EXPONENTIAL CONCENTRATION INEQUALITIES FOR INDEPENDENT RANDOM VECTORS UNDER SUBLINEAR EXPECTATIONS

BY NAHOM SEYOUM[1,a]

[1]*Department of Statistics and Data Science, Yale University,* [a]*nahom.seyoum@yale.edu*

Li and Hu recently established variance-type $O(1/n)$ bounds for the sample mean of independent random vectors under sublinear expectations. We extend their results to the exponential concentration regime. For bounded, independent $\mathbb{R}^d$-valued random vectors $\{X_i\}_{i=1}^n$ under a regular sublinear expectation $\hat{\mathbb{E}}$, we prove: (i) an Azuma–Hoeffding-type inequality showing that the distance from the sample mean to the Minkowski average of the expectation sets has sub-Gaussian tails; (ii) a sharper Bernstein-type inequality incorporating the variance parameter of Li and Hu; (iii) a dimension-free bound for identically distributed vectors via the matrix Freedman inequality; and (iv) an explicit construction demonstrating the optimality of the sub-Gaussian rate.

1. Introduction. Let $(\Omega, \mathcal{H}, \hat{\mathbb{E}})$ be a regular sublinear expectation space with $\mathcal{H} = C_{b.\mathrm{Lip}}(\Omega)$. By the representation theorems of Denis, Hu and Peng [3] and Hu and Peng [7], there exists a convex and weakly compact set $\mathcal{P}$ of probability measures on $(\Omega, \mathcal{B}(\Omega))$ such that

$$\hat{\mathbb{E}}[X] = \sup_{P \in \mathcal{P}} E_P[X] \quad \text{for } X \in \mathcal{H}. \tag{1.1}$$

Under this framework, the "expectation" of a random vector $X_i \in L^1(\Omega; \mathbb{R}^d)$ is not a single point but a convex, compact set $\Theta_i = \{E_P[X_i] : P \in \mathcal{P}\} \subset \mathbb{R}^d$.

For independent random vectors $\{X_i\}_{i=1}^n$ under $\hat{\mathbb{E}}$ (Definition 2.2), Li and Hu [8] recently established the moment inequality

$$\hat{\mathbb{E}}\left[\rho_\Theta^2\left(\frac{1}{n}\sum_{i=1}^n X_i\right)\right] \le \frac{\bar{\sigma}_n^2}{n}, \tag{1.2}$$

where $\Theta = \left\{\frac{1}{n}\sum_{i=1}^n \theta_i : \theta_i \in \Theta_i\right\}$ is the Minkowski average of the expectation sets, $\rho_\Theta(x) = \inf_{\theta \in \Theta} |x - \theta|$, and $\bar{\sigma}_n^2 = \sup_{i \le n} \inf_{\theta_i \in \Theta_i} \hat{\mathbb{E}}[|X_i - \theta_i|^2]$. This generalizes classical variance bounds for sample means to the sublinear setting and removes a convex polytope assumption on Θ_i required in earlier work of Fang et al. [4].

The bound (1.2), via Markov's inequality, yields only polynomial tail decay:

$$\hat{V}\left(\rho_\Theta\left(\frac{1}{n}\sum_{i=1}^n X_i\right) > t\right) \le \frac{\bar{\sigma}_n^2}{nt^2},$$

where $\hat{V}(A) := \sup_{P \in \mathcal{P}} P(A)$ is the upper capacity. Yet when each X_i is almost surely bounded, classical intuition suggests that *sub-Gaussian* tails of the form $\exp(-cnt^2)$ should be attainable, as they are under a single probability measure [1, 12]. Closing the gap between polynomial and exponential concentration under distributional uncertainty is the central objective of this paper.

MSC2020 subject classifications: Primary 60E15, 60G42; secondary 60F10, 60G65.

Keywords and phrases: Sublinear expectation, concentration inequality, exponential inequality, martingale difference, Azuma–Hoeffding, Bernstein inequality, independent random vectors.

1.1. *Main results.* We prove four results, stated informally here and made precise in Sections 4–6.

