Probability, Uncertainty and Quantitative Risk
Vol. 10, No. 1, 2025, 135–158

https://www.aimsciences.org/PUQR
DOI:10.3934/puqr.2025007

Deviation bounds for the norm of a random vector under exponential moment conditions with applications

Vladimir Spokoiny

Weierstrass Institute, HU Berlin, and HSE Moscow, Mohrenstr. 39, 10117 *Berlin, Germany*

Email: spokoiny@wias-berlin.de

Abstract Hanson-Wright inequality provides a powerful tool for bounding the norm $\|\boldsymbol{\xi}\|$ of a centered stochastic vector $\boldsymbol{\xi}$ with independent entries and sub-gaussian behavior. This paper extends the bounds to the case when $\boldsymbol{\xi}$ only has bounded exponential moments of the form $\log E\exp\langle V^{-1}\boldsymbol{\xi},u\rangle \leqslant \|u\|^2/2$, where $V^2 \geqslant \operatorname{Var}(\boldsymbol{\xi})$ and $\|u\| \leqslant g$ for some fixed g. For a linear mapping Q, we present an upper quantile function $z_c(B,x)$ ensuring $P(\|Q\xi\| > z_c(B,x)) \leqslant 3e^{-x}$ with $B = Q\,V^2Q^T$. The obtained results exhibit a phase transition effect: with a value x_c depending on g and B, for $x \leqslant x_c$, the function $z_c(B,x)$ replicates the case of a Gaussian vector ξ, that is, $z_c^2(B,x) = \operatorname{tr}(B)+2\sqrt{x\operatorname{tr}(B^2)} + 2x\|B\|$. For $x > x_c$, the function $z_c(B,x)$ grows linearly in x. The results are specified to the case of Bernoulli vector sums and to covariance estimation in Frobenius norm.

Keywords Upper quantiles, Phase transition, Vector Bernoulli sums, Frobenius loss

2020 Mathematics Subject Classification 62E15, 62E10

1. Introduction

Many statistical results rely on some deviation bounds on the Euclidean norm of a zero mean random vector $\boldsymbol{\xi} \in \mathbb{R}^p$. Modern applications motivate further studies of this problem under nonclassical setups with small or moderate n and large p. If $\boldsymbol{\xi}$ is a zero mean Gaussian vector and $Q\colon \mathbb{R}^p \to \mathbb{R}^q$ is a linear mapping, then $Q\boldsymbol{\xi}$ is also zero mean Gaussian with the variance $B = \operatorname{Var}(Q\boldsymbol{\xi}) = Q\operatorname{Var}(\boldsymbol{\xi})Q^\top$. For the squared norm $\|Q\boldsymbol{\xi}\|^2$, it holds $\mathbb{E}\|Q\boldsymbol{\xi}\|^2 = \operatorname{tr}(B)$, $\operatorname{Var}(\|Q\boldsymbol{\xi}\|^2) = \operatorname{tr}(B^2)$. Moreover, the upper and lower quantile functions for $\|Q\boldsymbol{\xi}\|^2$ can be written explicitly in terms of $\operatorname{tr}(B)$, $\operatorname{tr}(B^2)$, and $\|B\|$: for any $\mathtt{x} > 0$,

$$\mathbb{P}\Big(\|Q\boldsymbol{\xi}\|^2 - \operatorname{tr}(B) > 2\sqrt{\mathtt{x}\operatorname{tr}(B^2)} + 2\|B\|\mathtt{x}\Big) \leqslant \mathrm{e}^{-\mathtt{x}},$$

$$\mathbb{P}\Big(\|Q\boldsymbol{\xi}\|^2 - \operatorname{tr}(B) < -2\sqrt{\mathtt{x}\operatorname{tr}(B^2)}\Big) \leqslant \mathrm{e}^{-\mathtt{x}},$$

see e.g. [8]. If the *effective trace* $\operatorname{tr}(B)/\|B\|$ is large, the squared norm $\|Q\boldsymbol{\xi}\|^2 - \operatorname{tr}(B)$ concentrates around its expectation $\operatorname{tr}(B)$. An extension of these bounds to a sub-gaussian case

---

Received 8 September 2023; Accepted 2 January 2025; Early access 4 March 2025

is discussed in [10], however, under limiting assumption of independent entries of $\boldsymbol{\xi}$. In the recent years, a number of new results were obtained in this direction. We refer to [7] for an extensive overview and advanced results on Hanson-Wright type concentration inequalities. However, sub-gaussian behavior of the vector $\boldsymbol{\xi}$ can be very restrictive in many applications. In this paper we consider a couple of typical examples.

The first one corresponds to generalized regression which includes models with binary, exponential, poissonian data etc. To be more specific, we focus on models with binary responses such as binary classification, logistic regression, and many others. Given feature vectors $\boldsymbol{\Psi}_i \in \mathbb{R}^p$, binary responses Y_i are modelled as $Y_i \sim \operatorname{Bernoulli}(\boldsymbol{\Psi}_i^\top \boldsymbol{\upsilon})$, where $\boldsymbol{\upsilon} \in \mathbb{R}^p$ is an unknown parameter vector to be estimated. The corresponding penalized maximum likelihood estimator (MLE) $\widetilde{\boldsymbol{\upsilon}}_G$ can be written as

$$\widetilde{\boldsymbol{\upsilon}}_G \stackrel{\text{def}}{=} \operatorname*{argmax}_{\boldsymbol{\upsilon}} \sum_{i=1}^{n} \bigl\{ Y_i \boldsymbol{\Psi}_i^\top \boldsymbol{\upsilon} - \phi(\boldsymbol{\Psi}_i^\top \boldsymbol{\upsilon}) \bigr\} - \operatorname{pen}_G(\boldsymbol{\upsilon}),$$

where $\phi(\theta) = -\log(1-\theta)$ and $\operatorname{pen}_G(\boldsymbol{\upsilon})$ is a penalty indexed by a tuning parameter G. [11] studied in details the case of a quadratic penalization $\operatorname{pen}_G(\boldsymbol{\upsilon}) = \|G\boldsymbol{\upsilon}\|^2/2$ for some matrix G^2 describing smoothness of $\boldsymbol{\upsilon}$, and argued that parametric inference for such logistic regression requires some possibly sharp bounds on the scaled vector $\boldsymbol{\xi}_G$ with

$$\boldsymbol{\xi}_G \stackrel{\text{def}}{=} \mathbb{D}_G^{-1} \sum_{i=1}^{n} (Y_i - \mathbb{E} Y_i) \boldsymbol{\Psi}_i, \quad \mathbb{D}_G^2 \stackrel{\text{def}}{=} \sum_{i=1}^{n} \phi''(\boldsymbol{\Psi}_i^\top \boldsymbol{\upsilon}^*) \boldsymbol{\Psi}_i \boldsymbol{\Psi}_i^\top + G^2.$$

Here $\boldsymbol{\upsilon}^*$ is the background truth under the correct model specification $Y_i \sim \operatorname{Bernoulli}(\boldsymbol{\Psi}_i^\top \boldsymbol{\upsilon}^*)$, otherwise it is the best parametric fit to this model. The main difficulty in obtaining some deviation bounds on $\|\boldsymbol{\xi}_G\|$ is that this vector is in general only sub-exponential, not sub-gaussian; see Section 3 for further details.

Another example is provided by empirical covariance. Let $\boldsymbol{X}_i \sim \mathcal{N}(0, \varSigma)$ be i.i.d. zero mean Gaussian vectors in $\mathbb{R}^p$ with a covariance matrix $\varSigma \in \mathfrak{M}_p$. By $\widehat{\varSigma}$ we denote the empirical covariance

$$\widehat{\varSigma} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^\top .$$

The well developed random matrix theory mainly focuses on the spectral or operator norm of $\widehat{\varSigma} - \varSigma$; see e.g. [13, 14] and references therein. The Frobenius loss $\|\widehat{\varSigma} - \varSigma\|_{\mathrm{Fr}}^2$ is much less studied. We mention [9] for the Gaussian case and [3] for $\boldsymbol{X}_i$ sub-gaussian. In some statistical problem like high dimensional random design regression [2, 5] or error-in-operator models [12], Frobenius norm of $\widehat{\varSigma} - \varSigma$ arises in a natural way. The difference $\widehat{\varSigma} - \varSigma$ is quadratic in the $\boldsymbol{X}_i$'s, hence, only sub-exponential. [5] applies Hanson-Wright approach and illustrates difficulties of studying $\|\widehat{\varSigma} - \varSigma\|_{\mathrm{Fr}}^2$.

This note presents some deviation bounds on the norm $\|Q\boldsymbol{\xi}\|$ for the case when the moment generating function $\mathbb{E} \exp\langle \boldsymbol{\xi}, \boldsymbol{u} \rangle$ is well defined on a sufficiently large but bounded set of vectors $\boldsymbol{u} \in \mathbb{R}^p$. The main challenge of the study is that the standard technique based on Markov inequality for $\|Q\boldsymbol{\xi}\|^2$ does not apply because the exponential moments of $\|Q\boldsymbol{\xi}\|^2$ diverge. We apply a kind of trimming technique combined with pilling device to obtain nearly sharp bounds in the form

$$\mathbb{P}\bigl(\|Q\boldsymbol{\xi}\| > z_c(B, \mathtt{x}) \bigr) \leqslant 3 \mathrm{e}^{-\mathtt{x}}$$

with a phase transition effect: for $\mathtt{x} \leqslant \mathtt{x}_c \approx \mathtt{g}^2/4$, the quantile function z_c is exactly as in the

Gaussian case: $z_c^2(B,\mathtt{x}) = \operatorname{tr}(B) + 2\sqrt{\mathtt{x}\operatorname{tr}(B^2)} + 2\|B\|\mathtt{x}$. This is important because the value $\mathtt{g}$ is typically large and for all moderate $\mathtt{x}$, we can apply nearly sharp Gaussian quantiles which only depend on the covariance matrix B and do not involve any other constant(s). The upper quantiles from Hanson-Wright or Bernstein type inequalities involve additional constants like $\|\boldsymbol{\xi}\|_{\psi_1}$ which can be quite large. For $\mathtt{x} > \mathtt{x}_c$, the function $z_c(B,\mathtt{x})$ grows linearly as $\mathtt{C}\mathtt{x}/\mathtt{g}$ for an absolute constant $\mathtt{C}$.

The paper is organized as follows. The main deviation bounds on $\|Q\boldsymbol{\xi}\|$ are collected in Section 2. Applications to weighted Bernoulli vector sums are given in Section 3. Sharp deviation bounds for empirical covariance matrix are given in Section 4. Some useful technical facts about Gaussian quadratic forms are collected in the Appendix A and Appendix B.

2. Deviation bounds under light exponential tails

Let $\boldsymbol{\xi}$ be a zero mean random vector in $\mathbb{R}^p$ with covariance $\operatorname{Var}(\boldsymbol{\xi})$ and let $Q\colon \mathbb{R}^p \to \mathbb{R}^q$ be a linear mapping. This section presents some deviation bounds on the norm $\|Q\boldsymbol{\xi}\|$ for the case of light exponential tails of $\boldsymbol{\xi}$. Namely,

$(\mathtt{g})$ *for some fixed* $\mathtt{g} > 0$ *and some self-adjoint operator* $\mathbb{V}^2$ *in* $\mathbb{R}^p$ *with* $\mathbb{V}^2 \geqslant \operatorname{Var}(\boldsymbol{\xi})$,

$$\phi(\boldsymbol{u}) \stackrel{\text{def}}{=} \log \mathbb{E} \exp\bigl(\langle \boldsymbol{u}, \mathbb{V}^{-1}\boldsymbol{\xi}\rangle\bigr) \leqslant \frac{\|\boldsymbol{u}\|^2}{2}, \quad \boldsymbol{u} \in \mathbb{R}^p, \quad \|\boldsymbol{u}\| \leqslant \mathtt{g}. \tag{2.1}$$

In fact, it is sufficient to assume that

$$\sup_{\|\boldsymbol{u}\| \leqslant \mathtt{g}} \mathbb{E} \exp\bigl(\langle \boldsymbol{u}, \mathbb{V}^{-1}\boldsymbol{\xi}\rangle\bigr) \leqslant \mathtt{C}. \tag{2.2}$$

The quantity $\mathtt{C}$ can be very large but it is not important. Indeed, the function $\phi(\boldsymbol{u})$ is analytic on the disk $\|\boldsymbol{u}\| \leqslant \mathtt{g}$, and condition (2.2) implies an analog of (2.1):

$$\phi(\boldsymbol{u}) \leqslant \frac{\|\boldsymbol{u}\|^2}{2} + \frac{\tau_3 \|\boldsymbol{u}\|^3}{6} \leqslant \frac{\|\boldsymbol{u}\|^2}{2}\Bigl(1 + \frac{\tau_3 \mathtt{g}}{3}\Bigr), \quad \|\boldsymbol{u}\| \leqslant \mathtt{g},$$

for a fixed value τ_3. Moreover, reducing $\mathtt{g}$ allows to take $\mathbb{V}^2$ equal or close to $\operatorname{Var}(\boldsymbol{\xi})$ and τ_3 close to zero. In the contrary to Hanson-Wright' type results, we do not specify the structure of the random vector $\boldsymbol{\xi}$ and do not require its entries to be independent. The next section presents our main results under $(\mathtt{g})$. The proofs are postponed until the end of the section.

2.1 Main results

Let a random vector $\boldsymbol{\xi}$ satisfy $\mathbb{E}\boldsymbol{\xi} = 0$ and $(\mathtt{g})$. The goal is to establish possibly sharp deviation bounds on $\|Q\boldsymbol{\xi}\|^2$ for a given linear mapping $Q\colon \mathbb{R}^p \to \mathbb{R}^q$. Define

$$\begin{gathered} B \stackrel{\text{def}}{=} Q\mathbb{V}^2 Q^\top, \quad \mathbb{p} \stackrel{\text{def}}{=} \operatorname{tr}(B), \quad \mathrm{v}^2 \stackrel{\text{def}}{=} \operatorname{tr}(B^2), \quad \lambda \stackrel{\text{def}}{=} \|B\|, \\ z^2(B,\mathtt{x}) \stackrel{\text{def}}{=} \operatorname{tr} B + 2\sqrt{\mathtt{x}\operatorname{tr}(B^2)} + 2\mathtt{x}\|B\| = \mathbb{p} + 2\mathrm{v}\sqrt{\mathtt{x}} + 2\mathtt{x}\lambda. \end{gathered} \tag{2.3}$$

Also fix some $\rho < 1$, a standard choice is $\rho = 1/2$. Our main result applies for all $\mathtt{x}$ satisfying the condition

$$z^2(B,\mathtt{x}) \leqslant \rho \left(\frac{\mathtt{g}\sqrt{\lambda}}{\mu(\mathtt{x})} - \sqrt{\frac{\mathbb{p}}{\mu(\mathtt{x})}} \right)^2 \tag{2.4}$$

with $z(B,\mathtt{x})$ from (2.3) and $\mu(\mathtt{x})$ defined by $\mu^{-1}(\mathtt{x}) = 1 + \frac{\mathrm{v}}{2\lambda\sqrt{\mathtt{x}}}$; see (B.3). One can see that the left hand-side of (2.4) increases with $\mathtt{x}$ while the right hand-side decreases. Therefore, there exists a unique root $\mathtt{x}_c$ such that with $\mu_c = \mu(\mathtt{x}_c)$,

$$z^2(B, \mathtt{x}_c) = \rho \left(\frac{\mathtt{g}\sqrt{\lambda}}{\mu_c} - \sqrt{\frac{\mathbb{p}}{\mu_c}} \right)^2 . \tag{2.5}$$

The value $\mathtt{x}_c$ is important, it describes the *phase transition* effect: the upper quantile function of $\|Q\boldsymbol{\xi}\|$ exhibits the Gaussian-like behavior for $\mathtt{x} \leqslant \mathtt{x}_c$, while it grows linearly with $\mathtt{x}/\mathtt{g}$ for $\mathtt{x} > \mathtt{x}_c$ as in a sub-exponential case.

Theorem 2.1 *Assume* $(\mathtt{g})$. *Fix* $\mathtt{x}_c$ *by* (2.5) *for some* $\rho \leqslant 1/2$. *It holds*

$$\mathbb{P}\bigl(\|Q\boldsymbol{\xi}\| \geqslant z(B, \mathtt{x})\bigr) \leqslant 3\mathrm{e}^{-\mathtt{x}}, \quad \mathtt{x} \leqslant \mathtt{x}_c. \tag{2.6}$$

For $\rho = 1/2$, *the value* $\mathtt{x}_c$ *from* (2.5) *fulfills*

$$\frac{1}{4}\left(\mathtt{g} - \sqrt{\frac{2\mathbb{p}}{\lambda}} \right)_+^2 \leqslant \mathtt{x}_c \leqslant \frac{\mathtt{g}^2}{4}. \tag{2.7}$$

If $\mathtt{g} > \sqrt{2\mathbb{p}/\lambda}$ *then* $z_c = z(B, \mathtt{x}_c)$ *follows*

$$\mathtt{g}\sqrt{\lambda/2} - (1 - 2^{-1/2})\sqrt{\mathbb{p}} \leqslant z_c \leqslant \mathtt{g}\sqrt{\lambda/2} + \sqrt{\mathbb{p}}. \tag{2.8}$$

The results of Theorem 2.1 state nearly Gaussian deviation bounds for the norm of the vector $Q\boldsymbol{\xi}$ satisfying $(\mathtt{g})$. Namely, the Gaussian deviation bound $\mathbb{P}\bigl(\|Q\boldsymbol{\xi}\| \geqslant z(B, \mathtt{x})\bigr) \leqslant \mathrm{e}^{-\mathtt{x}}$ from Theorem B.1 applies with the additional factor 3 for all $\mathtt{x} \leqslant \mathtt{x}_c$. Condition $\mathtt{g} \gg \sqrt{\mathbb{p}/\lambda}$ is important. Otherwise, the value $\mathtt{x}_c$ is not significantly large and the zone $\mathtt{x} \leqslant \mathtt{x}_c$ with Gaussian-like quantiles is too narrow. It turns out that out of this range, the norm $\|Q\boldsymbol{\xi}\|$ exhibits a sub-exponential behavior.

