Weierstraß-Institut
für Angewandte Analysis und Stochastik

Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 2198-5855

Bernstein-type and Bennett-type inequalities for unbounded matrix martingales

Alexey Kroshnin, Alexandra Suvorikova

submitted: November 28, 2024

Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: alexei.kroshnin@wias-berlin.de
alexandra.suvorikova@wias-berlin.de

No. 3146
Berlin 2024

2020 *Mathematics Subject Classification.* 60E15.

Key words and phrases. Concentration, Bernstein inequality, Bennett inequality, martingales, Orlicz norm, empirical Bernstein inequality, McDiarmid inequality.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Bernstein-type and Bennett-type inequalities for unbounded matrix martingales

Alexey Kroshnin, Alexandra Suvorikova

Abstract

We derive explicit Bernstein-type and Bennett-type concentration inequalities for matrix-valued supermartingale processes with unbounded observations. Specifically, we assume that the ψ_α-Orlicz (quasi-)norms of their difference process are bounded for some $\alpha > 0$. As corollaries, we prove an empirical version of Bernstein's inequality and an extension of the bounded differences inequality, also known as McDiarmid's inequality.

1 Introduction

Non-asymptotic concentration inequalities play an essential role in a wide variety of fields, including probability theory, statistics [Arcones, 1995], graph theory [Krebs, 2018], machine learning [Lopez-Paz et al., 2014], theoretical computer science [Tolstikhin and Seldin, 2013], quantum statistics [Girotti et al., 2023], etc. These inequalities provide crucial probabilistic bounds that facilitate rigorous analysis in both theoretical and applied contexts. Key references for comprehensive surveys include the works by Ledoux and Talagrand [1991], Koltchinskii [2011], Boucheron et al. [2013].

This paper explores Bernstein-type and Bennett-type inequalities, which are pivotal in various research domains. These concentration inequalities play a crucial role in the analysis of weakly dependent observations [Merlevède et al., 2009, Banna et al., 2016], martingales [Dzhaparidze and Van Zanten, 2001, Tropp, 2011], stochastic and empirical processes [Bechar, 2009, Baraud, 2010, Hang and Steinwart, 2017], and the concentration of matrices and operators [Mackey et al., 2014, Minsker, 2017].

1.1 Related works

In the rest of the literature survey, we aim to highlight the significant milestones in the development of Bernstein-type bounds. The survey is structured chronologically, providing a comprehensive understanding of the field's evolution.

The early results, dating back to the beginning of the 20th century, deal mainly with bounded observations.

The celebrated Bernstein's inequality—formulated by Sergei Bernstein in the late 1920s [Bernstein, 1927]—stands as a cornerstone in the theory of concentration inequalities. It guarantees an exponential decay rate for the tail probabilities of the sum of independent bounded random variables.

Proposition 1.1 (Bernstein's inequality (bounded case))**.** *Let $X_1, \dots, X_n$ be independent random variables such that*

$$\mathbb{E}\, X_i = 0, \quad X_i \le K \text{ a.s.}, \quad \sigma^2 := \sum_{i=1}^n \mathbb{E}\, X_i^2.$$

DOI 10.20347/WIAS.PREPRINT.3146

Then for all $t > 0$

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq t\right) \leq \exp\left\{-\frac{t^2}{2\left(\sigma^2 + \frac{Kt}{3}\right)}\right\}.$$

Note that Bernstein also proposed going beyond the bounded case and considered the following moment bounds,

$$\mathbb{E}\, X_i^p \leq \frac{p!}{2} U_i^{p-2} \sigma_i^2, \quad p = 2, 3, \dots \tag{1}$$

with $\sigma_i^2 = \mathbb{E}\, X_i^2$, and $U_i > 0$ being some constant. This assumption is now known as Bernstein's moment condition. It ensures the sub-gamma behavior of X_i, see Corollary 5.2 in Zhang and Chen [2020]. However, further research on the unbounded case did not attract much attention until the beginning of the 21st century.

The next famous result concerning bounded observations—derived by George Bennett in 1962 [Bennett, 1962]—presents a sharper version of Proposition 1.1.

Proposition 1.2 (Bennett's inequality). *Under assumptions of Proposition 1.1, it holds for all $t > 0$ that*

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq t\right) \leq \exp\left\{-\frac{\sigma^2}{K^2} h\left(\frac{Kt}{\sigma^2}\right)\right\},$$

where $h(x) := (1 + x)\ln(1 + x) - x$ for all $x \geq 0$.

Alongside independent observations, the dependent case also gained attention. So, in 1975, David Freedman [Freedman, 1975] derived the famous martingale extension of Proposition 1.1.

Almost in parallel, in 1976, Vadim Yurinskii generalized Proposition 1.1 to the case of random variables in Banach spaces. He assumed the norm of observations to satisfy Bernstein's moment condition (1) [Yurinskii, 1976].

Joel Tropp, in 2011, generalized Freedman's result to the case of matrix-valued martingales (see Proposition 1.3). One year later, he got a result similar to Yurinsky's one. Namely, he applied assumption (1) to the matrix-valued case (see Proposition 1.4).

In the following, we denote as $\lambda_{\max}(\mathbf{X})$ the largest eigenvalue of $\mathbf{X}$, as $\|X\|$ the operator norm of X, and as $\mathbb{H}(d)$ the space of $d \times d$ Hermitian matrices.

Proposition 1.3 (Theorem 3.1, Tropp [2011]). *Let $\mathbf{X}_1, \dots, \mathbf{X}_n \in \mathbb{H}(d)$ be a sequence adapted to a filtration $\mathrm{F}_1 \subset \mathrm{F}_2 \subset \dots \subset \mathrm{F}_n$. Let $\mathbf{X}_k$ satisfy for all $k = 1, \dots, n$*

$$\mathbb{E}[\mathbf{X}_k | \mathrm{F}_{k-1}] = 0, \quad \lambda_{\max}(\mathbf{X}_k) \leq K \text{ a.s.}$$

Set for all $k = 1, \dots, n$

$$\mathbf{S}_k := \sum_{i=1}^{k} \mathbf{X}_i, \quad \mathbf{\Sigma}_k := \sum_{i=1}^{k} \mathbb{E}\left[\mathbf{X}_i^2 \middle| \mathrm{F}_{i-1}\right].$$

Then, for all $t \geq 0$ and $\sigma^2 > 0$,

$$\mathbb{P}\left(\exists k \geq 0 : \lambda_{\max}(\mathbf{S}_k) \geq t \text{ and } \|\mathbf{\Sigma}_k\| \leq \sigma^2\right) \leq d \exp\left\{-\frac{\sigma^2}{K^2} h\left(\frac{Kt}{\sigma^2}\right)\right\}.$$

Proposition 1.4 (Theorem 6.2, Tropp [2012]). *Let random matrices $\mathbf{X}_1, \dots, \mathbf{X}_n \in \mathbb{H}(d)$ satisfy for all $k = 1, \dots, n$, all $p = 2, 3, \dots$, with some $K > 0$ and positive-definite matrices $\mathbf{\Sigma}_k$*

$$\mathbb{E}\,\mathbf{X}_k = 0, \quad \mathbb{E}\,\mathbf{X}_k^p \preccurlyeq \frac{p!}{2} K^{p-2} \mathbf{\Sigma}_k.$$

Let

$$\sigma^2 := \left\| \sum_{k=1}^n \mathbf{\Sigma}_k \right\|.$$

Then for all $t \geq 0$

$$\mathbb{P}\left(\lambda_{\max} \left(\sum_{k=1}^n \mathbf{X}_k \right) \geq t \right) \leq d \exp\left\{ -\frac{t^2}{2(\sigma^2 + Kt)} \right\}.$$

All the above results deal with bounded random variables or those satisfying Bernstein's moment condition. As one can see, the moment condition is too strong, especially in the case of random matrices.

The current study focuses on the unbounded case. To introduce the setting, we briefly recall the concept of the Orlicz norm. The Orlicz function we use is

$$\psi_\alpha(x) := e^{x^\alpha} - 1, \quad \alpha > 0.$$

The ψ_α-Orlicz (quasi-)norm of a real-valued random variable X is

$$\|X\|_{\psi_\alpha} := \inf\left\{ t > 0 : \mathbb{E}\,\psi_\alpha\left(\frac{|X|}{t} \right) \leq 1 \right\}. \tag{2}$$

If $\alpha \geq 1$, $\|X\|_{\psi_\alpha}$ is a norm. In particular, if $\|X\|_{\psi_1} < \infty$, X is sub-exponential, and if $\|X\|_{\psi_2} < \infty$, X is sub-Gaussian. Moreover, if $0 < \alpha < 1$, $\|X\|_{\psi_\alpha}$ is a quasi-norm.

In 2008, Radoslaw Adamczak got the concentration result for unbounded empirical processes [Adamczak, 2008]. Being applied to a particular case of a sum of independent observations, it yields a Bernstein-type deviation bound. The result holds under the assumption that the summands have finite ψ_α-Orlicz (quasi-)norm for $0 < \alpha \leq 1$.

In 2011, Vladimir Koltchinskii obtained an extension of Proposition 1.1 for a sum of independent Hermitian matrices with bounded ψ_α-Orlicz norm.

Proposition 1.5 (Theorem 2.7, Koltchinskii [2011]). *Let $\mathbf{X}_1, \dots, \mathbf{X}_n \in \mathbb{H}(d)$ be independent random matrices. Fix $\alpha \geq 1$. Suppose, for all $k = 1, \dots, n$, and some $K > 0$,*

$$\mathbb{E}\,\mathbf{X}_k = 0, \quad \max\left(\| \|\mathbf{X}_k\| \|_{\psi_\alpha}, 2\sqrt{\mathbb{E}\|\mathbf{X}_k\|^2} \right) \leq K. \tag{3}$$

Set

$$\sigma^2 := \left\| \sum_{k=1}^n \mathbb{E}\,\mathbf{X}_k^2 \right\|.$$

Then, there exists an absolute constant $C > 0$ such that, for all $t \geq 0$,

$$\mathbb{P}\left(\left\| \sum_{k=1}^n \mathbf{X}_k \right\| \geq t \right) \leq 2d \exp\left\{ -\frac{1}{C} \frac{t^2}{\sigma^2 + tK \left(\log \frac{nK^2}{\sigma^2} \right)^{1/\alpha}} \right\}.$$

Many excellent results deal with the Bernstei-type bounds under different settings, e.g., [van de Geer and Lederer, 2013, Gao et al., 2014]; we recommend Boucheron et al. [2013] for other references. However, they are beyond the scope of the current study.

1.2 Main results

In this work, we consider a matrix-valued supermartingale difference sequence $\mathbf{0} \equiv \mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_n \in \mathbb{H}(d)$ adapted to a filtration $(\mathrm{F}_i)_{i=0}^n$ ($\mathrm{F}_0 := \{\Omega, \emptyset\}$ is the trivial σ-algebra), i.e., $\mathbb{E}\|\mathbf{X}_i\| < \infty$ and $\mathbb{E}\left[\mathbf{X}_i | \mathrm{F}_{i-1}\right] \preccurlyeq \mathbf{0}$ for all $i = 1, \ldots, n$. Set for any $k = 1, \ldots, n$

$$\mathbf{S}_k := \sum_{i=1}^{k} \mathbf{X}_i.$$

Clearly, $(\mathbf{S}_i)_{i=1}^n$ is a matrix supermartingale adapted to $(\mathrm{F}_i)_{i=1}^n$.

The main result, Theorem 2.1, shows that one can obtain a combination of Bernstein- and Bennett-type deviation bounds on $\max_{k\in[n]} \lambda_{\max}(\mathbf{S}_k)$. All constants are computed explicitly. We note that the result can be extended to the case of rectangular matrices using dilations (see, e.g., [Paulsen, 2002]).

The validity of the result depends on assumptions about the behavior of the observations $\mathbf{X}_i$. Specifically, we assume the *conditioned Orlicz norm*, $\|\lambda_{\max}(\mathbf{X}_i)_+ | \mathrm{F}_{i-1}\|_{\psi_\alpha}$, is bounded. As far as we know, this is a new concept, so we explain it below.

Conditioned Orlicz norm. Let us fix a probability space $(\Omega, \mathrm{F}, \mathbb{P})$. We denote by $\mathbb{I}\{E\}$ the indicator of an event $E \in \mathrm{F}$. Given a sub-σ-algebra of events $\mathcal{G} \subset \mathrm{F}$ and a random variable $X \in \mathbb{R}$, let $\mu_{X|\mathcal{G}}$ be a conditional distribution of X w.r.t. $\mathcal{G}$; i.e. $\mu_{X|\mathcal{G}}$ is a $\mathcal{G}$-measurable random measure on $\mathbb{R}$ such that for any Borel set $A \subset \mathbb{R}$

$$\mu_{X|\mathcal{G}}(A) = \mathbb{P}\left[X \in A | \mathcal{G}\right] := \mathbb{E}\left[\mathbb{I}\{X \in A\} | \mathcal{G}\right] \text{ a.s.},$$

see Chapter 2, §7 in the book by Shiryaev [2016].

We define the conditional Orlicz norm of X as the norm of the conditional distribution $\mu_{X|\mathrm{F}}$,

$$\|X|\mathrm{F}\|_{\psi_\alpha} := \left\|\mu_{X|\mathrm{F}}\right\|_{\psi_\alpha},$$

i.e. this is a $\mathcal{G}$-measurable random variable $\omega \mapsto \|\mu_{X|\mathrm{F}}(\omega)\|_{\psi_\alpha}$. Here, abusing notations, we denote by the norm of a measure μ the norm of a r.v. distributed according to μ. It can be explicitly written, e.g., as

$$\|X|\mathrm{F}\|_{\psi_\alpha} = \sup_{t\in\mathbb{Q},\, t\geq 0} t\, \mathbb{I}\left\{\mathbb{E}\left[\psi_\alpha\left(\frac{|X|}{t}\right)\middle|\mathrm{F}\right] > 1\right\} \text{ a.s.},$$

where $\mathbb{Q}$ is the set of rational numbers. As $\mathbb{Q}$ is countable and dense in $\mathbb{R}$, one can see that this is indeed a random variable, and it coincides with $\|\mu_{X|\mathrm{F}}(\omega)\|_{\psi_\alpha}$ a.s.

Corollaries. To demonstrate the applicability of Theorem 2.1, we derive two corollaries: an empirical Bernstein inequality and a version of McDiarmid inequality.

Empirical Bernstein-type bound. Bernstein-type bounds rely on the true variance of the observations, while empirical Bernstein-type inequalities, in contrast, incorporate a data-driven variance estimator [Peel et al., 2010, Martinez-Taboada and Ramdas, 2024]. The latter ones play an essential role in the theoretical analysis of machine learning algorithms [Audibert et al., 2007, Mnih et al., 2008, Maurer and Pontil, 2009, Shivaswamy and Jebara, 2010, Tolstikhin and Seldin, 2013].

Corollary 3.3 presents an empirical Bernstein-type bound for the case of i.i.d. matrix-valued observations. To the best of our knowledge, this result is novel.

McDiarmid's inequality. McDiarmid's inequality provides a powerful tool for bounding the deviations of functions of independent random variables from their expected values. Specifically, it addresses the functions satisfying the bounded property.

Let $\mathcal{Y}$ be a measurable space and let $f\colon \mathcal{Y}^n \to \mathbb{R}$ be such that there exist $U_1, \dots U_n \in \mathbb{R}_+$ satisfying

$$\sup_{y_i' \in \mathcal{Y}} |f(y_1, \dots, y_{i-1}, y_i, y_{i+1}, \dots, y_n) - f(y_1, \dots, y_{i-1}, y_i', y_{i+1}, \dots, y_n)| \le U_i$$

for all $y \in \mathcal{Y}^n$.