(i) *Azuma–Hoeffding inequality* (Corollary 4.2). For $|X_i| \le M$ a.s.,

$$\hat{V}\left(\rho_\Theta\left(\bar{X}_n\right) > t\right) \le 2 \cdot 5^d \exp\left(-\frac{nt^2}{32M^2}\right).$$

The prefactor 5^d arises from an ε-net covering of S^{d-1}.

(ii) *Bernstein inequality* (Corollary 4.3). Under the same condition,

$$\hat{V}\left(\rho_\Theta\left(\bar{X}_n\right) > t\right) \le 2 \cdot 5^d \exp\left(-\frac{nt^2}{8\bar{\sigma}_n^2 + \frac{8Mt}{3}}\right).$$

This interpolates between the sub-Gaussian regime ($t \ll \bar{\sigma}_n^2/M$) and the sub-exponential regime ($t \gg \bar{\sigma}_n^2/M$), and recovers the Li–Hu bound (1.2) upon integration.

(iii) *Dimension-free bound* (Theorem 5.1). We bypass the covering argument via the matrix Freedman inequality [11]:

$$\hat{V}\left(\rho_\Theta\left(\bar{X}_n\right) > t\right) \le (d+1) \exp\left(-\frac{nt^2}{2\bar{\sigma}_n^2 + \frac{4Mt}{3}}\right).$$

The prefactor is polynomial in d rather than exponential, the constants in the exponent are tighter, and no identical distribution assumption is required.

(iv) *Optimality* (Theorem 6.3). We construct an explicit sublinear expectation space in which the tail probability decays as $\exp(-cnt^2)$ and no faster, so the sub-Gaussian rate is sharp.

1.2. *Related work.* Concentration under sublinear expectations has been studied primarily in the scalar case. Fang et al. [4] obtained rates of convergence for Peng's law of large numbers, and Hu, Li and Li [6] improved these rates. Zhang [13] established exponential inequalities for scalar sublinear expectations. Peng [10] developed the foundational theory of G-expectations and the associated central limit theorem. For an overview of classical concentration, see Boucheron, Lugosi and Massart [1]. To the best of our knowledge, the present work constitutes the first exponential concentration inequalities for the multivariate sample mean under sublinear expectations in terms of the set-valued distance ρ_Θ.

2. Preliminaries. We work throughout with a complete separable metric space Ω and set $\mathcal{H} = C_{b.\mathrm{Lip}}(\Omega)$. A functional $\hat{\mathbb{E}} : \mathcal{H} \to \mathbb{R}$ is a *regular sublinear expectation* if it satisfies monotonicity, constant preservation, subadditivity, positive homogeneity, and the regularity property that $X_n \downarrow 0$ implies $\hat{\mathbb{E}}[X_n] \downarrow 0$. We refer to Peng [10] for full details.

THEOREM 2.1 (Representation [3, 7]). *There exists a convex and weakly compact set of probability measures $\mathcal{P}$ on $(\Omega, \mathcal{B}(\Omega))$ such that (1.1) holds. The associated* upper capacity *is $\hat{V}(A) := \sup_{P \in \mathcal{P}} P(A)$.*

The space $L^p(\Omega)$ is the completion of $\mathcal{H}$ under $\|X\|_p := (\hat{\mathbb{E}}[|X|^p])^{1/p}$. Hölder's inequality gives $L^p(\Omega) \subset L^1(\Omega)$ for $p \ge 1$. The representation (1.1) extends to $L^1(\Omega)$.

DEFINITION 2.2 (Independence [10]). A sequence $\{X_i\}_{i=1}^n \subset L^1(\Omega; \mathbb{R}^d)$ is *independent* under $\hat{\mathbb{E}}$ if for each $1 \le i \le n-1$ and every $\psi \in C_{b.\mathrm{Lip}}(\mathbb{R}^{d(i+1)})$,

$$\hat{\mathbb{E}}[\psi(X_1, \ldots, X_i, X_{i+1})] = \hat{\mathbb{E}}\Big[\hat{\mathbb{E}}[\psi(x_1, \ldots, x_i, X_{i+1})]\big|_{(x_1,\ldots,x_i)=(X_1,\ldots,X_i)}\Big].$$

