}\big(\|\boldsymbol{\xi}\|\le x\big)\big|\lesssim\varkappa(\varSigma_{\boldsymbol{\xi}})\|\boldsymbol{a}\|^2. \tag{3.7}$$

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 $\varSigma_{\boldsymbol{\xi}}=\boldsymbol{\lambda}_{\boldsymbol{\xi}}$, $\varSigma_{\boldsymbol{\eta}}=\boldsymbol{\lambda}_{\boldsymbol{\eta}}$, where $\boldsymbol{\lambda}_{\boldsymbol{\xi}}\stackrel{\text{def}}{=}\operatorname{diag}(\lambda_{1\boldsymbol{\xi}},\lambda_{2\boldsymbol{\xi}},\dots)$, $\boldsymbol{\lambda}_{\boldsymbol{\eta}}\stackrel{\text{def}}{=}\operatorname{diag}(\lambda_{1\boldsymbol{\eta}},\lambda_{2\boldsymbol{\eta}},\dots)$ and $\lambda_{1\boldsymbol{\xi}}\ge\lambda_{1\boldsymbol{\xi}}\ge\dots$ 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 $\lambda_1(s)\ge\lambda_2(s)\ge\dots$ be the eigenvalues of $\mathbf{V}(s)$ and introduce the diagonal resolvent operator $\mathbf{G}(t,s)\stackrel{\text{def}}{=}(\mathbb{I}-2it\mathbf{V}(s))^{-1}$. Recall that $\|Z(s)\|^2\stackrel{\mathsf{d}}{=}\sum_{j=1}^n\lambda_j(s)Z_j^2$, where Z_j, $j\ge 1$, 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

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

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

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

It is well known, see, for example, Chung [13], §6.2, p. 168, that for a continuous d.f. $F(x)$ with c.f. $f(t)$ we may write

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

where V.P. stands for the principal value of integral. Let us fix an arbitrary $x>0$. Then

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

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

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

Changing the order of integration we get

$$\begin{aligned} &\mathbb{P}\big(\|\boldsymbol{\xi}\|^2 < x\big) - \mathbb{P}\big(\|\boldsymbol{\eta}\|^2 < x\big) \\ &\quad = -\frac{1}{2\pi}\int_0^1 \int_{-\infty}^{\infty} \mathrm{tr}\big\{(\boldsymbol{\lambda}_{\xi} - \boldsymbol{\lambda}_{\eta})\mathbf{G}(t,s)\big\} f(t,s) e^{-itx}\, dt\, ds. \end{aligned} \tag{3.8}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\sup_{x\geq 0} p(x,s) \lesssim \frac{1}{(\Lambda_1(s)\Lambda_2(s))^{1/2}},$$

where $\Lambda_k(s)$, $k=1,2$, are defined by (2.1) if one replace $\boldsymbol{\xi}$ by $Z(s)$. It remains to integrate over s and use (2.3) to obtain

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

This bounds concludes the proof of (3.6).

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 $\overline{\boldsymbol{a}}(s) \stackrel{\text{def}}{=} \overline{\boldsymbol{a}}\sqrt{s}$. Introduce the diagonal operator $\mathbf{G}(t) \stackrel{\text{def}}{=} (\mathbb{I} - 2it\Lambda_{\boldsymbol{\xi}})^{-1}$. 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

$$f\big(t, \overline{\boldsymbol{a}}(s)\big) = \mathbb{E}\exp\big\{it\big\|\boldsymbol{\xi} - \overline{\boldsymbol{a}}(s)\big\|^2\big\} = \exp\bigg\{it\bigg(s\big\langle\mathbf{G}(t)\overline{\boldsymbol{a}}, \overline{\boldsymbol{a}}\big\rangle - \frac{1}{2it}\operatorname{tr}\log(\mathbb{I} - 2it\Lambda_{\boldsymbol{\xi}})\bigg)\bigg\}.$$

