ISSN 1066-5307, Mathematical Methods of Statistics, 2013, Vol. 22, No. 2, pp. 100–113. © Allerton Press, Inc., 2013.

Sharp Deviation Bounds for Quadratic Forms

V. Spokoiny[1*] **and M. Zhilova**[2**]

[1]Moscow Inst. of Physics and Technology, Weierstrass-Inst. and Humboldt Univ. Berlin, Germany

[2]Moscow Inst. of Physics and Technology and Weierstrass-Inst. Berlin, Germany

Received February 3, 2013; in final form, May 12, 2013

Abstract—This paper presents sharp inequalities for deviation probability of a general quadratic form of a random vector $\boldsymbol{\xi}$ with finite exponential moments. The obtained deviation bounds are similar to the case of a Gaussian random vector. The results are stated under general conditions and do not suppose any special structure of the vector $\boldsymbol{\xi}$. The obtained bounds are exact (non-asymptotic), all constants are explicit and the leading terms in the bounds are sharp.

Keywords: quadratic forms, deviation bounds.

2000 Mathematics Subject Classification: primary 60F10; secondary 62F10.

DOI: 10.3103/S1066530713020026

1. INTRODUCTION

This paper presents a number of deviation probability bounds for a quadratic form $\|\boldsymbol{\xi}\|^2$ or more generally $\|I\!\!B\,\boldsymbol{\xi}\|^2$ of a random p-vector $\boldsymbol{\xi}$ satisfying a general exponential moment condition. Such quadratic forms arise in many applications. Baraud (2010) lists some statistical tasks relying on such deviation bounds including hypothesis testing for linear models or linear model selection. We also refer to Massart (2007) for an extensive overview and numerous results on probability bounds and their applications in statistical model selection. Limit theorems for quadratic forms can be found, e.g., in Götze and Tikhomirov (1999) and Horváth and Shao (1999). Some concentration bounds for U-statistics are available in Bretagnolle (1999), Giné et al. (2000), Houdré and Reynaud-Bouret (2003). Most of results assume that the components of the vector $\boldsymbol{\xi}$ are independent and bounded.

Hsu, Kakade and Zhang (2012) study the tail behavior of the quadratic form under the condition of sub-Gaussianity of the random vector $\boldsymbol{\xi}$ and show that the deviation probabilities are essentially the same as in the Gaussian case. However, the assumption that the vector $\boldsymbol{\xi}$ has finite exponential moments of arbitrary order is quite strict and is not fulfilled in many applications. A particular example is given by the Poisson and exponential cases. In the present work we only suppose that some exponential moments of $\boldsymbol{\xi}$ are finite. This makes the problem much more involved and requires new approaches and tools.

If $\boldsymbol{\xi}$ is standard normal then $\|\boldsymbol{\xi}\|^2$ is chi-squared with p degrees of freedom. We aim to extend this behavior to the case of a general vector $\boldsymbol{\xi}$ satisfying the following exponential moment condition:

$$\log I\!\!E \exp\bigl(\boldsymbol{\gamma}^\top \boldsymbol{\xi}\bigr) \le \|\boldsymbol{\gamma}\|^2/2, \qquad \boldsymbol{\gamma} \in I\!\!R^p, \quad \|\boldsymbol{\gamma}\| \le \mathtt{g}. \tag{1.1}$$

Here $\mathtt{g}$ is a positive constant which appears to be very important in our results. Namely, it determines the frontier between the Gaussian and non-Gaussian type deviation bounds. Our first result shows that under (1.1) the deviation bounds for the quadratic form $\|\boldsymbol{\xi}\|^2$ are essentially the same as in the Gaussian case, if the value $\mathtt{g}^2$ exceeds $\mathtt{C}p$ for a fixed constant $\mathtt{C}$. Further we extend the result to the case of a more general form $\|I\!\!B\boldsymbol{\xi}\|^2$. An important advantage of the approach of this paper which differs it from all the previous studies is that there are no additional conditions on the structure or origin of the vector $\boldsymbol{\xi}$. For instance, we do not assume that $\boldsymbol{\xi}$ is a sum of independent or weakly dependent random variables,

*E-mail: spokoiny@wias-berlin.de
**E-mail: zhilova@wias-berlin.de

or components of $\boldsymbol{\xi}$ are independent. The results are exact, stated in a non-asymptotic fashion, all the constants are explicit and the leading terms are sharp.

As a motivating example, we consider a linear regression model $\boldsymbol{Y} = \Psi^\top \boldsymbol{\theta}^* + \boldsymbol{\varepsilon}$ in which $\boldsymbol{Y}$ is an n-vector of observations, $\boldsymbol{\varepsilon}$ is the vector of errors with zero mean, and Ψ is a $p \times n$ design matrix. The ordinary least squares estimator $\tilde{\boldsymbol{\theta}}$ for the parameter vector $\boldsymbol{\theta}^* \in I\!R^p$ reads as

$$\tilde{\boldsymbol{\theta}} = \bigl(\Psi\Psi^\top\bigr)^{-1}\Psi\boldsymbol{Y}$$

and it can be viewed as the maximum likelihood estimator in a Gaussian linear model with a diagonal covariance matrix, that is, $\boldsymbol{Y} \sim \mathcal{N}(\Psi^\top\boldsymbol{\theta}, \sigma^2 I\!I_n)$. Define the $p \times p$ matrix

$$D_0^2 \stackrel{\text{def}}{=} \Psi\Psi^\top.$$

Then

$$D_0(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) = D_0^{-1}\boldsymbol{\zeta}$$

with $\boldsymbol{\zeta} \stackrel{\text{def}}{=} \Psi\boldsymbol{\varepsilon}$. The likelihood ratio test statistic for this problem is exactly $\|D_0^{-1}\boldsymbol{\zeta}\|^2/2$. Similarly, the model selection procedure is based on comparing such quadratic forms for different matrices D_0; see, e.g., Baraud (2010).

Now we indicate how this situation can be reduced to a bound for a vector $\boldsymbol{\xi}$ satisfying condition (1.1). Suppose for simplicity that the entries ε_i of the error vector $\boldsymbol{\varepsilon}$ are independent and have exponential moments.

$(\boldsymbol{e_1})$ *There exist some constants ν_0 and $\mathtt{g}_1 > 0$, and for every i a constant $\mathfrak{s}_i$ such that $I\!E\bigl(\varepsilon_i/\mathfrak{s}_i\bigr)^2 \le 1$ and*

$$\log I\!E \exp\bigl(\lambda\varepsilon_i/\mathfrak{s}_i\bigr) \le \nu_0^2\lambda^2/2, \qquad |\lambda| \le \mathtt{g}_1. \tag{1.2}$$

Here $\mathtt{g}_1$ is a fixed positive constant. One can show that if this condition is fulfilled for some $\mathtt{g}_1 > 0$ and a constant $\nu_0 \ge 1$, then one can get a similar condition with ν_0 arbitrary close to one and $\mathtt{g}_1$ slightly decreased. A natural candidate for $\mathfrak{s}_i$ is σ_i where $\sigma_i^2 = I\!E\varepsilon_i^2$ is the variance of ε_i. Under (1.2), introduce a $p \times p$ matrix V_0 defined by

$$V_0^2 \stackrel{\text{def}}{=} \sum \mathfrak{s}_i^2 \Psi_i \Psi_i^\top,$$

where $\Psi_1, \ldots, \Psi_n \in I\!R^p$ are the columns of the matrix Ψ. Define also

$$\boldsymbol{\xi} = V_0^{-1}\Psi\boldsymbol{\varepsilon}, \qquad N^{-1/2} \stackrel{\text{def}}{=} \max_i \sup_{\boldsymbol{\gamma} \in I\!R^p} \frac{\mathfrak{s}_i|\Psi_i^\top\boldsymbol{\gamma}|}{\|V_0\boldsymbol{\gamma}\|}.$$

A simple calculation shows that for $\|\boldsymbol{\gamma}\| \le \mathtt{g} = \mathtt{g}_1 N^{1/2}$

$$\log I\!E \exp\bigl(\boldsymbol{\gamma}^\top\boldsymbol{\xi}\bigr) \le \nu_0^2\|\boldsymbol{\gamma}\|^2/2, \qquad \boldsymbol{\gamma} \in I\!R^p, \quad \|\boldsymbol{\gamma}\| \le \mathtt{g}.$$

We conclude that (1.1) is nearly fulfilled under (e_1) and moreover, the value $\mathtt{g}^2$ is proportional to the effective sample size N. The results of the paper allow us to get a nearly χ^2-behavior of the test statistic $\|\boldsymbol{\xi}\|^2$ which is a finite sample version of the famous Wilks phenomenon; see, e.g., Fan et al. (2001); Fan and Huang (2005), Boucheron and Massart (2011).

The paper is organized as follows. Section 2 recalls the classical results about deviation probability of a Gaussian quadratic form. These results are presented only for comparison and to make the paper self-contained.

Section 3 studies the probability of the form $I\!P\bigl(\|\boldsymbol{\xi}\| > \mathtt{y}\bigr)$ under the condition

$$\log I\!E \exp\bigl(\boldsymbol{\gamma}^\top\boldsymbol{\xi}\bigr) \le \nu_0^2\|\boldsymbol{\gamma}\|^2/2, \qquad \boldsymbol{\gamma} \in I\!R^p, \quad \|\boldsymbol{\gamma}\| \le \mathtt{g}.$$

The general case can be reduced to $\nu_0 = 1$ by rescaling $\boldsymbol{\xi}$ and $\mathtt{g}$:

$$\log I\!E \exp\bigl(\boldsymbol{\gamma}^\top\boldsymbol{\xi}/\nu_0\bigr) \le \|\boldsymbol{\gamma}\|^2/2, \qquad \boldsymbol{\gamma} \in I\!R^p, \quad \|\boldsymbol{\gamma}\| \le \nu_0\mathtt{g},$$

that is, $\nu_0^{-1}\boldsymbol{\xi}$ fulfills (1.1) with a slightly increased $\mathtt{g}$.

The obtained result is extended to the case of a general quadratic form in Section 4. Some more extensions motivated by different statistical problems are given in Sections 6 and 7. They include the bound with sup-norm constraint and the bound under Bernstein conditions. Among the statistical problems demanding such bounds is estimation of the regression model with Poissonian or bounded random noise. More examples can be found in Baraud (2010). All the proofs are collected in the Appendix.

2. GAUSSIAN CASE

Our benchmark will be a deviation bound for $\|\boldsymbol{\xi}\|^2$ for a standard Gaussian vector $\boldsymbol{\xi}$. The ultimate goal is to show that under (1.1) the norm of the vector $\boldsymbol{\xi}$ exhibits behavior expected for a Gaussian vector, at least in the region of moderate deviations. For the reason of comparison, we begin by stating the result for a Gaussian vector $\boldsymbol{\xi}$. We use the notation $a \vee b$ for the maximum of a and b, while $a \wedge b = \min\{a, b\}$.