McDiarmid's inequality (see, e.g., (1.3) in McDiarmid et al. [1989]) ensures that if $Y_1, \dots, Y_n \in \mathcal{Y}$ are independent and if f satisfies the above properties than

$$\mathbb{P}\left(f(Y_1, \dots, Y_n) - \mathbb{E}\, f(Y_1, \dots, Y_n) \ge t\right) \le \exp\left\{-\frac{2t^2}{\sum_i U_i^2}\right\}.$$

Many works develop McDiarmid inequalities under extended settings [Kutin, 2002, Rio, 2013, Zhang et al., 2019]. Among the recent results on the concentration of dependent and unbounded observation, one should mention the work by Maurer and Pontil [2021]. The authors propose a Bernstein-type generalization of McDiarmid's inequality for functions with sub-exponential differences.

Corollary 3.4 and Corollary 3.5 present Bernstein-type McDiarmid inequalities for functions whose differences have bounded ψ_α-Orlicz norm. We compare the results with those by Maurer and Pontil [2021].

Organization of the paper

Section 2 presents the main result and several straightforward corollaries. It also examines the tail behavior of the bounds and compares the main result with those discussed in the Introduction. Section 3 contains all corollaries. Finally, Section 4 collects the main proofs. Auxiliary results are collected in the Appendix.

Accepted notations

Spaces and sets. We denote as $\mathbb{H}(d)$ the space of all d-dimensional Hermitian matrices. $\mathbb{H}_+(d) \subset \mathbb{H}(d)$ is the set of positive semi-definite Hermitian matrices. $\mathbb{H}_{++}(d) \subset \mathbb{H}(d)$ is the set of positive-definite Hermitian matrices. Further, we denote the integer indices as $[n] = \{1, \dots, n\}$.

Norms. Let $\|\boldsymbol{A}\|$ be the operator norm of a matrix $\boldsymbol{A}$.

Functions. From now on, we set for any $x \in \mathbb{R}$

$$\underline{\log}\, x := \max(\ln x, 1), \quad x_+ := \max(x, 0).$$

Let $\mathbb{I}[E]$ be the indicator of an event E (i.e., $\mathbb{I}[E]$ is a random variable). Respectively, $\mathbb{I}_E$ denotes the indicator function of a set E.

Further, we define functions $\phi\colon \mathbb{R} \to \mathbb{R}$ and $h\colon (-1, \infty) \to \mathbb{R}$ as

$$\phi(t) := e^t - 1 - t, \quad h(x) := (1 + x)\ln(1 + x) - x. \tag{4}$$

Note that h is the convex conjugate of ϕ.

Now, let f be a scalar function. For any $d \times d$ diagonal matrix $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \dots, \lambda_d)$, we define

$$f(\mathbf{\Lambda}) := \mathrm{diag}\big(f(\lambda_1), \dots, f(\lambda_d)\big).$$

Respectively, given a matrix $\mathbf{A} \in \mathbb{H}(d)$ with a spectral decomposition $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^*$, we set

$$f(\mathbf{A}) := \mathbf{U} f(\mathbf{\Lambda}) \mathbf{U}^*.$$

We also recall the transfer rule. Let for any $x \in I \subset \mathbb{R}$, $f(x) \le g(x)$. If all eigenvalues of $\mathbf{A}$ belong to I, then $f(\mathbf{A}) \preccurlyeq g(\mathbf{A})$.

2 Bernstein- and Bennett-type inequalities for unbounded matrix martingales

This section presents the core Bernstein- and Bennett-type bounds for matrix supermartingales. Recall that we formulate them in terms of a difference sequence defined in Section 1.2.

Theorem 2.1. *Let $0 \equiv \mathbf{X}_0, \mathbf{X}_1, \dots, \mathbf{X}_n \in \mathbb{H}(d)$ be a supermartingale difference sequence adapted to a filtration $(\mathrm{F}_i)_{i=0}^n$ with $\mathrm{F}_0 = \{\Omega, \emptyset\}$.*

Fix $\alpha > 0$ and set for all $i \in [n]$

$$\mathbf{\Sigma}_i := \mathbb{E}\left[\mathbf{X}_i^2 \middle| \mathrm{F}_{i-1}\right], \quad U_i := \left\| \lambda_{\max}(\mathbf{X}_i)_+ \middle| \mathrm{F}_{i-1} \right\|_{\psi_\alpha}.$$

Let $\sigma > 0$, $U \ge K > 0$ be such that, with probability at least $1 - p$,

$$\lambda_{\max}\left(\sum_{i=1}^n \mathbf{\Sigma}_i\right) \le \sigma^2, \quad \sum_{i=1}^n U_i^2 \le U^2, \quad \max_{i \in [n]} U_i \le K, \tag{5}$$

Let

$$z := \begin{cases} \left(4 \,\underline{\log}\, \frac{eU}{\sigma}\right)^{1/\alpha} & \textit{if } \alpha \ge 1, \\ \left[\frac{4}{\alpha} \ln \frac{e}{\alpha} + 4\left(\ln \frac{U}{\sigma}\right)_+\right]^{1/\alpha}, & \textit{if } \alpha < 1. \end{cases} \tag{6}$$

Then for any $\mathrm{x} > 0$, it holds, with probability at least $1 - p - de^{-\mathrm{x}} - e^{-\mathrm{x}}\, \mathbb{I}[\alpha < 1]$, that

$$\begin{aligned} \max_{k \in [n]} \lambda_{\max}(\mathbf{S}_k) \le \sigma\sqrt{2\mathrm{x}} &+ \frac{4Kz\mathrm{x}}{\min\left\{2\alpha z^\alpha,\, \underline{\log}\left(\left(\frac{Kz}{\sigma}\right)^2 \mathrm{x}\right)\right\}} \\ &+ \frac{3K}{\alpha}\left(2\mathrm{x} + 2\ln\left(\frac{4U}{K}\right) + \frac{4}{\alpha}\ln\left(\frac{4}{\alpha e}\right)\right)^{\frac{1-\alpha}{\alpha}} \mathbb{I}[\alpha < 1] \end{aligned} \tag{Ben}$$

and, moreover,

$$\begin{aligned} \max_{k \in [n]} \lambda_{\max}(\mathbf{S}_k) \le \sigma\sqrt{2\mathrm{x}} &+ \frac{3}{4}Kz\mathrm{x} \\ &+ \frac{3K}{\alpha}\left(2\mathrm{x} + 2\ln\left(\frac{4U}{K}\right) + \frac{4}{\alpha}\ln\left(\frac{4}{\alpha e}\right)\right)^{\frac{1-\alpha}{\alpha}} \mathbb{I}[\alpha < 1]. \end{aligned} \tag{Ber}$$

We postpone the proof to Section 4.

Remark 2.2. *Note that both bounds can be non-monotone w.r.t. σ. Yet, as σ is just an upper bound, one can improve them.*

For simplicity, we consider the case $\alpha \geq 1$. Setting $z' := \left(4\,\underline{\log}\,\frac{eU}{\sigma'}\right)^{1/\alpha}$, we get

$$\max_{k\in[n]} \lambda_{\max}(\mathbf{S}_k) \leq \inf_{\sigma'\geq\sigma} \left\{ \sigma'\sqrt{2\mathrm{x}} + \frac{4Kz'\mathrm{x}}{\min\left\{2\alpha(z')^{\alpha},\ \underline{\log}\left(\left(\frac{Kz'}{\sigma'}\right)^2\mathrm{x}\right)\right\}} \right\},$$

$$\max_{k\in[n]} \lambda_{\max}(\mathbf{S}_k) \leq \inf_{\sigma'\geq\sigma} \left\{ \sigma'\sqrt{2\mathrm{x}} + \frac{3}{4}Kz'\mathrm{x} \right\}.$$

The same holds for $0 < \alpha < 1$ with the corresponding substitution of σ by σ' to z.

Tail regimes. In this section we consider $\alpha \geq 1$. The bound (Ben) has three tail regimes: sub-Gaussian, sub-Poisson, and sub-exponential.

Sub-Gaussian. If $\left(\frac{Kz}{\sigma}\right)^2\mathrm{x} < e$,

$$\max_k \lambda_{\max}(\mathbf{S}_k) \leq 6\sigma\sqrt{2\mathrm{x}}. \tag{7}$$

Sub-Poisson. If $e \leq \left(\frac{Kz}{\sigma}\right)^2\mathrm{x} \leq e^{2\alpha z^{\alpha}}$,

$$\max_k \lambda_{\max}(\mathbf{S}_k) \leq 8\frac{Kz\mathrm{x}}{\ln\left(\left(\frac{Kz}{\sigma}\right)^2\mathrm{x}\right)}. \tag{8}$$

Sub-exponential. If $\left(\frac{Kz}{\sigma}\right)^2\mathrm{x} \geq e^{2\alpha z^{\alpha}}$,

$$\max_k \lambda_{\max}(\mathbf{S}_k) \leq \frac{6Kz}{\alpha z^{\alpha}}\mathrm{x}. \tag{9}$$

The proofs are postponed to Appendix A.

2.1 Comparison with other results

Bennett-type result for $\alpha \to \infty$, Proposition 1.3. Recall that the classical Bennett's bound corresponds to the case $\alpha = +\infty$ (bounded case). Consider (Ben). If $\alpha \to \infty$, then $z \to 1$. This yields $\alpha z^{\alpha} \to \infty$. Thus, one gets the sub-Poisson tail behavior (8) that coincides with the Bennett-type bound from Proposition 1.3 up to a multiplicative constant.

Bernstein for sub-gamma matrices, Proposition 1.4. Bernstein's moment condition is equivalent to

$$\mathbb{E}\,\phi\left(\frac{|X_i|}{U_i}\right) \lesssim \frac{\sigma_i^2}{U_i^2},$$

up to multiplicative constants (see p.103 in [Van Der Vaart et al., 1996]). Moreover, the proof of Lemma 4.6 ensures that a bound on the Orlicz norm yields Bernstein's moment condition. This, in turn, yields Bernstein's concentration inequality in the scalar case.

This approach requires two-sided bounds on X_i while Theorem 2.1 requires only a one-sided one. Furthermore, in the matrix setting, it is not apparent how to obtain Bernstein's condition for non-isotropic $\mathbf{\Sigma}_i$ (i.e., for $\mathbf{\Sigma}_i$ with large condition number), except in the case of bounded or commutative random matrices $\mathbf{X}_i$.

Adamczak's result ($\alpha \le 1$) [Adamczak, 2008]. This work focuses on empirical processes. The proof technique primarily builds upon the Klein–Rio bound [Klein and Rio, 2005], Hoffman–Jørgensen, and Talagrand inequalities. As a result, the derived bound includes an additional multiplicative term that arises naturally from these methods and is standard in the setting of empirical processes.

To illustrate the difference between our results and those of Adamczak, we consider a sum of independent scalar random variables $X_1, \dots, X_n$. The author uses truncation of X_i at a certain constant level. This yields a bound on a quantile of the tail term. The bound is proportional to $\left\|\max_{i\in[n]} |X_i|\right\|_{\psi_\alpha} \mathrm{x}^{1/\alpha}$ (see equation (11) in [Adamczak, 2008]). This bound is comparable to the quantile threshold τ given by (25) in the proof of Theorem 2.1.

Koltchinkii's bound ($\alpha \ge 1$), Proposition 1.5. This setting by Vladimir Koltchinskii is the closest to the current study setting. However, there are several differences. First, Theorem 2.1 handles dependent observations, while Proposition 1.5 assumes their independence. Further, our result is one-sided: (3) requires $\big\| \|\mathbf{X}_k\| \big\|_{\psi_\alpha}$ to be bounded, while our result relies only on the boundedness of the $\|\lambda_{\max}(\mathbf{X}_k)_+\|_{\psi_\alpha}$. Moreover, (3) depends on n, as it uses the upper bound nK^2 instead of U^2. Finally, Koltchinskii derives only a Bernstein-type bound, while the current study presents a mixed bound.

3 Corollaries

3.1 Straightforward corollaries

Theorem 2.1 entails two trivial corollaries. For the sake of brevity, we provide them only for $\alpha \ge 1$.

Corollary 3.1. *Let $\mathbf{X}_1, \dots, \mathbf{X}_n \in \mathbb{H}(d)$ be i.i.d. random matrices and*

$$\mathbb{E}\,\mathbf{X}_1 \preccurlyeq \mathbf{0}, \quad \lambda_{\max}(\mathbb{E}\,\mathbf{X}_1^2) \le \sigma^2, \quad \|\lambda_{\max}(\mathbf{X}_1)_+\|_{\psi_\alpha} \le K, \quad z := \left(4 \underline{\log} \frac{eK}{\sigma}\right)^{1/\alpha}.$$

Then, with probability at least $1 - de^{-\mathrm{x}}$,

$$\lambda_{\max}\left(\frac{1}{n}\sum_i \mathbf{X}_i\right) \le \sigma\sqrt{\frac{2\mathrm{x}}{n}} + \frac{4Kz}{\min\left\{2\alpha z^\alpha,\, \underline{\log}\left(\left(\frac{Kz}{\sigma}\right)^2 \frac{\mathrm{x}}{n}\right)\right\}} \frac{\mathrm{x}}{n}.$$

The same holds for (Ber).

Proof. The proof is trivial. The assumptions of the Corollary ensure the validity of assumptions from Theorem 2.1 with $p = 0$ and $n\sigma^2$, nK^2 instead of σ^2, U^2. □

Corollary 3.2. *Let $X_1, \dots, X_n$ be scalar random variables satisfying assumptions of Theorem 2.1 with*

$$\sigma_i^2 := \mathbb{E}[X_i^2 | \mathrm{F}_{i-1}], \quad \sigma^2 := \sum_i \sigma_i^2, \quad U_i := \|(X_i)_+ | \mathrm{F}_{i-1}\|_{\psi_\alpha}.$$

Then with probability at least $1 - p - e^{-\mathrm{x}}$

$$\max_{k\in[n]} \sum_{i=1}^{k} X_i \le \sigma\sqrt{2\mathrm{x}} + \frac{4Kz\mathrm{x}}{\min\left\{2\alpha z^\alpha,\ \underline{\log}\left(\left(\frac{Kz}{\sigma}\right)^2 \mathrm{x}\right)\right\}}. \tag{10}$$

The same holds for (Ber)*.*

3.2 Empirical Bernstein bounds

This section presents modified bound (Ber). Namely, we replace σ with its data-driven estimator $\hat{\sigma}$.

For the sake of simplicity, we focus on the case $\alpha \ge 1$. However, one can obtain a similar result for $\alpha < 1$.

Corollary 3.3. *Let $\mathbf{X}_1, \dots, \mathbf{X}_n \in \mathbb{H}(d)$ be i.i.d. with $\mathbf{\Sigma} := \mathbb{E}(\mathbf{X}_1 - \mathbb{E}\,\mathbf{X}_1)^2$ and*

$$\lambda_{\max}(\mathbf{\Sigma}) \le \sigma^2, \quad \vert\vert\vert \mathbf{X}_1 - \mathbb{E}\,\mathbf{X}_1 \vert\vert\vert_{\psi_\alpha} \le K.$$

Denote

$$\overline{\mathbf{X}} := \frac{1}{n}\sum_i \mathbf{X}_i, \quad \hat{\mathbf{\Sigma}} := \frac{1}{n}\sum_i (\mathbf{X}_i - \overline{\mathbf{X}})^2, \quad \hat{\sigma}^2 := \lambda_{\max}(\hat{\mathbf{\Sigma}}),$$

and define $\hat{z} := z(K, \hat{\sigma}; \alpha) := \left(4\,\underline{\log}\,\frac{Ke}{\hat{\sigma}}\right)^{1/\alpha}$.