Repeating the arguments from the proof of Theorem 2.1 we obtain (compare with (3.8))

$$\mathbb{P}\big(\|\boldsymbol{\xi} - \boldsymbol{a}\|^2 < x\big) - \mathbb{P}\big(\|\boldsymbol{\xi}\|^2 < x\big) = -\frac{1}{2\pi}\int_0^1\int_{-\infty}^{\infty}\big\langle\mathbf{G}(t)\overline{\boldsymbol{a}}, \overline{\boldsymbol{a}}\big\rangle f\big(t, \overline{\boldsymbol{a}}(s)\big)e^{-itx}\,dt\,ds.$$

Moreover, we may rewrite the last equation as follows

$$\mathbb{P}\big(\|\boldsymbol{\xi} - \boldsymbol{a}\|^2 < x\big) - \mathbb{P}\big(\|\boldsymbol{\xi}\|^2 < x\big) = -\sum_{j=1}^{\infty}[\overline{a}_j]^2\int_0^1 p_j\big(x, \overline{\boldsymbol{a}}(s)\big)\,ds,$$

where $p_j(x, \overline{\boldsymbol{a}}(s))$ is p.d.f of $\overline{Z}_j + \|\boldsymbol{\xi} - \overline{\boldsymbol{a}}(s)\|^2$. Here $\overline{Z}_j$ is a random variable with exponential distribution $\operatorname{Exp}(2\lambda_{j\boldsymbol{\xi}})$. It remains to apply Theorem 2.6 and integrate over s. We conclude (3.7). □

Appendix A: Technical results

Lemma A.1. *It holds*

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

and

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

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 $a_0 \in [1, 2]$ and an integer $k \ge 0$, we use that

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

This proves (A.2). □

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

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

Define

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

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 (cf. Bobkov and Chistyakov [6], Lemma 5). By the Hölder inequality

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

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

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

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

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

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

$$\tau \geq \frac{1}{4}\left(\Lambda^2 - 2\lambda_1^2\right) \geq \frac{1}{4}\left(\Lambda^2 - \frac{2\Lambda^2}{3}\right) \gtrsim \Lambda^2.$$

Now it follows from (A.4) that

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

as required. $\square$

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 $\Sigma_{\boldsymbol{\xi}}$. Denote by $p_{\boldsymbol{\xi}}(x, \boldsymbol{a})$ the p.d.f. of $\|\boldsymbol{\xi} - \boldsymbol{a}\|^2$. Then for any $\boldsymbol{a} \in \mathbb{H}$ and all $\lambda > \lambda_{1\boldsymbol{\xi}}$*

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

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 operator 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

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

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

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

and

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

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

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

Moreover,

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

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

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

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

$$\begin{aligned} g(3,x) &= \int_0^x g(2,x-y)g_3(y)\,dy \\ &\le \frac{1}{2\sqrt{\lambda_1\lambda_2}\sqrt{2\pi\lambda_3}}\exp\bigl(-\bigl(x^{1/2}-\bigl(a_1^2+a_2^2+a_3^2\bigr)^{1/2}\bigr)^2/(2\lambda)\bigr) \\ &\quad\times\int_0^x \frac{d_3(\lambda\lambda_3/(\lambda-\lambda_3),y)}{y^{1/2}}\,dy \\ &\le \frac{1}{2\sqrt{\lambda_1\lambda_2}}\exp\bigl(-\bigl(x^{1/2}-\bigl(a_1^2+a_2^2+a_3^2\bigr)^{1/2}\bigr)^2/(2\lambda)\bigr)\left(1-\frac{\lambda_3}{\lambda}\right)^{-1/2}. \end{aligned}$$

By induction, we get

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

Now take an arbitrary $\varepsilon > 0$ and any integer $m > 0$. Let $0 < \mu < 1/(2\lambda_j)$ for all $j \ge m+1$. By Markov's inequality and using the moment generating function of χ^2 we obtain