Theorem 2.1. *Let $\boldsymbol{\xi}$ be a standard normal vector in $I\!R^p$. Then for any $u > 0$, it holds*

$$I\!P\bigl(\|\boldsymbol{\xi}\|^2 > p + u\bigr) \le \exp\bigl\{-(p/2)\phi(u/p)\bigr]\bigr\}$$

with

$$\phi(t) \stackrel{\text{def}}{=} t - \log(1+t).$$

Let $\phi^{-1}(\cdot)$ stand for the inverse of $\phi(\cdot)$. For any $\mathtt{x}$,

$$I\!P\bigl(\|\boldsymbol{\xi}\|^2 > p + p\,\phi^{-1}(2\mathtt{x}/p)\bigr) \le \exp(-\mathtt{x}).$$

This particularly yields with $\kappa = 6.6$

$$I\!P\bigl(\|\boldsymbol{\xi}\|^2 > p + \sqrt{\kappa \mathtt{x} p} \vee (\kappa \mathtt{x})\bigr) \le \exp(-\mathtt{x}).$$

This is a simple version of a well-known result and we present it only for comparison with the non-Gaussian case. The message of this result is that the squared norm of the Gaussian vector $\boldsymbol{\xi}$ concentrates around the value p and its deviation over the level $p + \sqrt{\mathtt{x}p}$ is exponentially small in $\mathtt{x}$.

A similar bound can be obtained for a norm of the vector $I\!B\boldsymbol{\xi}$, where $I\!B$ is some given deterministic matrix. For notational simplicity we assume that $I\!B$ is symmetric. Otherwise one should replace it with $(I\!B^\top I\!B)^{1/2}$.

Theorem 2.2. *Let $\boldsymbol{\xi}$ be standard normal in $I\!R^p$. Then for every $\mathtt{x} > 0$ and any symmetric matrix $I\!B$, it holds with $\mathtt{p} = \operatorname{tr}(I\!B^2)$, $\mathtt{v}^2 = 2\operatorname{tr}(I\!B^4)$, and $a^* = \|I\!B^2\|_\infty$*

$$I\!P\bigl(\|I\!B\boldsymbol{\xi}\|^2 > \mathtt{p} + (2\mathtt{v}\mathtt{x}^{1/2}) \vee (6a^*\mathtt{x})\bigr) \le \exp(-\mathtt{x}).$$

Below we establish similar bounds for a non-Gaussian vector $\boldsymbol{\xi}$ obeying (1.1).

3. A BOUND FOR THE ℓ_2-NORM

This section presents a general exponential bound for the probability $I\!P\bigl(\|\boldsymbol{\xi}\| > \mathtt{y}\bigr)$ under (1.1). The main result tells us that if $\mathtt{y}$ is not too large, namely if $\mathtt{y} \le \mathtt{y}_c$ with $\mathtt{y}_c^2 \asymp \mathtt{g}^2$, then the deviation probability is essentially the same as in the Gaussian case.

To describe the value $\mathtt{y}_c$, introduce the following notation. Given $\mathtt{g}$ and p, define the values $w_0 = \mathtt{g}p^{-1/2}$ and w_c by the equation

$$\frac{w_c(1+w_c)}{(1+w_c^2)^{1/2}} = w_0 = \mathtt{g}p^{-1/2}. \tag{3.1}$$

It is easy to see that $w_0/\sqrt{2} \le w_c \le w_0$. Further define

$$\mu_c \stackrel{\text{def}}{=} w_c^2/(1+w_c^2), \qquad \mathtt{y}_c \stackrel{\text{def}}{=} \sqrt{(1+w_c^2)p}, \qquad \mathtt{x}_c \stackrel{\text{def}}{=} 0.5p\bigl[w_c^2 - \log\bigl(1+w_c^2\bigr)\bigr]. \tag{3.2}$$

Note that for $\mathtt{g}^2 \ge p$, the quantities $\mathtt{y}_c$ and $\mathtt{x}_c$ can be evaluated as $\mathtt{y}_c^2 \ge w_c^2 p \ge \mathtt{g}^2/2$ and $\mathtt{x}_c \gtrsim pw_c^2/2 \ge \mathtt{g}^2/4$.

Theorem 3.1. *Let* $\boldsymbol{\xi} \in I\!R^p$ *fulfill* (1.1). *Then it holds for each* $\mathtt{x} \le \mathtt{x}_c$

$$I\!P\bigl(\|\boldsymbol{\xi}\|^2 > p + \sqrt{\kappa \mathtt{x} p} \vee (\kappa \mathtt{x}),\ \|\boldsymbol{\xi}\| \le \mathtt{y}_c\bigr) \le 2\exp(-\mathtt{x}),$$

where $\kappa = 6.6$. *Moreover, for* $\mathtt{y} \ge \mathtt{y}_c$, *it holds with* $\mathtt{g}_c = \mathtt{g} - \sqrt{\mu_c p} = \mathtt{g} w_c/(1 + w_c)$

$$\begin{aligned} I\!P\bigl(\|\boldsymbol{\xi}\| > \mathtt{y}\bigr) &\le 8.4\exp\bigl\{-\mathtt{g}_c \mathtt{y}/2 - (p/2)\log(1 - \mathtt{g}_c/\mathtt{y})\bigr\} \\ &\le 8.4\exp\bigl\{-\mathtt{x}_c - \mathtt{g}_c(\mathtt{y} - \mathtt{y}_c)/2\bigr\}. \end{aligned}$$

The statements of Theorem 4.1 can be simplified under the assumption $\mathtt{g}^2 \ge p$.

Corollary 3.2. *Let* $\boldsymbol{\xi}$ *fulfill* (1.1) *and* $\mathtt{g}^2 \ge p$. *Then it holds for* $\mathtt{x} \le \mathtt{x}_c$

$$I\!P\bigl(\|\boldsymbol{\xi}\|^2 \ge \mathfrak{z}(\mathtt{x}, p)\bigr) \le 2\mathrm{e}^{-\mathtt{x}} + 8.4\mathrm{e}^{-\mathtt{x}_c}, \tag{3.3}$$

$$\mathfrak{z}(\mathtt{x}, p) \stackrel{\text{def}}{=} \begin{cases} p + \sqrt{\kappa \mathtt{x} p}, & \mathtt{x} \le p/\kappa, \\ p + \kappa \mathtt{x}, & p/\kappa < \mathtt{x} \le \mathtt{x}_c, \end{cases} \tag{3.4}$$

with $\kappa = 6.6$. *For* $\mathtt{x} > \mathtt{x}_c$

$$I\!P\bigl(\|\boldsymbol{\xi}\|^2 \ge \mathfrak{z}_c(\mathtt{x}, p)\bigr) \le 8.4\mathrm{e}^{-\mathtt{x}}, \qquad \mathfrak{z}_c(\mathtt{x}, p) \stackrel{\text{def}}{=} \bigl|\mathtt{y}_c + 2(\mathtt{x} - \mathtt{x}_c)/\mathtt{g}_c\bigr|^2.$$

This result implicitly assumes that $p \le \kappa \mathtt{x}_c$, which is fulfilled if $w_0^2 = \mathtt{g}^2/p \ge 1$:

$$\kappa \mathtt{x}_c = 0.5\kappa\bigl[w_0^2 - \log(1 + w_0^2)\bigr]p \ge 3.3\bigl[1 - \log(2)\bigr]p > p.$$

For $\mathtt{x} \le \mathtt{x}_c$, the function $\mathfrak{z}(\mathtt{x}, p)$ mimics the quantile behavior of the chi-squared distribution χ_p^2 with p degrees of freedom. Moreover, increase of the value $\mathtt{g}$ yields a growth of the sub-Gaussian zone. In particular, for $\mathtt{g} = \infty$, a general quadratic form $\|\boldsymbol{\xi}\|^2$ has under (1.1) the same tail behavior as in the Gaussian case.

Finally, in the large deviation zone $\mathtt{x} > \mathtt{x}_c$ the deviation probability decays as $\mathrm{e}^{-c\mathtt{x}^{1/2}}$ for some fixed c. However, if the constant $\mathtt{g}$ in condition (1.1) is sufficiently large relative to p, then $\mathtt{x}_c$ is large as well and the large deviation zone $\mathtt{x} > \mathtt{x}_c$ can be ignored at a small price of $8.4\mathrm{e}^{-\mathtt{x}_c}$ and one can focus on the deviation bound described by (3.3) and (3.4).

4. A BOUND FOR A QUADRATIC FORM

Now we extend the result to more general bound for $\|I\!B\boldsymbol{\xi}\|^2 = \boldsymbol{\xi}^\top I\!B^2 \boldsymbol{\xi}$ with a given matrix $I\!B$ and a vector $\boldsymbol{\xi}$ obeying condition (1.1). Similarly to the Gaussian case we assume that $I\!B$ is symmetric. Define important characteristics of $I\!B$

$$\mathtt{p} = \operatorname{tr}(I\!B^2), \qquad \mathtt{v}^2 = 2\operatorname{tr}(I\!B^4), \qquad \lambda^* \stackrel{\text{def}}{=} \|I\!B^2\|_\infty \stackrel{\text{def}}{=} \lambda_{\max}(I\!B^2).$$

For simplicity of formulation we suppose that $\lambda^* = 1$, otherwise one has to replace $\mathtt{p}$ and $\mathtt{v}^2$ with $\mathtt{p}/\lambda^*$ and $\mathtt{v}^2/\lambda^*$.

Let $\mathtt{g}$ be shown in (1.1). Similarly to the ℓ_2-case define w_c by the equation

$$\frac{w_c(1 + w_c)}{(1 + w_c^2)^{1/2}} = \mathtt{g}\mathtt{p}^{-1/2}.$$

Define also $\mu_c = w_c^2/(1 + w_c^2) \wedge 2/3$. Note that $w_c^2 \ge 2$ implies $\mu_c = 2/3$. Further, define

$$\mathtt{y}_c^2 = (1 + w_c^2)\mathtt{p}, \qquad 2\mathtt{x}_c = \mu_c \mathtt{y}_c^2 + \log\det\{I\!\!I_p - \mu_c I\!B^2\}. \tag{4.1}$$

Similarly to the case with $I\!B = I\!\!I_p$, under the condition $\mathtt{g}^2 \ge \mathtt{p}$, one can bound $\mathtt{y}_c^2 \ge \mathtt{g}^2/2$ and $\mathtt{x}_c \gtrsim \mathtt{g}^2/4$.