Then for any $\mathrm{x} > 0$ such that $n \ge 8\mathrm{x}$, with probability at least $1 - 3de^{-\mathrm{x}}$,

$$\left\|\overline{\mathbf{X}} - \mathbb{E}\,\mathbf{X}_1\right\| \le \hat{\sigma}\sqrt{2\frac{\mathrm{x}}{n}} + 15K\hat{z}\frac{\mathrm{x}}{n}.$$

The proof is in Section 4.2.

3.3 McDiarmid inequality

In the following, we consider $\alpha \ge 1$. Let $\mathcal{Y}$ be a measurable space and $Y_1, \dots, Y_n \in \mathcal{Y}$ be independent random variables. Denote $Y := (Y_1, \dots, Y_n)$ and set $Y' := (Y_1', \dots, Y_n')$ to be an i.i.d. copy of Y.

For all $k \in [n]$ define σ-algebras

$$\mathrm{F}_{-k} := \sigma\left(Y_1, \dots, Y_{k-1}, Y_{k+1}, \dots, Y_n\right).$$

Let $f\colon \mathcal{Y}^n \to \mathbb{R}$ be a measurable function. Denoting as $y = (y_1, \dots, y_n) \in \mathcal{Y}^n$ a non-random vector, we define

$$f_i(y) := f(y) - \mathbb{E}'\, f(y_1, \dots, y_{i-1}, Y_i', y_{i+1}, \dots, y_n),$$

where $\mathbb{E}'$ is the expectation w.r.t. Y'.

Corollary 3.4. *Set*

$$U_i := \left\| f_i^2(Y) \mid \mathrm{F}_{-i} \right\|_{\psi_\alpha}, \quad \sigma_i^2 := \mathbb{E}\left[f_i^2(Y) \mid \mathrm{F}_{-i} \right],$$

and let the following inequalities hold a.s.:

$$\max_k U_k \le K \le U, \quad \sum_k U_k^2 \le U^2, \quad \sum_k \sigma_k^2 \le \sigma^2.$$

Then, with probability at least $1 - e^{-\mathrm{x}}$*,*

$$f(Y) - \mathbb{E}\, f(Y) \le \sigma \sqrt{2\mathrm{x} \frac{n+1}{n}} + \frac{4Kz\mathrm{x}}{\min\left\{ 2\alpha z^\alpha,\ \underline{\log} \left(\left(\frac{Kz}{\sigma \frac{n+1}{n}} \right)^2 \mathrm{x} \right) \right\}}. \tag{11}$$

Moreover, with probability at least $1 - e^{-\mathrm{x}}$*,*

$$f(Y) - \mathbb{E}\, f(Y) \le \sigma \sqrt{2\mathrm{x} \frac{n+1}{n}} + \frac{3}{4} K z \mathrm{x} \frac{n+1}{n}. \tag{12}$$

The proof is postponed to Section 4.3. The next corollary specifies the result for the case of $\mathcal{Y}$ being a normed space.

Corollary 3.5. *Let* $(\mathcal{Y}, \|\cdot\|)$ *be a normed space and assume* $Y_1, \dots, Y_n$ *are independent random variables. Set*

$$K := \max_{i \in [n]} \| \|Y_i\| \|_{\psi_\alpha}, \quad U^2 := \sum_{i \in [n]} \| \|Y_i\| \|_{\psi_\alpha}^2, \quad \sigma^2 := \sum_{i \in [n]} \mathbb{E} \|Y_i\|^2, \quad z := \left(4 \,\underline{\log} \frac{eU}{\sigma} \right)^{1/\alpha}$$

Then, with probability at least $1 - e^{-\mathrm{x}}$*, the bounds* (11) *and* (12) *hold for* $f(y) := \frac{1}{2} \|\sum_i y_i\|$*.*

The proof is in Section 4.3. For completeness, we provide the results by Maurer and Pontil [2021].

Proposition 3.6 (Theorem 4 in [Maurer and Pontil, 2021])**.** *In the setting of Corollary 3.4 it holds for* $\alpha = 1$*, with probability at least* $1 - e^{-\mathrm{x}}$*, that*

$$f(Y) - \mathbb{E}\, f(Y) \le 2eU\sqrt{\mathrm{x}} + 2eK\mathrm{x}.$$

Proposition 3.7 (Proposition 7 (i) in [Maurer and Pontil, 2021])**.** *In the setting of Corollary 3.5 it holds for* $\alpha = 1$*, with probability at least* $1 - e^{-\mathrm{x}}$*, that*

$$\left\| \sum_i Y_i \right\| - \mathbb{E} \left\| \sum_i Y_i \right\| \le 4eU\sqrt{\mathrm{x}} + 4eK\mathrm{x}.$$

Note that the bounds do not depend on σ. Specifically, U is used as a proxy for σ.

4 Proofs

4.1 Proof of Theorem 2.1

The proof relies on the Chernoff method for (super-)martingales introduced in Freedman [1975] and further generalized by Tropp [2011] to the matrix case. For the sake of completeness, let us provide it here with the proofs. We start with the following simple generalization of the Markov inequality to supermartingales.

Lemma 4.1. *Let $X_0, \dots, X_n$ be a non-negative supermartingale adapted to the filtration $(\mathrm{F}_i)_{i=0}^n$. Then for any $t > 0$*

$$\mathbb{P}\left\{\max_{i\in[n]} X_i \geq t\right\} \leq \frac{\mathbb{E}\, X_0}{t}.$$

Proof. Define the events

$$A_k := \{X_k \geq t\}, \quad B_k := \bigcup_{i=1}^{k} A_i, \quad C_k := A_k \setminus B_{k-1}$$

and the stopping time

$$\tau(\omega) := n \wedge \min\{k \in [n] : \omega \in A_k\}, \quad \omega \in \Omega$$

(with convention $\min \emptyset = \infty$). Since (a) $(X_i)_{i=0}^n$ is a supermartingale, (b) $C_1, \dots, C_n$ are disjoint, and (c) $X_\tau\, \mathbb{I}[C_k] = X_k\, \mathbb{I}[C_k] \geq t\, \mathbb{I}[C_k]$ by the definition of C_k,

$$\begin{aligned}\mathbb{E}\, X_0 &\overset{(a)}{\geq} \mathbb{E}\, X_\tau \overset{(b)}{\geq} \mathbb{E}\, X_\tau \sum_{k=1}^{n} \mathbb{I}[C_k] \overset{(c)}{=} \mathbb{E} \sum_{k=1}^{n} X_k\, \mathbb{I}[C_k] \\ &\overset{(c)}{\geq} \mathbb{E} \sum_{k=1}^{n} t\, \mathbb{I}[C_k] = t\mathbb{P}(B_n) = t\mathbb{P}\left\{\max_{i\in[n]} X_i \geq t\right\}.\end{aligned}$$

□

The next proposition is a master bound in the core of the proof. It is a simplified version of [Tropp, 2011, Theorem 2.3].

Proposition 4.2. *Let $(\mathbf{Y}_i)_{i=1}^n \subset \mathbb{H}(d)$ be a matrix-valued stochastic process adapted to the filtration $(\mathrm{F}_i)_{i=1}^n$. Define*

$$\mathbf{V}_i := \ln \mathbb{E}\left[e^{\mathbf{Y}_i} \middle| \mathrm{F}_{i-1}\right] \in \mathbb{H}(d), \quad i = 1, \dots, n,$$
$$\mathbf{Z}_k := \sum_{i=1}^{k} (\mathbf{Y}_i - \mathbf{V}_i), \quad k = 0, \dots, n.$$

Then for any $\mathbf{A} \in \mathbb{H}(d)$ it holds that

$$\mathbb{E}\left[\operatorname{tr}\exp\{\mathbf{Z}_k - \mathbf{A}\} | \mathrm{F}_{k-1}\right] \leq \operatorname{tr}\exp\{\mathbf{Z}_{k-1} - \mathbf{A}\} \quad \textit{a.s. for all } k \in [n], \tag{13}$$

and

$$\mathbb{P}\left\{\max_{k\in[n]} \lambda_{\max}(\mathbf{Z}_k - \mathbf{A}) \geq 0\right\} \leq \operatorname{tr} e^{-\mathbf{A}}.$$

Proof. By Lieb's theorem the function $\mathbf{X} \mapsto \operatorname{tr}\exp\{\mathbf{H} + \ln \mathbf{X}\}$ is concave on $\mathbb{H}_{++}(d)$ for any fixed $\mathbf{H} \in \mathbb{H}(d)$ [Lieb, 1973, Theorem 6], thus by Jensen's inequality for all $k \in [n]$

$$\begin{aligned}\mathbb{E}\left[\operatorname{tr}\exp\{\mathbf{Z}_k - \mathbf{A}\} | \mathrm{F}_{k-1}\right] &= \mathbb{E}\left[\operatorname{tr}\exp\left\{\mathbf{Z}_{k-1} - \mathbf{V}_k + \ln e^{\mathbf{Y}_k} - \mathbf{A}\right\} \middle| \mathrm{F}_{k-1}\right] \\ &\leq \operatorname{tr}\exp\left\{\mathbf{Z}_{k-1} - \mathbf{V}_k + \ln \mathbb{E}\left[e^{\mathbf{Y}_k} \middle| \mathrm{F}_{k-1}\right] - \mathbf{A}\right\} \\ &= \operatorname{tr}\exp\{\mathbf{Z}_{k-1} - \mathbf{A}\} \quad \text{a.s.}\end{aligned}$$

[see Tropp, 2011, Corollary 1.5]. Then by Lemma 4.1

$$\mathbb{P}\left\{\max_{k\in[n]} \operatorname{tr} e^{\mathbf{Z}_k-\mathbf{A}} \geq 1\right\} \leq \mathbb{E} \operatorname{tr} e^{\mathbf{Z}_0-\mathbf{A}} = \operatorname{tr} e^{-\mathbf{A}}.$$

Finally, if $\lambda_{\max}(\mathbf{Z}_k - \mathbf{A}) \geq 0$, then $\operatorname{tr} e^{\mathbf{Z}_k-\mathbf{A}} \geq 1$, thus the claim follows. □

In the next lemmata, we often use the following simple fact [see Freedman, 1975, Lemma 3.1].

Proposition 4.3. *Function $\frac{\phi(t)}{t^2}$, extended at 0 by continuity to $\frac{1}{2}$, is analytic and increasing on $\mathbb{R}$.*

Lemma 4.4. *Fix $\lambda > 0$, $\alpha > 0$. Define the function*

$$\rho_{\lambda,\alpha}(x) := \left(\phi(\lambda x) - \frac{(\lambda x)^2}{2}\right) \exp\{-x^\alpha\}. \tag{14}$$

Then for $x > 0$ it holds that

$$\operatorname{sign} \rho'_{\lambda,\alpha}(x) = \operatorname{sign}\left(\upsilon(\lambda x) - \alpha x^\alpha\right),$$

where

$$\upsilon(t) := \frac{t\phi(t)}{\phi(t) - t^2/2}. \tag{15}$$

We postpone the proof to the appendix.

Lemma 4.5. *Fix $\lambda, \alpha > 0$. Let $y > 0$ satisfy*

$$\upsilon(\lambda y) \leq \alpha y^\alpha. \tag{16}$$

If $\alpha < 1$, let $\tau > 0$ satisfy (16) *as well, i.e., $\upsilon(\lambda\tau) \leq \alpha\tau^\alpha$.*

Then for all $x \in \mathbb{R}$ in case $\alpha \geq 1$ and for all $x \leq \tau$ in case $\alpha < 1$ it holds that

$$\phi(\lambda x) \leq x^2 \frac{\phi(\lambda y)}{y^2} + \rho_{\lambda,\alpha}(y) \exp\left\{x_+^\alpha\right\} \mathbb{I}[x > y],$$

where $\rho_{\lambda,\alpha}(x)$ is defined by (14).

Proof. By the monotonicity of $\frac{\phi(t)}{t^2}$,

$$\phi(\lambda x)\, \mathbb{I}[x \leq y] \leq (\lambda x)^2 \frac{\phi(\lambda y)}{(\lambda y)^2}\, \mathbb{I}[x \leq y] = x^2 \frac{\phi(\lambda y)}{y^2}\, \mathbb{I}[x \leq y].$$

Case $\alpha \geq 1$. Consider $x \geq y$. The monotonicity of $\frac{\phi(t)}{t^2}$ yields that

$$0 < \frac{\upsilon(\lambda x)}{\lambda x} = \frac{1}{1 - \frac{(\lambda x)^2}{2\phi(\lambda x)}} \leq \frac{1}{1 - \frac{(\lambda y)^2}{2\phi(\lambda y)}} = \frac{\upsilon(\lambda y)}{\lambda y}.$$

Thus,

$$\upsilon(\lambda x) - \alpha x^\alpha \leq \frac{x}{y}\left(\upsilon(\lambda y) - \alpha y^\alpha\right) \leq 0.$$

Therefore, by Lemma 4.4, $\rho_{\lambda,\alpha}$ is decreasing on $[y,\infty)$. Since $\frac{\phi(\lambda y)}{(\lambda y)^2} \geq \lim_{t\to 0} \frac{\phi(t)}{t^2} = \frac{1}{2}$, we get

$$\begin{aligned}\phi(\lambda x)\,\mathbb{I}[x>y] &= \left(\phi(\lambda x) - \frac{(\lambda x)^2}{2}\right)\mathbb{I}[x>y] + \frac{(\lambda x)^2}{2}\,\mathbb{I}[x>y] \\ &= \rho_{\lambda,\alpha}(x)\exp\left\{x_+^\alpha\right\}\mathbb{I}[x>y] + \frac{(\lambda x)^2}{2}\,\mathbb{I}[x>y] \\ &\leq \rho_{\lambda,\alpha}(y)\exp\left\{x_+^\alpha\right\}\mathbb{I}[x>y] + x^2\frac{\phi(\lambda y)}{y^2}\,\mathbb{I}[x>y].\end{aligned}$$

Combining the above inequalities, we obtain the first result.

Case $0<\alpha<1$**.** If $y \geq \tau$, then we have a simple bound

$$\phi(\lambda x) \leq x^2\frac{\phi(\lambda\tau)}{\tau^2} \leq x^2\frac{\phi(\lambda y)}{y^2}, \quad x \leq \tau.$$

Now, if $0<y<\tau$, then, due to the convexity of $\upsilon(x)$ (Lemma B.2) and the concavity of x^α,

$$\upsilon(\lambda x) - \alpha x^\alpha \leq 0, \quad y \leq x \leq \tau.$$

Thus, $\rho_{\lambda,\alpha}$ is non-increasing on $[y,\tau]$, and the second claim follows the same way as the first one. □

The next lemma ensures a bound on a matrix moment-generating function.

Lemma 4.6. *Let* $\mathbf{X}\in\mathbb{H}(d)$ *be a random matrix such that*

$$\mathbb{E}\,\mathbf{X} \preccurlyeq 0, \quad \mathbb{E}\,\mathbf{X}^2 = \mathbf{\Sigma}, \quad \|\lambda_{\max}(\mathbf{X})_+\|_{\psi_\alpha} = u < +\infty,$$

for some $\alpha>0$.