Theorem 2.2 *Assume* $(\mathtt{g})$. *With* $\mathtt{x}_c$ *from* (2.5) *and* $z_c = z(B, \mathtt{x}_c)$, *set* $\varkappa = \frac{\sqrt{\rho}\,\mathtt{g}}{(2+\sqrt{\rho})\sqrt{\lambda}}$. *It holds*

$$\begin{aligned} \mathbb{P}\bigl(\|Q\boldsymbol{\xi}\| > z_c + \varkappa^{-1}(\mathtt{x} - \mathtt{x}_c)\bigr) &\leqslant 3\mathrm{e}^{-\mathtt{x}}, & \mathtt{x} \geqslant \mathtt{x}_c, \\ \mathbb{P}\bigl(\|Q\boldsymbol{\xi}\| > z\bigr) &\leqslant 3\exp\{-\mathtt{x}_c - \varkappa(z - z_c)\}, & z \geqslant z_c. \end{aligned} \tag{2.9}$$

The obtained deviation bounds of Theorem 2.1 and Theorem 2.2 can be fused into one. To be more specific, we fix $\rho = 1/2$.

Corollary 2.3 *Assume* $(\mathtt{g})$. *Let* $\mathtt{x}_c$ *be defined by* (2.5) *with* $\rho = 1/2$. *For all* $\mathtt{x} > 0$,

$$\mathbb{P}\bigl(\|Q\boldsymbol{\xi}\| > z_c(B, \mathtt{x})\bigr) \leqslant 3\mathrm{e}^{-\mathtt{x}}, \tag{2.10}$$

where with $\varkappa \stackrel{\mathrm{def}}{=} \frac{\mathtt{g}}{(\sqrt{8}+1)\sqrt{\lambda}}$ *and* $\mathtt{x} \wedge \mathtt{x}_c \stackrel{\mathrm{def}}{=} \min\{\mathtt{x}, \mathtt{x}_c\}$,

$$z_c(B, \mathtt{x}) \stackrel{\mathrm{def}}{=} z(B, \mathtt{x} \wedge \mathtt{x}_c) + \varkappa^{-1}(\mathtt{x} - \mathtt{x}_c)_+ = \begin{cases} z(B, \mathtt{x}), & \mathtt{x} \leqslant \mathtt{x}_c, \\ z(B, \mathtt{x}_c) + \dfrac{\mathtt{x} - \mathtt{x}_c}{\varkappa}, & \mathtt{x} > \mathtt{x}_c. \end{cases} \tag{2.11}$$

Moreover, $\mathtt{x}_c$ *follows* (2.7) *and* $z_c = z(B, \mathtt{x}_c)$ *satisfies* (2.8) *provided* $\mathtt{g} \geqslant \sqrt{2\mathbb{p}/\lambda}$.

If $\mathtt{g} \gg \sqrt{\mathbb{p}/\lambda}$ then $\mathtt{x}_c$ is large and $z_c(B, \mathtt{x}) = z(B, \mathtt{x}) \leqslant \sqrt{\mathbb{p}} + \sqrt{2\mathtt{x}\lambda}$ for all reasonable $\mathtt{x}$. For $\mathtt{g} < \sqrt{2\mathbb{p}/\lambda}$, the accurate bound (2.11) can be simplified by a linear majorant which does not involve $\mathtt{x}_c$.

Theorem 2.4 *Assume* $(\mathtt{g})$. *Fix* $\varkappa = \frac{\mathtt{g}}{(\sqrt{8}+1)\sqrt{\lambda}}$. *Then* (2.10) *applies with*

$$z_c(B, \mathtt{x}) \leqslant \sqrt{\mathbb{p}} + \frac{\varkappa}{\sqrt{2}} + \varkappa^{-1}\mathtt{x}.$$

The next result provides some upper bounds on the exponential moments of $\|Q\boldsymbol{\xi}\|$. We distinguish between zones $z \leqslant z_c$ and $z > z_c$ with $z_c = z(B, \mathtt{x}_c)$; see (2.5).

Theorem 2.5 *Assume* (g). *Let* $\mathtt{x}_c$ *fulfill* (2.5) *and* $z_c = z(B, \mathtt{x}_c)$. *For any* $z \in [\sqrt{\mathtt{p}}, z_c]$ *and any* $\nu \leqslant \frac{z-\sqrt{\mathtt{p}}}{2\sqrt{\lambda}}$, *it holds*

$$\mathbb{E}\mathrm{e}^{\nu\|Q\boldsymbol{\xi}\|}\,\mathbb{I}(\|Q\boldsymbol{\xi}\| \geqslant z) \leqslant 6\exp\Bigl\{\nu z - \frac{(z-\sqrt{\mathtt{p}})^2}{2\lambda}\Bigr\}. \tag{2.12}$$

Further, for any $\nu < \varkappa \stackrel{\mathrm{def}}{=} \frac{\mathtt{g}\sqrt{\rho}}{\sqrt{\lambda}\,(2+\sqrt{\rho})}$,

$$\mathbb{E}\mathrm{e}^{\nu\|Q\boldsymbol{\xi}\|}\,\mathbb{I}(\|Q\boldsymbol{\xi}\| > z_c) \leqslant \frac{3\varkappa}{\varkappa-\nu}\,\exp\Bigl\{\nu z_c - \frac{(z_c-\sqrt{\mathtt{p}})^2}{2\lambda}\Bigr\}. \tag{2.13}$$

Moreover, for $z \geqslant z_c$,

$$\mathbb{E}\mathrm{e}^{\nu\|Q\boldsymbol{\xi}\|}\,\mathbb{I}(\|Q\boldsymbol{\xi}\| > z) \leqslant \frac{3\varkappa}{\varkappa-\nu}\,\exp\Bigl\{\nu z_c - \frac{(z_c-\sqrt{\mathtt{p}})^2}{2\lambda} - (\varkappa-\nu)(z-z_c)\Bigr\}. \tag{2.14}$$

2.2 Proof of Theorem 2.1

By normalization, one can easily reduce the study to the case $\|B\| = 1$. Moreover, replacing $\boldsymbol{\xi}$ with $\mathbb{V}^{-1}\boldsymbol{\xi}$ and Q with $Q\mathbb{V}$ reduces the proof to the situation with $\mathbb{V} = \mathbb{I}_p$. This will be assumed later on. For $\mu \in (0,1)$ and $\mathfrak{z}(\mu) = \mathtt{g}/\mu - \sqrt{\mathtt{p}/\mu} > 0$, define trimming $t_\mu(\boldsymbol{u})$ of $\boldsymbol{u} \in \mathbb{R}^p$ as

$$t_\mu(\boldsymbol{u}) \stackrel{\mathrm{def}}{=} \begin{cases} \boldsymbol{u}, & \text{if } \|\boldsymbol{u}\| \leqslant \mathfrak{z}(\mu), \\ \dfrac{\mathfrak{z}(\mu)}{\|\boldsymbol{u}\|}\,\boldsymbol{u}, & \text{otherwise.} \end{cases} \tag{2.15}$$

By construction $\|t_\mu(\boldsymbol{u})\| \leqslant \mathfrak{z}(\mu)$ for all $\boldsymbol{u} \in \mathbb{R}^p$.

Lemma 2.6 *Assume* (g) *and let* $\|B\| = 1$. *Fix* $\mu \in (0,1)$ *s.t.* $\mathfrak{z}(\mu) = \mathtt{g}/\mu - \sqrt{\mathtt{p}/\mu} > 0$. *Then with* $t_\mu(\cdot)$ *from* (2.15)

$$\mathbb{E}\exp\Bigl\{\frac{\mu}{2}\,t_\mu^2(Q\boldsymbol{\xi})\Bigr\} \leqslant 2\exp\{\varPhi(\mu)\}, \tag{2.16}$$

where

$$\varPhi(\mu) \stackrel{\mathrm{def}}{=} \frac{\mu^2\mathtt{v}^2}{4(1-\mu)} + \frac{\mu\,\mathtt{p}}{2}. \tag{2.17}$$

Furthermore, for any $\mathfrak{z} < \mathfrak{z}(\mu)$,

$$\mathbb{P}\bigl(\|Q\boldsymbol{\xi}\| > \mathfrak{z}, \|Q\boldsymbol{\xi}\| \leqslant \mathfrak{z}(\mu)\bigr) \leqslant 2\exp\Bigl\{-\frac{\mu\,\mathfrak{z}^2}{2} + \varPhi(\mu)\Bigr\}. \tag{2.18}$$

Proof Let us fix any value of $\boldsymbol{\xi}$. We intend to show that

$$\exp\Bigl\{\frac{\mu}{2}\,\|t_\mu(Q\boldsymbol{\xi})\|^2\Bigr\} \leqslant 2\mathbb{E}_{\boldsymbol{\gamma}}\,\exp\{\mu^{1/2}\boldsymbol{\gamma}^\top t_\mu(Q\boldsymbol{\xi})\}. \tag{2.19}$$

Here $\mathbb{E}_{\boldsymbol{\gamma}}$ means conditional expectation w.r.t. $\boldsymbol{\gamma} \sim \mathcal{N}(0, \mathbb{I}_p)$ given $\boldsymbol{\xi}$. Obviously, with $A = \{\boldsymbol{u}\colon \mu^{1/2}\|Q^\top\boldsymbol{u}\| \leqslant \mathtt{g}\}$, it suffices to check that

$$\mathcal{I}_\mu(\boldsymbol{\xi}) \stackrel{\mathrm{def}}{=} \mathbb{E}_{\boldsymbol{\gamma}}\exp\Bigl\{\mu^{1/2}\boldsymbol{\gamma}^\top t_\mu(Q\boldsymbol{\xi}) - \frac{\mu}{2}\|t_\mu(Q\boldsymbol{\xi})\|^2\Bigr\}\,\mathbb{I}(\boldsymbol{\gamma} \in A) \geqslant 1/2. \tag{2.20}$$

With $\mathtt{C}_p = (2\pi)^{-p/2}$, it holds

$$\begin{aligned}\mathcal{I}_\mu(\boldsymbol{\xi}) &= \mathtt{C}_p \int_A \exp\Bigl(\mu^{1/2}\boldsymbol{u}^\top t_\mu(Q\boldsymbol{\xi}) - \frac{\mu}{2}\|t_\mu(Q\boldsymbol{\xi})\|^2 - \frac{1}{2}\|\boldsymbol{u}\|^2\Bigr)\mathrm{d}\boldsymbol{u} \\ &= \mathtt{C}_p \int_A \exp\Bigl(-\frac{1}{2}\|\boldsymbol{u} - \mu^{1/2}t_\mu(Q\boldsymbol{\xi})\|^2\Bigr)\mathrm{d}\boldsymbol{u} = \mathbb{P}_{\boldsymbol{\gamma}}(\boldsymbol{\gamma} - \mu^{1/2}t_\mu(Q\boldsymbol{\xi}) \in A).\end{aligned}$$

The definition of A and the condition $\|t_\mu(Q\boldsymbol{\xi})\| \leqslant \mathfrak{z}(\mu)$ imply in view of $\|Q\| \leqslant 1$,

$$\begin{aligned}\mathbb{P}_{\boldsymbol{\gamma}}(\boldsymbol{\gamma} - \mu^{1/2}t_\mu(Q\boldsymbol{\xi}) \in A) &= \mathbb{P}_{\boldsymbol{\gamma}}\bigl(\|Q^\top(\boldsymbol{\gamma} - \mu^{1/2}t_\mu(Q\boldsymbol{\xi}))\| \leqslant \mathtt{g}/\mu^{1/2}\bigr) \\ &\geqslant \mathbb{P}_{\boldsymbol{\gamma}}\bigl(\|Q^\top\boldsymbol{\gamma}\| \leqslant \mathtt{g}/\mu^{1/2} - \mu^{1/2}\mathfrak{z}(\mu)\bigr) \geqslant \mathbb{P}_{\boldsymbol{\gamma}}\bigl(\|Q^\top\boldsymbol{\gamma}\| \leqslant \sqrt{\mathtt{p}}\bigr) \geqslant 1/2\end{aligned}$$

and (2.20) follows. Taking expectation for both sides of (2.19) and the use of Fubini's theorem yield

$$\mathbb{E}\exp\Bigl\{\frac{\mu}{2}\,\|t_\mu(Q\boldsymbol{\xi})\|^2\Bigr\} \leqslant 2\mathbb{E}_{\boldsymbol{\gamma}}\bigl\{\mathbb{E}\exp\{\mu^{1/2}\boldsymbol{\gamma}^\top t_\mu(Q\boldsymbol{\xi})\}\ \mathbb{1}(\mu^{1/2}\|Q^\top\boldsymbol{\gamma}\| \leqslant \mathtt{g})\bigr\}.$$

Obviously, for any $\boldsymbol{u} \in \mathbb{R}^p$,

$$\exp\{\boldsymbol{u}^\top t_\mu(Q\boldsymbol{\xi})\} + \exp\{-\boldsymbol{u}^\top t_\mu(Q\boldsymbol{\xi})\} \leqslant \exp\{\boldsymbol{u}^\top Q\boldsymbol{\xi}\} + \exp\{-\boldsymbol{u}^\top Q\boldsymbol{\xi}\}$$

and by (2.1)

$$\begin{aligned}\mathbb{E}\exp\Bigl\{\frac{\mu}{2}\,\|t_\mu(Q\boldsymbol{\xi})\|^2\Bigr\} &\leqslant 2\mathbb{E}_{\boldsymbol{\gamma}}\Bigl\{\exp\Bigl(\frac{1}{2}\,\|\mu^{1/2}\boldsymbol{\gamma}^\top Q\|^2\Bigr)\ \mathbb{1}(\mu^{1/2}\|Q^\top\boldsymbol{\gamma}\| \leqslant \mathtt{g})\Bigr\} \\ &\leqslant 2\mathbb{E}_{\boldsymbol{\gamma}}\ \exp\Bigl(\frac{1}{2}\,\|\mu^{1/2}\boldsymbol{\gamma}^\top Q\|^2\Bigr) = 2\det(\mathbb{I}_p - \mu Q^\top Q)^{-1/2}.\end{aligned}$$

We also use that for any $\mu > 0$ by (B.4),

$$\log\det\bigl(\mathbb{I} - \mu B\bigr)^{-1/2} \leqslant \frac{\mu\operatorname{tr}(B)}{2} + \frac{\mu^2\operatorname{tr}(B^2)}{4(1-\mu)} = \varPhi(\mu),$$

and the first statement follows. Moreover, by Markov's inequality

$$\mathbb{P}\bigl(\|Q\boldsymbol{\xi}\| > \mathfrak{z}, \|Q\boldsymbol{\xi}\| \leqslant \mathfrak{z}(\mu)\bigr) \leqslant \mathrm{e}^{-\mu\mathfrak{z}^2/2}\mathbb{E}\exp\Bigl\{\frac{\mu}{2}\,\|t_\mu(Q\boldsymbol{\xi})\|^2\Bigr\} \leqslant 2\exp\Bigl\{-\frac{\mu\mathfrak{z}^2}{2} + \varPhi(\mu)\Bigr\},$$

and (2.18) follows as well. □

The use of $\mu = \mu(\mathtt{x})$ from (B.3) in (2.16) yields

$$-\frac{\mu z^2(B,\mathtt{x})}{2} + \varPhi(\mu) = -\mathtt{x}, \tag{2.21}$$

and similarly to the proof of Theorem B.1

$$\mathbb{P}\Bigl(\|Q\boldsymbol{\xi}\|^2 > z^2(B,\mathtt{x}),\ \|Q\boldsymbol{\xi}\| \leqslant \mathfrak{z}(\mu)\Bigr) \leqslant 2\mathrm{e}^{-\mathtt{x}}. \tag{2.22}$$

It remains to consider the probability of large deviation $\mathbb{P}\bigl(\|Q\boldsymbol{\xi}\| > \mathfrak{z}(\mu)\bigr)$.