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

Let us choose $\mu \overset{\text{def}}{=} 1/(4\sum_{j=m+1}^\infty \lambda_j)$. Taking logarithm and using $\log(1+x) \ge x/(x+1)$ for $x > -1$ we obtain

$$0 < -\frac{1}{2}\log\prod_{j=m+1}^\infty(1-2\mu\lambda_j) \le \frac{\mu}{1-2\mu\lambda_{m+1}}\sum_{j=m+1}^\infty \lambda_j < \frac{1}{4(1-1/2)} = 1/2,$$

It follows now that $\prod_{j=m+1}^\infty(1-2\mu\lambda_j)^{-1/2} \le \sqrt{e} < 2$. Hence,

$$\mathbb{P}\left(\sum_{j=m+1}^\infty \xi_j^2 \ge \varepsilon^2\right) \le 2\exp\left\{-\varepsilon^2\left(4\sum_{j=m+1}^\infty \lambda_j\right)^{-1}\right\}.$$

Since $\|\Sigma\|_1 < \infty$ it follows that $\sum_{j=m+1}^{\infty} \lambda_j$ tends to zero as m goes to infinity. Hence, there exists $M_1 = M_1(\varepsilon)$ such that for all $m \ge M_1$

$$\mathbb{P}\left(\sum_{j=m+1}^{\infty} \xi_j^2 \ge \varepsilon^2\right) \le \varepsilon^2.$$

For any $m \ge M_1$ we obtain

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

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

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

The last inequality implies

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

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

Acknowledgements

The authors would like to thank the Associate Editor and the Reviewer for the careful reading of the manuscript and pertinent comments. Their constructive feedback helped to improve the quality of this work and shape its final form.

We are grateful to Prof. Sergey Bobkov and Yuri Tavyrikov for helpful discussions on the subject of this paper.

The work has been funded by the Russian Academic Excellence Project '5-100'. Results of Section 2 have been obtained under support of the RSF grant No. 18-11-00132. Research of F. Götze and V. Spokoiny has also been supported by DFG through the Collaborative Research Centres 1283 and 1294, respectively.

References

[1] Ball, K. (1993). The reverse isoperimetric problem for Gaussian measure. *Discrete Comput. Geom.* **10** 411–420. MR1243336

[2] Barsov, S.S. and Ulyanov, V.V. (1987). Estimates of the proximity of Gaussian measures. *Soviet Math. Dokl.* **34** 462–466. MR866375

[3] Belitser, E. (2017). On coverage and local radial rates of credible sets. *Ann. Statist.* **45** 1124–1151. MR3662450

[4] Bentkus, V. (2003). On the dependence of the Berry–Esseen bound on dimension. *J. Statist. Plann. Inference* **113** 385–402. MR1965117

[5] Bentkus, V. (2005). A Lyapunov-type bound in R^d. *Theory Probab. Appl.* **49** 311–323. MR2144310

[6] Bobkov, S.G. and Chistyakov, G.P. (2014). Bounds for the maximum of the density of the sum of independent random variables. *J. Math. Sci.* **199** 100–106. MR3032208

[7] Bobkov, S.G., Chistyakov, G.P. and Götze, F. (2014). Berry–Esseen bounds in the entropic central limit theorem. *Probab. Theory Related Fields* **159** 435–478. MR3230000

[8] Bogachev, V.I. (1998). *Gaussian Measures. Mathematical Surveys and Monographs* **62**. Providence, RI: Amer. Math. Soc. MR1642391

[9] Bontemps, D. (2011). Bernstein–von Mises theorems for Gaussian regression with increasing number of regressors. *Ann. Statist.* **39** 2557–2584. MR2906878

[10] Castillo, I. (2012). A semiparametric Bernstein–von Mises theorem for Gaussian process priors. *Probab. Theory Related Fields* **152** 53–99. MR2875753

[11] Castillo, I. and Nickl, R. (2013). Nonparametric Bernstein–von Mises theorems in Gaussian white noise. *Ann. Statist.* **41** 1999–2028. MR3127856