Fix $\lambda>0$. *Let* $y>0$ *satisfy*

$$\upsilon(\lambda y) \leq \alpha\left(\frac{y}{u}\right)^\alpha. \tag{17}$$

If $\alpha<1$*, let* $\lambda_{\max}(\mathbf{X}) \leq \tau$ *a.s. for some* $\tau>0$ *satisfying* (17)*, i.e.* $\upsilon(\lambda\tau) \leq \alpha\left(\frac{\tau}{u}\right)^\alpha$. *Then*

$$\mathbb{E}\exp\{\lambda\mathbf{X}\} \preccurlyeq \mathbf{I} + \frac{\phi(\lambda y)}{y^2}\mathbf{\Sigma} + 2\left(\phi(\lambda y) - \frac{\lambda^2y^2}{2}\right)\exp\left\{-\left(\frac{y}{u}\right)^\alpha\right\}\mathbf{I}.$$

Proof. By rescaling, it is enough to consider the case $u=1$. First, we recall that the moment-generating function satisfies

$$\mathbb{E}\exp\{\lambda\mathbf{X}\} = \mathbf{I} + \lambda\,\mathbb{E}\,\mathbf{X} + \mathbb{E}\left(e^{\lambda\mathbf{X}} - \mathbf{I} - \lambda\mathbf{X}\right) \preccurlyeq \mathbf{I} + \mathbb{E}\,\phi(\lambda\mathbf{X}).$$

Further, we can apply Lemma 4.5 because its conditions are fulfilled,

$$\begin{aligned}\mathbb{E}\,\phi(\lambda\mathbf{X}) &\preccurlyeq \mathbb{E}\left(\frac{\phi(\lambda y)}{y^2}\mathbf{X}^2 + \rho_{\lambda,\alpha}(y)\exp\left\{\mathbf{X}_+^\alpha\right\}\right) \\ &\preccurlyeq \frac{\phi(\lambda y)}{y^2}\mathbf{\Sigma} + \rho_{\lambda,\alpha}(y)\,\mathbb{E}\exp\left\{\lambda_{\max}(\mathbf{X})_+^\alpha\right\}\mathbf{I} \\ &\preccurlyeq \frac{\phi(\lambda y)}{y^2}\mathbf{\Sigma} + 2\rho_{\lambda,\alpha}(y)\mathbf{I}\end{aligned}$$

By replacing $y \to \frac{y}{u}$, $\mathbf{\Sigma} \to \frac{1}{u^2}\mathbf{\Sigma}$ and $\lambda \to \lambda u$, we get the result. □

Lemma 4.7. *Let $X_1, \ldots, X_n$ be non-negative r.v. adapted to the filtration $(\mathrm{F}_i)_{i=1}^n$, $\alpha > 0$, and $U_i := \left\|X_i \mid \mathrm{F}_{i-1}\right\|_{\psi_\alpha} \in [0, \infty]$. Fix $U \geq K > 0$ and set*

$$E_n := \left\{ \sum_{i=1}^n U_i^2 \leq U^2,\ \max_{i \in [n]} U_i \leq K \right\}.$$

Then for any $\tau \geq K$

$$\mathbb{P}\left(\{\max_{i \in [n]} X_i \geq \tau\} \cap E_n \right) \leq 2 \left(\frac{4}{\alpha e} \right)^{\frac{2}{\alpha}} \frac{U^2}{\tau^2} e^{-\frac{1}{2}\left(\frac{\tau}{K}\right)^\alpha}. \tag{18}$$

Moreover, if

$$\tau \geq K \left(2\mathrm{x} + 2 \ln \left(\frac{4U}{K} \right) + \frac{4}{\alpha} \ln \left(\frac{4}{\alpha e} \right) \right)^{1/\alpha}, \tag{19}$$

then

$$\mathbb{P}\left(\{\max_i X_i \geq \tau\} \cap E_n \right) \leq e^{-\mathrm{x}}. \tag{20}$$

Proof. First, we derive the key ingredient of the lemma, a bound on the indicator function $\mathbb{I}[s \geq t]$ for any $t > 0$ and $s \geq 0$.

Lemma B.3 ensures $e^{t^\alpha} \geq \left(\frac{\alpha e}{4}\right)^{\frac{4}{\alpha}} t^4$. Thus, for any $s \geq t > 0$

$$e^{s^\alpha} \geq e^{t^\alpha} \geq \left(\frac{\alpha e}{4} \right)^{\frac{2}{\alpha}} t^2 e^{\frac{t^\alpha}{2}} \quad \Rightarrow \quad \left(\frac{4}{\alpha e} \right)^{\frac{2}{\alpha}} \frac{1}{t^2} e^{-\frac{t^\alpha}{2}} e^{s^\alpha} \geq 1.$$

This entails for any $t > 0$ and $s \geq 0$,

$$\mathbb{I}[s \geq t] \leq \left(\frac{4}{\alpha e} \right)^{\frac{2}{\alpha}} \frac{1}{t^2} e^{-\frac{t^\alpha}{2}} e^{s^\alpha}. \tag{21}$$

Now define auxiliary events

$$E_k := \left\{ \sum_{i=1}^k U_i^2 \leq U^2,\ \max_{i \in [k]} U_i \leq K \right\} \in \mathrm{F}_k, \quad k \in [n]. \tag{22}$$

Note that $\Omega =: E_0 \supset E_1 \supset \cdots \supset E_n \supset E_{n+1} := \varnothing$.

The union bound ensures

$$\mathbb{P}\left(\{\max_i X_i \geq \tau\} \cap E_n \right) \leq \sum_k \mathbb{P}\left(\{X_i \geq \tau\} \cap E_n \right) \leq \sum_i \mathbb{P}\left(\{X_i \geq \tau\} \cap E_i \right).$$

Now, we are to bound $\mathbb{P}\left(\{X_i \geq \tau\} \cap E_i \right)$. In the following bound w.l.o.g., we consider all $U_i > 0$ a.s.

Otherwise, one could consider instead of U_i the upper bound $U_i + \varepsilon$ and let $\varepsilon \to 0$. Notice that

$$\begin{aligned}
\mathbb{P}\left(\{X_i \geq \tau\} \cap E_i\right) &= \mathbb{E}\,\mathbb{I}\left[X_i \geq \tau\right] \cdot \mathbb{I}\left[E_i\right] \\
&\overset{\text{by (21)}}{\leq} \mathbb{E}\left(\frac{4}{\alpha e}\right)^{\frac{2}{\alpha}} \left(\frac{U_i}{\tau}\right)^2 e^{-\frac{1}{2}\left(\frac{\tau}{U_i}\right)^\alpha} e^{\left(\frac{X_i}{U_i}\right)^\alpha}\, \mathbb{I}\left[E_i\right] \\
&= \mathbb{E}\,\mathbb{E}\left[\left(\frac{4}{\alpha e}\right)^{\frac{2}{\alpha}} \left(\frac{U_i}{\tau}\right)^2 e^{-\frac{1}{2}\left(\frac{\tau}{U_i}\right)^\alpha} e^{\left(\frac{X_i}{U_i}\right)^\alpha}\, \mathbb{I}\left[E_i\right] \middle| \mathrm{F}_{k-1}\right] \\
&= \mathbb{E}\left(\left(\frac{4}{\alpha e}\right)^{\frac{2}{\alpha}} \left(\frac{U_i}{\tau}\right)^2 e^{-\frac{1}{2}\left(\frac{\tau}{U_i}\right)^\alpha}\, \mathbb{I}\left[E_i\right] \cdot \mathbb{E}\left[e^{\left(\frac{X_i}{U_i}\right)^\alpha} \middle| \mathrm{F}_{k-1}\right]\right) \\
&\leq \left(\frac{4}{\alpha e}\right)^{\frac{2}{\alpha}} \frac{2}{\tau^2} e^{-\frac{1}{2}\left(\frac{\tau}{K}\right)^\alpha}\, \mathbb{E}\, U_i^2\, \mathbb{I}\left[E_i\right]. \qquad (23)
\end{aligned}$$

The last inequality holds because $U_i \leq K$ on E_i and $\mathbb{E}\left[\exp\left\{\left(\frac{X_i}{U_i}\right)^\alpha\right\} \middle| \mathrm{F}_{k-1}\right] \leq 2.$

Now, we consider

$$\begin{aligned}
\sum_k \mathbb{E}\, U_k^2\, \mathbb{I}\left[E_k\right] &= \mathbb{E} \sum_{k=1}^n U_k^2 \sum_{i=k}^n \left(\mathbb{I}\left[E_i\right] - \mathbb{I}\left[E_{i+1}\right]\right) \\
&= \mathbb{E} \sum_{i=1}^n \left[\left(\mathbb{I}\left[E_i\right] - \mathbb{I}\left[E_{i+1}\right]\right) \sum_{k=1}^i U_k^2\right] \\
&\overset{(a)}{\leq} U^2\, \mathbb{E} \sum_{i=1}^n \left(\mathbb{I}\left[E_i\right] - \mathbb{I}\left[E_{i+1}\right]\right) = U^2\, \mathbb{E}\,\mathbb{I}\left[E_1\right] \leq U^2, \qquad (24)
\end{aligned}$$

where (a) holds because on any event E_k it holds $\sum_{i=1}^k U_i^2 \leq U^2$. Combining (23) and (24), we get the first result (18).

To get (20), one has to find τ s.t.

$$-\mathrm{x} \geq \frac{2}{\alpha} \ln\left(\frac{4}{\alpha e}\right) + \ln(2) + 2\ln\left(\frac{U}{\tau}\right) - \frac{1}{2}\left(\frac{\tau}{K}\right)^\alpha.$$

Thus,

$$\left(\frac{\tau}{K}\right)^\alpha - 4\ln\left(\frac{U}{\tau}\right) \geq 2\mathrm{x} + \ln 4 + \frac{4}{\alpha} \ln\left(\frac{4}{\alpha e}\right)$$

This is equivalent to

$$\left(\frac{\tau}{K}\right)^\alpha + 4\ln\left(\frac{\tau}{K}\right) - 4\ln\left(\frac{U}{K}\right) \geq 2\mathrm{x} + \ln 4 + \frac{4}{\alpha} \ln\left(\frac{4}{\alpha e}\right).$$

As $\ln 4 > 1, U \geq K$, and $\tau \geq K$, the second claim follows. □

Now, we are ready to prove the main result.

Proof of Theorem 2.1. Define

$$\tau := \begin{cases} K \max\left\{z,\ \left(2\mathrm{x} + 2\ln\left(\frac{4U}{K}\right) + \frac{4}{\alpha}\ln\left(\frac{4}{\alpha e}\right)\right)^{1/\alpha}\right\}, & \alpha < 1, \\ Kz, & \alpha \geq 1. \end{cases} \qquad (25)$$

Let E_p be the event, where (5) holds. Set $Y_i \coloneqq U_i z \le Kz$ on E_p. Let

$$\lambda_0 \coloneqq \frac{2\alpha}{3K}\left(\frac{\tau}{K}\right)^{\alpha-1} \le \frac{2\alpha}{3K} z^{\alpha-1}, \tag{26}$$

and consider $0 \le \lambda \le \lambda_0$. Then by Lemma B.2

$$\upsilon(\lambda Y_i) \le \upsilon(\lambda K z) < \min\{4, 1.5\lambda K z\} \le \alpha z^\alpha = \alpha\left(\frac{Y_i}{U_i}\right)^\alpha \text{ on } E_p.$$

Now, we set

$$\widetilde{\mathbf{X}}_i \coloneqq \begin{cases} \mathbf{X}_i, & \text{if } \alpha \ge 1,\\ \mathbf{X}_i \, \mathbb{I}_{(-\infty,\tau]}(\mathbf{X}_i), & \text{otherwise.}\end{cases}$$

By construction, for all $i \in [n]$ one has $\widetilde{\mathbf{X}}_i \preccurlyeq \mathbf{X}_i$ and $\widetilde{\mathbf{X}}_i^2 \preccurlyeq \mathbf{X}_i^2$, thus

$$\mathbb{E}\,\widetilde{\mathbf{X}}_i^2 \preccurlyeq \mathbb{E}\,\mathbf{X}_i^2, \quad \left\|\lambda_{\max}(\widetilde{\mathbf{X}}_i)_+\right\|_{\psi_\alpha} \le U_i.$$

Moreover, if $\alpha < 1$, then $\lambda_{\max}(\widetilde{\mathbf{X}}_i)_+ \le \tau$ by construction.

Denote

$$\widetilde{\mathbf{S}}_k = \sum_{i=1}^k \widetilde{\mathbf{X}}_i.$$

By Lemma 4.6, using the monotonicity of $\frac{\phi(t)}{t^2}$ and the fact that $\ln(\mathbf{X})$ is a monotone map on the cone of positive-definite matrices (see, e.g., (2.8) in Tropp [2012]), we obtain

$$\begin{aligned}
\mathbf{V}_i(\lambda) &\coloneqq \ln \mathbb{E}\left[\exp\left\{\lambda \widetilde{\mathbf{X}}_i\right\}\middle|\mathrm{F}_{i-1}\right]\\
&\preccurlyeq \frac{\phi(\lambda Y_i)}{Y_i^2}\boldsymbol{\Sigma}_i + 2\left(\phi(\lambda Y_i) - \frac{\lambda^2 Y_i^2}{2}\right)\exp\left\{-\left(\frac{Y_i}{U_i}\right)^\alpha\right\}\mathbf{I}\\
&= \frac{\phi(\lambda Y_i)}{Y_i^2}\boldsymbol{\Sigma}_i + 2U_i^2 z^2\left(\frac{\phi(\lambda Y_i)}{Y_i^2} - \frac{\lambda^2}{2}\right)\exp\{-z^\alpha\}\mathbf{I}\\
&\preccurlyeq \frac{\phi(\lambda K z)}{(Kz)^2}\boldsymbol{\Sigma}_i + 2\frac{U_i^2}{K^2}\left(\phi(\lambda K z) - \frac{(\lambda K z)^2}{2}\right)\exp\{-z^\alpha\}\mathbf{I} \text{ on } E_p
\end{aligned}$$

The last inequality is due to $Y_i = U_i z \le Kz$ on E_p.