Lemma 2.7 *Assume* $\|B\| = 1$. *Given* $\mathtt{x} > 0$, *fix* $\mu = \mu(\mathtt{x})$ *and* $\mathfrak{z}(\mu) = \mathtt{g}/\mu - \sqrt{\mathtt{p}/\mu}$. *Assume* (2.4) *for some* $\rho \leqslant 1/2$. *Then*

$$\mathbb{P}\bigl(\|Q\boldsymbol{\xi}\| > \mathfrak{z}(\mu)\bigr) \leqslant \mathrm{e}^{-\mathtt{x}}. \tag{2.23}$$

Proof Denote $\eta = \|Q\boldsymbol{\xi}\|$. By (2.22)

$$\mathbb{P}\Bigl(\eta > z(B,\mathtt{x}),\ \eta \leqslant \mathfrak{z}(\mu)\Bigr) \leqslant 2\mathrm{e}^{-\mathtt{x}}. \tag{2.24}$$

For $\mu = \mu(\mathtt{x})$, it holds (2.21) with $\varPhi(\mu)$ given by (2.17). Bounding the tails of η in the region $\eta > \mathfrak{z}(\mu)$ requires another choice of μ. Namely, we apply (2.18) with $\rho\mu$ instead of μ yielding

$$\mathbb{P}\bigl(\eta > \mathfrak{z}(\mu), \eta \leqslant \mathfrak{z}(\rho\mu)\bigr) \leqslant 2\exp\Bigl\{-\frac{\rho\mu\,\mathfrak{z}^2(\mu)}{2} + \varPhi(\rho\mu)\Bigr\}.$$

In a similar way, applying (2.24) with $\rho^2\mu$ in place of μ and using that

$$\rho\,\mathfrak{z}(\rho\mu) = \mathtt{g}/\mu - \sqrt{\rho\,\mathbb{p}/\mu} \leqslant \mathfrak{z}(\mu) \tag{2.25}$$

yields

$$\begin{aligned}\mathbb{P}\bigl(\eta > \mathfrak{z}(\rho\mu), \eta \leqslant \mathfrak{z}(\rho^2\mu)\bigr) &\leqslant 2\exp\Bigl\{-\frac{\rho^2\mu\,\mathfrak{z}^2(\rho\mu)}{2} + \varPhi(\rho^2\mu)\Bigr\} \\ &\leqslant 2\exp\Bigl\{-\frac{\mu\,\mathfrak{z}^2(\mu)}{2} + \varPhi(\rho^2\mu)\Bigr\}.\end{aligned}$$

This trick can be applied again and again yielding in view of (2.25)

$$\begin{aligned}\mathbb{P}\bigl(\eta > \mathfrak{z}(\mu)\bigr) &\leqslant \sum_{k=0}^{\infty} \mathbb{P}\bigl(\eta > \mathfrak{z}(\rho^k\mu), \eta \leqslant \mathfrak{z}(\rho^{k+1}\mu)\bigr) \\ &\leqslant \sum_{k=0}^{\infty} 2\exp\bigl\{-\rho^{k+1}\mu\,\mathfrak{z}^2(\rho^k\mu)/2 + \varPhi(\rho^{k+1}\mu)\bigr\} \\ &\leqslant \sum_{k=0}^{\infty} 2\exp\bigl\{-\rho^{-k+1}\mu\,\mathfrak{z}^2(\mu)/2 + \varPhi(\rho^{k+1}\mu)\bigr\}.\end{aligned}$$

Condition $\rho\,\mathfrak{z}^2(\mu) \geqslant z^2(B,\mu)/2$ and (2.21) ensure for $\rho \leqslant 1/2$,

$$\begin{aligned}\mathbb{P}\bigl(\eta > \mathfrak{z}(\mu)\bigr) &\leqslant \sum_{k=0}^{\infty} 2\exp\bigl\{-\rho^{-k}\mu\, z^2(B,\mu)/2 + \varPhi(\rho^{k+1}\mu)\bigr\} \\ &\leqslant 2\sum_{k=0}^{\infty} \exp\bigl\{\varPhi(\rho^{k+1}\mu) - \rho^{-k}\varPhi(\mu) - \rho^{-k}\mathtt{x}\bigr\} \leqslant \mathrm{e}^{-\mathtt{x}}.\end{aligned}$$

This yields (2.23). □

Putting together (2.22) and (2.23) yields (2.6).

Now we check (2.7). Normalization by λ reduces the proof to the case $\|B\| = \|Q\mathbb{V}^2Q^\top\| = 1$. We use the simplified bounds $z(B,\mathtt{x}) \leqslant \sqrt{\mathbb{p}} + \sqrt{2\mathtt{x}}$ and $\mu^{-1} = 1 + \sqrt{\mathbb{p}/(4\mathtt{x})}$. Now (2.4) with $\rho = 1/2$ can be rewritten as

$$\mathtt{g} \geqslant \sqrt{\mu\,\mathbb{p}} + \mu\sqrt{2}\bigl(\sqrt{\mathbb{p}} + \sqrt{2\mathtt{x}}\bigr). \tag{2.26}$$

The use of $\mu = \sqrt{4\mathtt{x}}/(\sqrt{4\mathtt{x}} + \sqrt{\mathbb{p}})$ yields

$$\mu\sqrt{2}\bigl(\sqrt{\mathbb{p}} + \sqrt{2\mathtt{x}}\bigr) = \sqrt{8\mathtt{x}}\,\frac{\sqrt{\mathbb{p}} + \sqrt{2\mathtt{x}}}{\sqrt{\mathbb{p}} + \sqrt{4\mathtt{x}}} \geqslant \sqrt{4\mathtt{x}},$$

and (2.26) is not possible for $\mathtt{x} > \mathtt{g}^2/4$. Further, with $\mathtt{y} = \sqrt{4\mathtt{x}}/\mathtt{g}$ and $\alpha = \sqrt{\mathbb{p}}/\mathtt{g}$,

$$\frac{\sqrt{\mu\,\mathbb{p}} + \mu\sqrt{2}\bigl(\sqrt{\mathbb{p}} + \sqrt{2\mathtt{x}}\bigr)}{\mathtt{g}} = \sqrt{\frac{\mathtt{y}\alpha^2}{\alpha + \mathtt{y}}} + \frac{\mathtt{y}(\sqrt{2}\alpha + \mathtt{y})}{\alpha + \mathtt{y}} \leqslant \alpha + \mathtt{y} + \frac{\mathtt{y}(\sqrt{2} - 1)\alpha}{\alpha + \mathtt{y}} \leqslant \mathtt{y} + \sqrt{2}\alpha.$$

Together with (2.26), this yields $\mathtt{y} \geqslant 1 - \sqrt{2}\alpha$ and (2.7) follows. For (2.8) we use $z_c \leqslant \sqrt{\mathbb{p}} + \sqrt{2\lambda\mathtt{x}_c}$ and $z_c \geqslant \sqrt{\mathbb{p}/2} + \sqrt{2\lambda\mathtt{x}_c}$.

2.3 Proof of Theorem 2.2

Assume w.l.o.g. $\lambda = 1$. First we present an accurate deviation bound, which, however, does not provide a closed form quantile function for $\|Q\boldsymbol{\xi}\|$. Then we show how it implies a rough linear upper bound on this quantile function. For $\mathtt{x}_c$ from (2.5) and $\mathtt{x} > \mathtt{x}_c$, fix μ by the relation

$$\frac{\rho\,\mu\,\mathfrak{z}^2(\mu)}{2} = \mathtt{x} + \varPhi(\mu) = \mathtt{x} + \frac{\mu\,\mathbb{p}}{2} + \frac{\mu^2\mathtt{v}^2}{4(1-\mu)}, \tag{2.27}$$

where $\mathfrak{z}(\mu) = \mathsf{g}/\mu - \sqrt{\mathbb{p}/\mu}$; cf. (2.21). It is easy to see that the solution μ exists and unique. Moreover, if $\mathtt{x} = \mathtt{x}_c$ then $\mu = \mu_c$ and $\mathfrak{z}^2(\mu_c) = z^2(B,\mathtt{x}_c)$; see (2.5). If $\mathtt{x} > \mathtt{x}_c$, then $\mu < \mu_c$ and $\mathfrak{z}^2(\mu) > z^2(B,\mathtt{x})$.

Lemma 2.8 *For $\mathtt{x} > \mathtt{x}_c$, define μ by* (2.27)*. Then with $\mathfrak{z}(\mu) = \mathsf{g}/\mu - \sqrt{\mathbb{p}/\mu}$,*

$$\mathbb{P}\bigl(\|Q\boldsymbol{\xi}\|^2 > \rho\,\mathfrak{z}^2(\mu)\bigr) \leqslant 3\mathrm{e}^{-\mathtt{x}}. \tag{2.28}$$

Proof We again apply Lemma 2.6, however, the choice $\mu = \mu(\mathtt{x})$ from (B.3) is not possible anymore in view of $z(B,\mathtt{x}) > \mathfrak{z}(\mu)$. More precisely, for $\mathtt{x}$ large, the value $\mu(\mathtt{x})$ approaches one and this choice of μ yields the value $\mathfrak{z}(\mu)$ smaller than we need. To cope with this problem, we apply (2.18) of Lemma 2.6 with a sub-optimal μ from (2.27) ensuring $\rho\mu\,\mathfrak{z}^2(\mu) - \varPhi(\mu) = \mathtt{x}$. By (2.18) of Lemma 2.6

$$\mathbb{P}\bigl(\|Q\boldsymbol{\xi}\| > \sqrt{\rho}\,\mathfrak{z}(\mu), \|Q\boldsymbol{\xi}\| \leqslant \mathfrak{z}(\mu)\bigr) \leqslant 2\exp\Bigl\{-\frac{\rho\mu\,\mathfrak{z}^2(\mu)}{2} + \varPhi(\mu)\Bigr\} = 2\mathrm{e}^{-\mathtt{x}}.$$

Repeating the arguments from the proof of Lemma 2.7 implies

$$\begin{aligned}
\mathbb{P}\bigl(\|Q\boldsymbol{\xi}\|^2 > \rho\,\mathfrak{z}^2(\mu)\bigr) &\leqslant \sum_{k=0}^{\infty} 2\exp\Bigl\{-\frac{1}{2}\rho^{k+1}\mu\,\mathfrak{z}^2(\rho^k\mu) + \varPhi(\rho^k\mu)\Bigr\} \\
&\leqslant \sum_{k=0}^{\infty} 2\exp\Bigl\{-\frac{1}{2}\rho^{-k+1}\mu\,\mathfrak{z}^2(\mu) + \varPhi(\rho^k\mu)\Bigr\} \\
&\leqslant 2\mathrm{e}^{-\mathtt{x}} + 2\mathrm{e}^{-\mathtt{x}}\sum_{k=1}^{\infty}\exp\Bigl\{-\frac{1}{2}(\rho^{-k}-1)\rho\mu\,\mathfrak{z}^2(\mu) + \varPhi(\rho^k\mu) - \varPhi(\mu)\Bigr\} \leqslant 3\mathrm{e}^{-\mathtt{x}}
\end{aligned}$$

as stated in (2.28). □

It remains to evaluate $\rho\,\mathfrak{z}^2(\mu)$ with μ from (2.27) and $\mathfrak{z}(\mu) = \mathsf{g}/\mu - \sqrt{\mathbb{p}/\mu}$. For $\mu \leqslant \mu_c$,

$$\frac{\rho}{2}\Bigl(\frac{\mathsf{g}}{\sqrt{\mu}} - \sqrt{\mathbb{p}}\Bigr)^2 = \mathtt{x} + \varPhi(\mu)$$

and

$$\frac{\sqrt{\rho}\,\mathsf{g}}{\sqrt{\mu}} = \sqrt{2\mathtt{x} + 2\varPhi(\mu)} + \sqrt{\rho\,\mathbb{p}}.$$

This results in

$$\begin{aligned}
\sqrt{\rho}\,\mathfrak{z}(\mu) = \frac{\sqrt{\rho}}{\sqrt{\mu}}\Bigl(\frac{\mathsf{g}}{\sqrt{\mu}} - \sqrt{\mathbb{p}}\Bigr) &\leqslant \frac{1}{\sqrt{\rho}\,\mathsf{g}}\bigl(\sqrt{2\mathtt{x} + 2\varPhi(\mu)} + \sqrt{\rho\,\mathbb{p}}\bigr)\sqrt{2\mathtt{x} + 2\varPhi(\mu)} \\
&\leqslant \frac{1}{\sqrt{\rho}\,\mathsf{g}}\bigl(2\mathtt{x} + 2\varPhi(\mu_c) + \sqrt{\rho\,\mathbb{p}(2\mathtt{x} + 2\varPhi(\mu_c))}\bigr) \stackrel{\mathrm{def}}{=} \bar{z}(\mathtt{x}).
\end{aligned}$$

By (2.5), this inequality becomes equality for $\mathtt{x} = \mathtt{x}_c$ and $\mu = \mu_c$ with $\sqrt{\rho}\,\mathfrak{z}(\mu_c) = \bar{z}(\mathtt{x}_c) = z(B, \mathtt{x}_c)$. Furthermore, the derivative of $\bar{z}(\mathtt{x})$ w.r.t. $\mathtt{x}$ satisfies

$$\frac{\mathrm{d}}{\mathrm{d}\mathtt{x}}\,\bar{z}(\mathtt{x}) = \frac{1}{\sqrt{\rho}\,\mathtt{g}}\biggl(2 + \frac{\sqrt{\rho}\,\mathtt{p}}{\sqrt{2\mathtt{x} + 2\varPhi(\mu_c)}}\biggr) \leqslant \frac{1}{\sqrt{\rho}\,\mathtt{g}}\biggl(2 + \frac{\sqrt{\rho}\,\mathtt{p}}{\sqrt{2\mathtt{x}_c + 2\varPhi(\mu_c)}}\biggr).$$

Moreover, $2\mathtt{x}_c + 2\varPhi(\mu_c) = z^2(B, \mathtt{x}_c)$ and

$$\frac{\mathrm{d}}{\mathrm{d}\mathtt{x}}\,\bar{z}(\mathtt{x}) \leqslant \frac{1}{\sqrt{\rho}\,\mathtt{g}}\biggl(2 + \frac{\sqrt{\rho}\,\mathtt{p}}{z(B, \mathtt{x}_c)}\biggr) \leqslant \frac{2 + \sqrt{\rho}}{\sqrt{\rho}\,\mathtt{g}}$$

yielding

$$\bar{z}(\mathtt{x}) \leqslant \bar{z}(\mathtt{x}_c) + \frac{2 + \sqrt{\rho}}{\sqrt{\rho}\,\mathtt{g}}(\mathtt{x} - \mathtt{x}_c) = z(B, \mathtt{x}_c) + \frac{2 + \sqrt{\rho}}{\sqrt{\rho}\,\mathtt{g}}(\mathtt{x} - \mathtt{x}_c),$$

and hence,

$$\sqrt{\rho}\,\mathfrak{z}(\mu) \leqslant z(B, \mathtt{x}_c) + \frac{2 + \sqrt{\rho}}{\sqrt{\rho}\,\mathtt{g}}(\mathtt{x} - \mathtt{x}_c) = z_c + \frac{\mathtt{x} - \mathtt{x}_c}{\varkappa}. \tag{2.29}$$

This implies (2.9).

2.4 Proof of Theorem 2.4

As previously, assume $\lambda = 1$. We use $z(B, \mathtt{x}_c) \leqslant \sqrt{\mathtt{p}} + \sqrt{2\mathtt{x}_c}$. Further, $\varkappa^{-1}\mathtt{x}_c - \sqrt{2\mathtt{x}_c} + \varkappa/\sqrt{2} \geqslant 0$ and thus,

$$\sqrt{2\mathtt{x}_c} - \varkappa^{-1}\mathtt{x}_c \leqslant \varkappa/\sqrt{2}.$$

Therefore, for $\mathtt{x} \geqslant \mathtt{x}_c$, it holds

$$z_c(B, \mathtt{x}) = z(B, \mathtt{x}_c) + \frac{\mathtt{x} - \mathtt{x}_c}{\varkappa} \leqslant \sqrt{\mathtt{p}} + \sqrt{2\mathtt{x}_c} - \frac{\mathtt{x}_c}{\varkappa} + \frac{\mathtt{x}}{\varkappa} \leqslant \sqrt{\mathtt{p}} + \frac{\varkappa}{\sqrt{2}} + \frac{\mathtt{x}}{\varkappa}.$$

In the zone $\mathtt{x} \leqslant \mathtt{x}_c$, it holds $z_c(B, \mathtt{x}) = z(B, \mathtt{x}) \leqslant \sqrt{\mathtt{p}} + \sqrt{2\mathtt{x}}$ and it remains to note that $\sqrt{2\mathtt{x}} \leqslant \varkappa/\sqrt{2} + \varkappa^{-1}\mathtt{x}$.

2.5 Proof of Theorem 2.5

Assume w.o.l.g. $\lambda = 1$. First consider $z \geqslant z_c$. By (2.29) of Theorem 2.2, it holds with $\varkappa = \mathtt{g}\sqrt{\rho}/(2 + \sqrt{\rho})$ and $\mathtt{x}_c = (z_c - \sqrt{\mathtt{p}})^2/2$,

$$\mathbb{P}\bigl(\|Q\boldsymbol{\xi}\| \geqslant z\bigr) = \mathbb{P}\bigl(\|Q\boldsymbol{\xi}\| \geqslant z_c + z - z_c\bigr) \leqslant 3\mathrm{e}^{-\mathtt{x}_c - \varkappa(z - z_c)}.$$

In particular, $\mathbb{P}(\|Q\boldsymbol{\xi}\| \geqslant z_c) \leqslant 3\mathrm{e}^{-\mathtt{x}_c}$. Integration by parts yields for $\nu < \varkappa$,

$$\begin{aligned}
\mathbb{E}\mathrm{e}^{\nu(\|Q\boldsymbol{\xi}\| - z_c)}\,\mathbb{I}(\|Q\boldsymbol{\xi}\| > z_c) &= -\int_{z_c}^{\infty} \mathrm{e}^{\nu(z - z_c)}\mathrm{d}\mathbb{P}(\|Q\boldsymbol{\xi}\| \geqslant z) \\
&= \mathbb{P}(\|Q\boldsymbol{\xi}\| \geqslant z_c) + \nu\int_{z_c}^{\infty} \mathrm{e}^{\nu(z - z_c)}\,\mathbb{P}(\|Q\boldsymbol{\xi}\| \geqslant z)\,\mathrm{d}z \\
&\leqslant 3\mathrm{e}^{-\mathtt{x}_c} + \nu\int_{z_c}^{\infty} \mathrm{e}^{\nu(z - z_c) - \mathtt{x}_c - \varkappa(z - z_c)}\,\mathrm{d}z = \biggl(3 + \frac{3\nu}{\varkappa - \nu}\biggr)\,\mathrm{e}^{-\mathtt{x}_c},
\end{aligned} \tag{2.30}$$

and (2.13) follows. Similarly, for $z \geqslant z_c$, we derive (2.14) as follows

$$\begin{aligned} \mathbb{E}\mathrm{e}^{\nu\|Q\boldsymbol{\xi}\|}\, \mathrm{1\!I}(\|Q\boldsymbol{\xi}\| > z) &= -\int_{z}^{\infty} \mathrm{e}^{\nu t} \mathrm{d}\mathbb{P}(\|Q\boldsymbol{\xi}\| \geqslant t) \\ &\leqslant 3\mathrm{e}^{\nu z_c - \mathtt{x}_c - \varkappa(z - z_c)} + \frac{3\nu}{\varkappa - \nu}\, \mathrm{e}^{\nu z_c - \mathtt{x}_c - \varkappa(z - z_c)} = \frac{3\varkappa}{\varkappa - \nu}\, \mathrm{e}^{\nu z_c - \mathtt{x}_c - (\varkappa - \nu)(z - z_c)}. \end{aligned}$$

Now fix $z_\circ$ with $z_\circ - \sqrt{\mathtt{p}} \geqslant 2\nu$ but $z_\circ \leqslant z_c$. Then

$$\begin{aligned} \mathbb{E}\mathrm{e}^{\nu\|Q\boldsymbol{\xi}\|}\, \mathrm{1\!I}(\|Q\boldsymbol{\xi}\| > z_\circ) &= -\int_{z_\circ}^{\infty} \mathrm{e}^{\nu z} \mathrm{d}\mathbb{P}(\|Q\boldsymbol{\xi}\| \geqslant z) \\ &= \mathrm{e}^{\nu z_\circ} \mathbb{P}(\|Q\boldsymbol{\xi}\| \geqslant z_\circ) + \nu \left(\int_{z_\circ}^{z_c} + \int_{z_c}^{\infty} \right) \mathrm{e}^{\nu z} \mathbb{P}(\|Q\boldsymbol{\xi}\| \geqslant z) \mathrm{d}z. \end{aligned}$$

By (2.6), for any $z \in [z_\circ, z_c]$, it holds in view of $z(B, \mathtt{x}) \leqslant \sqrt{\mathtt{p}} + \sqrt{2\mathtt{x}}$,

$$\mathbb{P}(\|Q\boldsymbol{\xi}\| \geqslant z) \leqslant 3\mathrm{e}^{-(z - \sqrt{\mathtt{p}})^2/2}.$$

As $(\nu z - (z - \sqrt{\mathtt{p}})^2/2)' = \nu - z + \sqrt{\mathtt{p}} \leqslant -\nu$ for $z - \sqrt{\mathtt{p}} \geqslant 2\nu$, it holds

$$\nu \int_{z_\circ}^{z_c} \mathrm{e}^{\nu z - (z - \sqrt{\mathtt{p}})^2/2} \mathrm{d}z \leqslant \mathrm{e}^{\nu z_\circ - (z_\circ - \sqrt{\mathtt{p}})^2/2}\, \nu \int_{z_\circ}^{z_c} \mathrm{e}^{-\nu(z - z_\circ)} \mathrm{d}z \leqslant \mathrm{e}^{\nu z_\circ - (z_\circ - \sqrt{\mathtt{p}})^2/2}$$

and also $\nu z_\circ - (z_\circ - \sqrt{\mathtt{p}})^2/2 > \nu z_c - (z_c - \sqrt{\mathtt{p}})^2/2$. Putting this together with the above bound on $\int_{z_c}^{\infty} \mathrm{e}^{\nu z} \mathbb{P}(\|Q\boldsymbol{\xi}\| \geqslant z) \mathrm{d}z$ as in (2.30) completes the proof of (2.12).