Theorem 4.1. *Let a random vector* $\boldsymbol{\xi}$ *in* $\mathbb{R}^p$ *fulfill* (1.1). *Then for each* $\mathtt{x} < \mathtt{x}_c$

$$\mathbb{P}\big(\|\mathbb{B}\boldsymbol{\xi}\|^2 > \mathtt{p} + (2\mathtt{v}\mathtt{x}^{1/2}) \vee (6\mathtt{x}),\ \|\mathbb{B}\boldsymbol{\xi}\| \le \mathtt{y}_c\big) \le 2\exp(-\mathtt{x}).$$

Moreover, for $\mathtt{y} \ge \mathtt{y}_c$, *with* $\mathtt{g}_c = \mathtt{g} - \sqrt{\mu_c \mathtt{p}} = \mathtt{g} w_c/(1 + w_c)$, *it holds*

$$\mathbb{P}\big(\|\mathbb{B}\boldsymbol{\xi}\| > \mathtt{y}\big) \le 8.4\exp\big(-\mathtt{x}_c - \mathtt{g}_c(\mathtt{y} - \mathtt{y}_c)/2\big).$$

Now we describe the value $\mathfrak{z}(\mathtt{x}, \mathbb{B})$ ensuring a small value for the large deviation probability $\mathbb{P}\big(\|\mathbb{B}\boldsymbol{\xi}\|^2 > \mathfrak{z}(\mathtt{x}, \mathbb{B})\big)$. For ease of formulation, we suppose that $\mathtt{g}^2 \ge 2\mathtt{p}$ yielding $\mu_c^{-1} \le 3/2$. The other case can be easily adjusted.

Corollary 4.2. *Let* $\boldsymbol{\xi}$ *fulfill* (1.1) *with* $\mathtt{g}^2 \ge 2\mathtt{p}$. *Then it holds for* $\mathtt{x} \le \mathtt{x}_c$ *with* $\mathtt{x}_c$ *from* (4.1):

$$\mathbb{P}\big(\|\mathbb{B}\boldsymbol{\xi}\|^2 \ge \mathfrak{z}(\mathtt{x}, \mathbb{B})\big) \le 2\mathrm{e}^{-\mathtt{x}} + 8.4\mathrm{e}^{-\mathtt{x}_c},$$

$$\mathfrak{z}(\mathtt{x}, \mathbb{B}) \stackrel{\text{def}}{=} \begin{cases} \mathtt{p} + 2\mathtt{v}\mathtt{x}^{1/2}, & \mathtt{x} \le \mathtt{v}/18, \\ \mathtt{p} + 6\mathtt{x}, & \mathtt{v}/18 < \mathtt{x} \le \mathtt{x}_c. \end{cases} \tag{4.2}$$

For $\mathtt{x} > \mathtt{x}_c$

$$\mathbb{P}\big(\|\mathbb{B}\boldsymbol{\xi}\|^2 \ge \mathfrak{z}_c(\mathtt{x}, \mathbb{B})\big) \le 8.4\mathrm{e}^{-\mathtt{x}}, \qquad \mathfrak{z}_c(\mathtt{x}, \mathbb{B}) \stackrel{\text{def}}{=} \big|\mathtt{y}_c + 2(\mathtt{x} - \mathtt{x}_c)/\mathtt{g}_c\big|^2.$$

5. RESCALING AND REGULARITY CONDITION

The result of Theorem 4.1 can be extended to a more general situation when condition (1.1) is fulfilled for a vector $\boldsymbol{\zeta}$ rescaled by a matrix V_0. More precisely, let the random p-vector $\boldsymbol{\zeta}$ fulfills for some $p \times p$ matrix V_0 the condition

$$\sup_{\boldsymbol{\gamma} \in \mathbb{R}^p} \log \mathbb{E} \exp\Big(\lambda \frac{\boldsymbol{\gamma}^\top \boldsymbol{\zeta}}{\|V_0 \boldsymbol{\gamma}\|}\Big) \le \nu_0^2 \lambda^2/2, \qquad |\lambda| \le \mathtt{g}, \tag{5.1}$$

with some constants $\mathtt{g} > 0$, $\nu_0 \ge 1$. Again, a simple change of variables reduces the case of an arbitrary $\nu_0 \ge 1$ to $\nu_0 = 1$. Our aim is to bound the squared norm $\|D_0^{-1}\boldsymbol{\zeta}\|^2$ of a vector $D_0^{-1}\boldsymbol{\zeta}$ for another $p \times p$ positive symmetric matrix D_0^2. Note that condition (5.1) implies (1.1) for the rescaled vector $\boldsymbol{\xi} = V_0^{-1}\boldsymbol{\zeta}$. This leads to bounding the quadratic form $\|D_0^{-1}V_0\boldsymbol{\xi}\|^2 = \|\mathbb{B}\boldsymbol{\xi}\|^2$ with $\mathbb{B}^2 = D_0^{-1}V_0^2 D_0^{-1}$. It obviously holds

$$\mathtt{p} = \operatorname{tr}(\mathbb{B}^2) = \operatorname{tr}(D_0^{-2}V_0^2).$$

Now we can apply the result of Corollary 4.2.

Corollary 5.1. *Let* $\boldsymbol{\zeta}$ *fulfill* (5.1) *with some* V_0 *and* $\mathtt{g}$. *Given* D_0, *define* $\mathbb{B}^2 = D_0^{-1}V_0^2 D_0^{-1}$, *and let* $\mathtt{g}^2 \ge 2\mathtt{p}$. *Then it holds for* $\mathtt{x} \le \mathtt{x}_c$ *with* $\mathtt{x}_c$ *from* (4.1):

$$\mathbb{P}\big(\|D_0^{-1}\boldsymbol{\zeta}\|^2 \ge \mathfrak{z}(\mathtt{x}, \mathbb{B})\big) \le 2\mathrm{e}^{-\mathtt{x}} + 8.4\mathrm{e}^{-\mathtt{x}_c}$$

with $\mathfrak{z}(\mathtt{x}, \mathbb{B})$ *from* (4.2). *For* $\mathtt{x} > \mathtt{x}_c$

$$\mathbb{P}\big(\|D_0^{-1}\boldsymbol{\zeta}\|^2 \ge \mathfrak{z}_c(\mathtt{x}, \mathbb{B})\big) \le 8.4\mathrm{e}^{-\mathtt{x}}, \qquad \mathfrak{z}_c(\mathtt{x}, \mathbb{B}) \stackrel{\text{def}}{=} \big|\mathtt{y}_c + 2(\mathtt{x} - \mathtt{x}_c)/\mathtt{g}_c\big|^2.$$

In the *regular* case with $D_0 \ge \mathfrak{a} V_0$ for some $\mathfrak{a} > 0$, one obtains $\|\mathbb{B}\|_\infty \le \mathfrak{a}^{-1}$ and

$$\mathtt{v}^2 = 2\operatorname{tr}(\mathbb{B}^4) \le 2\mathfrak{a}^{-2}\mathtt{p}.$$

6. A CHI-SQUARED BOUND WITH NORM-CONSTRAINTS

This section extends the results to the case when the bound (1.1) requires some other conditions than the ℓ_2-norm of the vector $\boldsymbol{\gamma}$. Namely, we suppose that

$$\log I\!\!E \exp\bigl(\boldsymbol{\gamma}^\top \boldsymbol{\xi}\bigr) \le \|\boldsymbol{\gamma}\|^2/2, \qquad \boldsymbol{\gamma} \in I\!\!R^p, \quad \|\boldsymbol{\gamma}\|_\circ \le \mathtt{g}_\circ, \tag{6.1}$$

where $\|\cdot\|_\circ$ is a norm which differs from the usual Euclidean norm. Our driving example is given by the sup-norm case with $\|\boldsymbol{\gamma}\|_\circ \equiv \|\boldsymbol{\gamma}\|_\infty$. We are interested to check whether the previous results of Section 3 still apply. The answer depends on how massive the set $\mathcal{A}(r) = \{\boldsymbol{\gamma} \colon \|\boldsymbol{\gamma}\|_\circ \le r\}$ is in terms of the standard Gaussian measure on $I\!\!R^p$. Recall that the quadratic norm $\|\boldsymbol{\varepsilon}\|^2$ of a standard Gaussian vector $\boldsymbol{\varepsilon}$ in $I\!\!R^p$ concentrates around p at least for p large. We need a similar concentration property for the norm $\|\cdot\|_\circ$. More precisely, we assume for a fixed r_* that

$$I\!\!P\bigl(\|\boldsymbol{\varepsilon}\|_\circ \le r_*\bigr) \ge 1/2, \qquad \boldsymbol{\varepsilon} \sim \mathcal{N}(0, I\!\!I_p). \tag{6.2}$$

This implies for any value $\mathtt{u}_\circ > 0$ and all $\boldsymbol{u} \in I\!\!R^p$ with $\|\boldsymbol{u}\|_\circ \le \mathtt{u}_\circ$ that

$$I\!\!P\bigl(\|\boldsymbol{\varepsilon} - \boldsymbol{u}\|_\circ \le r_* + \mathtt{u}_\circ\bigr) \ge 1/2, \qquad \boldsymbol{\varepsilon} \sim \mathcal{N}(0, I\!\!I_p).$$

For each $\mathfrak{z} > p$, consider

$$\mu(\mathfrak{z}) = (\mathfrak{z} - p)/\mathfrak{z}.$$

Given $\mathtt{u}_\circ$, denote by $\mathfrak{z}_\circ = \mathfrak{z}_\circ(\mathtt{u}_\circ)$ the root of the equation

$$\frac{\mathtt{g}_\circ}{\mu(\mathfrak{z}_\circ)} - \frac{r_*}{\mu^{1/2}(\mathfrak{z}_\circ)} = \mathtt{u}_\circ. \tag{6.3}$$

One can easily see that this value exists and is unique if $\mathtt{u}_\circ \ge \mathtt{g}_\circ - r_*$ and it can be defined as the largest $\mathfrak{z}$ for which

$$\frac{\mathtt{g}_\circ}{\mu(\mathfrak{z})} - \frac{r_*}{\mu^{1/2}(\mathfrak{z})} \ge \mathtt{u}_\circ.$$

Let $\mu_\circ = \mu(\mathfrak{z}_\circ)$ be the corresponding μ-value. Define also $\mathtt{x}_\circ$ by

$$2\mathtt{x}_\circ = \mu_\circ \mathfrak{z}_\circ + p \log(1 - \mu_\circ).$$

If $\mathtt{u}_\circ < \mathtt{g}_\circ - r_*$, then set $\mathfrak{z}_\circ = \infty$, $\mathtt{x}_\circ = \infty$.