Lemma B.4 ensures for all $\exp\{z^\alpha\} \ge \frac{e^4}{16}\left(\frac{Uz}{\sigma}\right)^2$. Therefore, for all $k \in [n]$

$$\begin{aligned}
\sum_{i=1}^k \mathbf{V}_i(\lambda) &\preccurlyeq \boldsymbol{\Sigma}\frac{\phi(\lambda K z)}{(Kz)^2} + 2\frac{U^2}{K^2}\left(\phi(\lambda K z) - \frac{(\lambda K z)^2}{2}\right)\exp\{-z^\alpha\}\mathbf{I}\\
&\preccurlyeq \left(\frac{\sigma}{Kz}\right)^2\left(\phi(\lambda K z) + \frac{2}{3}\left(\phi(\lambda K z) - \frac{(\lambda K z)^2}{2}\right)\right)\mathbf{I} \text{ on } E_p.
\end{aligned} \tag{27}$$

Note that for any $a > 0$

$$\phi(t) + a\left(\phi(t) - \frac{t^2}{2}\right) = \frac{t^2}{2} + (1+a)\sum_{k=3}^\infty \frac{t^k}{k!} \le \frac{1}{(1+a)^2}\sum_{k=2}^\infty \frac{((1+a)t)^k}{k!} = \frac{\phi((1+a)t)}{(1+a)^2},$$

thus for all $k \in [n]$

$$\sum_{i=1}^{k} \mathbf{V}_i(\lambda) \preccurlyeq \left(\frac{\sigma}{M}\right)^2 \phi(\lambda M)\mathbf{I} \text{ on } E_p, \text{ where } M := \frac{5}{3}Kz. \tag{28}$$

Proposition 4.2 with $\mathbf{Y}_i = \lambda\widetilde{\mathbf{X}}_i$ and $\mathbf{A} = \Big(\lambda t - \left(\frac{\sigma}{M}\right)^2 \phi(\lambda M)\Big)\,\mathbf{I}$ yields that

$$\begin{aligned}\mathbb{P}\left\{\max_{k\in[n]} \lambda_{\max}\left(\lambda\widetilde{\mathbf{S}}_k - \sum_{i=1}^{k} \mathbf{V}_i(\lambda) - \mathbf{A}\right) \geq 0\right\} &\leq \operatorname{tr}\exp\left\{-\mathbf{A}\right\} \\ &= d\exp\left\{\left(\frac{\sigma}{M}\right)^2 \phi(\lambda M) - \lambda t\right\}.\end{aligned}$$

Note that

$$\lambda_{\max}(\lambda\widetilde{\mathbf{S}}_k) \leq \lambda_{\max}\left(\lambda\widetilde{\mathbf{S}}_k - \sum_{i=1}^{k} \mathbf{V}_i(\lambda) - \mathbf{A}\right) + \lambda_{\max}\left(\sum_{i=1}^{k} \mathbf{V}_i(\lambda) + \mathbf{A}\right).$$

Moreover, (28) yields on E_p

$$\max_{k\in[n]} \lambda_{\max}\left(\sum_{i=1}^{k} \mathbf{V}_i(\lambda) + \mathbf{A}\right) \leq \lambda t,$$

thus

$$\mathbb{P}\left(\left\{\max_{k\in[n]} \lambda_{\max}(\widetilde{\mathbf{S}}_k) \geq t\right\} \cap E_p\right) \leq d\exp\left\{\left(\frac{\sigma}{M}\right)^2 \phi\left(\lambda M\right) - \lambda t\right\}. \tag{29}$$

If $\alpha \geq 1$, $\widetilde{\mathbf{X}}_i = \mathbf{X}_i$, and we immediately get $\widetilde{\mathbf{S}}_k = \mathbf{S}_k$. Thus,

$$\mathbb{P}\left\{\max_{k\in[n]} \lambda_{\max}(\mathbf{S}_k) \geq t\right\} \leq \mathbb{P}\{E_p\} + d\exp\left\{\left(\frac{\sigma}{M}\right)^2 \phi\left(\lambda M\right) - \lambda t\right\}.$$

Consider $0 < \alpha < 1$. To get the bound on $\mathbf{S}_k$, one has to estimate the probability that $\widetilde{\mathbf{X}}_i = \mathbf{X}_i$ for all i and the event E_p is true. Note that, by construction, $\widetilde{\mathbf{X}}_i \neq \mathbf{X}_i$ iff $\lambda_{\max}(\mathbf{X}_i) > \tau$. Recall that $E_p \subset E_n$ with E_n coming from (22). Thus, due to the choice of τ (25), by Lemma 4.7

$$\begin{aligned}\mathbb{P}\left(\{\exists k \in [n] : \widetilde{\mathbf{S}}_k \neq \mathbf{S}_k\} \cap E_p\right) &= \mathbb{P}\left(\{\exists i \in [n] : \widetilde{\mathbf{X}}_i \neq \mathbf{X}_i\} \cap E_p\right) \\ &= \mathbb{P}\left(\{\exists i \in [n] : \lambda_{\max}(\mathbf{X}_i) > \tau\} \cap E_p\right) \\ &\leq \mathbb{P}\left(\{\max_{i\in[n]} \lambda_{\max}(\mathbf{X}_i)_+ > \tau\} \cap E_n\right) \leq e^{-\mathrm{x}}.\end{aligned}$$

Combining this bound with (29), we get for $\alpha < 1$

$$\mathbb{P}\left\{\max_{k\in[n]} \lambda_{\max}(\mathbf{S}_k) \geq t\right\} \leq \mathbb{P}\{E_p\} + e^{-\mathrm{x}} + d\exp\left\{\left(\frac{\sigma}{M}\right)^2 \phi\left(\lambda M\right) - \lambda t\right\}$$

Optimization over λ. We have to minimize $\exp\left\{\left(\frac{\sigma}{M}\right)^2 \phi\left(\lambda M\right) - \lambda t\right\}$ w.r.t. λ. Let $\xi_0 := \lambda_0 M$, then

$$\min_{0 \le \lambda \le \lambda_0} \left(\frac{\sigma}{M}\right)^2 \phi\left(\lambda M\right) - \lambda t = \left(\frac{\sigma}{M}\right)^2 \min_{0 \le \xi \le \xi_0} \phi(\xi) - \left(\frac{M}{\sigma}\right)^2 \frac{\xi}{M} t = -\left(\frac{\sigma}{M}\right)^2 g_{\xi_0}\left(\frac{Mt}{\sigma^2}\right),$$

with g_{ξ_0} coming from Lemma B.7.

According to the previous bounds, it is enough to find $t = t(\mathrm{x})$, s.t.

$$\left(\frac{\sigma}{M}\right)^2 g_{\xi_0}\left(\frac{Mt(\mathrm{x})}{\sigma^2}\right) \ge \mathrm{x},$$

that is equivalent to

$$t(\mathrm{x}) \ge \frac{\sigma^2}{M} g_{\xi_0}^{-1}\left(\frac{M^2 \mathrm{x}}{\sigma^2}\right).$$

In view of Lemma B.7, we choose

$$t(\mathrm{x}) := \begin{cases} \frac{\sigma^2}{M} h^{-1}\left(\frac{M^2 \mathrm{x}}{\sigma^2}\right), & \text{if } \mathrm{x} \le \mathrm{x}_0, \\ \frac{2}{\lambda_0} \mathrm{x}, & \text{if } \mathrm{x} > \mathrm{x}_0, \end{cases}$$

with $\mathrm{x}_0 := \xi_0 \phi'(\xi_0) - \phi(\xi_0)$.

Substituting λ_0 defined in (26), we get for $\mathrm{x} > \mathrm{x}_0$

$$t(\mathrm{x}) = \frac{3K}{\alpha}\left(\frac{\tau}{K}\right)^{1-\alpha} \mathrm{x} \le \frac{3Kz}{\alpha z^\alpha} + \frac{3K}{\alpha}\left(2\mathrm{x} + 2\ln\left(\frac{4U}{K}\right) + \frac{4}{\alpha}\ln\left(\frac{4}{\alpha e}\right)\right)^{\frac{1-\alpha}{\alpha}} \mathbb{I}[\alpha < 1]. \quad (30)$$

If $\mathrm{x} \le \mathrm{x}_0$, the bound (35) on $h^{-1}(\cdot)$ yields

$$\begin{aligned} t(\mathrm{x}) &\le \frac{\sigma^2}{M}\left(\sqrt{2\left(\frac{M}{\sigma}\right)^2 \mathrm{x}} + \frac{2\left(\frac{M}{\sigma}\right)^2 \mathrm{x}}{\underline{\log}\left(2\left(\frac{M}{\sigma}\right)^2 \mathrm{x}\right)}\right) = \sigma\sqrt{2\mathrm{x}} + \frac{2M\mathrm{x}}{\underline{\log}\left(2\left(\frac{M}{\sigma}\right)^2 \mathrm{x}\right)} \\ &= \sigma\sqrt{2\mathrm{x}} + \frac{\frac{10}{3} Kz\mathrm{x}}{\underline{\log}\left(2\left(\frac{5Kz}{3\sigma}\right)^2 \mathrm{x}\right)} \le \sigma\sqrt{2\mathrm{x}} + \frac{4Kz\mathrm{x}}{\underline{\log}\left(\left(\frac{Kz}{\sigma}\right)^2 \mathrm{x}\right)}, \end{aligned}$$

where the last inequality holds due to $M = 5/3Kz$. We get (Ben) by combining this bound with (30).

Finally, let us prove (Ber). First, we consider the case $\mathrm{x} \le \mathrm{x}_0$ and apply the well-known bound on $h^{-1}(\cdot)$,

$$h^{-1}(u) \le \sqrt{2u} + \frac{u}{3}.$$

This yields for $\mathrm{x} \le \mathrm{x}_0$

$$t(\mathrm{x}) \le \sigma\sqrt{2\mathrm{x}} + \frac{M}{3}\mathrm{x} = \sigma\sqrt{2\mathrm{x}} + \frac{5Kz}{9}\mathrm{x}.$$

Combining this bound with (30) for the case $\mathrm{x} > \mathrm{x}_0$ and using the inequality $\alpha z^\alpha \ge 4$, we get (Ber). □

4.2 Proof of Corollary 3.3

Lemma 4.8. *Under assumptions of Corollary 3.3, for any* $\mathrm{x} > 0$ *and* $n \geq 2\mathrm{x}$ *it holds, with probability at least* $1 - 3de^{-\mathrm{x}}$*, that*

$$\sigma \leq \left(\hat{\sigma} + \left\|\overline{\mathbf{X}} - \mathbb{E}\,\mathbf{X}_1\right\|\right)\left(1 + \frac{2\mathrm{x}}{3n}\right) + \frac{4}{3}K\hat{z}\sqrt{\frac{\mathrm{x}}{n}}.$$

Proof. The statement is trivial if $\sigma \leq \hat{\sigma}$. Now, consider the case $\sigma > \hat{\sigma}$. In the following, we will construct the bound of the form $\sigma \leq (\hat{\sigma} + \left\|\overline{\mathbf{X}} - \mathbb{E}\,\mathbf{X}_1\right\|)(1 + C_1\frac{\mathrm{x}}{n}) + C_2\sqrt{\frac{\mathrm{x}}{n}}$, with $C_1, C_2 > 0$ being some constants.

To construct this bound, we will use the square-root trick. Let $\boldsymbol{Q}_i := -(\mathbf{X}_i - \mathbb{E}\,\mathbf{X}_i)^2$. First, we notice that

$$\hat{\mathbf{\Sigma}} = \frac{1}{n}\sum(\mathbf{X}_i - \mathbb{E}\,\mathbf{X}_1)^2 - (\overline{\mathbf{X}} - \mathbb{E}\,\mathbf{X}_1)^2 = -\frac{1}{n}\sum_i \boldsymbol{Q}_i - (\overline{\mathbf{X}} - \mathbb{E}\,\mathbf{X}_1)^2.$$

Thus,

$$\mathbf{\Sigma} = -\,\mathbb{E}\,\boldsymbol{Q}_1 = \hat{\mathbf{\Sigma}} + \frac{1}{n}\sum_i \boldsymbol{Q}_i - \mathbb{E}\,\boldsymbol{Q}_1 + (\overline{\mathbf{X}} - \mathbb{E}\,\mathbf{X}_1)^2.$$

This yields

$$\begin{aligned}\sigma^2 := \lambda_{\max}(\mathbf{\Sigma}) &\leq \lambda_{\max}(\hat{\mathbf{\Sigma}}) + \lambda_{\max}\left(\frac{1}{n}\sum_i \boldsymbol{Q}_i - \mathbb{E}\,\boldsymbol{Q}_1\right) + \lambda_{\max}\left((\overline{\mathbf{X}} - \mathbb{E}\,\mathbf{X}_1)^2\right) \\ &= \hat{\sigma}^2 + \lambda_{\max}\left(\frac{1}{n}\sum_i \boldsymbol{Q}_i - \mathbb{E}\,\boldsymbol{Q}_1\right) + \left\|\overline{\mathbf{X}} - \mathbb{E}\,\mathbf{X}_1\right\|^2. \qquad (31)\end{aligned}$$

Now we have to bound $\lambda_{\max}\left(\frac{1}{n}\sum_i \boldsymbol{Q}_i - \mathbb{E}\,\boldsymbol{Q}_1\right)$. We will use Theorem 1.4 from Tropp [2012]. Its conditions are fulfilled, because all $\boldsymbol{Q}_i$ are i.i.d., $\boldsymbol{Q}_i \preccurlyeq 0$, and $\mathbb{E}\,\boldsymbol{Q}_i = -\mathbf{\Sigma}$, thus

$$\lambda_{\max}\left(\boldsymbol{Q}_i - \mathbb{E}\,\boldsymbol{Q}_1\right) \leq \lambda_{\max}(\mathbf{\Sigma}) := \sigma^2.$$

Theorem 1.4 in [Tropp, 2012] ensures that, with probability at least $1 - de^{-\mathrm{x}}$,

$$\begin{aligned}\lambda_{\max}\left(\frac{1}{n}\sum_i \boldsymbol{Q}_i - \mathbb{E}\,\boldsymbol{Q}_1\right) &\leq \sqrt{2\lambda_{\max}\left(\mathbb{E}\left(\boldsymbol{Q}_1 - \mathbb{E}\,\boldsymbol{Q}_1\right)^2\right)\frac{\mathrm{x}}{n}} + \frac{2}{3}\sigma^2\frac{\mathrm{x}}{n} \\ &\leq 2\sigma K z\sqrt{\frac{\mathrm{x}}{n}} + \frac{2}{3}\sigma^2\frac{\mathrm{x}}{n}. \qquad (32)\end{aligned}$$

The last inequality follows from Lemma B.5 that ensures the bound

$$\mathbb{E}(\boldsymbol{Q}_1 - \mathbb{E}\,\boldsymbol{Q}_1)^2 = \mathbb{E}\,\boldsymbol{Q}_1^2 - (\mathbb{E}\,\boldsymbol{Q}_1)^2 \preccurlyeq \mathbb{E}(\mathbf{X}_1 - \mathbb{E}\,\mathbf{X}_1)^4 \preccurlyeq \frac{5}{3}(\sigma K z)^2\mathbf{I}.$$

Combining (31) and (32), we get

$$\sigma^2 \leq \hat{\sigma}^2 + \left\|\overline{\mathbf{X}} - \mathbb{E}\,\mathbf{X}_1\right\|^2 + 2\sigma K z\sqrt{\frac{\mathrm{x}}{n}} + \frac{2}{3}\sigma^2\frac{\mathrm{x}}{n}.$$

thus

$$\left(1 - \frac{2\mathrm{x}}{3n}\right)\sigma^2 \le \hat{\sigma}^2 + \left\|\overline{\mathbf{X}} - \mathbb{E}\,\mathbf{X}_1\right\|^2 + 2\sigma K z\sqrt{\frac{\mathrm{x}}{n}}.$$

Using a bound on the roots of a square inequality w.r.t. σ, we get

$$\begin{aligned}\sigma &\le \sqrt{\frac{1}{1-\frac{2\mathrm{x}}{3n}}\left(\hat{\sigma}^2 + \left\|\overline{\mathbf{X}} - \mathbb{E}\,\mathbf{X}_1\right\|^2\right)} + \frac{2Kz}{1-\frac{2\mathrm{x}}{3n}}\sqrt{\frac{\mathrm{x}}{n}} \\ &\le \frac{\hat{\sigma} + \left\|\overline{\mathbf{X}} - \mathbb{E}\,\mathbf{X}_1\right\|}{\sqrt{1-\frac{2\mathrm{x}}{3n}}} + \frac{2Kz}{1-\frac{2\mathrm{x}}{3n}}\sqrt{\frac{\mathrm{x}}{n}}.\end{aligned}$$

Lemma's condition ensures $\frac{2\mathrm{x}}{3n} \le \frac{1}{3}$, thus

$$\sqrt{\frac{1}{1-\frac{2\mathrm{x}}{3n}}} \le 1 + \frac{2\mathrm{x}}{3n} \quad \text{and} \quad \frac{1}{1-\frac{2\mathrm{x}}{3n}} \le \frac{3}{2}.$$

Then

$$\sigma \le \left(\hat{\sigma} + \left\|\overline{\mathbf{X}} - \mathbb{E}\,\mathbf{X}_1\right\|\right)\left(1 + \frac{2\mathrm{x}}{3n}\right) + \frac{4}{3}Kz\sqrt{\frac{\mathrm{x}}{n}}$$

Finally, $\hat{\sigma} \le \sigma$ yields that $z \le \hat{z}$. Thus, we get the result. □

Proof of Corollary 3.3. To bound $\left\|\overline{\mathbf{X}} - \mathbb{E}\,\mathbf{X}_1\right\|$, we use (in two sides) Corollary 3.1 and Remark 2.2. This yields that, with probability at least $1 - 2de^{-\mathrm{x}}$,

$$\left\|\overline{\mathbf{X}} - \mathbb{E}\,\mathbf{X}_1\right\| \le \inf_{\sigma' \ge \sigma}\left\{\sigma'\sqrt{2\frac{\mathrm{x}}{n}} + \frac{3}{4}Kz(K, \sigma'; \alpha)\frac{\mathrm{x}}{n}\right\}. \tag{33}$$

In the case $\sigma \le \hat{\sigma}$ we immediately obtain

$$\left\|\overline{\mathbf{X}} - \mathbb{E}\,\mathbf{X}_1\right\| \le \hat{\sigma}\sqrt{2\frac{\mathrm{x}}{n}} + \frac{3}{4}K\hat{z}\frac{\mathrm{x}}{n}.$$

This ensures the result.