3. Deviation bounds for Bernoulli vector sums

Let Y_i be independent Bernoulli(θ_i^*), $i = 1, \ldots, n$. We denote $\boldsymbol{Y} = (Y_i) \in \mathbb{R}^n$. Weighted sums of the Y_i naturally appear in various statistical tasks including classification, binary response models, logistic regression etc. Recent applications include e.g. stochastic block modeling; see e.g. [1, 6] and references therein, or ranking from pairwise comparison [4] among many others. We show how the general bounds of Section 2 can be used for vector sums of Bernoulli r.v.s. For a linear mapping $\boldsymbol{\Psi} \colon \mathbb{R}^n \to \mathbb{R}^p$, define $\boldsymbol{\xi} = \boldsymbol{\Psi}(\boldsymbol{Y} - \mathbb{E}\boldsymbol{Y})$. Below we state some deviation bounds on the squared norm $\|\boldsymbol{\xi}\|^2$ starting from the univariate case.

3.1 Weighted sums of Bernoulli r.v.'s: Univariate case

Given a collections of weights (w_i), define

$$\begin{aligned} S &= \sum_{i=1}^{n} Y_i w_i, \\ V^2 &= \operatorname{Var}(S) = \sum_{i=1}^{n} \theta_i^*(1 - \theta_i^*) w_i^2, \\ w^* &= \max_i |w_i|. \end{aligned}$$

First, we state a deviation bound for a centered sum $S - \mathbb{E}S$.

Proposition 3.1 *Let* Y_i *be independent* Bernoulli(θ_i^*) *and* $w_i \in \mathbb{R}$, $i = 1, \ldots, n$. *Then* $S = \sum_{i=1}^{n} Y_i w_i$ *satisfies*

$$\log \mathbb{E} \exp\Bigl\{ \frac{\lambda(S - \mathbb{E}S)}{V} \Bigr\} \leqslant \lambda^2, \quad \lambda \leqslant \frac{\log(2) V}{w^*}. \tag{3.1}$$

Furthermore, suppose that given $\mathtt{x} \geqslant 0$,

$$V \geqslant \frac{3}{2}\, w^* \sqrt{\mathtt{x}}. \tag{3.2}$$

Then

$$\mathbb{P}\bigl(V^{-1}|S - \mathbb{E}S| \geqslant 2\sqrt{\mathtt{x}}\bigr) \leqslant 2\mathrm{e}^{-\mathtt{x}}. \tag{3.3}$$

Without (3.2)*, the bound* (3.3) *applies with* V *replaced by* $V_{\mathtt{x}} = V \vee (3\, w^* \sqrt{\mathtt{x}}\,/2)$*.*

Proof Without loss of generality assume $w^* = 1$, otherwise just rescale all the weights by the factor $1/w^*$. We use that

$$f(u) \stackrel{\mathrm{def}}{=} \log \mathbb{E} \exp\Bigl\{ u(S - \mathbb{E}S) \Bigr\} = \sum_{i=1}^{N} \Bigl[\log\bigl(\theta_i^* \mathrm{e}^{u w_i} + 1 - \theta_i^*\bigr) - u w_i \theta_i^* \Bigr].$$

This is an analytic function of u for $|u| \leqslant \log 2$ satisfying $f(0) = 0$, $f'(0) = 0$, and, with $\upsilon_i^* = \log\theta_i^* - \log(1 - \theta_i^*)$,

$$f''(u) = \sum_{i=1}^{N} \frac{w_i^2\, \theta_i^* (1 - \theta_i^*)\, \mathrm{e}^{u w_i}}{(\theta_i^* \mathrm{e}^{u w_i} + 1 - \theta_i^*)^2} = \sum_{i=1}^{N} \frac{w_i^2\, \mathrm{e}^{\upsilon_i^* + u w_i}}{(\mathrm{e}^{\upsilon_i^* + u w_i} + 1)^2} = \sum_{i=1}^{N} \theta_i(u) \bigl\{ 1 - \theta_i(u) \bigr\} w_i^2$$

for $\theta_i(u) = \mathrm{e}^{\upsilon_i^* + u w_i} / (\mathrm{e}^{\upsilon_i^* + u w_i} + 1)$. Clearly $\theta_i(u)$ and thus, $\theta_i(u)\bigl\{1 - \theta_i(u)\bigr\}$ monotonously increases with u and it holds for $\theta_i^* = \theta_i(0)$,

$$\theta_i(u)\bigl\{1 - \theta_i(u)\bigr\} \leqslant \mathrm{e}^{|u|}\, \theta_i^*\, (1 - \theta_i^*) \leqslant 2\, \theta_i^*\, (1 - \theta_i^*), \quad |u| \leqslant \log 2.$$

This yields

$$f(u) \leqslant V^2\, u^2, \qquad |u| \leqslant \log 2.$$

As $\mathtt{x} \leqslant 4V^2/9$, the value $\lambda = \sqrt{\mathtt{x}}$ fulfills $\lambda/V = \sqrt{\mathtt{x}}/V \leqslant \log 2 \leqslant 2^{-1/2}$. Now by the exponential Chebyshev inequality

$$\begin{aligned} \mathbb{P}\Bigl(V^{-1}(S - \mathbb{E}S) \geqslant 2\sqrt{\mathtt{x}} \Bigr) &\leqslant \exp\bigl\{ -2\lambda\sqrt{\mathtt{x}} + f(\lambda/V) \bigr\} \\ &\leqslant \exp\bigl(-2\lambda\sqrt{\mathtt{x}} + \lambda^2 \bigr) = \mathrm{e}^{-\mathtt{x}}. \end{aligned}$$

Similarly one can bound $\mathbb{E}S - S$. □

3.2 Deviation bounds for Bernoulli vector sums

Now we present an upper bound on the norm of a vector $\boldsymbol{\xi} = \boldsymbol{\Psi}(\boldsymbol{Y} - \mathbb{E}\boldsymbol{Y})$, where $\boldsymbol{\Psi}$ is a linear mapping $\boldsymbol{\Psi}\colon \mathbb{R}^n \to \mathbb{R}^p$. It holds

$$\operatorname{Var}(\boldsymbol{\xi}) = \operatorname{Var}(\boldsymbol{\Psi}\boldsymbol{Y}) = \boldsymbol{\Psi} \operatorname{Var}(\boldsymbol{Y}) \boldsymbol{\Psi}^\top.$$

We aim at bounding the squared norm $\|Q\boldsymbol{\xi}\|^2$ for another linear mapping $Q\colon \mathbb{R}^p \to \mathbb{R}^q$.

Theorem 3.2 *Let* $Y_i \sim \mathrm{Bernoulli}(\theta_i^*)$, $i = 1, \ldots, n$. *Consider* $\boldsymbol{\xi} = \boldsymbol{\Psi}(\boldsymbol{Y} - \mathbb{E}\boldsymbol{Y})$, *and let* $\mathbb{V}^2 \geqslant 2\operatorname{Var}(\boldsymbol{\xi})$*. Define*

$$w^* = \max_{i \leqslant n} \|\mathbb{V}^{-1}\boldsymbol{\Psi}_i\|, \quad \mathsf{g} = \log(2)/w^*.$$

Then with $B = Q\mathbb{V}^2 Q^\top$ *and* $z_c(B, \mathtt{x})$ *from* (2.11)*, it holds*

$$\mathbb{P}\bigl(\|Q\boldsymbol{\xi}\| \geqslant z_c(B, \mathtt{x}) \bigr) \leqslant 3\mathrm{e}^{-\mathtt{x}}.$$

Proof We apply the general result of Corollary 2.3 under conditions (2.1). For any vector $\boldsymbol{u}$, consider the scalar product $\langle \mathbb{V}^{-1}\boldsymbol{\xi}, \boldsymbol{u}\rangle = \langle \mathbb{V}^{-1}\boldsymbol{\Psi}(\boldsymbol{Y} - \mathbb{E}\boldsymbol{Y}), \boldsymbol{u}\rangle$. It is obviously a weighted centered sum of the Bernoulli r.v.'s $Y_i - \theta_i^*$ with

$$\operatorname{Var}\langle \mathbb{V}^{-1}\boldsymbol{\xi}, \boldsymbol{u}\rangle \leqslant \|\boldsymbol{u}\|^2/2.$$

One can write with $\epsilon_i = Y_i - \theta_i^*$ and $\boldsymbol{\varepsilon} = (\epsilon_i)$,

$$\langle \mathbb{V}^{-1}\boldsymbol{\xi}, \boldsymbol{u}\rangle = \bigl\langle \boldsymbol{\varepsilon}, \boldsymbol{\Psi}^\top \mathbb{V}^{-1}\boldsymbol{u}\bigr\rangle.$$

By the Cauchy-Schwarz inequality, it holds

$$\|\boldsymbol{\Psi}^\top \mathbb{V}^{-1}\boldsymbol{u}\|_\infty = \max_i \bigl|(\mathbb{V}^{-1}\boldsymbol{\Psi}_i)^\top \boldsymbol{u}\bigr| \leqslant w^* \|\boldsymbol{u}\|.$$

Bound (3.1) of Proposition 3.1 on the exponential moments of $\langle \mathbb{V}^{-1}\boldsymbol{\xi}, \boldsymbol{u}\rangle$ implies

$$\log \mathbb{E} \exp\bigl\{\langle \mathbb{V}^{-1}\boldsymbol{\xi}, \boldsymbol{u}\rangle\bigr\} \leqslant \|\boldsymbol{u}\|^2/2, \quad \|\boldsymbol{u}\| \leqslant \log(2)/w^*.$$

Therefore, (2.1) is fulfilled with $\mathsf{g} = \log(2)/w^*$. The deviation bound (2.10) of Corollary 2.3 yields the assertion. $\square$

4. Frobenius norm losses for empirical covariance

Let $\boldsymbol{X}_i \sim \mathcal{N}(0, \Sigma)$ be i.i.d. zero mean Gaussian vectors in $\mathbb{R}^p$ with a covariance matrix $\Sigma \in \mathfrak{M}_p$. By $\widehat{\Sigma}$ we denote the empirical covariance

$$\widehat{\Sigma} \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i \boldsymbol{X}_i^\top .$$

Our goal is to establish sharp dimension free deviation bounds on the squared Frobenius norm $\|\widehat{\Sigma} - \Sigma\|_{\mathrm{Fr}}^2$:

$$\|\widehat{\Sigma} - \Sigma\|_{\mathrm{Fr}}^2 = \operatorname{tr}(\widehat{\Sigma} - \Sigma)^2.$$

We demonstrate how the general results of Section 2 can be used for obtaining accurate deviation bounds for $\|\widehat{\Sigma} - \Sigma\|_{\mathrm{Fr}}^2$ and for supporting the concentration phenomenon.

4.1 Upper bounds

First we establish a tight upper bound on $\|\widehat{\Sigma} - \Sigma\|_{\mathrm{Fr}}^2$. We identify the matrix $\widehat{\Sigma}$ with the vector in the linear subspace of $\mathbb{R}^{p\times p}$ composed by symmetric matrices. Our aim is in showing that the quantiles of $\|\widehat{\Sigma} - \Sigma\|_{\mathrm{Fr}}^2$ mimic well similar quantiles of $\|\widetilde{\Sigma} - \Sigma\|_{\mathrm{Fr}}^2$ for a Gaussian matrix $\widetilde{\Sigma}$ with the same covariance structure as $\widehat{\Sigma}$. Define

$$\mathbb{p}(\Sigma) = (\operatorname{tr}\Sigma)^2 + \operatorname{tr}\Sigma^2, \quad \mathtt{v}^2(\Sigma) = \bigl(\operatorname{tr}\Sigma^2\bigr)^2 + \operatorname{tr}\Sigma^4. \tag{4.1}$$

Later we show that $\mathbb{p}(\Sigma) = \mathbb{E}\|\widehat{\Sigma} - \Sigma\|_{\mathrm{Fr}}^2 = \operatorname{tr}\operatorname{Var}(\widetilde{\Sigma})$ and $\mathtt{v}^2(\Sigma) = \operatorname{tr}\{\operatorname{Var}(\widetilde{\Sigma})\}^2$ while $\lambda(\Sigma) = \|\operatorname{Var}(\widetilde{\Sigma})\| = 2\|\Sigma\|^2$. In our results we implicitly assume a high dimensional situation with $\mathbb{p}(\Sigma)$ large. The presented bounds also require that $n \gg \mathbb{p}(\Sigma)$.

Theorem 4.1 *Assume $\|\Sigma\| = 1$ and $\mathbb{p}(\Sigma) < n/8$. Given $\mathtt{x}$ with $4\sqrt{\mathtt{x}} < \sqrt{n/8} - \sqrt{\mathbb{p}(\Sigma)}$, fix $\rho < 1$ by*

$$\rho(1-\rho)\sqrt{n/8} = \sqrt{\mathbb{p}(\Sigma)} + 4\sqrt{\mathtt{x}}. \tag{4.2}$$

Then

$$\mathbb{P}\Bigl(n\|\widehat{\varSigma}-\varSigma\|_{\mathrm{Fr}}^2 > \frac{1}{1-\rho}\bigl\{\mathtt{p}(\varSigma)+2\mathtt{v}(\varSigma)\sqrt{\mathtt{x}}+4\mathtt{x}\bigr\}\Bigr) \leqslant 3\mathrm{e}^{-\mathtt{x}}. \tag{4.3}$$

4.2 Lower bounds

This section presents a lower bound on the Frobenius norm of $\widehat{\varSigma}-\varSigma$. Later in Section 4.3 we state the concentration phenomenon for $\|\widehat{\varSigma}-\varSigma\|_{\mathrm{Fr}}^2$.

Theorem 4.2 *Let* $\|\varSigma\|=1$ *and* $\mathtt{p}(\varSigma)$ *and* $\mathtt{v}(\varSigma)$ *be defined by* (4.1). *For* $\mathtt{x}>0$ *with* $2\sqrt{\mathtt{x}}\leqslant \mathtt{p}(\varSigma)/\mathtt{v}(\varSigma)$, *define* $\mu=\mu(\mathtt{x})=2\sqrt{\mathtt{x}}/\mathtt{v}(\varSigma)$ *and assume that there is* $\alpha<1/2$ *satisfying*

$$\alpha\sqrt{\frac{1-2\alpha}{1-\alpha}} \geqslant \sqrt{\frac{\mu(\mathtt{x})}{n}}\Bigl(\sqrt{2\mathtt{p}(\varSigma)}+\frac{\sqrt{2}\,\mathtt{p}(\varSigma)}{\mathtt{v}(\varSigma)}\Bigr). \tag{4.4}$$

Then

$$\mathbb{P}\Bigl(n\|\widehat{\varSigma}-\varSigma\|_{\mathrm{Fr}}^2 < \frac{1-2\alpha}{1-\alpha}\,\mathtt{p}(\varSigma)-2\mathtt{v}(\varSigma)\sqrt{\mathtt{x}}\Bigr) \leqslant 2\mathrm{e}^{-\mathtt{x}}.$$

4.3 Concentration of the Frobenius loss

Putting together Theorem 4.1 and Theorem 4.2 yields the following corollary.