Theorem 6.1. *Let a random vector $\boldsymbol{\xi}$ in $I\!\!R^p$ fulfill* (6.1). *Suppose* (6.2) *and let, given $\mathtt{u}_\circ$, the value $\mathfrak{z}_\circ$ be defined by* (6.3). *Then it holds for any $u > 0$*

$$I\!\!P\bigl(\|\boldsymbol{\xi}\|^2 > p + u, \|\boldsymbol{\xi}\|_\circ \le \mathtt{u}_\circ\bigr) \le 2\exp\bigl\{-(p/2)\phi(u)\bigr]\bigr\} \tag{6.4}$$

yielding for $\mathtt{x} \le \mathtt{x}_\circ$

$$I\!\!P\bigl(\|\boldsymbol{\xi}\|^2 > p + \sqrt{\kappa \mathtt{x} p} \vee (\kappa \mathtt{x}),\ \|\boldsymbol{\xi}\|_\circ \le \mathtt{u}_\circ\bigr) \le 2\exp(-\mathtt{x}), \tag{6.5}$$

where $\kappa = 6.6$. Moreover, for $\mathfrak{z} \ge \mathfrak{z}_\circ$, it holds

$$\begin{aligned} I\!\!P\bigl(\|\boldsymbol{\xi}\|^2 > \mathfrak{z},\ \|\boldsymbol{\xi}\|_\circ \le \mathtt{u}_\circ\bigr) &\le 2\exp\bigl\{-\mu_\circ \mathfrak{z}/2 - (p/2)\log(1 - \mu_\circ)\bigr\} \\ &= 2\exp\bigl\{-\mathtt{x}_\circ - \mathtt{g}_\circ(\mathfrak{z} - \mathfrak{z}_\circ)/2\bigr\}. \end{aligned}$$

It is easy to check that the result continues to hold for the norm of $\Pi\boldsymbol{\xi}$ for a given sub-projector Π in $I\!\!R^p$ satisfying $\Pi = \Pi^\top$, $\Pi^2 \le \Pi$. As above, denote $\mathtt{p} \stackrel{\mathrm{def}}{=} \operatorname{tr}(\Pi^2)$, $\mathtt{v}^2 \stackrel{\mathrm{def}}{=} 2\operatorname{tr}(\Pi^4)$. Let r_* be fixed to ensure

$$I\!\!P\bigl(\|\Pi\boldsymbol{\varepsilon}\|_\circ \le r_*\bigr) \ge 1/2, \qquad \boldsymbol{\varepsilon} \sim \mathcal{N}(0, I\!\!I_p).$$

The next result is stated for $\mathtt{g}_\circ \ge r_* + \mathtt{u}_\circ$, which simplifies the formulation.

Theorem 6.2. *Let a random vector $\boldsymbol{\xi}$ in $I\!\!R^p$ fulfill* (6.1) *and let Π follows $\Pi = \Pi^\top$, $\Pi^2 \le \Pi$. Let some $\mathtt{u}_\circ$ be fixed. Then for any $\mu_\circ \le 2/3$ with $\mathtt{g}_\circ \mu_\circ^{-1} - r_* \mu_\circ^{-1/2} \ge \mathtt{u}_\circ$,*

$$I\!\!E \exp\Bigl\{ \frac{\mu_\circ}{2} (\|\Pi \boldsymbol{\xi}\|^2 - \mathtt{p}) \Bigr\} \mathbb{1}\bigl(\|\Pi^2 \boldsymbol{\xi}\|_\circ \le \mathtt{u}_\circ \bigr) \le 2 \exp(\mu_\circ^2 \mathtt{v}^2/4), \tag{6.6}$$

where $\mathtt{v}^2 = 2\operatorname{tr}(\Pi^4)$. Moreover, if $\mathtt{g}_\circ \ge r_* + \mathtt{u}_\circ$, then for any $\mathfrak{z} \ge 0$

$$I\!\!P\bigl(\|\Pi \boldsymbol{\xi}\|^2 > \mathfrak{z}, \|\Pi^2 \boldsymbol{\xi}\|_\circ \le \mathtt{u}_\circ \bigr) \le I\!\!P\bigl(\|\Pi \boldsymbol{\xi}\|^2 > \mathtt{p} + (2\mathtt{v}\mathtt{x}^{1/2}) \vee (6\mathtt{x}),\ \|\Pi^2 \boldsymbol{\xi}\|_\circ \le \mathtt{u}_\circ \bigr) \le 2 \exp(-\mathtt{x}).$$

7. A BOUND FOR THE ℓ_2-NORM UNDER BERNSTEIN CONDITIONS

For comparison, we specify the results to the case considered recently in Baraud (2010). Let $\boldsymbol{\zeta}$ be a random vector in $I\!\!R^n$ whose components ζ_i are independent and satisfy the Bernstein type conditions: for all $|\lambda| < c^{-1}$

$$\log I\!\!E e^{\lambda \zeta_i} \le \frac{\lambda^2 \sigma^2}{1 - c|\lambda|}. \tag{7.1}$$

Denote $\boldsymbol{\xi} = \boldsymbol{\zeta}/(2\sigma)$ and consider $\|\boldsymbol{\gamma}\|_\circ = \|\boldsymbol{\gamma}\|_\infty$. Fix $\mathtt{g}_\circ = \sigma/c$. If $\|\boldsymbol{\gamma}\|_\circ \le \mathtt{g}_\circ$, then $1 - c\gamma_i/(2\sigma) \ge 1/2$ and

$$\log I\!\!E \exp\bigl(\boldsymbol{\gamma}^\top \boldsymbol{\xi} \bigr) \le \sum_i \log I\!\!E \exp\Bigl(\frac{\gamma_i \zeta_i}{2\sigma} \Bigr) \le \sum_i \frac{|\gamma_i/(2\sigma)|^2 \sigma^2}{1 - c\gamma_i/(2\sigma)} \le \|\boldsymbol{\gamma}\|^2/2.$$

Let also S be some linear subspace of $I\!\!R^n$ with dimension $\mathtt{p}$ and Π_S denote the projector on S. For applying the result of Theorem 6.1, the value r_* has to be fixed. We use that the infinity norm $\|\boldsymbol{\varepsilon}\|_\infty$ concentrates around $\sqrt{2\log p}$.

Lemma 7.1. *It holds for a standard normal vector $\boldsymbol{\varepsilon} \in I\!\!R^p$ with $r_* = \sqrt{2\log p}$*

$$I\!\!P\bigl(\|\boldsymbol{\varepsilon}\|_\circ \le r_* \bigr) \ge 1/2.$$

Indeed

$$I\!\!P\bigl(\|\boldsymbol{\varepsilon}\|_\circ > r_* \bigr) \le I\!\!P\bigl(\|\boldsymbol{\varepsilon}\|_\infty > \sqrt{2\log p} \bigr) \le p\, I\!\!P\bigl(|\varepsilon_1| > \sqrt{2\log p} \bigr) \le 1/2.$$

Now the general bound of Theorem 6.1 is applied to bounding the norm of $\|\Pi_S \boldsymbol{\xi}\|$. For simplicity of formulation we assume that $\mathtt{g}_\circ \ge \mathtt{u}_\circ + r_*$.

Theorem 7.2. *Let S be some linear subspace of $I\!\!R^n$ with dimension $\mathtt{p}$. Let $\mathtt{g}_\circ \ge \mathtt{u}_\circ + r_*$. If the coordinates ζ_i of $\boldsymbol{\zeta}$ are independent and satisfy* (7.1)*, then for all $\mathtt{x}$,*

$$I\!\!P\bigl((4\sigma^2)^{-1} \|\Pi_S \boldsymbol{\zeta}\|^2 > \mathtt{p} + \sqrt{\kappa \mathtt{x} \mathtt{p}} \vee (\kappa \mathtt{x}),\ \|\Pi_S \boldsymbol{\zeta}\|_\infty \le 2\sigma \mathtt{u}_\circ \bigr) \le 2 \exp(-\mathtt{x}).$$

The bound of Baraud (2010) reads

$$I\!\!P\Bigl(\|\Pi_S \boldsymbol{\zeta}\|_2 > \bigl(3\sigma \ \vee \ \sqrt{6cu} \bigr) \sqrt{\mathtt{x} + 3\mathtt{p}},\ \|\Pi_S \boldsymbol{\zeta}\|_\infty \le 2\sigma \mathtt{u}_\circ \Bigr) \le e^{-\mathtt{x}}.$$

As expected, in the region $\mathtt{x} \le \mathtt{x}_c$ of Gaussian approximation, the bound of Baraud is not sharp and actually quite rough.

APPENDIX: A. PROOF OF THEOREM 2.1

The proof utilizes the following well-known fact, which can be obtained by straightforward calculus: for $\mu < 1$

$$\log \mathbb{E} \exp\bigl(\mu \|\boldsymbol{\xi}\|^2/2\bigr) = -0.5p \log(1-\mu).$$

Now consider any $u > 0$. By the exponential Chebyshev inequality

$$\begin{aligned} \mathbb{P}\bigl(\|\boldsymbol{\xi}\|^2 > p + u\bigr) &\le \exp\bigl\{-\mu(p+u)/2\bigr\} \mathbb{E} \exp\bigl(\mu \|\boldsymbol{\xi}\|^2/2\bigr) \\ &= \exp\bigl\{-\mu(p+u)/2 - (p/2)\log(1-\mu)\bigr\}. \end{aligned} \tag{A.1}$$

It is easy to see that the value $\mu = u/(u+p)$ maximizes $\mu(p+u) + p\log(1-\mu)$ w.r.t. μ yielding

$$\mu(p+u) + p\log(1-\mu) = u - p\log(1+u/p).$$

Further we use that $x - \log(1+x) \ge a_0 x^2$ for $x \le 1$ and $x - \log(1+x) \ge a_0 x$ for $x > 1$ with $a_0 = 1 - \log(2) \ge 0.3$. This implies with $x = u/p$ for $u = \sqrt{\kappa \mathtt{x} p}$ or $u = \kappa \mathtt{x}$ and $\kappa = 2/a_0 < 6.6$ that

$$\mathbb{P}\bigl(\|\boldsymbol{\xi}\|^2 \ge p + \sqrt{\kappa \mathtt{x} p} \vee (\kappa \mathtt{x})\bigr) \le \exp(-\mathtt{x})$$

as required. □

B. PROOF OF THEOREM 2.2

The matrix $\mathbb{B}^2$ can be represented as $U^\top \operatorname{diag}(a_1, \ldots, a_p) U$ for an orthogonal matrix U. The vector $\tilde{\boldsymbol{\xi}} = U\boldsymbol{\xi}$ is also standard normal and $\|\mathbb{B}\boldsymbol{\xi}\|^2 = \tilde{\boldsymbol{\xi}}^\top U \mathbb{B}^2 U^\top \tilde{\boldsymbol{\xi}}$. This means that one can reduce the situation to the case of a diagonal matrix $\mathbb{B}^2 = \operatorname{diag}(a_1, \ldots, a_p)$. We can also assume without loss of generality that $a_1 \ge a_2 \ge \ldots \ge a_p$. The expressions for the quantities $\mathtt{p}$ and $\mathtt{v}^2$ simplify to

$$\begin{aligned} \mathtt{p} &= \operatorname{tr}(\mathbb{B}^2) = a_1 + \ldots + a_p, \\ \mathtt{v}^2 &= 2\operatorname{tr}(\mathbb{B}^4) = 2(a_1^2 + \ldots + a_p^2). \end{aligned}$$

Moreover, rescaling the matrix $\mathbb{B}^2$ by a_1 reduces the situation to the case with $a_1 = 1$.