PROPOSITION 2.3 ([6, Proposition 2.1]). *Let $\{X_i\}_{i=1}^n$ be independent in $L^2(\Omega;\mathbb{R}^d)$ under $\hat{\mathbb{E}}$. Then for each $P \in \mathcal{P}$ and $\varphi \in C_{\mathrm{Lip}}(\mathbb{R}^d)$,*

$$E_P[\varphi(X_i) \mid \mathcal{F}_{i-1}] \le \hat{\mathbb{E}}[\varphi(X_i)], \quad P\text{-a.s., for } i \le n,$$

where $\mathcal{F}_i = \sigma(X_1, \ldots, X_i)$ and $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

For each $i \le n$, the support function $g_i(p) := \hat{\mathbb{E}}[\langle p, X_i\rangle]$ is sublinear on $\mathbb{R}^d$ and defines the convex compact set

$$\Theta_i = \{\theta \in \mathbb{R}^d : \langle \theta, p\rangle \le g_i(p) \text{ for all } p \in \mathbb{R}^d\}. \tag{2.1}$$

We recall two key results from Li and Hu [8].

THEOREM 2.4 ([8, Theorem 3.1]). *Let $\{X_i\}_{i=1}^n$ be independent in $L^2(\Omega;\mathbb{R}^d)$ under $\hat{\mathbb{E}}$. Then:*

(a) $\Theta_i = \{E_P[X_i] : P \in \mathcal{P}\}$ *for each $i \le n$.*
(b) $E_P[X_i \mid \mathcal{F}_{i-1}] \in \Theta_i$, *$P$-a.s., for each $P \in \mathcal{P}$ and $i \le n$.*

We define the Minkowski average, the distance function and the variance parameter:

$$\Theta = \left\{ \frac{1}{n}\sum_{i=1}^n \theta_i : \theta_i \in \Theta_i \right\}, \quad \rho_\Theta(x) = \inf_{\theta \in \Theta} |x - \theta|, \quad \bar{\sigma}_n^2 = \sup_{i \le n} \inf_{\theta_i \in \Theta_i} \hat{\mathbb{E}}[|X_i - \theta_i|^2]. \tag{2.2}$$

3. Martingale reduction. We begin with a conditional domination principle that makes the squared-deviation function compatible with the sublinear calculus.

LEMMA 3.1 (Conditional domination). *Let $\{X_i\}_{i=1}^n$ be independent under $\hat{\mathbb{E}}$ with $|X_i| \le M$ a.s. Then for each $P \in \mathcal{P}$, each $i \le n$, and each $\theta_i \in \Theta_i$,*

$$E_P\big[|X_i - \theta_i|^2 \mid \mathcal{F}_{i-1}\big] \le \hat{\mathbb{E}}\big[|X_i - \theta_i|^2\big], \quad P\text{-a.s.}$$

PROOF. Since $\theta_i \in \Theta_i = \{E_P[X_i] : P \in \mathcal{P}\}$ and $|X_i| \le M$ a.s., Jensen's inequality gives $|\theta_i| = |E_P[X_i]| \le E_P[|X_i|] \le M$ for every $P \in \mathcal{P}$. In particular, $|X_i - \theta_i| \le 2M$ on the effective support $\{|x| \le M\}$.

Define the truncation

$$\tilde{\varphi}(x) = \min\big(|x - \theta_i|^2, \, (M + |\theta_i|)^2\big).$$

Since $|\theta_i| \le M$, the cutoff satisfies $(M + |\theta_i|)^2 \ge (2M)^2 \ge |X_i - \theta_i|^2$ a.s., so the truncation is inactive: $\tilde{\varphi}(X_i) = |X_i - \theta_i|^2$ a.s. Moreover, $\tilde{\varphi}$ is bounded and Lipschitz on $\mathbb{R}^d$ with constant at most $2(M + |\theta_i|) \le 4M$. Proposition 2.3 applied to $\tilde{\varphi}$ gives the claim. □

REMARK 3.2 (Role of the truncation). Proposition 2.3 requires a bounded Lipschitz test function, but the map $x \mapsto |x - \theta_i|^2$ is neither bounded nor Lipschitz on all of $\mathbb{R}^d$. The truncation $\tilde{\varphi}$ resolves both issues simultaneously. On the effective support $\{|x| \le M\}$ it agrees with the squared deviation, while globally it has bounded gradient. This device is necessary even though $|X_i| \le M$, because the conditional expectation in Proposition 2.3 is stated for all of $\mathbb{R}^d$, not merely on the support.