If $\sigma \ge \hat{\sigma}$, it holds $z \le \hat{z}$. Moreover, we notice that because of Lemma B.3

$$\begin{aligned}\sigma^2 &= \lambda_{\max}(\mathbf{\Sigma}) = \lambda_{\max}\left(\mathbb{E}(\mathbf{X}_1 - \mathbb{E}\,X_1)^2\right) \\ &\le \mathbb{E}\|\mathbf{X}_1 - \mathbb{E}\,\mathbf{X}_1\|^2 \le 2\left(\frac{2}{\alpha e}\right)^{2/\alpha}\|\|\mathbf{X}_1 - \mathbb{E}\,\mathbf{X}_1\|\|_{\psi_\alpha}^2 \le 2K^2.\end{aligned}$$

This ensures $\hat{\sigma} \le \sqrt{2}K \le \sqrt{2}K\hat{z}$.

Now, we use Lemma 4.8. Keeping in mind Lemma's condition $\frac{\mathrm{x}}{n} \le \frac{1}{8}$, we get with probability at least $1 - 3de^{-\mathrm{x}}$

$$\begin{aligned}\sigma &\le \left(1 + \frac{2\mathrm{x}}{3n}\right)\left(\hat{\sigma} + \left\|\overline{\mathbf{X}} - \mathbb{E}\,\mathbf{X}_1\right\|\right) + \frac{4}{3}K\hat{z}\sqrt{\frac{\mathrm{x}}{n}} \\ &\le \hat{\sigma} + \sqrt{2}K\hat{z}\cdot\frac{2\mathrm{x}}{3n} + \frac{13}{12}\left\|\overline{\mathbf{X}} - \mathbb{E}\,\mathbf{X}_1\right\| + \frac{4}{3}K\hat{z}\sqrt{\frac{\mathrm{x}}{n}} \\ &\le \hat{\sigma} + \frac{13}{12}\left\|\overline{\mathbf{X}} - \mathbb{E}\,\mathbf{X}_1\right\| + \left(\frac{4}{3} + \frac{2}{3}\sqrt{2\frac{\mathrm{x}}{n}}\right)K\hat{z}\sqrt{\frac{\mathrm{x}}{n}} \\ &\le \hat{\sigma} + \frac{13}{12}\left\|\overline{\mathbf{X}} - \mathbb{E}\,\mathbf{X}_1\right\| + \frac{5}{3}K\hat{z}\sqrt{\frac{\mathrm{x}}{n}}\end{aligned}$$

Now, using the above result and (Ber), we get

$$\begin{aligned}
\left\|\overline{\mathbf{X}} - \mathbb{E}\,\mathbf{X}_1\right\| &\le \sigma\sqrt{2\frac{\mathrm{x}}{n}} + \frac{3}{4}Kz\frac{\mathrm{x}}{n} \\
&\le \hat{\sigma}\sqrt{2\frac{\mathrm{x}}{n}} + \left\|\overline{\mathbf{X}} - \mathbb{E}\,\mathbf{X}_1\right\| \cdot \frac{13}{12} \cdot \sqrt{2\frac{\mathrm{x}}{n}} + \frac{5\sqrt{2}}{3}K\hat{z}\frac{\mathrm{x}}{n} + \frac{3}{4}K\hat{z}\frac{\mathrm{x}}{n} \\
&\le \hat{\sigma}\sqrt{2\frac{\mathrm{x}}{n}} + \left\|\overline{\mathbf{X}} - \mathbb{E}\,\mathbf{X}_1\right\| \cdot \frac{13}{12} \cdot \sqrt{2\frac{\mathrm{x}}{n}} + \frac{20\sqrt{2}+9}{12}K\hat{z}\frac{\mathrm{x}}{n} \\
&\le \hat{\sigma}\sqrt{2\frac{\mathrm{x}}{n}} + \left\|\overline{\mathbf{X}} - \mathbb{E}\,\mathbf{X}_1\right\| \cdot \frac{13}{12} \cdot \sqrt{2\frac{\mathrm{x}}{n}} + \frac{7}{2}K\hat{z}\frac{\mathrm{x}}{n}
\end{aligned}$$

Thus,

$$\left(1 - \frac{13}{12}\sqrt{2\frac{\mathrm{x}}{n}}\right)\left\|\overline{\mathbf{X}} - \mathbb{E}\,\mathbf{X}_1\right\| \le \hat{\sigma}\sqrt{2\frac{\mathrm{x}}{n}} + \frac{7}{2}K\hat{z}\frac{\mathrm{x}}{n}.$$

Further, we recall that $n \ge 8\mathrm{x}$ and let $t \coloneqq \sqrt{2\mathrm{x}/n} \le \frac{1}{2}$.

$$\left(1 - \frac{13}{12}t\right)^{-1} = 1 + \frac{\frac{13}{12}t}{1 - \frac{13}{12}t} \le 1 + \frac{\frac{13}{12}}{1 - \frac{13}{24}}t \le 1 + \frac{26}{11}t \le \frac{24}{11}$$

Thus, we get

$$\begin{aligned}
\left\|\overline{\mathbf{X}} - \mathbb{E}\,\mathbf{X}_1\right\| &\le \left(1 + \frac{26}{11}\sqrt{2\frac{\mathrm{x}}{n}}\right)\hat{\sigma}\sqrt{2\frac{\mathrm{x}}{n}} + \frac{24}{11}\cdot\frac{7}{2}K\hat{z}\frac{\mathrm{x}}{n} \\
&\le \hat{\sigma}\sqrt{2\frac{\mathrm{x}}{n}} + \frac{52}{11}\cdot\sqrt{2}K\hat{z}\cdot\frac{\mathrm{x}}{n} + \frac{84}{11}K\hat{z}\frac{\mathrm{x}}{n} \\
&\le \hat{\sigma}\sqrt{2\frac{\mathrm{x}}{n}} + 15K\hat{z}\frac{\mathrm{x}}{n}.
\end{aligned}$$

□

4.3 Proofs of Corollaries 3.4 and 3.5

Let $g(x_1, \dots, x_n)$ be a measurable function on $\mathcal{X}^n$, where $\mathcal{X}$ is a measurable space.

Let $I \subset [n]$ and $\overline{I} \coloneqq [n] \setminus I$. We denote

$$g(x_I, y_{\overline{I}}) \coloneqq g(z), \quad z_i \coloneqq \begin{cases} x_i & \text{if } i \in I, \\ y_i & \text{if } i \in \overline{I}. \end{cases}$$

Let $\mathit{\Pi}_n$ be the set of all permutations of $[n]$, $\pi \in \mathit{\Pi}_n \colon [n] \to [n]$. We denote

$$\pi(I) \coloneqq \{\pi(i) : i \in I\}, \quad I \subset [n].$$

Lemma 4.9. *Let $Y = (Y_1, \dots, Y_n)$ be a set of i.i.d. random variables on a measurable space $\mathcal{Y}$. Let $g_i \colon \mathcal{Y}^n \to \mathbb{R}_+$, $i \in [n]$, be measurable and integrable functions, such that each $g_i(y_1, \dots, y_n)$ does not depend on y_i and*

$$\sum_{i \in [n]} g_i(Y) \le M \;\; \text{a.s.}$$

Let $I \subset [n]$ and define

$$\mathrm{F}_I := \sigma\left(\{Y_i : i \in I\}\right).$$

Let Π_n be the set of permutations of $[n]$. Then it holds that

$$\frac{1}{n!}\sum_{\pi\in\Pi_n}\sum_{i=1}^{n}\mathbb{E}[g_{\pi(i)}(Y)|\mathrm{F}_{\pi([1,i-1])}] \le \frac{n+1}{n}M \text{ a.s.}$$

The proof is postponed to the Appendix C. Now, we are ready to prove the corollaries.

Proof of Corollary 3.4. We set

$$X_i := \mathbb{E}[f_i(Y)|\mathrm{F}_i] = \mathbb{E}[f(Y)|\mathrm{F}_i] - \mathbb{E}[f(Y)|\mathrm{F}_{i-1}],$$

so that

$$\sum_{i=1}^{n} X_i = f(Y) - \mathbb{E}\, f(Y).$$

Let $\lambda > 0$. We consider an arbitrary permutation π and set

$$V_i^{\pi}(\lambda) := \ln \mathbb{E}\left[\exp\left\{\lambda f(Y) - \mathbb{E}\left[f(Y)|\mathrm{F}_{\pi(i)}\right]\right\}|\mathrm{F}_{\pi([1,i-1])}\right].$$

By Proposition 4.2 it holds

$$\mathbb{E}\exp\left\{\lambda f(Y) - \sum_i V_i^{\pi}(\lambda)\right\} \le 1.$$

Thus, Jensen's inequality yields

$$\mathbb{E}\exp\left\{\lambda f(Y) - \frac{1}{n!}\sum_{\pi}\sum_{i=1}^{n} V_i^{\pi}(\lambda)\right\} \le \frac{1}{n!}\sum_{\pi}\mathbb{E}\exp\left\{\lambda f(y) - \sum_{i=1}^{n} V_i^{\pi}(\lambda)\right\} \le 1.$$

Moreover, (27) ensures

$$\sum_i V_i^{\pi}(\lambda) \le \frac{\phi(\lambda K z)}{(Kz)^2}\sum_i \mathbb{E}\left[\sigma^2_{\pi(i)}\middle|\mathrm{F}_{\pi(i-1)}\right] + \frac{2}{K^2}\left(\phi(\lambda K z) - \frac{(\lambda K z)^2}{2}\right)e^{-z^{\alpha}}\sum_i \mathbb{E}\left[U^2_{\pi(i)}\middle|\mathrm{F}_{\pi(i-1)}\right].$$

Notice that the definitions of σ_i^2 and U_i are equivalent to

$$\begin{aligned}\sigma_i^2 = \sigma_i^2(Y), \quad \sigma_i^2(y) &:= \mathbb{E}\, f_i^2(y_1,\dots,y_{i-1},Y_i,y_{i+1},\dots,y_n),\\ U_i = U_i(Y), \quad U_i(y) &:= \|f_i(y_1,\dots,y_{i-1},Y_i,y_{i+1},\dots,y_n)\|_{\psi_\alpha}.\end{aligned}$$

Now we apply Lemma 4.9 first setting $g_i = \sigma_i^2$, and then setting $g_i = U_i^2$, and get

$$\begin{aligned}\frac{1}{n!}\sum_{\pi}\sum_{i}\mathbb{E}[\sigma^2_{\pi(i)}|\mathrm{F}_{\pi(i-1)}] &\le \frac{n+1}{n}\sigma^2 \text{ a.s.},\\ \frac{1}{n!}\sum_{\pi}\sum_{i}\mathbb{E}[U^2_{\pi(i)}|\mathrm{F}_{\pi(i-1)}] &\le \frac{n+1}{n}U^2 \text{ a.s.}\end{aligned}$$

Thus,

$$\frac{1}{n!}\sum_{\pi}\sum_{i}V_i^{\pi}(\lambda) \le \frac{\phi(\lambda K z)}{(Kz)^2}\frac{n+1}{n}\sigma^2 + \frac{2}{K^2}\left(\phi(\lambda K z) - \frac{(\lambda K z)^2}{2}\right)e^{-z^{\alpha}}\frac{n+1}{n}U^2 \text{ a.s.}$$

The result follows immediately from Corollary 3.2. □

Proof of Corollary 3.5. We set $f(y) = \|\sum_i y_i\|$ and notice, that

$$f_i(Y) = \left\|\sum_i Y_i\right\| - \mathbb{E}'\left\|\sum_{j\neq i} Y_j + Y_i'\right\|.$$

Applying the triangle inequality, we get

$$|f_i(Y)| \leq \mathbb{E}'\|Y_i - Y_i'\| \leq \|Y_i\| + \mathbb{E}\|Y_i\|.$$

Thus,

$$\|f_i(Y)|\mathrm{F}_{-i}\|_{\psi_\alpha} \leq \|\|Y_i\| + \mathbb{E}\|Y_i\|\|_{\psi_\alpha} \leq \|\|Y_i\|\|_{\psi_\alpha} + \mathbb{E}\|Y_i\| \leq 2\|\|Y_i\|\|_{\psi_\alpha} \quad \text{a.s.}$$

Similarly,

$$\mathbb{E}\left[f_i^2(Y)|\mathrm{F}_{-i}\right] \leq 4\,\mathbb{E}\|Y_i\|^2 \quad \text{a.s.}$$

The result follows from Corollary 3.4. □

References

Miguel A Arcones. A Bernstein-type inequality for u-statistics and u-processes. *Statistics & probability letters*, 22(3):239–247, 1995.

Johannes TN Krebs. A Bernstein inequality for exponentially growing graphs. *Communications in Statistics-Theory and Methods*, 47(20):5097–5106, 2018.

David Lopez-Paz, Suvrit Sra, Alex Smola, Zoubin Ghahramani, and Bernhard Schölkopf. Randomized nonlinear component analysis. In *International conference on machine learning*, pages 1359–1367. PMLR, 2014.

Ilya O Tolstikhin and Yevgeny Seldin. PAC-Bayes-empirical-Bernstein inequality. *Advances in Neural Information Processing Systems*, 26, 2013.

Federico Girotti, Juan P Garrahan, and Mădălin Guţă. Concentration inequalities for output statistics of quantum Markov processes. In *Annales Henri Poincaré*, pages 1–34. Springer, 2023.

Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces: isoperimetry and processes*, volume 23. Springer Science & Business Media, 1991.

Vladimir Koltchinskii. *Oracle inequalities in empirical risk minimization and sparse recovery problems: École D'Été de Probabilités de Saint-Flour XXXVIII-2008*, volume 2033. Springer Science & Business Media, 2011.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence.* Oxford university press, 2013.

Florence Merlevède, Magda Peligrad, Emmanuel Rio, et al. Bernstein inequality and moderate deviations under strong mixing conditions. *High dimensional probability V: the Luminy volume*, 5:273–292, 2009.

Marwa Banna, Florence Merlevède, and Pierre Youssef. Bernstein-type inequality for a class of dependent random matrices. *Random Matrices: Theory and Applications*, 5(02):1650006, 2016.

Kacha Dzhaparidze and JH Van Zanten. On Bernstein-type inequalities for martingales. *Stochastic processes and their applications*, 93(1):109–117, 2001.

Joel Tropp. Freedman's inequality for matrix martingales. *Electronic Communications in Probability*, 16 (none):262 – 270, 2011.