Lemma B.1. *It holds*

$$\mathbb{E}\|\mathbb{B}\boldsymbol{\xi}\|^2 = \operatorname{tr}(\mathbb{B}^2), \qquad \operatorname{Var}\bigl(\|\mathbb{B}\boldsymbol{\xi}\|^2\bigr) = 2\operatorname{tr}(\mathbb{B}^4).$$

Moreover, for $\mu < 1$

$$\mathbb{E}\exp\bigl\{\mu\|\mathbb{B}\boldsymbol{\xi}\|^2/2\bigr\} = \det(1 - \mu\mathbb{B}^2)^{-1/2} = \prod_{i=1}^{p}(1 - \mu a_i)^{-1/2}. \tag{B.1}$$

Proof. If $\mathbb{B}^2$ is diagonal, then $\|\mathbb{B}\boldsymbol{\xi}\|^2 = \sum_i a_i \xi_i^2$ and the summands $a_i\xi_i^2$ are independent. It remains to note that $\mathbb{E}(a_i\xi_i^2) = a_i$, $\operatorname{Var}(a_i\xi_i^2) = 2a_i^2$, and for $\mu a_i < 1$,

$$\mathbb{E}\exp\bigl\{\mu a_i \xi_i^2/2\bigr\} = (1 - \mu a_i)^{-1/2}$$

yielding (B.1). □

Given u, fix $\mu < 1$. The exponential Markov inequality yields

$$\begin{aligned} \mathbb{P}\bigl(\|\mathbb{B}\boldsymbol{\xi}\|^2 > \mathtt{p} + u\bigr) &\le \exp\Bigl\{-\frac{\mu(\mathtt{p}+u)}{2}\Bigr\} \mathbb{E}\exp\Bigl(\frac{\mu\|\mathbb{B}\boldsymbol{\xi}\|^2}{2}\Bigr) \\ &\le \exp\Bigl\{-\frac{\mu u}{2} - \frac{1}{2}\sum_{i=1}^{p}\bigl[\mu a_i + \log\bigl(1 - \mu a_i\bigr)\bigr]\Bigr\}. \end{aligned}$$

We start with the case when $\mathtt{x}^{1/2} \le \mathtt{v}/3$. Then $u = 2\mathtt{x}^{1/2}\mathtt{v}$ fulfills $u \le 2\mathtt{v}^2/3$. Define $\mu = u/\mathtt{v}^2 \le 2/3$ and use that $t + \log(1-t) \ge -t^2$ for $t \le 2/3$. This implies

$$I\!\!P\bigl(\|I\!\!B\boldsymbol{\xi}\|^2 > \mathtt{p} + u\bigr) \le \exp\Bigl\{-\frac{\mu u}{2} + \frac{1}{2}\sum_{i=1}^{p}\mu^2 a_i^2\Bigr\} = \exp\bigl(-u^2/(4\mathtt{v}^2)\bigr) = e^{-\mathtt{x}}. \tag{B.2}$$

Next, let $\mathtt{x}^{1/2} > \mathtt{v}/3$. Set $\mu = 2/3$. It holds similarly to the above

$$\sum_{i=1}^{p}\bigl[\mu a_i + \log\bigl(1 - \mu a_i\bigr)\bigr] \ge -\sum_{i=1}^{p}\mu^2 a_i^2 \ge -2\mathtt{v}^2/9 \ge -2\mathtt{x}.$$

Now, for $u = 6\mathtt{x}$ and $\mu u/2 = 2\mathtt{x}$, (B.2) implies

$$I\!\!P\bigl(\|I\!\!B\boldsymbol{\xi}\|^2 > \mathtt{p} + u\bigr) \le \exp\bigl\{-\bigl(2\mathtt{x} - \mathtt{x}\bigr)\bigr\} = \exp(-\mathtt{x})$$

as required. $\square$

C. PROOF OF THEOREM 3.1

The main step of the proof is the following exponential bound.

Lemma C.1. *Suppose* (1.1). *For any* $\mu < 1$ *with* $\mathtt{g}^2 > p\mu$, *it holds*

$$I\!\!E\exp\Bigl(\frac{\mu\|\boldsymbol{\xi}\|^2}{2}\Bigr)\, 1\!\!\mathrm{I}\Bigl(\|\boldsymbol{\xi}\| \le \mathtt{g}/\mu - \sqrt{p/\mu}\Bigr) \le 2(1-\mu)^{-p/2}. \tag{C.1}$$

Proof. Let $\boldsymbol{\varepsilon}$ be a standard normal vector in $I\!\!R^p$ and $\boldsymbol{u} \in I\!\!R^p$. The bound $I\!\!P\bigl(\|\boldsymbol{\varepsilon}\|^2 > p\bigr) \le 1/2$ and the triangle inequality imply for any vector $\boldsymbol{u}$ and any $\mathtt{r}$ with $\mathtt{r} \ge \|\boldsymbol{u}\| + p^{1/2}$ that $I\!\!P\bigl(\|\boldsymbol{u} + \boldsymbol{\varepsilon}\| \le \mathtt{r}\bigr) \ge 1/2$. Let us fix some $\boldsymbol{\xi}$ with $\|\boldsymbol{\xi}\| \le \mathtt{g}/\mu - \sqrt{p/\mu}$ and denote by $I\!\!P_{\boldsymbol{\xi}}$ the conditional probability given $\boldsymbol{\xi}$. The previous arguments yield:

$$I\!\!P_{\boldsymbol{\xi}}\bigl(\|\boldsymbol{\varepsilon} + \mu^{1/2}\boldsymbol{\xi}\| \le \mu^{-1/2}\mathtt{g}\bigr) \ge 0.5.$$

It holds with $c_p = (2\pi)^{-p/2}$

$$\begin{aligned}
&c_p \int \exp\Bigl(\boldsymbol{\gamma}^\top\boldsymbol{\xi} - \frac{\|\boldsymbol{\gamma}\|^2}{2\mu}\Bigr)\, 1\!\!\mathrm{I}(\|\boldsymbol{\gamma}\| \le \mathtt{g})\, d\boldsymbol{\gamma} \\
&\quad = c_p \exp\bigl(\mu\|\boldsymbol{\xi}\|^2/2\bigr) \int \exp\Bigl(-\frac{1}{2}\bigl\|\mu^{-1/2}\boldsymbol{\gamma} - \mu^{1/2}\boldsymbol{\xi}\bigr\|^2\Bigr)\, 1\!\!\mathrm{I}(\mu^{-1/2}\|\boldsymbol{\gamma}\| \le \mu^{-1/2}\mathtt{g})\, d\boldsymbol{\gamma} \\
&\quad = \mu^{p/2}\exp\bigl(\mu\|\boldsymbol{\xi}\|^2/2\bigr)\, I\!\!P_{\boldsymbol{\xi}}\bigl(\|\boldsymbol{\varepsilon} + \mu^{1/2}\boldsymbol{\xi}\| \le \mu^{-1/2}\mathtt{g}\bigr) \ge 0.5\mu^{p/2}\exp\bigl(\mu\|\boldsymbol{\xi}\|^2/2\bigr)
\end{aligned}$$

because $\|\mu^{1/2}\boldsymbol{\xi}\| + p^{1/2} \le \mu^{-1/2}\mathtt{g}$. This implies in view of $p < \mathtt{g}^2/\mu$ that

$$\exp\bigl(\mu\|\boldsymbol{\xi}\|^2/2\bigr)\, 1\!\!\mathrm{I}\bigl(\|\boldsymbol{\xi}\|^2 \le \mathtt{g}/\mu - \sqrt{p/\mu}\bigr) \le 2\mu^{-p/2} c_p \int \exp\Bigl(\boldsymbol{\gamma}^\top\boldsymbol{\xi} - \frac{\|\boldsymbol{\gamma}\|^2}{2\mu}\Bigr)\, 1\!\!\mathrm{I}(\|\boldsymbol{\gamma}\| \le \mathtt{g})\, d\boldsymbol{\gamma}.$$

Further, by (1.1)

$$\begin{aligned}
&c_p I\!\!E \int \exp\Bigl(\boldsymbol{\gamma}^\top\boldsymbol{\xi} - \frac{1}{2\mu}\|\boldsymbol{\gamma}\|^2\Bigr)\, 1\!\!\mathrm{I}(\|\boldsymbol{\gamma}\| \le \mathtt{g})\, d\boldsymbol{\gamma} \\
&\quad \le c_p \int \exp\Bigl(-\frac{\mu^{-1} - 1}{2}\|\boldsymbol{\gamma}\|^2\Bigr)\, 1\!\!\mathrm{I}(\|\boldsymbol{\gamma}\| \le \mathtt{g})\, d\boldsymbol{\gamma} \\
&\quad \le c_p \int \exp\Bigl(-\frac{\mu^{-1} - 1}{2}\|\boldsymbol{\gamma}\|^2\Bigr)\, d\boldsymbol{\gamma} \le (\mu^{-1} - 1)^{-p/2}
\end{aligned}$$

and (C.1) follows. $\square$

Due to this result, the scaled squared norm $\mu\|\boldsymbol{\xi}\|^2/2$ after a proper truncation possesses the same exponential moments as in the Gaussian case. A straightforward implication is the probability bound $I\!P\bigl(\|\boldsymbol{\xi}\|^2 > p + u\bigr)$ for moderate values u. Namely, given $u > 0$, define $\mu = u/(u+p)$. This value optimizes inequality (A.1) in the Gaussian case. Now we can apply a similar bound under the constraints $\|\boldsymbol{\xi}\| \le \mathtt{g}/\mu - \sqrt{p/\mu}$. Therefore the bound is only meaningful if $\sqrt{u+p} \le \mathtt{g}/\mu - \sqrt{p/\mu}$ with $\mu = u/(u+p)$ or with $w = \sqrt{u/p} \le w_c$, see (3.1).