Corollary 4.3 *Under conditions of Theorem 4.1 and Theorem 4.2, it holds for any* $\mathtt{x}$ *resolving* (4.2) *and* (4.4) *on a random set* $\varOmega(\mathtt{x})$ *with* $\mathbb{P}\bigl(\varOmega(\mathtt{x})\bigr)\geqslant 1-5\mathrm{e}^{-\mathtt{x}}$,

$$\frac{1-2\alpha}{1-\alpha}\,\mathtt{p}(\varSigma)-2\mathtt{v}(\varSigma)\sqrt{\mathtt{x}} \leqslant n\|\widehat{\varSigma}-\varSigma\|_{\mathrm{Fr}}^2 \leqslant \frac{1}{1-\rho}\bigl\{\mathtt{p}(\varSigma)+2\mathtt{v}(\varSigma)\sqrt{\mathtt{x}}+4\mathtt{x}\bigr\}. \tag{4.5}$$

This result mimics similar bound of Theorem B.1 for $\widehat{\varSigma}$ Gaussian and of Theorem 2.1 for $\widehat{\varSigma}$ sub-Gaussian. However, the empirical covariance $\widehat{\varSigma}$ is quadratic in the $\boldsymbol{X}_i$'s and thus, only sub-exponential. We pay an additional factor $(1-\rho)^{-1}$ in the upper quantile function and the factor $\frac{1-2\alpha}{1-\alpha}$ in the lower quantile function for this extension.

Further we discuss the concentration phenomenon for the Frobenius error $n\|\widehat{\varSigma}-\varSigma\|_{\mathrm{Fr}}^2$ around its expectation $\mathtt{p}(\varSigma)$. Even in the Gaussian case, it meets only in high-dimensional situation with $\mathtt{p}(\varSigma)$ large. As $\mathtt{v}^2(\varSigma)\leqslant \mathtt{p}(\varSigma)\lambda(\varSigma)=2\mathtt{p}(\varSigma)$, this also implies $\mathtt{v}(\varSigma)\ll \mathtt{p}(\varSigma)$. Statement (4.5) can be rewritten as

$$-\frac{\alpha\,\mathtt{p}(\varSigma)}{1-\alpha}-2\mathtt{v}(\varSigma)\sqrt{\mathtt{x}} \leqslant n\|\widehat{\varSigma}-\varSigma\|_{\mathrm{Fr}}^2-\mathtt{p}(\varSigma) \leqslant \frac{\rho\,\mathtt{p}(\varSigma)}{1-\rho}+\frac{2\mathtt{v}(\varSigma)\sqrt{\mathtt{x}}+4\mathtt{x}}{1-\rho}.$$

Therefore, concentration effect of the loss $n\|\widehat{\varSigma}-\varSigma\|_{\mathrm{Fr}}^2$ requires $\mathtt{p}(\varSigma)$ large and α and ρ small. Then for $\mathtt{x}\ll\mathtt{p}(\varSigma)$, quantiles of $n\|\widehat{\varSigma}-\varSigma\|_{\mathrm{Fr}}^2-\mathtt{p}(\varSigma)$ are smaller in order than $\mathtt{p}(\varSigma)$. Definition (4.2) of ρ ensures $\rho\asymp\sqrt{\mathtt{p}(\varSigma)/n}$, and hence, " $\rho\ll 1$ " is equivalent to " $\mathtt{p}(\varSigma)\ll n$ ". Condition ensuring $\alpha\ll 1$ is similar. To see this, assume $\mathtt{v}^2(\varSigma)\asymp\mathtt{p}(\varSigma)$. Then $\mathtt{x}\ll\mathtt{p}(\varSigma)$ yields $\mu(\mathtt{x})=2\sqrt{\mathtt{x}}/\mathtt{v}(\varSigma)\ll 1$ and definition (4.4) of α implies

$$\alpha \lesssim \sqrt{\frac{\mu}{n}}\Bigl(\sqrt{2\mathtt{p}(\varSigma)}+\frac{\sqrt{2}\,\mathtt{p}(\varSigma)}{\mathtt{v}(\varSigma)}\Bigr) \lesssim \sqrt{\frac{\mathtt{p}(\varSigma)}{n}}.$$

4.4 Weighted frobenius norm

The result can be easily extended to the case of a weighted Frobenius norm. Consider for any linear mapping $A\colon \mathbb{R}^p \to \mathbb{R}^q$ the value $n\|A(\widehat{\Sigma}-\Sigma)A^\top\|_{\mathrm{Fr}}^2$.

Theorem 4.4 *Let* $\|\Sigma\| = 1$ *and* $A\colon \mathbb{R}^p \to \mathbb{R}^q$ *be a linear operator with* $\|A\| = \|A^\top A\| = 1$. *Define* $\Sigma_A \stackrel{\mathrm{def}}{=} A\Sigma A^\top$,

$$\mathbb{p}_A \stackrel{\mathrm{def}}{=} \mathbb{p}(\Sigma_A) = \operatorname{tr}^2(\Sigma_A) + \operatorname{tr}(\Sigma_A)^2, \quad \mathtt{v}_A^2 \stackrel{\mathrm{def}}{=} \mathtt{v}^2(\Sigma_A) = \bigl\{\operatorname{tr}(\Sigma_A^2)\bigr\}^2 + \operatorname{tr}(\Sigma_A)^4,$$

and assume $\mathbb{p}_A < n/8$. *The the statements of Theorem 4.1 and Theorem 4.2 apply to* $n\|A(\widehat{\Sigma}-\Sigma)A^\top\|_{\mathrm{Fr}}^2$ *after replacing* $\mathbb{p}(\Sigma)$ *and* $\mathtt{v}(\Sigma)$ *with* $\mathbb{p}_A$ *and* $\mathtt{v}_A$.

Proof We can represented

$$\sqrt{n}\, A(\widehat{\Sigma}-\Sigma)A^\top = A\,\Sigma^{1/2}\mathcal{E}\,\Sigma^{1/2}A^\top$$

with $\mathcal{E}$ from (4.6). This reduces the result to the previous case with $\Sigma_A = A\Sigma A^\top$ in place of Σ. □

4.5 Proof of Theorem 4.1

Each vector $\boldsymbol{\gamma}_i = \Sigma^{-1/2}\boldsymbol{X}_i$ is standard normal. Define

$$\mathcal{E} = \frac{1}{n^{1/2}} \sum_{i=1}^n (\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^\top - \mathbb{I}_p). \tag{4.6}$$

We will use the representation $\widehat{\Sigma} - \Sigma = n^{-1/2}\Sigma^{1/2}\,\mathcal{E}\,\Sigma^{1/2}$ and

$$n\|\widehat{\Sigma}-\Sigma\|_{\mathrm{Fr}}^2 = \operatorname{tr}(\Sigma^{1/2}\mathcal{E}\,\Sigma\,\mathcal{E}\,\Sigma^{1/2}) = \|\Sigma^{1/2}\mathcal{E}\,\Sigma^{1/2}\|_{\mathrm{Fr}}^2.$$

The main step is in applying Theorem 2.1 to the quadratic form $\|Q\mathcal{E}\|_{\mathrm{Fr}}^2$ with $Q\mathcal{E} = \Sigma^{1/2}\,\mathcal{E}\,\Sigma^{1/2}$. First check (2.1) for $\boldsymbol{\xi} = \mathcal{E}$.

Lemma 4.5 *For any symmetric* $\Gamma \in \mathfrak{M}_p$ *with* $\|\Gamma\|_{\mathrm{Fr}} \leqslant \mathtt{g} < \sqrt{n}/2$, *it holds*

$$\begin{aligned} \mathbb{E}\langle \Gamma, \mathcal{E}\rangle^2 &= 2\|\Gamma\|_{\mathrm{Fr}}^2, \\ \log \mathbb{E}\exp\langle \Gamma, \mathcal{E}\rangle &\leqslant \frac{1}{1-2n^{-1/2}\|\Gamma\|}\,\|\Gamma\|_{\mathrm{Fr}}^2 \leqslant \frac{1}{1-2n^{-1/2}\mathtt{g}}\,\|\Gamma\|_{\mathrm{Fr}}^2. \end{aligned} \tag{4.7}$$

Proof Let us fix any symmetric $\Gamma \in \mathfrak{M}_p$ with $\|\Gamma\|_{\mathrm{Fr}} \leqslant \mathtt{g}$. For the scalar product $\langle \Gamma, \mathcal{E}\rangle$, we use the representation

$$\langle \Gamma, \mathcal{E}\rangle = \operatorname{tr}(\Gamma\mathcal{E}) = \frac{1}{n^{1/2}} \sum_{i=1}^n \bigl\{\boldsymbol{\gamma}_i^\top \Gamma \boldsymbol{\gamma}_i - \mathbb{E}(\boldsymbol{\gamma}_i^\top \Gamma \boldsymbol{\gamma}_i)\bigr\}.$$

Then by independence of the $\boldsymbol{\gamma}_i$'s and Lemma A.1, it holds

$$\mathbb{E}\langle \Gamma, \mathcal{E}\rangle^2 = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\bigl\{\boldsymbol{\gamma}_i^\top \Gamma \boldsymbol{\gamma}_i - \mathbb{E}(\boldsymbol{\gamma}_i^\top \Gamma \boldsymbol{\gamma}_i)\bigr\}^2 = 2\operatorname{tr}\Gamma^2.$$

Now consider the exponential moment of $\langle \Gamma, \mathcal{E}\rangle$. Again, independence of the $\boldsymbol{\gamma}_i$'s yields

$$\begin{aligned} \log \mathbb{E}\exp\langle \Gamma, \mathcal{E}\rangle &= \sum_{i=1}^n \log \mathbb{E}\exp\frac{\boldsymbol{\gamma}_i^\top \Gamma \boldsymbol{\gamma}_i}{\sqrt{n}} - \sqrt{n}\,\operatorname{tr}\Gamma \\ &= \frac{n}{2}\log\det\bigl(\mathbb{I}_p - \frac{2}{\sqrt{n}}\,\Gamma\bigr) - \sqrt{n}\,\operatorname{tr}\Gamma \end{aligned}$$

provided that $2\varGamma < \sqrt{n}\mathbb{I}_p$. Moreover, by Lemma A.2

$$\Bigl| \frac{n}{2} \log\det(\mathbb{I}_p - 2n^{-1/2}\varGamma) - \sqrt{n}\,\operatorname{tr}\varGamma \Bigr| \leqslant \frac{\operatorname{tr}\varGamma^2}{1 - 2n^{-1/2}\|\varGamma\|} = \frac{\|\varGamma\|_{\mathrm{Fr}}^2}{1 - 2n^{-1/2}\|\varGamma\|},$$

and the assertion follows in view of $\|\varGamma\| \leqslant \|\varGamma\|_{\mathrm{Fr}} \leqslant \mathtt{g}$. □

We now fix $\mathtt{g} = \rho\sqrt{n}/2$. Then the random matrix $\boldsymbol{\xi} = \mathcal{E}$ follows condition (2.1) with $\mathbb{V}^2 = 2(1-\rho)^{-1}\mathbb{I}$. This enables us to apply Theorem 2.1 to the quadratic form $\|Q\mathcal{E}\|_{\mathrm{Fr}}^2$ for $Q\mathcal{E} = \varSigma^{1/2}\,\mathcal{E}\,\varSigma^{1/2}$. By (4.7), it holds $\operatorname{Var}(\mathcal{E}) = 2\mathbb{I}$. Now introduce a Gaussian element $\widetilde{\mathcal{E}}$ with the same covariance structure. One can use $\widetilde{\mathcal{E}} = (\boldsymbol{\zeta} + \boldsymbol{\zeta}^\top)/\sqrt{2}$, where $\boldsymbol{\zeta} = (\zeta_{ij})$ is a random p-matrix with i.i.d. standard normal entries ζ_{ij}. Indeed, for any symmetric p-matrix $\varGamma$,

$$\mathbb{E}\langle \widetilde{\mathcal{E}}, \varGamma \rangle^2 = 2\mathbb{E}\langle \boldsymbol{\zeta}, \varGamma \rangle^2 = 2.$$

Statement (2.6) of Theorem 2.1 yields nearly the same deviation bounds for $\|Q\mathcal{E}\|_{\mathrm{Fr}}^2$ as for $\|Q\widetilde{\mathcal{E}}\|_{\mathrm{Fr}}^2$ with $\widetilde{\mathcal{E}} \sim \mathcal{N}(0, \operatorname{Var}(\mathcal{E}))$. Theorem B.1 claims

$$\mathbb{P}\bigl(\|Q\widetilde{\mathcal{E}}\|_{\mathrm{Fr}}^2 > z^2(\widetilde{B}, \mathtt{x})\bigr) \leqslant \mathrm{e}^{-\mathtt{x}},$$

where $\widetilde{B} = \operatorname{Var}(Q\widetilde{\mathcal{E}})$ and the quantile $z(B, \mathtt{x})$ is defined as

$$z^2(B, \mathtt{x}) = \operatorname{tr} B + 2\sqrt{\mathtt{x}\operatorname{tr}(B^2)} + 2\mathtt{x}\|B\|. \tag{4.8}$$

Lemma 4.6 *Let* $\widetilde{\mathcal{E}} = (\boldsymbol{\zeta} + \boldsymbol{\zeta}^\top)/\sqrt{2}$, *where* $\boldsymbol{\zeta} = (\zeta_{ij})$ *is a random p-matrix with i.i.d. standard normal entries* ζ_{ij}. *Consider* $Q\widetilde{\mathcal{E}} = \varSigma^{1/2}\,\widetilde{\mathcal{E}}\,\varSigma^{1/2}$. *It holds for* $\widetilde{B} = \operatorname{Var}(Q\widetilde{\mathcal{E}})$,

$$\operatorname{tr}\widetilde{B} = \mathbb{p}(\varSigma), \quad \operatorname{tr}\widetilde{B}^2 = \mathtt{v}^2(\varSigma), \quad \|\widetilde{B}\| = 2.$$

Proof We may assume $\varSigma = \operatorname{diag}\{\lambda_1, \ldots, \lambda_p\}$. Then it holds by Lemma A.1

$$\|Q\widetilde{\mathcal{E}}\|_{\mathrm{Fr}}^2 = \|\varSigma^{1/2}\,\widetilde{\mathcal{E}}\,\varSigma^{1/2}\|_{\mathrm{Fr}}^2 = \frac{1}{2}\sum_{i,j=1}^{p} \lambda_i\,\lambda_j\,(\zeta_{ij} + \zeta_{ji})^2 \overset{\mathrm{d}}{=} \ 2\sum_{i \leqslant j} \lambda_i\,\lambda_j\,\zeta_{ij}^2 \tag{4.9}$$

and thus

$$\operatorname{tr}\widetilde{B} = \mathbb{E}\|Q\widetilde{\mathcal{E}}\|_{\mathrm{Fr}}^2 = 2\sum_{i \leqslant j} \lambda_i\,\lambda_j = \Bigl(\sum_{i=1}^{p} \lambda_i\Bigr)^2 + \sum_{i=1}^{p} \lambda_i^2 = \mathbb{p}(\varSigma).$$

Further we compute $\mathtt{v}^2(\varSigma) = \operatorname{tr}\widetilde{B}^2$. Note that $\operatorname{Var}(\|Q\widetilde{\mathcal{E}}\|_{\mathrm{Fr}}^2) \neq \operatorname{Var}(\|Q\mathcal{E}\|_{\mathrm{Fr}}^2)$. Due to Lemma A.1, it holds $\mathtt{v}^2(\varSigma) = \operatorname{Var}(\|Q\widetilde{\mathcal{E}}\|_{\mathrm{Fr}}^2)/2$ yielding by (4.9)

$$\mathtt{v}^2(\varSigma) = 2\sum_{i \leqslant j} \lambda_i^2\lambda_j^2 \operatorname{Var}(\zeta_{ij}^2) = 2\sum_{i \neq j} \lambda_i^2\lambda_j^2 + 2\sum_{i=1}^{p} \lambda_i^4 = \bigl(\operatorname{tr}\varSigma^2\bigr)^2 + \operatorname{tr}\varSigma^4.$$

Finally, $\operatorname{Var}(\mathcal{E}) = 2\mathbb{I}$ and $\|\varSigma\| = 1$ implies $\lambda(\varSigma) = \|Q\operatorname{Var}(\mathcal{E})Q^\top\| = 2$. □

Now we apply Theorem 2.1 to $n\|\widehat{\varSigma} - \varSigma\|_{\mathrm{Fr}}^2 = \|Q\mathcal{E}\|_{\mathrm{Fr}}^2$. Following to Lemma 4.5, define $B = (1-\nu)^{-1}\widetilde{B}$. Then with $z^2(B, \mathtt{x})$ from (4.8)

$$\mathbb{P}\Bigl(n\|\widehat{\varSigma} - \varSigma\|_{\mathrm{Fr}}^2 > z^2(B, \mathtt{x})\Bigr) = \mathbb{P}\bigl(\|Q\mathcal{E}\|_{\mathrm{Fr}}^2 > z^2(B, \mathtt{x})\bigr) \leqslant 3\mathrm{e}^{-\mathtt{x}}, \quad \mathtt{x} \leqslant \mathtt{x}_c,$$

and assertion (4.3) follows in view of Lemma 4.6 and $z^2(B, \mathtt{x}) = (1-\nu)^{-1}z^2(\widetilde{B}, \mathtt{x})$. However, it is still necessary to check that the upper bound (4.3) applies for a given $\mathtt{x}$. (2.7) provides a sufficient condition $\mathtt{g}/\lambda \geqslant \sqrt{\mathtt{p}/\lambda} + \sqrt{8\mathtt{x}}$ with $\mathtt{p} = \mathtt{p}(\varSigma)/(1-\rho)$ and $\lambda = 2/(1-\rho)$ for $\mathtt{g} = \rho\sqrt{n}/2$. By (4.2)

$$\frac{\mathtt{g}}{\lambda} - \sqrt{\frac{\mathtt{p}}{\lambda}} = \frac{\rho\sqrt{n}}{2\lambda} - \sqrt{\frac{\mathtt{p}(\varSigma)}{2}} \geqslant \frac{\rho(1-\rho)\sqrt{n}}{4} - \sqrt{\frac{\mathtt{p}(\varSigma)}{2}} > \sqrt{8\mathtt{x}}$$

and the result follows.