Ikhlef Bechar. A Bernstein-type inequality for stochastic processes of quadratic forms of Gaussian variables. *arXiv preprint arXiv:0909.3595*, 2009.

Yannick Baraud. A Bernstein-type inequality for suprema of random processes with applications to model selection in non-Gaussian regression. *Bernoulli*, 16(4):1064 – 1085, 2010. doi:10.3150/09-BEJ245. URL `https://doi.org/10.3150/09-BEJ245`.

Hanyuan Hang and Ingo Steinwart. A Bernstein-type inequality for some mixing processes and dynamical systems with an application to learning. *The Annals of Statistics*, 45(2):708 – 743, 2017.

Lester Mackey, Michael I Jordan, Richard Y Chen, Brendan Farrell, and Joel A Tropp. Matrix concentration inequalities via the method of exchangeable pairs. *The Annals of Probability*, pages 906–945, 2014.

Stanislav Minsker. On some extensions of Bernstein's inequality for self-adjoint operators. *Statistics & Probability Letters*, 127:111–119, 2017.

Sergei Bernstein. Theory of probability. *Moscow. MR0169758*, 1927.

Huiming Zhang and Song Xi Chen. Concentration inequalities for statistical inference. *arXiv preprint arXiv:2011.02258*, 2020.

George Bennett. Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, 57(297):33–45, 1962.

David A Freedman. On tail probabilities for martingales. *the Annals of Probability*, pages 100–118, 1975.

Vadim Yurinskii. Exponential inequalities for sums of random vectors. *Journal of Multivariate Analysis*, 6(4):473–499, 1976. ISSN 0047-259X.

Joel Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12:389–434, 2012.

Radoslaw Adamczak. A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electronic Journal of Probability*, 13(none):1000 – 1034, 2008. doi:10.1214/EJP.v13-521. URL `https://doi.org/10.1214/EJP.v13-521`.

Sara van de Geer and Johannes Lederer. The Bernstein–Orlicz norm and deviation inequalities. *Probability theory and related fields*, 157(1):225–250, 2013.

Fuqing Gao, Arnaud Guillin, and Liming Wu. Bernstein-type concentration inequalities for symmetric Markov processes. *Theory of Probability & Its Applications*, 58(3):358–382, 2014.

Vern Paulsen. *Completely bounded maps and operator algebras*, volume 78. Cambridge University Press, 2002.

Albert N Shiryaev. *Probability-1*, volume 95. Springer, 2016.

Thomas Peel, Sandrine Anthoine, and Liva Ralaivola. Empirical Bernstein inequalities for u-statistics. *Advances in Neural Information Processing Systems*, 23, 2010.

Diego Martinez-Taboada and Aaditya Ramdas. Empirical Bernstein in smooth Banach spaces. *arXiv preprint arXiv:2409.06060*, 2024.

Jean-Yves Audibert, Rémi Munos, and Csaba Szepesvári. Tuning bandit algorithms in stochastic environments. In *International conference on algorithmic learning theory*, pages 150–165. Springer, 2007.

Volodymyr Mnih, Csaba Szepesvári, and Jean-Yves Audibert. Empirical Bernstein stopping. In *Proceedings of the 25th international conference on Machine learning*, pages 672–679, 2008.

Andreas Maurer and Massimiliano Pontil. Empirical Bernstein bounds and sample variance penalization. *arXiv preprint arXiv:0907.3740*, 2009.

Pannagadatta Shivaswamy and Tony Jebara. Empirical Bernstein boosting. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 733–740. JMLR Workshop and Conference Proceedings, 2010.

Colin McDiarmid et al. On the method of bounded differences. *Surveys in combinatorics*, 141(1): 148–188, 1989.

Samuel Kutin. Extensions to mcDiarmid's inequality when differences are bounded with high probability. *Dept. Comput. Sci., Univ. Chicago, Chicago, IL, USA, Tech. Rep. TR-2002-04*, 2002.

Emmanuel Rio. On mcDiarmid's concentration inequality. *Electronic Communications in Probability*, 18 (none):1 – 11, 2013. doi:10.1214/ECP.v18-2659. URL `https://doi.org/10.1214/ECP.v18-2659`.

Rui Ray Zhang, Xingwu Liu, Yuyi Wang, and Liwei Wang. McDiarmid-type inequalities for graph-dependent variables and stability bounds. *Advances in Neural Information Processing Systems*, 32, 2019.

Andreas Maurer and Massimiliano Pontil. Concentration inequalities under sub-gaussian and sub-exponential conditions. *Advances in Neural Information Processing Systems*, 34:7588–7597, 2021.

Aad W Van Der Vaart, Jon A Wellner, Aad W van der Vaart, and Jon A Wellner. *Weak convergence*. Springer, 1996.

T. Klein and E. Rio. Concentration around the mean for maxima of empirical processes. *The Annals of Probability*, 33(3):1060 – 1077, 2005. doi:10.1214/009117905000000044. URL `https://doi.org/10.1214/009117905000000044`.

Elliott H Lieb. Convex trace functions and the Wigner–Yanase–Dyson conjecture. *Les rencontres physiciens-mathématiciens de Strasbourg-RCP25*, 19:0–35, 1973.

Bodhisattva Sen. A gentle introduction to empirical process theory and applications. *Lecture Notes, Columbia University*, 11:28–29, 2018.

Appendix A Tail regimes

To get (7), we recall that $\alpha z^\alpha \geq 4$. The result holds due to

$$\max_k \lambda_{\max}(\mathbf{S}_k) \leq \sigma\sqrt{2\mathrm{x}} + \frac{3}{4}Kz\mathrm{x} \leq \sigma\sqrt{2\mathrm{x}}\left(1 + 2\sqrt{2\left(\frac{Kz}{\sigma}\right)^2\mathrm{x}}\right) \leq \sigma\sqrt{2\mathrm{x}}\left(1 + 2\sqrt{2e}\right) \leq 6\sigma\sqrt{2\mathrm{x}}.$$

Bound (8) holds due to

$$\begin{aligned}\max_k \lambda_{\max}(\mathbf{S}_k) &\leq \sigma\sqrt{2\mathrm{x}} + 4Kz\frac{\mathrm{x}}{\ln\left(\left(\frac{Kz}{\sigma}\right)^2\mathrm{x}\right)} \\ &\leq \frac{\sigma^2}{Kz}\left(2\sqrt{\left(\frac{Kz}{\sigma}\right)^2\mathrm{x}} + \frac{4\left(\frac{Kz}{\sigma}\right)^2\mathrm{x}}{\ln\left(\left(\frac{Kz}{\sigma}\right)^2\mathrm{x}\right)}\right) \\ &\leq 8\frac{\sigma^2}{Kz}\frac{\left(\frac{Kz}{\sigma}\right)^2\mathrm{x}}{\ln\left(\left(\frac{Kz}{\sigma}\right)^2\mathrm{x}\right)} = 8\frac{Kz\mathrm{x}}{\ln\left(\left(\frac{Kz}{\sigma}\right)^2\mathrm{x}\right)}.\end{aligned}$$

The last inequality is due to $\frac{1}{2}\ln x = \ln\sqrt{x} \leq \sqrt{x} - 1$.

Bound (9) holds due to

$$\max_k \lambda_{\max}(\mathbf{S}_k) \leq \sigma\sqrt{2\mathrm{x}} + \frac{2Kz}{\alpha z^\alpha}\mathrm{x} \leq \frac{\sigma^2}{Kz}\left(\sqrt{2\left(\frac{Kz}{\sigma}\right)^2\mathrm{x}} + \frac{1}{\alpha z^\alpha}4\left(\frac{Kz}{\sigma}\right)^2\mathrm{x}\right) \leq \frac{6Kz}{\alpha z^\alpha}\mathrm{x}.$$

The last inequality follows from

$$\sqrt{2x} + \frac{4x}{\alpha z^\alpha} = \frac{\sqrt{2x}}{\alpha z^\alpha}\left(\alpha z^\alpha + 2\sqrt{2x}\right) \leq \frac{6x}{\alpha z^\alpha}.$$

The last inequality is due to $\sqrt{x} \geq \exp\{\alpha z^\alpha\} \geq e\alpha z^\alpha$, since $x \geq \exp\{2\alpha z^\alpha\}$.

Appendix B Auxiliary results

Proof of Lemma 4.4. We consider only $x > 0$. Recall that $\phi'(t) = e^t - 1 = \phi(t) + t$, thus

$$\begin{aligned}\rho'_{\lambda,\alpha}(x) &= \lambda\left(\phi'(\lambda x) - \lambda x\right)\exp\{-x^\alpha\} - \alpha x^{\alpha-1}\left(\phi(\lambda x) - \frac{(\lambda x)^2}{2}\right)\exp\{-x^\alpha\} \\ &= \left(\lambda\phi(\lambda x) - \alpha x^{\alpha-1}\left(\phi(\lambda x) - \frac{(\lambda x)^2}{2}\right)\right)\exp\{-x^\alpha\} \\ &= \left(\frac{\lambda\phi(\lambda x)}{\phi(\lambda x) - (\lambda x)^2/2} - \alpha x^{\alpha-1}\right)\rho_{\lambda,\alpha}(x).\end{aligned}$$

Since $\phi(\lambda x) > \frac{(\lambda x)^2}{2}$, $\rho_{\lambda,\alpha}(x) > 0$, and hence

$$\operatorname{sign}\rho'_{\lambda,\alpha}(x) = \operatorname{sign}\left(\frac{\lambda\phi(\lambda x)}{\phi(\lambda x) - (\lambda x)^2/2} - \alpha x^{\alpha-1}\right) = \operatorname{sign}\left(\frac{\lambda x\phi(\lambda x)}{\phi(\lambda x) - (\lambda x)^2/2} - \alpha x^\alpha\right).$$

□

Lemma B.1. *The function* $f(t) := \frac{t\phi'(t)}{\phi(t)}$*, extended at* 0 *by continuity as* $f(0) = 2$*, is increasing on* $\mathbb{R}$

Proof. To prove the monotonicity of f, one has to show that f is continuous at 0, and $f'(t) > 0$ for all $t \neq 0$.

Note that

$$f(t) = \frac{t(e^t - 1)}{e^t - t - 1} = \frac{t\,(t + o(t))}{\frac{t^2}{2} + o(t^2)} = 2 + o(1), \;\; t \to 0,$$

hence it can be continuously extended at 0 as $f(0) = 2$.

The first derivative is

$$f'(t) = \frac{1 + e^{2t} - e^t(2 + t^2)}{\phi^2(t)}.$$

Consider $u(t) = 1 + e^{2t} - e^t(2 + t^2)$, its derivative is

$$u'(t) = e^t(2e^t - 2t - 2 - t^2) = 2e^t\left(\phi(t) - \frac{t^2}{2}\right).$$

By Proposition 4.3, $\operatorname{sign}\left(\phi(t) - \frac{t^2}{2}\right) = \operatorname{sign}(t)$, thus u attains a global minimum at $t = 0$ and $u(t) > u(0) = 0$ for all $t \neq 0$. Therefore, $f'(t) > 0$ for all $t \neq 0$. □

Lemma B.2. *The function* $\upsilon(t)$*, extended at* 0 *by continuity as* $\upsilon(0) = 3$*, is increasing and strictly convex on* $\mathbb{R}$*. Moreover, for any* $t \in \mathbb{R}$

$$\upsilon(t) \leq \max\{4, 1.5t\}.$$

Proof. Note that

$$\upsilon(t) = \frac{t\phi(t)}{\phi(t) - \frac{t^2}{2}} = \frac{t(e^t - 1 - t)}{e^t - 1 - t - \frac{t^2}{2}}.$$

To prove the strict convexity of υ, one has to show that υ and υ' are continuous at 0, and $\upsilon''(t) > 0$ for all $t \neq 0$. Note that

$$\upsilon(t) = \frac{t\left(\frac{t^2}{2} + o(t^2)\right)}{\frac{t^3}{6} + o(t^3)} = 3 + o(1), \quad t \to 0,$$

hence it can be continuously extended at 0 with $\upsilon(0) = 3$. The first derivative is

$$\begin{aligned}
\upsilon'(t) &= \frac{\phi(t) + t\phi'(t)}{\phi(t) - \frac{t^2}{2}} - \frac{t\phi(t)^2}{\left(\phi(t) - \frac{t^2}{2}\right)^2} \\
&= \frac{\left(\frac{t^2}{2} + \frac{t^3}{6} + t(t + \frac{t^2}{2}) + o(t^3)\right)\left(\frac{t^3}{6} + \frac{t^4}{24} + o(t^4)\right) - t\left(\frac{t^2}{2} + \frac{t^3}{6} + o(t^3)\right)^2}{\left(\frac{t^3}{6} + o(t^3)\right)^2} \\
&= \frac{\left(\frac{3t^2}{2} + \frac{2t^3}{3} + o(t^3)\right)\left(\frac{t^3}{6} + \frac{t^4}{24} + o(t^4)\right) - t\left(\frac{t^4}{4} + \frac{t^5}{6} + o(t^5)\right)}{\frac{t^6}{36} + o(t^6)} \\
&= \frac{\left(\frac{3}{2}\frac{1}{24} + \frac{2}{3}\frac{1}{6} - \frac{1}{6}\right)t^6 + o(t^6)}{\frac{t^6}{36} + o(t^6)} = \frac{t^6 + o(t^6)}{4t^6 + o(t^6)} = \frac{1}{4} + o(1), \quad t \to 0,
\end{aligned}$$

thus it is also continuous at 0 with $\upsilon'(0) = \frac{1}{4}$.

The second derivative is

$$v''(t) = \frac{te^t}{4\left(\phi(t) - \frac{t^2}{2}\right)^3} \underbrace{\left(t^4 + 8t^2 - 24 + 4(t^2+6)\cosh t - 24t \sinh t\right)}_{u(t)}.$$

By Proposition 4.3, $\operatorname{sign}\left(\phi(t) - \frac{t^2}{2}\right) = \operatorname{sign}(t)$, thus the first term is positive for any $t \neq 0$.

Now, we show that u is positive as well. We explicitly compute

$$\begin{aligned}
u^{(1)}(t) &= 4t\left(t^2 + 4 - 4\cosh t + t\sinh t\right),\\
u^{(2)}(t) &= 4(3t^2 + 4 + (t^2-4)\cosh t - 2t\sinh t),\\
u^{(3)}(t) &= 4(6t + (t^2-6)\sinh t),\\
u^{(4)}(t) &= 4(6 + (t^2-6)\cosh t + 2t\sinh t),\\
u^{(5)}(t) &= 4(4t\cosh t + (t^2-4)\sinh t),\\
u^{(6)}(t) &= 4t(t\cosh t + 6\sinh t) = 4t^2\cosh t + 24t\sinh t.
\end{aligned}$$

Since $t\sinh t > 0$ and $\cosh t > 1$ for all $t \neq 0$, one immediately obtains that $u^{(6)}(t) > 0$, $t \neq 0$. Moreover, $u^{(i)}(0) = 0$ for all $i = 0, \dots, 5$. By Taylor's theorem, this ensures that $u(t) \geq 0$ for all t, with the only global minimum $u(0) = 0$. Thus, $v''(t) > 0$ for all $t \neq 0$, and therefore v is strictly convex.

Finally,

$$\lim_{t\to-\infty} v(t) = \lim_{t\to-\infty} \frac{-t^2 + o(t^2)}{-\frac{t^2}{2} + o(t^2)} = 2,$$

hence $\lim_{t\to-\infty} v'(t) = 0$ and f is strictly increasing.

To get the last result, we consider $f(t) = \frac{\phi(t)}{\phi(t) - t^2/2}$. Note that f is decreasing by Proposition 4.3 and v is increasing by Lemma B.2.