The largest value u for which this constraint is still valid, is given by $p + u = \mathtt{y}_c^2$. Hence (C.1) yields for $p + u \le \mathtt{y}_c^2$

$$\begin{aligned} I\!P\bigl(\|\boldsymbol{\xi}\|^2 > p + u, \|\boldsymbol{\xi}\| \le \mathtt{y}_c\bigr) &\le \exp\Bigl\{-\frac{\mu(p+u)}{2}\Bigr\} I\!E \exp\Bigl(\frac{\mu\|\boldsymbol{\xi}\|^2}{2}\Bigr) \mathbb{I}\bigl(\|\boldsymbol{\xi}\| \le \mathtt{g}/\mu - \sqrt{p/\mu}\bigr) \\ &\le 2\exp\bigl\{-0.5\bigl[\mu(p+u) + p\log(1-\mu)\bigr]\bigr\} \\ &= 2\exp\bigl\{-0.5\bigl[u - p\log(1 + u/p)\bigr]\bigr\}. \end{aligned}$$

Similarly to the Gaussian case, this implies with $\kappa = 6.6$ that

$$I\!P\bigl(\|\boldsymbol{\xi}\| \ge p + \sqrt{\kappa \mathtt{x} p} \vee (\kappa \mathtt{x}), \|\boldsymbol{\xi}\| \le \mathtt{y}_c\bigr) \le 2\exp(-\mathtt{x}).$$

The Gaussian case means that (1.1) holds with $\mathtt{g} = \infty$ yielding $\mathtt{y}_c = \infty$. In the non-Gaussian case with a finite $\mathtt{g}$, we have to accompany the moderate deviation bound with a large deviation bound $I\!P\bigl(\|\boldsymbol{\xi}\| > \mathtt{y}\bigr)$ for $\mathtt{y} \ge \mathtt{y}_c$. This is done by combining the bound (C.1) with the standard slicing arguments.

Lemma C.2. *Let* $\mu_0 \le \mathtt{g}^2/p$. *Define* $\mathtt{y}_0 = \mathtt{g}/\mu_0 - \sqrt{p/\mu_0}$ *and* $\mathtt{g}_0 = \mu_0 \mathtt{y}_0 = \mathtt{g} - \sqrt{\mu_0 p}$. *It holds for* $\mathtt{y} \ge \mathtt{y}_0$

$$I\!P\bigl(\|\boldsymbol{\xi}\| > \mathtt{y}\bigr) \le 8.4(1 - \mathtt{g}_0/\mathtt{y})^{-p/2} \exp\bigl(-\mathtt{g}_0\mathtt{y}/2\bigr) \tag{C.2}$$

$$\le 8.4\exp\bigl\{-\mathtt{x}_0 - \mathtt{g}_0(\mathtt{y} - \mathtt{y}_0)/2\bigr\} \tag{C.3}$$

with $\mathtt{x}_0$ *defined by*

$$2\mathtt{x}_0 = \mu_0 \mathtt{y}_0^2 + p\log(1-\mu_0) = \mathtt{g}^2/\mu_0 - p + p\log(1-\mu_0).$$

Proof. Consider the growing sequence $\mathtt{y}_k$ with $\mathtt{y}_1 = \mathtt{y}$ and $\mathtt{g}_0\mathtt{y}_{k+1} = \mathtt{g}_0\mathtt{y} + k$. Define also $\mu_k = \mathtt{g}_0/\mathtt{y}_k$. In particular, $\mu_k \le \mu_1 = \mathtt{g}_0/\mathtt{y}$. Obviously

$$I\!P\bigl(\|\boldsymbol{\xi}\| > \mathtt{y}\bigr) = \sum_{k=1}^{\infty} I\!P\bigl(\|\boldsymbol{\xi}\| > \mathtt{y}_k, \|\boldsymbol{\xi}\| \le \mathtt{y}_{k+1}\bigr).$$

Now we try to evaluate every slicing probability in this expression. We use that

$$\mu_{k+1}\mathtt{y}_k^2 = \frac{(\mathtt{g}_0\mathtt{y} + k - 1)^2}{\mathtt{g}_0\mathtt{y} + k} \ge \mathtt{g}_0\mathtt{y} + k - 2,$$

and also $\mathtt{g}/\mu_k - \sqrt{p/\mu_k} \ge \mathtt{y}_k$ because $\mathtt{g} - \mathtt{g}_0 = \sqrt{\mu_0 p} > \sqrt{\mu_k p}$ and

$$\mathtt{g}/\mu_k - \sqrt{p/\mu_k} - \mathtt{y}_k = \mu_k^{-1}(\mathtt{g} - \sqrt{\mu_k p} - \mathtt{g}_0) \ge 0.$$

Hence by (C.1)

$$\begin{aligned} I\!P\bigl(\|\boldsymbol{\xi}\| > \mathtt{y}\bigr) &= \sum_{k=1}^{\infty} I\!P\bigl(\|\boldsymbol{\xi}\| > \mathtt{y}_k, \|\boldsymbol{\xi}\| \le \mathtt{y}_{k+1}\bigr) \\ &\le \sum_{k=1}^{\infty} \exp\Bigl(-\frac{\mu_{k+1}\mathtt{y}_k^2}{2}\Bigr) I\!E \exp\Bigl(\frac{\mu_{k+1}\|\boldsymbol{\xi}\|^2}{2}\Bigr) \mathbb{I}\bigl(\|\boldsymbol{\xi}\| \le \mathtt{y}_{k+1}\bigr) \\ &\le \sum_{k=1}^{\infty} 2\bigl(1 - \mu_{k+1}\bigr)^{-p/2} \exp\Bigl(-\frac{\mu_{k+1}\mathtt{y}_k^2}{2}\Bigr) \le 2\bigl(1-\mu_1\bigr)^{-p/2} \sum_{k=1}^{\infty} \exp\Bigl(-\frac{\mathtt{g}_0\mathtt{y} + k - 2}{2}\Bigr) \end{aligned}$$

$$= 2e^{1/2}(1-e^{-1/2})^{-1}(1-\mu_1)^{-p/2}\exp\bigl(-\mathtt{g}_0\mathtt{y}/2\bigr) \le 8.4(1-\mu_1)^{-p/2}\exp\bigl(-\mathtt{g}_0\mathtt{y}/2\bigr)$$

and the first assertion follows. For $\mathtt{y} = \mathtt{y}_0$, it holds

$$\mathtt{g}_0\mathtt{y}_0 + p\log(1-\mu_0) = \mu_0\mathtt{y}_0^2 + p\log(1-\mu_0) = 2\mathtt{x}_0$$

and (C.2) implies $I\!\!P\bigl(\|\boldsymbol{\xi}\| > \mathtt{y}_0\bigr) \le 8.4\exp(-\mathtt{x}_0)$. Now observe that the function $f(\mathtt{y}) = \mathtt{g}_0\mathtt{y}/2 + (p/2)\log\bigl(1-\mathtt{g}_0/\mathtt{y}\bigr)$ fulfills $f(\mathtt{y}_0) = \mathtt{x}_0$ and $f'(\mathtt{y}) \ge \mathtt{g}_0/2$ yielding $f(\mathtt{y}) \ge \mathtt{x}_0 + \mathtt{g}_0(\mathtt{y}-\mathtt{y}_0)/2$. This implies (C.3). □

The statements of the theorem are obtained by applying the lemmas with $\mu_0 = \mu_c = w_c^2/(1+w_c^2)$. This also implies $\mathtt{y}_0 = \mathtt{y}_c$, $\mathtt{x}_0 = \mathtt{x}_c$, and $\mathtt{g}_0 = \mathtt{g}_c = \mathtt{g} - \sqrt{\mu_c p}$, cf. (3.2). □

D. PROOF OF THEOREM 4.1

The main steps of the proof are similar to the proof of Theorem 3.1.

Lemma D.1. *Suppose* (1.1). *For any* $\mu < 1$ *with* $\mathtt{g}^2/\mu \ge \mathtt{p}$, *it holds*

$$I\!\!E\exp\bigl(\mu\|I\!\!B\boldsymbol{\xi}\|^2/2\bigr)\,\mathbb{1}\bigl(\|I\!\!B^2\boldsymbol{\xi}\| \le \mathtt{g}/\mu - \sqrt{\mathtt{p}/\mu}\bigr) \le 2\det(I\!\!I_p - \mu I\!\!B^2)^{-1/2}. \qquad \text{(D.1)}$$

Proof. With $c_p(I\!\!B) = \bigl(2\pi\bigr)^{-p/2}\det(I\!\!B^{-1})$

$$\begin{aligned}
&c_p(I\!\!B)\int\exp\Bigl(\boldsymbol{\gamma}^\top\boldsymbol{\xi} - \frac{1}{2\mu}\|I\!\!B^{-1}\boldsymbol{\gamma}\|^2\Bigr)\,\mathbb{1}(\|\boldsymbol{\gamma}\| \le \mathtt{g})\,d\boldsymbol{\gamma} \\
&\quad= c_p(I\!\!B)\exp\Bigl(\frac{\mu\|I\!\!B\boldsymbol{\xi}\|^2}{2}\Bigr)\int\exp\Bigl(-\frac{1}{2}\bigl\|\mu^{1/2}I\!\!B\boldsymbol{\xi} - \mu^{-1/2}I\!\!B^{-1}\boldsymbol{\gamma}\bigr\|^2\Bigr)\,\mathbb{1}(\|\boldsymbol{\gamma}\| \le \mathtt{g})\,d\boldsymbol{\gamma} \\
&\quad= \mu^{p/2}\exp\Bigl(\frac{\mu\|I\!\!B\boldsymbol{\xi}\|^2}{2}\Bigr)I\!\!P_{\boldsymbol{\xi}}\bigl(\|\mu^{-1/2}I\!\!B\boldsymbol{\varepsilon} + I\!\!B^2\boldsymbol{\xi}\| \le \mathtt{g}/\mu\bigr),
\end{aligned}$$

where $\boldsymbol{\varepsilon}$ denotes a standard normal vector in $I\!\!R^p$ and $I\!\!P_{\boldsymbol{\xi}}$ means the conditional probability given $\boldsymbol{\xi}$. Moreover, for any $\boldsymbol{u} \in I\!\!R^p$ and $\mathtt{r} \ge \mathtt{p}^{1/2} + \|\boldsymbol{u}\|$, it holds in view of $I\!\!P\bigl(\|I\!\!B\boldsymbol{\varepsilon}\|^2 > \mathtt{p}\bigr) \le 1/2$

$$I\!\!P\bigl(\|I\!\!B\boldsymbol{\varepsilon} - \boldsymbol{u}\| \le \mathtt{r}\bigr) \ge I\!\!P\bigl(\|I\!\!B\boldsymbol{\varepsilon}\| \le \sqrt{\mathtt{p}}\bigr) \ge 1/2.$$