4.6 Proof of Theorem 4.2

As in the proof of the upper bound, we apply Markov's inequality

$$\mathbb{P}\bigl(n\|\widehat{\varSigma} - \varSigma\|_{\mathrm{Fr}}^2 < z\bigr) \leqslant \mathrm{e}^{\mu z/2}\mathbb{E}\exp\Bigl(-\frac{\mu}{2}n\|\widehat{\varSigma} - \varSigma\|_{\mathrm{Fr}}^2\Bigr). \tag{4.10}$$

However, now we are free to choose any positive μ. Later we evaluate the exponential moments of $-n\|\widehat{\varSigma} - \varSigma\|_{\mathrm{Fr}}^2$ for all $\mu > 0$ and then, given $\mathtt{x}$, fix μ and z similarly to the Gaussian case to ensure the prescribed deviation probability $\mathrm{e}^{-\mathtt{x}}$.

Denote by $\boldsymbol{\zeta} = (\zeta_{ij})$ a random $p \times p$ matrix with i.i.d. standard Gaussian entries ζ_{ij} and $\bar{\boldsymbol{\zeta}} \stackrel{\mathrm{def}}{=} (\boldsymbol{\zeta} + \boldsymbol{\zeta}^{\top})/2$. Then for any $\mu > 0$,

$$\exp\bigl(-\mu n\|\widehat{\varSigma} - \varSigma\|_{\mathrm{Fr}}^2/2\bigr) = \mathbb{E}_{\zeta}\exp\bigl\{\mathtt{i}\sqrt{\mu n}\,\langle\widehat{\varSigma} - \varSigma, \boldsymbol{\zeta}\rangle\bigr\} = \mathbb{E}_{\zeta}\exp\bigl\{\mathtt{i}\sqrt{\mu n}\,\langle\widehat{\varSigma} - \varSigma, \bar{\boldsymbol{\zeta}}\rangle\bigr\}.$$

Therefore, by independence of the $\boldsymbol{X}_i$'s,

$$\begin{aligned}
\mathbb{E}\exp\bigl(-\mu n\|\widehat{\varSigma} - \varSigma\|_{\mathrm{Fr}}^2/2\bigr) &= \mathbb{E}_{\zeta}\mathbb{E}\exp\bigl(\mathtt{i}\sqrt{\mu n}\,\langle\widehat{\varSigma} - \varSigma, \bar{\boldsymbol{\zeta}}\rangle\bigr) \\
&= \mathbb{E}_{\zeta}\Bigl\{\mathbb{E}\exp\bigl(\mathtt{i}\sqrt{\mu/n}\,\langle\boldsymbol{X}_1\boldsymbol{X}_1^{\top} - \varSigma, \bar{\boldsymbol{\zeta}}\rangle\bigr)\Bigr\}^n \\
&= \mathbb{E}_{\zeta}\Bigl\{\mathbb{E}\exp\bigl(\mathtt{i}\sqrt{\mu/n}\,\langle\boldsymbol{\gamma}\boldsymbol{\gamma}^{\top} - \mathbb{I}_p, \varSigma^{1/2}\,\bar{\boldsymbol{\zeta}}\,\varSigma^{1/2}\rangle\bigr)\Bigr\}^n.
\end{aligned}$$

Further, by Lemma A.2, with $\mathcal{B} = \varSigma^{1/2}\,\bar{\boldsymbol{\zeta}}\,\varSigma^{1/2}$,

$$\begin{aligned}
&\Bigl\{\mathbb{E}\exp\bigl(\mathtt{i}\sqrt{\mu/n}\,\langle\boldsymbol{\gamma}\boldsymbol{\gamma}^{\top} - \mathbb{I}_p, \mathcal{B}\rangle\bigr)\Bigr\}^n \\
&= \exp\bigl\{n\log\det\bigl(\mathbb{I}_p - 2\mathtt{i}\sqrt{\mu/n}\,\mathcal{B}\bigr)^{-1/2} - \mathtt{i}\sqrt{\mu n}\,\mathrm{tr}(\mathcal{B})\bigr\}.
\end{aligned} \tag{4.11}$$

Let some $\mathtt{x} > 0$ and some $\alpha \in (0, 1/2)$ be fixed. Define

$$\mu \stackrel{\mathrm{def}}{=} \frac{2\sqrt{\mathtt{x}}}{\mathtt{v}(\varSigma)}, \quad \mu_{\alpha} \stackrel{\mathrm{def}}{=} \frac{1-\alpha}{1-2\alpha}\,\mu = \frac{1-\alpha}{1-2\alpha}\,\frac{2\sqrt{\mathtt{x}}}{\mathtt{v}(\varSigma)}, \tag{4.12}$$

and introduce a random set $\varOmega(\alpha)$ with

$$\varOmega(\alpha) \stackrel{\mathrm{def}}{=} \bigl\{\boldsymbol{\zeta}\colon 2\sqrt{\mu_{\alpha}/n}\,\|\mathcal{B}\| \leqslant \alpha\bigr\}, \quad \mathcal{B} = \varSigma^{1/2}\,(\boldsymbol{\zeta} + \boldsymbol{\zeta}^{\top})\,\varSigma^{1/2}/2. \tag{4.13}$$

It holds on $\varOmega(\alpha)$ by (4.11) similarly to (A.4) of Lemma A.2

$$\begin{aligned}
\mathbb{E}^n\exp\bigl\{\mathtt{i}\sqrt{\mu_{\alpha}/n}\,\langle\boldsymbol{\gamma}\boldsymbol{\gamma}^{\top} - \mathbb{I}_p, \mathcal{B}\rangle\bigr\} &\leqslant \exp\Bigl(-\mu_{\alpha}\,\mathrm{tr}(\mathcal{B}^2) + \frac{\mu_{\alpha}\,\alpha\,\mathrm{tr}(\mathcal{B}^2)}{1-\alpha}\Bigr) \\
&= \exp\Bigl(-\frac{1-2\alpha}{1-\alpha}\,\mu_{\alpha}\,\mathrm{tr}(\mathcal{B}^2)\Bigr) = \exp\bigl(-\mu\,\mathrm{tr}(\mathcal{B}^2)\bigr).
\end{aligned} \tag{4.14}$$

Exponential moments of $\mathrm{tr}(\mathcal{B}^2)$ from (4.15) under $\mathbb{P}_{\zeta}$ can be easily computed. We proceed assuming $\varSigma = \mathrm{diag}\{\lambda_j\}$ and using that $\zeta_{ij} + \zeta_{ji} \sim \mathcal{N}(0, 2)$ for $i \neq j$, and all $\zeta_{ij} + \zeta_{ji}$ are mutually independent for $i \leqslant j$. This implies

$$\operatorname{tr}(\mathcal{B}^2) = \frac{1}{4} \sum_{i,j=1}^{p} \lambda_i \, \lambda_j \, (\zeta_{ij} + \zeta_{ji})^2 \overset{\mathrm{d}}{=} \sum_{i \leqslant j} \lambda_i \, \lambda_j \, \zeta_{ij}^2 \tag{4.15}$$

and

$$\begin{gathered} \mathbb{E}_{\zeta} \operatorname{tr}(\mathcal{B}^2) = \sum_{i \leqslant j} \lambda_i \, \lambda_j = \frac{\mathbb{p}(\varSigma)}{2}, \\ \mathbb{E}_{\zeta} \exp\{-\mu \operatorname{tr}(\mathcal{B}^2)\} = \mathbb{E}_{\zeta} \exp\Bigl(-\mu \sum_{i \leqslant j} \lambda_i \, \lambda_j \, \zeta_{ij}^2\Bigr) = \exp\Bigl(-\frac{1}{2} \sum_{i \leqslant j} \log(1 + 2\mu \, \lambda_i \, \lambda_j)\Bigr). \end{gathered} \tag{4.16}$$

The latter expression can be evaluated by using (A.3) of Lemma A.2:

$$\mathbb{E}_{\zeta} \exp\{-\mu \operatorname{tr}(\mathcal{B}^2)\} \leqslant \exp\Bigl(-\mu \sum_{i \leqslant j} \lambda_i \, \lambda_j + \mu^2 \sum_{i \leqslant j} \lambda_i^2 \, \lambda_j^2\Bigr) = \exp\Bigl(-\frac{\mu \, \mathbb{p}(\varSigma)}{2} + \frac{\mu^2 \mathtt{v}^2(\varSigma)}{4}\Bigr).$$

This and (4.14) yield

$$\mathbb{E} \exp\Bigl(-\frac{\mu_{\alpha}}{2} n \|\widehat{\varSigma} - \varSigma\|_{\mathrm{Fr}}^2\Bigr) \leqslant \mathbb{P}_{\zeta}\bigl(\varOmega(\alpha)^c\bigr) + \exp\Bigl(-\frac{\mu \, \mathbb{p}(\varSigma)}{2} + \frac{\mu^2 \mathtt{v}^2(\varSigma)}{4}\Bigr)$$

and for any z by Markov's inequality (4.10)

$$\mathbb{P}\bigl(n \|\widehat{\varSigma} - \varSigma\|_{\mathrm{Fr}}^2 < z\bigr) \leqslant \mathrm{e}^{\mu_{\alpha} z/2} \mathbb{P}_{\zeta}\bigl(\varOmega(\alpha)^c\bigr) + \exp\Bigl(\frac{\mu_{\alpha} z}{2} - \frac{\mu \, \mathbb{p}(\varSigma)}{2} + \frac{\mu^2 \mathtt{v}^2(\varSigma)}{4}\Bigr).$$

With $\mu = 2\sqrt{\mathtt{x}}/\mathtt{v}(\varSigma)$, we define z by

$$\mu_{\alpha} \, z = \mu\{\mathbb{p}(\varSigma) - 2\mathtt{v}(\varSigma)\sqrt{\mathtt{x}}\} = \frac{2\sqrt{\mathtt{x}} \, \mathbb{p}(\varSigma)}{\mathtt{v}(\varSigma)} - 4\mathtt{x} \tag{4.17}$$

yielding

$$\frac{\mu_{\alpha} z}{2} - \frac{\mu \, \mathbb{p}(\varSigma)}{2} + \frac{\mu^2 \mathtt{v}^2(\varSigma)}{4} = \frac{\mu}{2}\bigl\{\mathbb{p}(\varSigma) - 2\mathtt{v}(\varSigma)\sqrt{\mathtt{x}}\bigr\} - \frac{\mu \, \mathbb{p}(\varSigma)}{2} + \frac{\mu^2 \mathtt{v}^2(\varSigma)}{4} = -\mathtt{x}$$

and

$$\mathbb{P}(n \|\widehat{\varSigma} - \varSigma\|_{\mathrm{Fr}}^2 < z) \leqslant \mathrm{e}^{-\mathtt{x}} + \mathrm{e}^{\mu_{\alpha} z/2} \mathbb{P}_{\zeta}\bigl(\varOmega(\alpha)^c\bigr)$$

where

$$z = \Bigl(1 - \frac{\alpha}{1 - \alpha}\Bigr)\bigl\{\mathbb{p}(\varSigma) - 2\mathtt{v}(\varSigma)\sqrt{\mathtt{x}}\bigr\} \geqslant \mathbb{p}(\varSigma) - \frac{\alpha}{1 - \alpha} \, \mathbb{p}(\varSigma) - 2\mathtt{v}(\varSigma)\sqrt{\mathtt{x}}.$$

For bounding the probability of the set $\varOmega(\alpha)^c$ from (4.13), one can apply the advanced results from the random matrix theory. To keep the proof self-contained, we use a simple bound $\|\mathcal{B}\|^2 \leqslant \|\mathcal{B}\|_{\mathrm{Fr}}^2 = \operatorname{tr}(\mathcal{B}^2)$. For any matrix $\varGamma$, it holds

$$\operatorname{Var}\langle \bar{\boldsymbol{\zeta}}, \varGamma \rangle = \frac{1}{4} \mathbb{E}\Bigl(\sum_{i,j=1}^{p} \varGamma_{ij} (\zeta_{ij} + \zeta_{ji})\Bigr)^2 = \|\varGamma\|_{\mathrm{Fr}}^2$$

yielding $\|\operatorname{Var}(\bar{\boldsymbol{\zeta}})\| \leqslant 1$ and $\|\operatorname{Var}(\mathcal{B})\| \leqslant 1$. Also by (4.16) $\mathbb{E}\|\mathcal{B}\|_{\mathrm{Fr}}^2 = \mathbb{p}(\varSigma)/2$. Therefore, by Theorem B.1 applied to $\|\mathcal{B}\|_{\mathrm{Fr}}^2$, it holds for any $\mathtt{x}_{\circ}$,

$$\mathbb{P}_{\zeta}\bigl(\|\mathcal{B}\|_{\mathrm{Fr}} > \sqrt{\mathbb{p}(\varSigma)/2} + \sqrt{2\mathtt{x}_{\circ}}\bigr) \leqslant \mathrm{e}^{-\mathtt{x}_{\circ}}.$$

By (4.12) and (4.17), it holds

$$\mathtt{x}_{\circ} \stackrel{\mathrm{def}}{=} \mathtt{x} + \frac{\mu_{\alpha} z}{2} \leqslant \frac{\mathbb{p}(\varSigma)\sqrt{\mathtt{x}}}{\mathtt{v}(\varSigma)} - \mathtt{x} \leqslant \frac{\mathbb{p}^{2}(\varSigma)}{4\mathtt{v}^{2}(\varSigma)}$$

and

$$\mathbb{P}_{\zeta}\biggl(\|\mathcal{B}\|_{\mathrm{Fr}} > \sqrt{\frac{\mathbb{p}(\varSigma)}{2}} + \frac{\mathbb{p}(\varSigma)}{\sqrt{2}\,\mathtt{v}(\varSigma)} \biggr) \leqslant \mathrm{e}^{-\mathtt{x}-\mu_{\alpha} z/2}.$$

Therefore, by definition (4.13) and condition (4.4)

$$\mathrm{e}^{\mu_{\alpha} z/2}\mathbb{P}_{\zeta}\bigl(\varOmega(\alpha)^{c}\bigr) \leqslant \mathrm{e}^{\mu_{\alpha} z/2}\mathbb{P}\biggl(\|\mathcal{B}\|_{\mathrm{Fr}} > \frac{\alpha\sqrt{n}}{2\sqrt{\mu_{\alpha}}} \biggr) \leqslant \mathrm{e}^{-\mathtt{x}}$$

and the result follows.

Appendix

A. Moments of a Gaussian quadratic form

Let $\boldsymbol{\gamma}$ be standard normal in $\mathbb{R}^{p}$ for $p \leqslant \infty$. Given a self-adjoint trace operator B, consider a quadratic form $\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \rangle$.

Lemma A.1 *It holds* $\mathbb{E}\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \rangle = \operatorname{tr} B$. *Moreover,*

$$\begin{aligned}
\mathbb{E}\bigl(\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \rangle - \operatorname{tr} B\bigr)^{2} &= 2 \operatorname{tr} B^{2}, \\
\mathbb{E}\bigl(\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \rangle - \operatorname{tr} B\bigr)^{3} &= 8 \operatorname{tr} B^{3}, \\
\mathbb{E}\bigl(\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \rangle - \operatorname{tr} B\bigr)^{4} &= 48 \operatorname{tr} B^{4} + 12 (\operatorname{tr} B^{2})^{2}, \\
\mathbb{E}\bigl(\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \rangle - \operatorname{tr} B\bigr)^{5} &= 512 \operatorname{tr} B^{5} + 32 \operatorname{tr} B^{2} \, \operatorname{tr} B^{3},
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \rangle^{2} &= (\operatorname{tr} B)^{2} + 2 \operatorname{tr} B^{2}, \\
\mathbb{E}\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \rangle^{3} &= (\operatorname{tr} B)^{3} + 6 \operatorname{tr} B \, \operatorname{tr} B^{2} + 8 \operatorname{tr} B^{3}, \\
\mathbb{E}\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \rangle^{4} &= (\operatorname{tr} B)^{4} + 12 (\operatorname{tr} B)^{2} \operatorname{tr} B^{2} + 32 (\operatorname{tr} B) \operatorname{tr} B^{3} + 48 \operatorname{tr} B^{4} + 12 (\operatorname{tr} B^{2})^{2}, \\
\operatorname{Var}\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \rangle^{2} &= 8 (\operatorname{tr} B)^{2} \operatorname{tr} B^{2} + 32 (\operatorname{tr} B) \operatorname{tr} B^{3} + 48 \operatorname{tr} B^{4} + 8 (\operatorname{tr} B^{2})^{2}.
\end{aligned}$$

Moreover, if $B \leqslant \mathbb{I}_{p}$ *and* $\mathbb{p} = \operatorname{tr} B$, *then* $\operatorname{tr} B^{m} \leqslant \mathbb{p} \|B\|^{m-1}$ *for* $m \geqslant 1$ *and*

$$\begin{aligned}
\mathbb{E}\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \rangle^{2} &\leqslant \mathbb{p}^{2} + 2\mathbb{p}\|B\| \leqslant (\mathbb{p} + \|B\|)^{2}, \\
\mathbb{E}\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \rangle^{3} &\leqslant \mathbb{p}^{3} + 6\mathbb{p}^{2}\|B\| + 8\mathbb{p}\|B\|^{2} \leqslant (\mathbb{p} + 2\|B\|)^{3}, \\
\mathbb{E}\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \rangle^{4} &\leqslant \mathbb{p}^{4} + 12\mathbb{p}^{3}\|B\| + 44\mathbb{p}^{2}\|B\|^{2} + 48\mathbb{p}\|B\|^{3} \leqslant (\mathbb{p} + 3\|B\|)^{4}, \\
\mathbb{E}\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \rangle^{5} &\leqslant \mathbb{p}^{5} + 20\mathbb{p}^{4}\|B\| + 140\mathbb{p}^{3}\|B\|^{2} + 272\mathbb{p}^{2}\|B\|^{3} + 512\mathbb{p}\|B\|^{4} \leqslant (\mathbb{p} + 4\|B\|)^{5}, \\
\operatorname{Var}\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \rangle^{2} &\leqslant 8\mathbb{p}^{3} + 40\mathbb{p}^{2}\|B\| + 48\mathbb{p}\|B\|^{2}.
\end{aligned}$$