If $t_0 = 2.68$, $v(t_0) < 1.5$ and $v(t_0) < 4$. Thus,

$$v(t) \leq \min\{v(t_0), tf(t_0)\} < \min\{4, 1.5t\}.$$

□

Lemma B.3. *For all $\alpha > 0$, $t \geq 0$, and $p > 0$ it holds*

$$t^p \leq \left(\frac{p}{\alpha e}\right)^{p/\alpha} e^{t^\alpha}.$$

Let t_1 be such that $\alpha t_1^\alpha \geq p$. Then for all $t_2 \geq t_1$

$$t_2^p e^{-t_2^\alpha} \leq t_1^p e^{-t_1^\alpha}.$$

Proof. Taking the derivative, one gets

$$\left(t^p e^{-t^\alpha}\right)' = (pt^{p-1} - \alpha t^{\alpha-1}t^p)e^{-t^\alpha} = (p - \alpha t^\alpha)t^{p-1}e^{-t^\alpha}.$$

Thus, the maximum of $t^p e^{-t^\alpha}$ is attained at $\alpha t^\alpha = p$, where

$$t^p e^{-t^\alpha} = \left(\frac{p}{\alpha}\right)^{p/\alpha} e^{-\frac{p}{\alpha}} = \left(\frac{p}{\alpha e}\right)^{\frac{p}{\alpha}}.$$

The first result follows.

The second result holds because $\left(t^p e^{-t^\alpha}\right)' \leq 0$ for $\alpha t^\alpha \geq p$. □

Lemma B.4. *Let $\alpha, u, \sigma > 0$. Then it holds for $z = z(u, \sigma; \alpha)$ coming from* (6) *that*

$$e^{z^\alpha} \geq \frac{e^4}{16}\left(\frac{uz}{\sigma}\right)^2 \geq 3\left(\frac{uz}{\sigma}\right)^2.$$

Proof. Let us define $A := \max\left\{\frac{u}{\sigma},\, 1\right\}$, so that

$$z = \left[\frac{4}{\min\{\alpha, 1\}} \ln \frac{e}{\min\{\alpha, 1\}} + 4\ln A\right]^{1/\alpha}.$$

Note that by Lemma B.3,

$$e^{z^\alpha} \geq \left(\frac{\alpha e}{4}\right)^{4/\alpha} z^4.$$

First, consider the case $\alpha < 1$. Then

$$e^{z^\alpha} = \exp\left\{\frac{4}{\alpha}\ln\frac{e}{\alpha} + 4\ln A\right\} = \left(\frac{e}{\alpha}\right)^{4/\alpha} A^4,$$

and combining two bounds, we get

$$e^{z^\alpha} \geq \left(\frac{\alpha e}{4}\right)^{2/\alpha} z^2 \cdot \left(\frac{e}{\alpha}\right)^{2/\alpha} A^2 = \left(\frac{e^2}{4}\right)^{2/\alpha} (Az)^2 \geq \left(\frac{e^2}{4}\right)^2 (Az)^2 = \frac{e^4}{16}(Az)^2.$$

Now, consider the case $\alpha \geq 1$:

$$e^{z^\alpha} = \exp\{4 + 4\ln A\} = e^4 A^4,$$

and thus

$$e^{z^\alpha} \geq \left(\frac{\alpha e}{4}\right)^{2/\alpha} z^2 \cdot e^2 A^2 \geq \left(\frac{e}{4}\right)^2 e^2 (Az)^2 = \frac{e^4}{16}(Az)^2.$$

The claim follows. □

Lemma B.5. *Fix $\alpha > 0$. Let a random matrix $\mathbf{X} \in \mathbb{H}(d)$ be such that $u := \left\|\|\mathbf{X}\|\right\|_{\psi_\alpha} < \infty$ and $\sigma^2 := \lambda_{\max}(\mathbb{E}\,\mathbf{X}^2)$. Then, with $z = z(u, \sigma; \alpha)$ defined by* (6)*,*

$$\lambda_{\max}(\mathbb{E}\,\mathbf{X}^4) \leq \frac{5}{3}(\sigma u z)^2.$$

Proof. W.l.o.g. we can assume $u = 1$. Notice that

$$\begin{aligned} x^4 &\leq x^2 z^2 + x^4\,\mathbb{I}[|x| \geq z] = x^2 z^2 + x^4 e^{-|x|^\alpha} e^{|x|^\alpha}\,\mathbb{I}[|x| \geq z] \\ &\leq x^2 z^2 + z^4 e^{-z^\alpha} e^{|x|^\alpha}\,\mathbb{I}[|x| \geq z], \end{aligned}$$

where the last inequality holds due to Lemma B.3 since $\alpha z^\alpha \geq 4$. Thus, for $\mathbf{\Sigma} := \mathbb{E}\,\mathbf{X}^2$, we get

$$\mathbb{E}\,\mathbf{X}^4 \preccurlyeq z^2\mathbf{\Sigma} + z^4 e^{-z^\alpha}\,\mathbb{E}\,e^{\|\mathbf{X}\|^\alpha}\mathbf{I} \preccurlyeq (z^2\sigma^2 + 2z^4 e^{-z^\alpha})\mathbf{I}. \tag{34}$$

Finally, Lemma B.4 ensures that

$$\lambda_{\max}(\mathbb{E}\,\mathbf{X}^4) \leq z^2\sigma^2 + 2z^4 e^{-z^\alpha} \leq z^2\sigma^2 + \frac{2}{3}z^4\left(\frac{\sigma}{z}\right)^2 = \frac{5}{3}(\sigma z)^2.$$

□

Lemma B.6. *The inverse function of* $h(x) = (x+1)\ln(x+1) - x$ *satisfies*

$$h^{-1}(u) \le \sqrt{2u} + \frac{2u}{\underline{\log} 2u}. \tag{35}$$

Proof. First, consider $0 \le u \le \frac{e^6}{2}$. Then

$$h^{-1}(u) \le \sqrt{2u} + \frac{u}{3} \le \sqrt{2u} + \frac{2u}{\underline{\log} 2u}.$$

The first inequality is well-known (see, e.g., Proposition 8 by Sen [2018]), and the second one trivially follows from the fact that $\underline{\log} 2u < 6$.

Now we consider $u > \frac{e^6}{2}$. As $h(\cdot)$ is increasing, the goal is to check

$$h\left(\sqrt{2u} + \frac{2u}{\underline{\log} 2u}\right) - u \ge 0.$$

Notice that

$$\ln\left(\frac{e^6}{6} + e^3 + 1\right) \ge 1.$$

Thus, using the definition of $h(\cdot)$ and the inequality $\underline{\log} 2u = \ln 2u \ge 6$, we get

$$\begin{aligned}
&\left(\frac{2u}{\ln 2u} + \sqrt{2u} + 1\right)\ln\left(\frac{2u}{\ln 2u} + \sqrt{2u} + 1\right) - \frac{2u}{\ln 2u} - \sqrt{2u} - u \\
&\ge \frac{2u}{\ln 2u}\ln\left(\frac{2u}{\ln 2u}\right) - \frac{2u}{\ln 2u} - u \\
&= u\left(\frac{2}{\ln 2u}(\ln 2u - \ln\ln 2u) - \frac{2}{\ln 2u} - 1\right) = u\left(1 - \frac{2}{\ln 2u}(1 + \ln\ln 2u)\right).
\end{aligned}$$

Due to the concavity of the logarithm,

$$\ln x \le \ln a + \frac{x-a}{a} \quad \forall x, a > 0,$$

and since $\ln 2u \ge 6$, we get

$$\frac{2}{\ln 2u}(1 + \ln\ln 2u) \le \frac{2}{\ln 2u}\left(1 + \ln 6 + \frac{\ln 2u - 6}{6}\right) = \frac{1}{3} + \frac{2\ln 6}{\ln 2u} \le \frac{1 + \ln 6}{3} < 1.$$

The claim follows. □

Lemma B.7. *Let*

$$g_{\lambda_0}(t) := \max_{0 \le \lambda \le \lambda_0}\{\lambda t - \phi(\lambda)\},$$

and set $x_0 := \lambda_0\phi'(\lambda_0) - \phi(\lambda_0)$.

Then, it holds that

$$g_{\lambda_0}^{-1}(x) = \begin{cases} h^{-1}(x), & \text{if } x \le x_0, \\ t_0 + \frac{x - x_0}{\lambda_0} \le \frac{2}{\lambda_0}x, & \text{if } x > x_0. \end{cases}$$

Proof. Notice that $\phi(\cdot)$ is strictly convex. Thus $u(\lambda) := \lambda t - \phi(\lambda)$ is strictly concave and its $\max$ is unique. Consider

$$u'(\lambda) = t - \phi'(\lambda) = t + 1 - e^{\lambda} = 0.$$

Thus, the global $\max$ is attained at $\ln(t+1)$.

Taking into account the condition $0 \le \lambda \le \lambda_0$, we get

$$g(t) = \lambda^* t - \phi(\lambda^*), \quad \lambda^* = \min\{\lambda_0, \ln(t+1)\}.$$

The critical point is $t_0 = e^{\lambda_0} - 1 = \phi'(\lambda_0)$.

If $t \le t_0$,

$$g(t) = t\ln(t+1) - \phi(\ln(t+1)) = h(t).$$

Thus, for all $x \le x_0$, $x_0 := h(t_0)$,

$$g^{-1}(x) = h^{-1}(x).$$

We also notice that substituting $t_0 = e^{\lambda_0} - 1 = \phi'(\lambda_0)$ to $h(t_0)$, one gets

$$x_0 := h(t_0) = \lambda_0 \phi'(\lambda_0) - \phi(\lambda_0).$$

Now consider $t > t_0$,

$$g(t) = \lambda_0 t - \phi(\lambda_0).$$

This yields

$$g_{\lambda_0}^{-1}(x) = \frac{\phi(\lambda_0) + x}{\lambda_0} \le \frac{\phi(\lambda_0)x/x_0 + x}{\lambda_0} = \frac{\phi(\lambda_0) + x_0}{x_0}\frac{x}{\lambda_0}.$$

Finally, notice that

$$\frac{x_0}{\phi(\lambda_0) + x_0} = \frac{\lambda_0\phi'(\lambda_0) - \phi(\lambda_0)}{\lambda_0\phi'(\lambda_0)} = 1 - \frac{\phi(\lambda_0)}{\lambda_0\phi'(\lambda_0)} \ge \frac{1}{2},$$

the last inequality follows from the bound $\frac{\lambda_0\phi'(\lambda_0)}{\phi(\lambda_0)} \ge 2$ due to Lemma B.1. □

Appendix C Proof of Lemma 4.9

Proof of Lemma 4.9. Let $Y' = (Y_1', \dots, Y_n')$ be an independent copy of Y (all Y_i' are i.i.d.). First, we notice that

$$\mathbb{E}_Y \left[g_{\pi(i)}(Y)|\mathrm{F}_{\pi([1,i-1])}\right] = \mathbb{E}_{Y'}\, g_{\pi(i)}\left(Y_{\pi([1,i-1])}, Y'_{\pi([i,n])}\right).$$

This yields

$$\frac{1}{n!}\sum_{\pi\in\Pi_n}\sum_{i=1}^{n}\mathbb{E}_Y[g_{\pi(i)}(Y)|\mathrm{F}_{\pi([1,i-1])}] = \mathbb{E}_{Y'}\frac{1}{n!}\sum_{i=1}^{n}\sum_{\pi\in\Pi_n} g_{\pi(i)}(Y_{\pi([1,i-1])}, Y'_{\pi([i,n])}).$$

Second, we notice that for $j \in I \subset [n]$

$$g_j(Y_I, Y'_{\overline{I}}) = g_j(Y_{I\setminus\{j\}}, Y'_{\overline{I}\bigcup\{j\}}),$$

since $g_j(x_1,\dots,x_n)$ does not depend on x_j. Thus, we get for a fixed $i\in[n]$

$$\begin{aligned}\sum_{\pi\in\Pi_n} g_{\pi(i)}\left(Y_{\pi([1,i])}, Y'_{\pi([i+1,n])}\right) &= (i-1)!(n-i)! \sum_{I\subset[n]:\ |I|=i}\ \sum_{j\in I} g_j(Y_I, Y'_{\overline{I}})\\ &= (i-1)!(n-i)! \sum_{J\subset[n]:\ |J|=i-1}\ \sum_{j\notin J} g_j(Y_J, Y'_{\overline{J}}).\end{aligned}$$

Now let $\alpha_i := \frac{i}{n}$ for $i\in[n]$. Combining the above results, we get

$$\begin{aligned}&\frac{1}{n!}\sum_{\pi\in\Pi_n}\sum_{i=1}^{n}\mathbb{E}\left[g_{\pi(i)}(Y)\middle|\mathrm{F}_{\pi([1,i-1])}\right]\\ &\quad= \frac{1}{n!}\sum_{i=1}^{n}(i-1)!(n-i)!\left(\alpha_i\sum_{|I|=i}\sum_{j\in I} g_j(Y_I,Y'_{\overline{I}}) + (1-\alpha_i)\sum_{|I|=i-1}\sum_{j\neq I} g_j(Y_I,Y'_{\overline{I}})\right)\\ &\quad= \frac{1}{n!}\sum_{i=1}^{n}\alpha_i(i-1)!(n-i)!\sum_{|I|=i}\sum_{j\in I} g_j(Y_I,Y'_{\overline{I}}) + \frac{1}{n!}\sum_{i=0}^{n-1}(1-\alpha_{i+1})i!(n-i-1)!\sum_{|I|=i}\sum_{j\notin I} g_j(Y_I,Y'_{\overline{I}}).\end{aligned}$$

Now we notice that

$$\alpha_i(i-1)!(n-i)! = \frac{i!(n-i)!}{n},\quad (1-\alpha_{i+1})i!(n-i-1)! \le \frac{i!(n-i)!}{n}.$$

Thus,

$$\begin{aligned}&\frac{1}{n!}\sum_{i=1}^{n}\alpha_i(i-1)!(n-i)!\sum_{|I|=i}\sum_{j\in I} g_j(Y_I,Y'_{\overline{I}}) + \frac{1}{n!}\sum_{i=0}^{n-1}(1-\alpha_{i+1})i!(n-i-1)!\sum_{|I|=i}\sum_{j\notin I} g_j(Y_I,Y'_{\overline{I}})\\ &\quad\le \sum_{i=1}^{n}\frac{i!(n-i)!}{n\cdot n!}\left(\sum_{|I|=i}\sum_{j\in I} g_j(Y_I,Y'_{\overline{I}}) + \sum_{|I|=i}\sum_{j\notin I} g_j(Y_I,Y'_{\overline{I}})\right)\\ &\quad= \sum_{i=0}^{n}\frac{i!(n-i)!}{n\cdot n!}\sum_{|I|=i}\sum_{j=1}^{n} g_j(Y_I,Y'_{\overline{I}}).\end{aligned}$$

Recall that by the Lemma's condition it holds that $\sum_{j=1}^{n} g_j(Y_I,Y'_{\overline{I}}) \le M$ a.s. Further, the number of subsets of cardinality i is $|\{I\subset[n] : |I|=i\}| = \binom{n}{i} = \frac{n!}{i!(n-i)!}$. Thus, we get

$$\frac{1}{n!}\sum_{\pi\in\Pi}\sum_{i=1}^{n}\mathbb{E}\left[g_{\pi(i)}(Y)\middle|\mathrm{F}_{\pi([1,i-1])}\right] \le \sum_{i=0}^{n}\frac{i!(n-i)!}{n\cdot n!}\binom{n}{i}M = \frac{n+1}{n}M.$$

□