This implies

$$\begin{aligned}
&\exp\bigl(\mu\|I\!\!B\boldsymbol{\xi}\|^2/2\bigr)\,\mathbb{1}\bigl(\|I\!\!B^2\boldsymbol{\xi}\| \le \mathtt{g}/\mu - \sqrt{\mathtt{p}/\mu}\bigr) \\
&\qquad\le 2\mu^{-p/2}c_p(I\!\!B)\int\exp\Bigl(\boldsymbol{\gamma}^\top\boldsymbol{\xi} - \frac{1}{2\mu}\|I\!\!B^{-1}\boldsymbol{\gamma}\|^2\Bigr)\,\mathbb{1}(\|\boldsymbol{\gamma}\| \le \mathtt{g})\,d\boldsymbol{\gamma}.
\end{aligned}$$

Further, by (1.1)

$$\begin{aligned}
&c_p(I\!\!B)I\!\!E\int\exp\Bigl(\boldsymbol{\gamma}^\top\boldsymbol{\xi} - \frac{1}{2\mu}\|I\!\!B^{-1}\boldsymbol{\gamma}\|^2\Bigr)\,\mathbb{1}(\|\boldsymbol{\gamma}\| \le \mathtt{g})\,d\boldsymbol{\gamma} \\
&\qquad\le c_p(I\!\!B)\int\exp\Bigl(\frac{\|\boldsymbol{\gamma}\|^2}{2} - \frac{1}{2\mu}\|I\!\!B^{-1}\boldsymbol{\gamma}\|^2\Bigr)\,d\boldsymbol{\gamma} \\
&\qquad\le \det(I\!\!B^{-1})\det(\mu^{-1}I\!\!B^{-2} - I\!\!I_p)^{-1/2} = \mu^{p/2}\det(I\!\!I_p - \mu I\!\!B^2)^{-1/2}
\end{aligned}$$

and (D.1) follows. □

Now we evaluate the probability $I\!\!P\bigl(\|I\!\!B\boldsymbol{\xi}\| > \mathtt{y}\bigr)$ for moderate values of $\mathtt{y}$.

Lemma D.2. *Let* $\mu_0 < 1 \wedge (\mathtt{g}^2/\mathtt{p})$. *With* $\mathtt{y}_0 = \mathtt{g}/\mu_0 - \sqrt{\mathtt{p}/\mu_0}$, *it holds for any* $u > 0$

$$I\!\!P\bigl(\|I\!\!B\boldsymbol{\xi}\|^2 > \mathtt{p} + u, \|I\!\!B^2\boldsymbol{\xi}\| \le \mathtt{y}_0\bigr) \le 2\exp\bigl\{-0.5\mu_0(\mathtt{p}+u) - 0.5\log\det(I\!\!I_p - \mu_0 I\!\!B^2)\bigr\}. \tag{D.2}$$

In particular, if $I\!\!B^2$ *is diagonal, that is,* $I\!\!B^2 = \operatorname{diag}\bigl(a_1, \ldots, a_p\bigr)$, *then*

$$I\!\!P\bigl(\|I\!\!B\boldsymbol{\xi}\|^2 > \mathtt{p} + u, \|I\!\!B^2\boldsymbol{\xi}\| \le \mathtt{y}_0\bigr) \le 2\exp\Bigl\{-\frac{\mu_0 u}{2} - \frac{1}{2}\sum_{i=1}^{p}\bigl[\mu_0 a_i + \log\bigl(1 - \mu_0 a_i\bigr)\bigr]\Bigr\}. \tag{D.3}$$

Proof. The exponential Chebyshev inequality and (D.1) imply

$$\begin{aligned}
&I\!\!P\bigl(\|I\!\!B\boldsymbol{\xi}\|^2 > \mathtt{p} + u, \|I\!\!B^2\boldsymbol{\xi}\| \le \mathtt{y}_0\bigr) \\
&\le \exp\Bigl\{-\frac{\mu_0(\mathtt{p}+u)}{2}\Bigr\} I\!\!E \exp\Bigl(\frac{\mu_0\|I\!\!B\boldsymbol{\xi}\|^2}{2}\Bigr) 1\!\!1\bigl(\|I\!\!B^2\boldsymbol{\xi}\| \le \mathtt{g}/\mu_0 - \sqrt{\mathtt{p}/\mu_0}\bigr) \\
&\le 2\exp\bigl\{-0.5\mu_0(\mathtt{p}+u) - 0.5\log\det(I\!\!I_p - \mu_0 I\!\!B^2)\bigr\}.
\end{aligned}$$

Moreover, the standard change-of-basis arguments allow us to reduce the problem to the case of a diagonal matrix $I\!\!B^2 = \operatorname{diag}\bigl(a_1, \ldots, a_p\bigr)$, where $1 = a_1 \ge a_2 \ge \ldots \ge a_p > 0$. Note that $\mathtt{p} = a_1 + \ldots + a_p$. Then the claim (D.2) can be written in the form (D.3). □

Now we evaluate a large deviation probability that $\|I\!\!B\boldsymbol{\xi}\| > \mathtt{y}$ for a large $\mathtt{y}$. Note that the condition $\|I\!\!B^2\|_\infty \le 1$ implies $\|I\!\!B^2\boldsymbol{\xi}\| \le \|I\!\!B\boldsymbol{\xi}\|$. So, the bound (D.2) continues to hold when $\|I\!\!B^2\boldsymbol{\xi}\| \le \mathtt{y}_0$ is replaced by $\|I\!\!B\boldsymbol{\xi}\| \le \mathtt{y}_0$.

Lemma D.3. *Let* $\mu_0 < 1$ *and* $\mu_0\mathtt{p} < \mathtt{g}^2$. *Define* $\mathtt{g}_0$ *by* $\mathtt{g}_0 = \mathtt{g} - \sqrt{\mu_0\mathtt{p}}$. *For any* $\mathtt{y} \ge \mathtt{y}_0 \stackrel{\mathrm{def}}{=} \mathtt{g}_0/\mu_0$, *it holds*

$$\begin{aligned}
I\!\!P\bigl(\|I\!\!B\boldsymbol{\xi}\| > \mathtt{y}\bigr) &\le 8.4\det\{I\!\!I_p - (\mathtt{g}_0/\mathtt{y})I\!\!B^2\}^{-1/2}\exp\bigl(-\mathtt{g}_0\mathtt{y}/2\bigr) \\
&\le 8.4\exp\bigl(-\mathtt{x}_0 - \mathtt{g}_0(\mathtt{y} - \mathtt{y}_0)/2\bigr),
\end{aligned} \tag{D.4}$$

where $\mathtt{x}_0$ *is defined by*

$$2\mathtt{x}_0 = \mathtt{g}_0\mathtt{y}_0 + \log\det\{I\!\!I_p - (\mathtt{g}_0/\mathtt{y}_0)I\!\!B^2\}.$$

Proof. The slicing arguments of Lemma C.2 apply here in the same manner. One has to replace $\|\boldsymbol{\xi}\|$ by $\|I\!\!B\boldsymbol{\xi}\|$ and $(1-\mu_1)^{-p/2}$ by $\det\{I\!\!I_p - (\mathtt{g}_0/\mathtt{y})I\!\!B^2\}^{-1/2}$. We omit the details. In particular, with $\mathtt{y} = \mathtt{y}_0 = \mathtt{g}_0/\mu$, this yields

$$I\!\!P\bigl(\|I\!\!B\boldsymbol{\xi}\| > \mathtt{y}_0\bigr) \le 8.4\exp(-\mathtt{x}_0).$$

Moreover, for the function $f(\mathtt{y}) = \mathtt{g}_0\mathtt{y} + \log\det\{I\!\!I_p - (\mathtt{g}_0/\mathtt{y})I\!\!B^2\}$, it holds $f'(\mathtt{y}) \ge \mathtt{g}_0$ and hence, $f(\mathtt{y}) \ge f(\mathtt{y}_0) + \mathtt{g}_0(\mathtt{y} - \mathtt{y}_0)$ for $\mathtt{y} > \mathtt{y}_0$. This implies (D.4). □

One important feature of the results of Lemmas D.2 and D.3 is that the value $\mu_0 < 1 \wedge (\mathtt{g}^2/\mathtt{p})$ can be selected arbitrarily. In particular, for $\mathtt{y} \ge \mathtt{y}_c$, Lemma D.3 with $\mu_0 = \mu_c$ yields the large deviation probability $I\!\!P\bigl(\|I\!\!B\boldsymbol{\xi}\| > \mathtt{y}\bigr)$. For bounding the probability $I\!\!P\bigl(\|I\!\!B\boldsymbol{\xi}\|^2 > \mathtt{p} + u,\ \|I\!\!B\boldsymbol{\xi}\| \le \mathtt{y}_c\bigr)$, we use the inequality $\log(1-t) \ge -t - t^2$ for $t \le 2/3$. It implies for $\mu \le 2/3$ that

$$\begin{aligned}
-\log I\!\!P\bigl(\|I\!\!B\boldsymbol{\xi}\|^2 > \mathtt{p} + u, \|I\!\!B\boldsymbol{\xi}\| \le \mathtt{y}_c\bigr) &\ge \mu(\mathtt{p}+u) + \sum_{i=1}^{p}\log\bigl(1 - \mu a_i\bigr) \\
&\ge \mu(\mathtt{p}+u) - \sum_{i=1}^{p}(\mu a_i + \mu^2 a_i^2) \ge \mu u - \mu^2\mathtt{v}^2/2.
\end{aligned} \tag{D.5}$$

Now we distinguish between $\mu_c = 2/3$ and $\mu_c < 2/3$ starting with $\mu_c = 2/3$. The bound (D.5) with $\mu = 2/3$ and with $u = (2\mathtt{v}\mathtt{x}^{1/2}) \vee (6\mathtt{x})$ yields

$$I\!\!P\bigl(\|I\!\!B\boldsymbol{\xi}\|^2 > \mathtt{p} + u,\, \|I\!\!B\boldsymbol{\xi}\| \le \mathtt{y}_c\bigr) \le 2\exp(-\mathtt{x});$$

see the proof of Theorem 2.2 for the Gaussian case.

Now consider $\mu_c < 2/3$. For $\mathtt{x}^{1/2} \le \mu_c \mathtt{v}/2$, use $u = 2\mathtt{v}\mathtt{x}^{1/2}$ and $\mu_0 = u/\mathtt{v}^2$. It holds $\mu_0 = u/\mathtt{v}^2 \le \mu_c$ and $u^2/(4\mathtt{v}^2) = \mathtt{x}$ yielding the desired bound by (D.5). For $\mathtt{x}^{1/2} > \mu_c\mathtt{v}/2$, we select again $\mu_0 = \mu_c$. It holds with $u = 4\mu_c^{-1}\mathtt{x}$ that $\mu_c u/2 - \mu_c^2\mathtt{v}^2/4 \ge 2\mathtt{x} - \mathtt{x} = \mathtt{x}$. This completes the proof. $\square$

E. PROOF OF THEOREM 6.1

The arguments behind the result are the same as in the one-norm case of Theorem 3.1. We only outline the main steps.