Finally,

$$\mathbb{E}(\boldsymbol{\gamma}\boldsymbol{\gamma}^{\top} - \mathbb{I}_{p}) B (\boldsymbol{\gamma}\boldsymbol{\gamma}^{\top} - \mathbb{I}_{p}) = B + \operatorname{tr}(B)\mathbb{I}_{p}$$

yielding

$$\mathbb{E}\|B(\boldsymbol{\gamma}\boldsymbol{\gamma}^{\top} - \mathbb{I}_{p})\|_{\mathrm{Fr}}^{2} = (\operatorname{tr} B)^{2} + \operatorname{tr} B^{2}. \tag{A.1}$$

Proof Let $\boldsymbol{\gamma}$ be standard normal in $\mathbb{R}^{p}$. The same holds for $\boldsymbol{U}\boldsymbol{\gamma}$ for any orthogonal transform

$\boldsymbol{U}$ in $\mathbb{R}^p$. The use of the spectral decomposition $B = \boldsymbol{U}^\top \mathit{\Lambda} \boldsymbol{U}$ with $\boldsymbol{U}$ orthonormal and $\mathit{\Lambda}$ diagonal enables us to represent $\bigl\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \bigr\rangle = \bigl\langle \mathit{\Lambda}\boldsymbol{U}\boldsymbol{\gamma}, \boldsymbol{U}\boldsymbol{\gamma} \bigr\rangle$ and thus, to reduce the statements to the case when B is diagonal: $B = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$. Then

$$\xi \overset{\text{def}}{=} \bigl\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \bigr\rangle - \operatorname{tr} B = \sum_{j=1}^{p} \lambda_j (\gamma_j^2 - 1),$$

where γ_j are i.i.d. standard normal. This easily yields with $\mathbb{p}_m = \operatorname{tr}(B^m)$,

$$\begin{aligned}
\mathbb{E}\xi^2 &= \sum_{j=1}^{p} \lambda_j^2 \mathbb{E}(\gamma_j^2 - 1)^2 = \mathbb{E}\chi^2 \ \operatorname{tr} B^2 = 2\mathbb{p}_2, \\
\mathbb{E}\xi^3 &= \sum_{j=1}^{p} \lambda_j^3 \mathbb{E}(\gamma_j^2 - 1)^3 = \mathbb{E}\chi^3 \ \operatorname{tr} B^3 = 8\mathbb{p}_3, \\
\mathbb{E}\xi^4 &= \sum_{j=1}^{p} \lambda_j^4 (\gamma_j^2 - 1)^4 + \sum_{i \neq j} \lambda_i^2 \lambda_j^2 \mathbb{E}(\gamma_i^2 - 1)^2 \mathbb{E}(\gamma_j^2 - 1)^2 \\
&= \bigl(\mathbb{E}\chi^4 - 3(\mathbb{E}\chi^2)^2\bigr) \operatorname{tr} B^4 + 3(\mathbb{E}\chi^2 \ \operatorname{tr} B^2)^2 = 48\mathbb{p}_4 + 12\mathbb{p}_2^2, \\
\mathbb{E}\xi^5 &= \sum_{j=1}^{p} \lambda_j^5 (\gamma_j^2 - 1)^5 + \sum_{i \neq j} \lambda_i^2 \lambda_j^3 \mathbb{E}(\gamma_i^2 - 1)^2 \mathbb{E}(\gamma_j^2 - 1)^3 \\
&= \bigl\{\mathbb{E}(\gamma^2 - 1)^5 - \mathbb{E}(\gamma^2 - 1)^2 \mathbb{E}(\gamma^2 - 1)^3\bigr\} \ \operatorname{tr} B^5 + \mathbb{E}(\gamma^2 - 1)^2 \mathbb{E}(\gamma^2 - 1)^3 \ \operatorname{tr} B^2 \ \operatorname{tr} B^3 \\
&= 512\mathbb{p}_5 + 32\mathbb{p}_2\, \mathbb{p}_3,
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\bigl\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \bigr\rangle^2 &= \bigl(\mathbb{E}\bigl\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \bigr\rangle\bigr)^2 + \mathbb{E}\xi^2 = \mathbb{p}^2 + 2\mathbb{p}_2, \\
\mathbb{E}\bigl\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \bigr\rangle^3 &= \mathbb{E}(\xi + \mathbb{p})^3 = \mathbb{p}^3 + \mathbb{E}\xi^3 + 3\mathbb{p}\ \mathbb{E}\xi^2 = \mathbb{p}^3 + 6\mathbb{p}\ \mathbb{p}_2 + 8\mathbb{p}_3, \\
\mathbb{E}\bigl\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \bigr\rangle^4 &= \mathbb{E}\bigl(\xi + \mathbb{p}\bigr)^4 = \mathbb{p}^4 + 6\mathbb{p}^2 \mathbb{E}\xi^2 + 4\mathbb{p}\, \mathbb{E}\xi^3 + \mathbb{E}\xi^4 \\
&= \mathbb{p}^4 + 12\mathbb{p}^2\, \mathbb{p}_2 + 32\mathbb{p}\, \mathbb{p}_3 + 48\mathbb{p}_4 + 12\mathbb{p}_2^2,
\end{aligned}$$

and

$$\begin{aligned}
\operatorname{Var}\bigl\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \bigr\rangle^2 &= \mathbb{E}(\xi + \mathbb{p})^4 - \bigl(\mathbb{p}^2 + 2\mathbb{p}_2\bigr)^2 \\
&= \mathbb{p}^4 + 6\mathbb{p}^2 \mathbb{E}\xi^2 + 4\mathbb{p}\, \mathbb{E}\xi^3 + \mathbb{E}\xi^4 - \bigl(\mathbb{p}^2 + 2\mathbb{p}_2\bigr)^2 = 8\mathbb{p}^2\, \mathbb{p}_2 + 32\mathbb{p}\, \mathbb{p}_3 + 48\mathbb{p}_4 + 8\mathbb{p}_2^2.
\end{aligned}$$

Also

$$\begin{aligned}
\mathbb{E}\bigl\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \bigr\rangle^5 &= \mathbb{E}\bigl(\xi + \mathbb{p}\bigr)^5 = \mathbb{p}^5 + 10\mathbb{p}^3 \mathbb{E}\xi^2 + 10\mathbb{p}^2\, \mathbb{E}\xi^3 + 5\mathbb{p}\mathbb{E}\xi^4 + \mathbb{E}\xi^5 \\
&= \mathbb{p}^5 + 20\mathbb{p}^3\, \mathbb{p}_2 + 80\mathbb{p}^2 \mathbb{p}_3 + 5\mathbb{p}(48\mathbb{p}_4 + 12\mathbb{p}_2^2) + 512\mathbb{p}_5 + 32\mathbb{p}_2\, \mathbb{p}_3.
\end{aligned}$$

Assume $\|B\| = 1$ yielding $\mathbb{p}_m \leqslant \mathbb{p}$. Then

$$\begin{aligned}
\mathbb{E}\bigl\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \bigr\rangle^2 &\leqslant \mathbb{p}^2 + 2\mathbb{p} \leqslant (\mathbb{p} + 1)^2, \\
\mathbb{E}\bigl\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \bigr\rangle^3 &\leqslant \mathbb{p}^3 + 6\mathbb{p}^2 + 8\mathbb{p} \leqslant (\mathbb{p} + 2)^3, \\
\mathbb{E}\bigl\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \bigr\rangle^4 &\leqslant \mathbb{p}^4 + 12\mathbb{p}^3 + 44\mathbb{p}^2 + 48\mathbb{p} \leqslant (\mathbb{p} + 3)^4, \\
\mathbb{E}\bigl\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \bigr\rangle^5 &\leqslant \mathbb{p}^5 + 20\mathbb{p}^4 + 140\mathbb{p}^3 + 272\mathbb{p}^2 + 512\mathbb{p} \leqslant (\mathbb{p} + 4)^5.
\end{aligned}$$

For the last result of the lemma, observe that with $B = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$,

$$\mathbb{E}\|B^{1/2}(\boldsymbol{\gamma}\boldsymbol{\gamma}^{\top} - \mathbb{I}_p)B^{1/2}\|_{\mathrm{Fr}}^2 = \sum_{i,j=1}^{p} \lambda_i \lambda_j \mathbb{E}(\gamma_i \gamma_j - \delta_{i,j})^2 = \biggl(\sum_{i=1}^{p} \lambda_i\biggr)^2 + \sum_{i=1}^{p} \lambda_i^2$$

and assertion (A.1) follows. ☐

Now we compute the exponential moments of centered and non-centered quadratic forms.

Lemma A.2 *Let $\|B\| = \lambda$ and $\boldsymbol{\gamma} \sim \mathcal{N}(0, \mathbb{I}_p)$. Then for any $\mu \in (0, \lambda^{-1})$,*

$$\mathbb{E}\exp\Bigl\{\frac{\mu}{2}\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma}\rangle\Bigr\} = \det(\mathbb{I}_p - \mu B)^{-1/2}.$$

Moreover, with $\mathbb{p} = \operatorname{tr} B$ and $\mathrm{v}^2 = \operatorname{tr} B^2$,

$$\log\mathbb{E}\exp\Bigl\{\frac{\mu}{2}\bigl(\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma}\rangle - \mathbb{p}\bigr)\Bigr\} \leqslant \frac{\mu^2 \mathrm{v}^2}{4(1 - \lambda\mu)}. \tag{A.2}$$

If B is positive semidefinite, $\lambda_j \geqslant 0$, then

$$\log\mathbb{E}\exp\Bigl\{-\frac{\mu}{2}\bigl(\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma}\rangle - \mathbb{p}\bigr)\Bigr\} \leqslant \frac{\mu^2 \mathrm{v}^2}{4}. \tag{A.3}$$

For any complex valued μ with $\lambda|\mu| < 1$,

$$\biggl|\log\mathbb{E}\exp\Bigl\{\frac{\mu}{2}\bigl(\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma}\rangle - \mathbb{p}\bigr) - \frac{\mu^2 \operatorname{tr} B^2}{4}\Bigr\}\biggr| \leqslant \frac{\lambda|\mu|^3 \mathrm{v}^2}{6(1 - \lambda|\mu|)}. \tag{A.4}$$

Proof W.l.o.g. assume $\lambda = 1$. Let λ_j be the eigenvalues of B, $|\lambda_j| \leqslant 1$. As in Lemma A.1, one can reduce the statement to the case of a diagonal matrix $B = \operatorname{diag}\bigl(\lambda_j\bigr)$. Then $\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma}\rangle = \sum_{j=1}^{p} \lambda_j \gamma_j^2$ and by independence of the γ_j 's,

$$\mathbb{E}\Bigl\{\frac{\mu}{2}\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma}\rangle\Bigr\} = \prod_{j=1}^{p} \mathbb{E}\exp\Bigl(\frac{\mu}{2}\lambda_j \epsilon_j^2\Bigr) = \prod_{j=1}^{p} \frac{1}{\sqrt{1 - \mu\lambda_j}} = \det\bigl(\mathbb{I}_p - \mu B\bigr)^{-1/2}.$$

Below we use the simple bounds:

$$-\log(1-u) - u = \sum_{k=2}^{\infty} \frac{u^k}{k} \leqslant \frac{u^2}{2}\sum_{k=0}^{\infty} u^k = \frac{u^2}{2(1-u)}, \quad u \in (0,1),$$

$$-\log(1-u) + u = \sum_{k=2}^{\infty} \frac{u^k}{k} \leqslant \frac{u^2}{2}, \quad u \in (-1,0).$$

Now it holds for $\mu > 0$,

$$\begin{aligned} \log\mathbb{E}\Bigl\{\frac{\mu}{2}\bigl(\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma}\rangle - \mathbb{p}\bigr)\Bigr\} &= \log\det(\mathbb{I}_p - \mu B)^{-1/2} - \frac{\mu\,\mathbb{p}}{2} \\ &= -\frac{1}{2}\sum_{j=1}^{p}\bigl\{\log(1 - \mu\lambda_j) + \mu\lambda_j\bigr\} \leqslant \sum_{j=1}^{p} \frac{\mu^2\lambda_j^2}{4(1 - \mu\lambda_j)} \leqslant \frac{\mu^2 \mathrm{v}^2}{4(1 - \mu\lambda)}. \end{aligned}$$

Similarly for any complex μ with $|\mu|\lambda < 1$,

$$\begin{aligned} &\biggl|\log\mathbb{E}\Bigl\{\frac{\mu}{2}\bigl(\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma}\rangle - \mathbb{p}\bigr) - \frac{\mu^2 \operatorname{tr} B^2}{4}\Bigr\}\biggr| = \biggl|\log\det(\mathbb{I}_p - \mu B)^{-1/2} - \frac{\mu\,\mathbb{p}}{2} - \frac{\mu^2 \operatorname{tr} B^2}{4}\biggr| \\ &= \frac{1}{2}\biggl|\sum_{j=1}^{p}\Bigl\{\log(1 - \mu\lambda_j) - \mu\lambda_j - \frac{\mu^2\lambda_j^2}{2}\Bigr\}\biggr| \leqslant \sum_{j=1}^{p} \frac{|\mu\lambda_j|^3}{6(1 - \lambda|\mu|)} = \frac{|\mu|^3 \lambda \mathrm{v}^2}{6(1 - \lambda|\mu|)}. \end{aligned}$$

Statement (A.3) can be proved similarly. ☐

Now we consider the case of a non-centered quadratic form $\langle B\boldsymbol{\gamma},\boldsymbol{\gamma}\rangle/2+\langle\boldsymbol{A},\boldsymbol{\gamma}\rangle$ for a fixed vector $\boldsymbol{A}$.

Lemma A.3 *Let $\|B\|=\lambda<1$. Then for any $\boldsymbol{A}$,*

$$\mathbb{E}\exp\Bigl\{\frac{1}{2}\langle B\boldsymbol{\gamma},\boldsymbol{\gamma}\rangle+\langle\boldsymbol{A},\boldsymbol{\gamma}\rangle\Bigr\}=\exp\Bigl\{\frac{\|(\mathbb{I}_p-B)^{-1/2}\boldsymbol{A}\|^2}{2}\Bigr\}\det(\mathbb{I}_p-B)^{-1/2}.$$

Moreover, for any $\mu\in(0,1)$,

$$\begin{aligned}&\log\mathbb{E}\exp\Bigl\{\frac{\mu}{2}\bigl(\langle B\boldsymbol{\gamma},\boldsymbol{\gamma}\rangle-\mathtt{p}\bigr)+\langle\boldsymbol{A},\boldsymbol{\gamma}\rangle\Bigr\}\\&=\frac{\|(\mathbb{I}_p-\mu B)^{-1/2}\boldsymbol{A}\|^2}{2}+\log\det(\mathbb{I}_p-\mu B)^{-1/2}-\mu\,\mathtt{p}\\&\leqslant\frac{\|(\mathbb{I}_p-\mu B)^{-1/2}\boldsymbol{A}\|^2}{2}+\frac{\mu^2\mathtt{v}^2}{4(1-\lambda\mu)}.\end{aligned}\tag{A.5}$$

Proof Denote $\boldsymbol{a}=(\mathbb{I}_p-B)^{-1/2}\boldsymbol{A}$. It holds by change of variables $(\mathbb{I}_p-B)^{1/2}\boldsymbol{x}=\boldsymbol{u}$ for $\mathtt{C}_p=(2\pi)^{-p/2}$,

$$\begin{aligned}\mathbb{E}\exp\Bigl\{\frac{1}{2}\langle B\boldsymbol{\gamma},\boldsymbol{\gamma}\rangle+\langle\boldsymbol{A},\boldsymbol{\gamma}\rangle\Bigr\}&=\mathtt{C}_p\int\exp\Bigl\{-\frac{1}{2}\langle(\mathbb{I}_p-B)\boldsymbol{x},\boldsymbol{x}\rangle+\langle\boldsymbol{A},\boldsymbol{x}\rangle\Bigr\}d\boldsymbol{x}\\=\mathtt{C}_p\det(\mathbb{I}_p-B)^{-1/2}&\int\exp\Bigl\{-\frac{1}{2}\|\boldsymbol{u}\|^2+\langle\boldsymbol{a},\boldsymbol{u}\rangle\Bigr\}d\boldsymbol{u}=\det(\mathbb{I}_p-B)^{-1/2}\,\mathrm{e}^{\|\boldsymbol{a}\|^2/2}.\end{aligned}$$

The last inequality (A.5) follows by (A.2). □

B. Deviation bounds for Gaussian quadratic forms

The next result explains the concentration effect of $\|Q\boldsymbol{\xi}\|^2$ for a centered Gaussian vector $\boldsymbol{\xi}\sim\mathcal{N}(0,\mathbb{V}^2)$ and a linear operator $Q\colon\mathbb{R}^p\to\mathbb{R}^q$, $p,q\leqslant\infty$. We use a version from [8]. For completeness, we present a simple proof.