[12] Christoph, G., Prokhorov, Y. and Ulyanov, V. (1996). On distribution of quadratic forms in Gaussian random variables. *Theory Probab. Appl.* **40** 250–260. MR1346468

[13] Chung, K.L. (2001). *A Course in Probability Theory*, 3rd ed. San Diego, CA: Academic Press. MR1796326

[14] Götze, F. and Zaitsev, A.Yu. (2018). New applications of Arak's inequalities to the Littlewood–Offord problem. *Eur. J. Math.* **4** 639–663. MR3799161

[15] Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions. Vol.* 1, 2nd ed. *Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. New York: Wiley. A Wiley-Interscience Publication. MR1299979

[16] Johnstone, I.M. (2010). High dimensional Bernstein–von Mises: Simple examples. In *Borrowing Strength: Theory Powering Applications—A Festschrift for Lawrence D. Brown. Inst. Math. Stat. (IMS) Collect.* **6** 87–98. Beachwood, OH: IMS. MR2798513

[17] Ledoux, M. and Talagrand, M. (2002). *Probability in Banach Spaces. Isoperimetry and Processes*. Berlin: Springer. MR2814399

[18] Li, W.V. and Shao, Q.-M. (2001). Gaussian processes: Inequalities, small ball probabilities and applications. In *Stochastic Processes: Theory and Methods. Handbook of Statist.* **19** 533–597. Amsterdam: North-Holland. MR1861734

[19] Lifshits, M. (2012). *Lectures on Gaussian Processes. SpringerBriefs in Mathematics*. Heidelberg: Springer. MR3024389

[20] Markus, A.S. (1964). The eigen- and singular values of the sum and product of linear operators. *Russian Math. Surveys* **19** 91–120. MR0169063

[21] Naumov, A.A., Spokoiny, V.G., Tavyrikov, Yu.E. and Ulyanov, V.V. (2018). Nonasymptotic estimates for the closeness of Gaussian measures on balls. *Dokl. Math.* **98**.

[22] Naumov, A.A., Spokoiny, V.G. and Ulyanov, V.V. (2018). Confidence sets for spectral projectors of covariance matrices. *Dokl. Math.* **98**.

[23] Panov, M. and Spokoiny, V. (2015). Finite sample Bernstein–von Mises theorem for semiparametric problems. *Bayesian Anal.* **10** 665–710. MR3420819

[24] Prokhorov, Y.V. and Ulyanov, V.V. (2013). Some approximation problems in statistics and probability. In *Limit Theorems in Probability, Statistics and Number Theory. Springer Proc. Math. Stat.* **42** 235–249. Heidelberg: Springer. MR3079145

[25] Rudelson, M. and Vershynin, R. (2008). The Littlewood–Offord problem and invertibility of random matrices. *Adv. Math*. **218** 600–633. MR2407948
[26] Sazonov, V.V. (1981). *Normal Approximation—Some Recent Advances*. *Lecture Notes in Math*. **879**. Berlin–New York: Springer. MR0643968
[27] Spokoiny, V. (2017). Penalized maximum likelihood estimation and effective dimension. *Ann*. *Inst*. *Henri Poincaré Probab. Stat*. **53** 389–429. MR3606746
[28] Spokoiny, V. and Zhilova, M. (2015). Bootstrap confidence sets under model misspecification. *Ann*. *Statist*. **43** 2653–2675. MR3405607
[29] Tropp, J.A. (2012). User-friendly tail bounds for sums of random matrices. *Found*. *Comput*. *Math*. **12** 389–434. MR2946459
[30] Ul'yanov, V. (1987). Asymptotic expansions for distributions of sums of independent random variables in H. *Theory Probab. Appl*. **31** 25–39. MR0836950
[31] Yurinsky, V. (1995). *Sums and Gaussian Vectors*. *Lecture Notes in Math*. **1617**. Berlin: Springer. MR1442713

Received March 2018 and revised August 2018