Lemma E.1. *Suppose* (6.1) *and* (6.2). *For any* $\mu < 1$ *with* $\mathtt{g}_\circ > \mu^{1/2} r_*$, *it holds*

$$I\!\!E \exp\bigl(\mu\|\boldsymbol{\xi}\|^2/2\bigr)\, \mathrm{I\!I}\bigl(\|\boldsymbol{\xi}\|_\circ \le \mathtt{g}_\circ/\mu - r_*/\mu^{1/2}\bigr) \le 2(1-\mu)^{-p/2}. \tag{E.1}$$

Proof. Let $\boldsymbol{\varepsilon}$ be a standard normal vector in $I\!\!R^p$ and $\boldsymbol{u} \in I\!\!R^p$. Let us fix some $\boldsymbol{\xi}$ with $\mu^{1/2}\|\boldsymbol{\xi}\|_\circ \le \mu^{-1/2}\mathtt{g}_\circ - r_*$ and denote by $I\!\!P_{\boldsymbol{\xi}}$ the conditional probability given $\boldsymbol{\xi}$. It holds by (6.2) with $c_p = (2\pi)^{-p/2}$

$$\begin{aligned} &c_p \int \exp\Bigl(\boldsymbol{\gamma}^\top\boldsymbol{\xi} - \frac{1}{2\mu}\|\boldsymbol{\gamma}\|^2\Bigr)\, \mathrm{I\!I}(\|\boldsymbol{\gamma}\|_\circ \le \mathtt{g}_\circ)\, d\boldsymbol{\gamma} \\ &\quad = c_p \exp\bigl(\mu\|\boldsymbol{\xi}\|^2/2\bigr) \int \exp\Bigl(-\frac{1}{2}\bigl\|\mu^{1/2}\boldsymbol{\xi} - \mu^{-1/2}\boldsymbol{\gamma}\bigr\|^2\Bigr)\, \mathrm{I\!I}(\|\mu^{-1/2}\boldsymbol{\gamma}\|_\circ \le \mu^{-1/2}\mathtt{g}_\circ)\, d\boldsymbol{\gamma} \\ &\quad = \mu^{p/2}\exp\bigl(\mu\|\boldsymbol{\xi}\|^2/2\bigr) I\!\!P_{\boldsymbol{\xi}}\bigl(\|\boldsymbol{\varepsilon} - \mu^{1/2}\boldsymbol{\xi}\|_\circ \le \mu^{-1/2}\mathtt{g}_\circ\bigr) \ge 0.5\mu^{p/2}\exp\bigl(\mu\|\boldsymbol{\xi}\|^2/2\bigr). \end{aligned}$$

This implies

$$\begin{aligned} &\exp\Bigl(\frac{\mu\|\boldsymbol{\xi}\|^2}{2}\Bigr)\, \mathrm{I\!I}\bigl(\|\boldsymbol{\xi}\|_\circ \le \mathtt{g}_\circ/\mu - r_*/\mu^{1/2}\bigr) \\ &\qquad \le 2\mu^{-p/2} c_p \int \exp\Bigl(\boldsymbol{\gamma}^\top\boldsymbol{\xi} - \frac{1}{2\mu}\|\boldsymbol{\gamma}\|^2\Bigr)\, \mathrm{I\!I}(\|\boldsymbol{\gamma}\|_\circ \le \mathtt{g}_\circ)\, d\boldsymbol{\gamma}. \end{aligned}$$

Further, by (6.1)

$$\begin{aligned} &c_p I\!\!E \int \exp\Bigl(\boldsymbol{\gamma}^\top\boldsymbol{\xi} - \frac{1}{2\mu}\|\boldsymbol{\gamma}\|^2\Bigr)\, \mathrm{I\!I}(\|\boldsymbol{\gamma}\|_\circ \le \mathtt{g}_\circ)\, d\boldsymbol{\gamma} \\ &\qquad \le c_p \int \exp\Bigl(-\frac{\mu^{-1}-1}{2}\|\boldsymbol{\gamma}\|^2\Bigr)\, d\boldsymbol{\gamma} \le (\mu^{-1}-1)^{-p/2} \end{aligned}$$

and (E.1) follows. $\square$

As in the Gaussian case, (E.1) implies for $\mathfrak{z} > p$ with $\mu = \mu(\mathfrak{z}) = (\mathfrak{z} - p)/\mathfrak{z}$ the bounds (6.4) and (6.5). Note that the value $\mu(\mathfrak{z})$ clearly grows with $\mathfrak{z}$ from zero to one, while $\mathtt{g}_\circ/\mu(\mathfrak{z}) - r_*/\mu^{1/2}(\mathfrak{z})$ is strictly decreasing. The value $\mathfrak{z}_\circ$ is defined exactly as the point where $\mathtt{g}_\circ/\mu(\mathfrak{z}) - r_*/\mu^{1/2}(\mathfrak{z})$ crosses $\mathtt{u}_\circ$, so that $\mathtt{g}_\circ/\mu(\mathfrak{z}) - r_*/\mu^{1/2}(\mathfrak{z}) \ge \mathtt{u}_\circ$ for $\mathfrak{z} \le \mathfrak{z}_\circ$.

For $\mathfrak{z} > \mathfrak{z}_\circ$, the choice $\mu = \mu(\mathtt{y})$ conflicts with $\mathtt{g}_\circ/\mu(\mathfrak{z}) - r_*/\mu^{1/2}(\mathfrak{z}) \ge \mathtt{u}_\circ$. So, we apply $\mu = \mu_\circ$ yielding by the Markov inequality

$$I\!\!P\bigl(\|\boldsymbol{\xi}\|^2 > \mathfrak{z},\, \|\boldsymbol{\xi}\|_\circ \le \mathtt{u}_\circ\bigr) \le 2\exp\bigl\{-\mu_\circ\mathfrak{z}/2 - (p/2)\log(1-\mu_\circ)\bigr\},$$

and the assertion follows. $\square$

F. PROOF OF THEOREM 6.2

Arguments from the proof of Lemmas D.1 and E.1 yield in view of $\mathtt{g}_\circ \mu_\circ^{-1} - r_* \mu_\circ^{-1/2} \geq \mathtt{u}_\circ$

$$\begin{aligned} &I\!\!E \exp\bigl\{\mu_\circ \|\Pi \boldsymbol{\xi}\|^2/2\bigr\} \, \mathrm{1\!I}\bigl(\|\Pi^2 \boldsymbol{\xi}\|_\circ \leq \mathtt{u}_\circ\bigr) \\ &\qquad \leq I\!\!E \exp\bigl(\mu_\circ \|\Pi \boldsymbol{\xi}\|^2/2\bigr) \, \mathrm{1\!I}\bigl(\|\Pi^2 \boldsymbol{\xi}\|_\circ \leq \mathtt{g}_\circ/\mu_\circ - \mathtt{p}/\mu_\circ^{1/2}\bigr) \leq 2\det(I\!\!I_p - \mu_\circ \Pi^2)^{-1/2}. \end{aligned}$$

The inequality $\log(1-t) \geq -t - t^2$ for $t \leq 2/3$ and symmetry of the matrix Π imply

$$-\log\det(I\!\!I_p - \mu_\circ \Pi^2) \leq \mu_\circ \mathtt{p} + \mu_\circ^2 \mathtt{v}^2/2$$

cf.(D.5); the assertion (6.6) follows. □

ACKNOWLEDGMENTS

The authors are supported by Laboratory for Structural Methods of Data Analysis in Predictive Modeling, MIPT, RF government grant, ag. 11.G34.31.0073.

Financial support by the German Research Foundation (DFG) through the Collaborative Research Center 649 “Economic Risk” is gratefully acknowledged

REFERENCES

1. Y. Baraud, “A Bernstein-Type Inequality for Suprema of Random Processes with Applications to Model Selection in Non-Gaussian Regression”, Bernoulli **16** (4), 1064–1085 (2010).
2. S. Boucheron and P, Massart, “A High-Dimensional Wilks Phenomenon”, Probab. Theory and Rel. Fields **150** (3), 405–433 (2011) 10.1007/s00440-010-0278-7.
3. J. Bretagnolle, “A New Large Deviation Inequality for U-Statistics of Order 2”, ESAIM, Probab. Statist. **3**, 151–162 (1999).
4. J. Fan and T. Huang, “Profile Likelihood Inferences on Semiparametric Varying-Coefficient Partially Linear Models”, Bernoulli **11** (6), 1031–1057 (2005).
5. J. Fan, C. Zhang, and J. Zhang, “Generalized Likelihood Ratio Statistics and Wilks Phenomenon”, Ann. Statist. **29** (1), 153–193 (2001).
6. E. Giné, R. Latała, and J. Zinn, “Exponential and Moment Inequalities for U-Statistics”, in *Birkhäuser Prog. Probab.*, Vol. 47: *High Dimensional Probability II. 2nd Internat. Conf., Univ. of Washington, DC, USA, August 1-6, 1999*, Ed. by E. Giné et al. (Birkhäuser, Boston, MA, 2000), pp. 13–38.
7. F. Götze and A. N. Tikhomirov, “Asymptotic Distribution of Quadratic Forms”, Ann. Statist. **27** (2), 1072–1098 (1999).
8. L. Horváth and Q.-M. Shao, “Limit Theorems for Quadratic Forms with Applications to Whittle’s Estimate”, Ann. Appl. Probab. **9** (1), 146–187 (1999).
9. C. Houdré and P. Reynaud-Bouret, “Exponential Inequalities, with Constants, for U-Statistics of Order Two”, in *Birkhäuser Prog. Probab.*, Vol. 56: *Stochastic Inequalities and Applications. Selected Papers Presented at the Euroconference on “Stochastic Inequalities and Their Applications", Barcelona, June 18–22, 2002*, Ed. by E. Giné et al. (Birkhäuser, Basel, 2003), pp. 55–69.
10. D. Hsu, S.M. Kakade and T. Zhang, “A Tail Inequality for Quadratic Forms of Subgaussian Random Vectors”, Electron. Commun. Probab. **17** (52), 6 (2012).
11. P. Massart, *Concentration Inequalities and Model Selection. Ecole d’Eté de Probabilités de Saint-Flour XXXIII-2003*, in *Lecture Notes in Mathematics* (Springer, 2007).