Theorem B.1 *Let $\boldsymbol{\xi}\sim\mathcal{N}(0,\mathbb{V}^2)$ be a Gaussian element in $\mathbb{R}^p$ and let $Q\colon\mathbb{R}^p\to\mathbb{R}^q$ be such that $B=Q\mathbb{V}^2Q^\top$ is a trace operator in $\mathbb{R}^q$. Then with $\mathtt{p}=\operatorname{tr}(B)$, $\mathtt{v}^2=\operatorname{tr}(B^2)$, and $\lambda=\|B\|$, it holds for any $\mathtt{x}\geqslant0$,*

$$\mathbb{P}\Bigl(\|Q\boldsymbol{\xi}\|^2-\mathtt{p}>2\mathtt{v}\sqrt{\mathtt{x}}+2\lambda\mathtt{x}\Bigr)\leqslant\mathrm{e}^{-\mathtt{x}},\tag{B.1}$$

$$\mathbb{P}\Bigl(\|Q\boldsymbol{\xi}\|^2-\mathtt{p}\leqslant-2\mathtt{v}\sqrt{\mathtt{x}}\Bigr)\leqslant\mathrm{e}^{-\mathtt{x}}.\tag{B.2}$$

It also implies

$$\mathbb{P}\bigl(\bigl|\|Q\boldsymbol{\xi}\|^2-\mathtt{p}\bigr|>z_2(B,\mathtt{x})\bigr)\leqslant2\mathrm{e}^{-\mathtt{x}},$$

with

$$z_2(B,\mathtt{x})\stackrel{\text{def}}{=}2\mathtt{v}\sqrt{\mathtt{x}}+2\lambda\mathtt{x}\,.$$

Proof We use the identity in distribution $\|Q\boldsymbol{\xi}\|^2\stackrel{\mathrm{d}}{=}\langle B\boldsymbol{\gamma},\boldsymbol{\gamma}\rangle$ with $\boldsymbol{\gamma}\sim\mathcal{N}(0,\mathbb{I}_q)$. Markov's inequality yields for any $\mu>0$,

$$\mathbb{P}\Bigl(\langle B\boldsymbol{\gamma},\boldsymbol{\gamma}\rangle-\mathtt{p}>z_2(B,\mathtt{x})\Bigr)\leqslant\mathbb{E}\exp\Bigl(\frac{\mu}{2}\bigl(\langle B\boldsymbol{\gamma},\boldsymbol{\gamma}\rangle-\mathtt{p}\bigr)-\frac{\mu\,z_2(B,\mathtt{x})}{2}\Bigr).$$

Given $\mathtt{x}>0$, fix $\mu<1/\lambda$ by the equation

$$\frac{\mu}{1-\lambda\mu} = \frac{2\sqrt{\mathtt{x}}}{\mathtt{v}} \quad \text{or} \quad \mu^{-1} = \lambda + \frac{\mathtt{v}}{2\sqrt{\mathtt{x}}}. \tag{B.3}$$

By (A.2)

$$\log \mathbb{E}\Bigl\{ \frac{\mu}{2} \bigl(\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \rangle - \mathbb{p} \bigr) \Bigr\} \leqslant \frac{\mu^2 \mathtt{v}^2}{4(1-\lambda\mu)}. \tag{B.4}$$

For (B.1), it remains to check that the choice μ by (B.3) yields

$$\frac{\mu^2 \mathtt{v}^2}{4(1-\lambda\mu)} - \frac{\mu\, z_2(B,\mathtt{x})}{2} = \frac{\mu^2 \mathtt{v}^2}{4(1-\lambda\mu)} - \mu\bigl(\mathtt{v}\sqrt{\mathtt{x}} + \lambda\mathtt{x}\bigr) = \mu\Bigl(\frac{\mathtt{v}\sqrt{\mathtt{x}}}{2} - \mathtt{v}\sqrt{\mathtt{x}} - \lambda\mathtt{x} \Bigr) = -\mathtt{x}.$$

The bound (B.2) is obtained similarly from Markov's inequality applied to $-\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \rangle + \mathbb{p}$ with $\mu = 2\mathtt{v}^{-1}\sqrt{\mathtt{x}}$. The use of (A.3) yields

$$\begin{aligned} \mathbb{P}\Bigl(\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \rangle - \mathbb{p} < -2\mathtt{v}\sqrt{\mathtt{x}} \Bigr) &\leqslant \mathbb{E} \exp\Bigl\{ \frac{\mu}{2} \bigl(-\langle B\boldsymbol{\gamma}, \boldsymbol{\gamma} \rangle + \mathbb{p} \bigr) - \mu\, \mathtt{v}\sqrt{\mathtt{x}} \Bigr\} \\ &\leqslant \exp\Bigl(\frac{\mu^2 \mathtt{v}^2}{4} - \mu\, \mathtt{v}\sqrt{\mathtt{x}} \Bigr) = \mathrm{e}^{-\mathtt{x}} \end{aligned}$$

as required. □

Corollary B.2 *Assume the conditions of Theorem B.1. Then for* $z > \mathtt{v}$*,*

$$\mathbb{P}\bigl(\bigl| \| Q\boldsymbol{\xi} \|^2 - \mathbb{p} \bigr| \geqslant z \bigr) \leqslant 2 \exp\Biggl\{ - \frac{z^2}{\bigl(\mathtt{v} + \sqrt{\mathtt{v}^2 + 2\lambda z} \bigr)^2} \Biggr\} \leqslant 2 \exp\Bigl(- \frac{z^2}{4\mathtt{v}^2 + 4\lambda z} \Bigr). \tag{B.5}$$

Proof Given z, define $\mathtt{x}$ by $2\mathtt{v}\sqrt{\mathtt{x}} + 2\lambda\mathtt{x} = z$ or $2\lambda\sqrt{\mathtt{x}} = \sqrt{\mathtt{v}^2 + 2\lambda z} - \mathtt{v}$. Then

$$\mathbb{P}\bigl(\| Q\boldsymbol{\xi} \|^2 - \mathbb{p} \geqslant z \bigr) \leqslant \mathrm{e}^{-\mathtt{x}} = \exp\Biggl\{ - \frac{\bigl(\sqrt{\mathtt{v}^2 + 2\lambda z} - \mathtt{v} \bigr)^2}{4\lambda^2} \Biggr\} = \exp\Biggl\{ - \frac{z^2}{\bigl(\mathtt{v} + \sqrt{\mathtt{v}^2 + 2\lambda z} \bigr)^2} \Biggr\}.$$

This yields (B.5) by direct calculus. □

Of course, bound (B.5) is sensible only if $z \gg \mathtt{v}$,

Corollary B.3 *Assume the conditions of Theorem B.1. If also* $B \geqslant 0$*, then*

$$\mathbb{P}\Bigl(\| Q\boldsymbol{\xi} \|^2 \geqslant z^2(B,\mathtt{x}) \Bigr) \leqslant \mathrm{e}^{-\mathtt{x}}$$

with

$$z^2(B,\mathtt{x}) \stackrel{\mathrm{def}}{=} \mathbb{p} + 2\mathtt{v}\sqrt{\mathtt{x}} + 2\lambda\mathtt{x} \leqslant \bigl(\sqrt{\mathbb{p}} + \sqrt{2\lambda\mathtt{x}} \bigr)^2.$$

Also

$$\mathbb{P}\Bigl(\| Q\boldsymbol{\xi} \|^2 - \mathbb{p} < -2\mathtt{v}\sqrt{\mathtt{x}} \Bigr) \leqslant \mathrm{e}^{-\mathtt{x}}.$$

Proof The definition implies $\mathtt{v}^2 \leqslant \mathbb{p}\lambda$ yielding the statement of the corollary. □

As a special case, we present a bound for the chi-squared distribution corresponding to $Q = \mathbb{V}^2 = \mathbb{I}_p$, $p < \infty$. Then $B = \mathbb{I}_p$, $\operatorname{tr}(B) = p$, $\operatorname{tr}(B^2) = p$ and $\lambda(B) = 1$.

Corollary B.4 *Let* $\boldsymbol{\gamma}$ *be a standard normal vector in* $\mathbb{R}^p$*. Then for any* $\mathtt{x} > 0$*,*

$$\begin{aligned} [ccl]\mathbb{P}\bigl(\| \boldsymbol{\gamma} \|^2 &\geqslant p + 2\sqrt{p\,\mathtt{x}} + 2\mathtt{x} \bigr) \leqslant \mathrm{e}^{-\mathtt{x}}, \\ \mathbb{P}\bigl(\| \boldsymbol{\gamma} \| \ &\geqslant \sqrt{p} + \sqrt{2\mathtt{x}} \bigr) \leqslant \mathrm{e}^{-\mathtt{x}}, \\ \mathbb{P}\bigl(\| \boldsymbol{\gamma} \|^2 &\leqslant p - 2\sqrt{p\,\mathtt{x}} \bigr) \leqslant \mathrm{e}^{-\mathtt{x}}. \end{aligned}$$

The bound of Theorem B.1 can be represented as a usual deviation bound.

Theorem B.5 *Assume the conditions of Theorem B.1. For* $\mathtt{y} > 0$*, define*

$$\mathtt{x}(\mathtt{y}) \overset{\mathrm{def}}{=} \frac{(\sqrt{\mathtt{y}+\mathbb{p}}-\sqrt{\mathbb{p}})^2}{4\lambda}.$$

Then

$$\mathbb{P}\bigl(\|Q\boldsymbol{\xi}\|^2 \geqslant \mathbb{p}+\mathtt{y}\bigr) \leqslant \mathrm{e}^{-\mathtt{x}(\mathtt{y})}, \tag{B.6}$$

$$\mathbb{E}\bigl\{(\|Q\boldsymbol{\xi}\|^2-\mathbb{p})\,\mathbb{1}\bigl(\|Q\boldsymbol{\xi}\|^2 \geqslant \mathbb{p}+\mathtt{y}\bigr)\bigr\} \leqslant 2\Bigl(\frac{\mathtt{y}+\mathbb{p}}{\lambda\,\mathtt{x}(\mathtt{y})}\Bigr)^{1/2}\,\mathrm{e}^{-\mathtt{x}(\mathtt{y})}. \tag{B.7}$$

Moreover, let $\mu > 0$ *fulfill* $\epsilon = \mu\lambda + \mu\sqrt{\lambda\mathbb{p}/\mathtt{x}(\mathtt{y})} < 1$*. Then*

$$\mathbb{E}\bigl\{\mathrm{e}^{\mu(\|Q\boldsymbol{\xi}\|^2-\mathbb{p})/2}\,\mathbb{1}(\|Q\boldsymbol{\xi}\|^2 \geqslant \mathbb{p}+\mathtt{y})\bigr\} \leqslant \frac{1}{1-\epsilon}\,\exp\{-(1-\epsilon)\mathtt{x}(\mathtt{y})\}. \tag{B.8}$$

Proof Normalizing by λ reduces the statements to the case with $\lambda = 1$. Define $\eta = \|Q\boldsymbol{\xi}\|^2 - \mathbb{p}$ and

$$z(\mathtt{x}) = 2\sqrt{\mathbb{p}\,\mathtt{x}} + 2\mathtt{x}. \tag{B.9}$$

Then by (B.1) $\mathbb{P}(\eta \geqslant z(\mathtt{x})) \leqslant \mathrm{e}^{-\mathtt{x}}$. Inverting the relation (B.9) yields

$$\mathtt{x}(z) = \frac{1}{4}\bigl(\sqrt{z+\mathbb{p}}-\sqrt{\mathbb{p}}\bigr)^2$$

and (B.6) follows by applying $z = \mathtt{y}$. Further,

$$\mathbb{E}\bigl\{\eta\,\mathbb{1}(\eta \geqslant \mathtt{y})\bigr\} = \int_{\mathtt{y}}^{\infty} \mathbb{P}(\eta \geqslant z)\,dz \leqslant \int_{\mathtt{y}}^{\infty} \mathrm{e}^{-\mathtt{x}(z)}\,dz = \int_{\mathtt{x}(\mathtt{y})}^{\infty} \mathrm{e}^{-\mathtt{x}}\,z'(\mathtt{x})\,d\mathtt{x}.$$

As $z'(\mathtt{x}) = 2 + \sqrt{\mathbb{p}/\mathtt{x}}$ monotonously decreases with $\mathtt{x}$, we derive

$$\mathbb{E}\bigl\{\eta\,\mathbb{1}(\eta \geqslant \mathtt{y})\bigr\} \leqslant z'(\mathtt{x}(\mathtt{y}))\mathrm{e}^{-\mathtt{x}(\mathtt{y})} = \frac{1}{\mathtt{x}'(\mathtt{y})}\,\mathrm{e}^{-\mathtt{x}(\mathtt{y})} = \frac{4\sqrt{\mathtt{y}+\mathbb{p}}}{\sqrt{\mathtt{y}+\mathbb{p}}-\sqrt{\mathbb{p}}}\,\mathrm{e}^{-\mathtt{x}(\mathtt{y})}$$

and (B.7) follows.

In a similar way, define $\mathsf{z}(\mathtt{x})$ from the relation $\mu^{-1}\log\mathsf{z}(\mathtt{x}) = \sqrt{\mathbb{p}\,\mathtt{x}} + \mathtt{x}$ yielding

$$\mathsf{z}(\mathtt{x}) = \exp\bigl(\mu\sqrt{\mathbb{p}\,\mathtt{x}} + \mu\,\mathtt{x}\bigr).$$

The inverse relation reads

$$\mathrm{x}_{\mathrm{e}}(\mathsf{z}) = \bigl(\sqrt{\mu^{-1}\log\mathsf{z} + \mathbb{p}/4} - \sqrt{\mathbb{p}/4}\bigr)^2.$$

Then with $\mathtt{x}(\mathtt{y}) = \mathrm{x}_{\mathrm{e}}(\mathrm{e}^{\mu\mathtt{y}/2}) = \bigl(\sqrt{\mathtt{y}+\mathbb{p}}-\sqrt{\mathbb{p}}\bigr)^2/4$,

$$\begin{aligned}\mathbb{E}\bigl\{\mathrm{e}^{\mu\eta/2}\,\mathbb{1}(\eta \geqslant \mathtt{y})\bigr\} &= \int_{\mathrm{e}^{\mu\mathtt{y}/2}}^{\infty} \mathbb{P}(\mathrm{e}^{\mu\eta/2} \geqslant \mathsf{z})\,d\mathsf{z} = \int_{\mathrm{e}^{\mu\mathtt{y}/2}}^{\infty} \mathbb{P}(\eta \geqslant 2\mu^{-1}\log\mathsf{z})\,d\mathsf{z}\\ &\leqslant \int_{\mathrm{e}^{\mu\mathtt{y}/2}}^{\infty} \mathrm{e}^{-\mathrm{x}_{\mathrm{e}}(\mathsf{z})}\,d\mathsf{z} = \int_{\mathtt{x}(\mathtt{y})}^{\infty} \mathrm{e}^{-\mathtt{x}}\,\mathsf{z}'(\mathtt{x})\,d\mathtt{x}.\end{aligned}$$

Further, in view of $\mu + 0.5\,\mu\sqrt{\mathbb{p}/\mathtt{x}} < \mu + \mu\sqrt{\mathbb{p}/\mathtt{x}(\mathtt{y})} = \epsilon < 1$ for $\mathtt{x} \geqslant \mathtt{x}(\mathtt{y})$, it holds

$$\mathsf{z}'(\mathtt{x}) = \bigl(\mu + 0.5\,\mu\sqrt{\mathbb{p}/\mathtt{x}}\bigr)\exp\bigl(\mu\sqrt{\mathbb{p}\,\mathtt{x}} + \mu\,\mathtt{x}\bigr) \leqslant \exp\bigl(\mu\,\mathtt{x}\sqrt{\mathbb{p}/\mathtt{x}(\mathtt{y})} + \mu\,\mathtt{x}\bigr) = \exp(\epsilon\,\mathtt{x})$$

and

$$\mathbb{E}\bigl\{e^{\mu\eta/2}\,\mathbb{I}(\eta \geqslant \mathtt{y})\bigr\} \leqslant \int_{\mathtt{x}(\mathtt{y})}^{\infty} e^{-(1-\epsilon)\mathtt{x}}\,d\mathtt{x} = \frac{1}{1-\epsilon}\,e^{-(1-\epsilon)\mathtt{x}(\mathtt{y})}$$

and (B.8) follows. □

Acknowledgements

Financial support by the German Research Foundation (DFG) through the Collaborative Research Center 1294 “Data assimilation” is gratefully acknowledged.

References

[1] Abbe, E., Community detection and stochastic block models: Recent developments, Journal of Machine Learning Research, 2018, 18(177): 1–86.

[2] Bartlett, P. L., Long, P. M., Lugosi, G. and Tsigler, A., Benign overfitting in linear regression, Proceedings of the National Academy of Sciences, 2020, 117(48): 30063–30070.

[3] Bunea, F. and Xiao, L., On the sample covariance matrix estimator of reduced effective rank population matrices, with applications to fPCA, Bernoulli, 2015, 21(2): 1200–1230.

[4] Chen, P., Gao, C. and Zhang, A. Y., Optimal full ranking from pairwise comparisons, The Annals of Statistics, 2022, 50(3): 1775–1805.

[5] Cheng, C. and Montanari, A., Dimension free ridge regression, arXiv: 2210.08571, 2022.

[6] Gao, C., Ma, Z., Zhang, A. Y. and Zhou, H. H., Achieving optimal misclassification proportion in stochastic block models, Journal of Machine Learning Research, 2017, 18(60): 1–45.

[7] Klochkov, Y. and Zhivotovskiy, N., Uniform Hanson-Wright type concentration inequalities for unbounded entries via the entropy method, Electronic Journal of Probability, 2020, 25: 1–30.

[8] Laurent, B. and Massart, P., Adaptive estimation of a quadratic functional by model selection, The Annals of Statistics, 2000, 28(5): 1302–1338.

[9] Lounici, K., High-dimensional covariance matrix estimation with missing observations, Bernoulli, 2014, 20(3): 1029–1058.

[10] Rudelson, M. and Vershynin, R., Hanson-Wright inequality and sub-Gaussian concentration, Electronic Communications in Probability, 2013, 18: 1–9.

[11] Spokoiny, V., Inexact Laplace approximation and the use of posterior mean in Bayesian inference, Bayesian Analysis, 2025, 20(1): 1303–1330.

[12] Spokoiny, V., Mixed Laplace approximation for marginal posterior and Bayesian inference in error-in-operator model, arXiv: 2305.09336, 2023.

[13] Tropp, J. A., An introduction to matrix concentration inequalities, Foundations and Trends in Machine Learning, 2015, 8(1–2): 1–230.

[14] Vershynin, R., High-Dimensional Probability: An Introduction with Applications in Data Science, Number 47 in Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2018.