The next lemma converts the sublinear problem into a classical one. For every prior $P \in \mathcal{P}$, the centred residuals form a martingale difference sequence whose conditional variance is controlled uniformly.

LEMMA 3.3 (Martingale reduction). *Under the assumptions of Lemma 3.1, fix $P \in \mathcal{P}$ and define*

$$Y_i := X_i - E_P[X_i \mid \mathcal{F}_{i-1}], \quad i = 1, \dots, n.$$

Then the following hold P-a.s.:

(a) $\{Y_i\}_{i=1}^n$ *is a martingale difference sequence with respect to* $(\mathcal{F}_i)_{i=0}^n$ *under* P.

(b) $\rho_\Theta\left(\frac{1}{n}\sum_{i=1}^n X_i\right) \le \left|\frac{1}{n}\sum_{i=1}^n Y_i\right|$.

(c) $|Y_i| \le 2M$.

(d) $E_P[|Y_i|^2 \mid \mathcal{F}_{i-1}] \le \bar{\sigma}_n^2$.

PROOF. (a) By construction, $E_P[Y_i \mid \mathcal{F}_{i-1}] = 0$.

(b) By Theorem 2.4(b), $E_P[X_i \mid \mathcal{F}_{i-1}] \in \Theta_i$ P-a.s. Setting $\theta_i = E_P[X_i \mid \mathcal{F}_{i-1}]$ gives $\frac{1}{n}\sum_{i=1}^n \theta_i \in \Theta$, so

$$\rho_\Theta\left(\frac{1}{n}\sum_{i=1}^n X_i\right) \le \left|\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n \theta_i\right| = \left|\frac{1}{n}\sum_{i=1}^n Y_i\right|.$$

(c) $|Y_i| \le |X_i| + |E_P[X_i \mid \mathcal{F}_{i-1}]| \le M + M = 2M$, since $|E_P[X_i \mid \mathcal{F}_{i-1}]| \le E_P[|X_i| \mid \mathcal{F}_{i-1}] \le M$ by Jensen's inequality.

(d) The conditional mean minimises the conditional L^2 distance, so for any $\theta_i \in \mathbb{R}^d$,

$$E_P[|Y_i|^2 \mid \mathcal{F}_{i-1}] \le E_P[|X_i - \theta_i|^2 \mid \mathcal{F}_{i-1}].$$

Restricting to $\theta_i \in \Theta_i$ and applying Lemma 3.1,

$$E_P[|Y_i|^2 \mid \mathcal{F}_{i-1}] \le \hat{\mathbb{E}}[|X_i - \theta_i|^2].$$

Taking the infimum over $\theta_i \in \Theta_i$ and then the supremum over i yields $E_P[|Y_i|^2 \mid \mathcal{F}_{i-1}] \le \bar{\sigma}_n^2$. $\square$

It remains to pass from scalar martingale tail bounds to vector-valued concentration. The following covering lemma is used repeatedly below.

LEMMA 3.4 (Covering transfer). *Let $Y_1, \dots, Y_n$ be $\mathbb{R}^d$-valued random vectors under a probability measure P. Suppose there exists a decreasing function $\Psi : (0, \infty) \to [0, 1]$ such that for every unit vector $p \in S^{d-1}$,*

$$P\left(\sum_{i=1}^n \langle p, Y_i\rangle \ge s\right) \le \Psi(s), \quad s > 0.$$

Then for every $t > 0$,

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n Y_i\right| > t\right) \le 2 \cdot 5^d\, \Psi\big(\tfrac{nt}{2}\big).$$

PROOF. Let $\mathcal{N}$ be a $\frac{1}{2}$-net of S^{d-1}. By a standard volumetric argument [12, Corollary 4.2.13], we may choose $|\mathcal{N}| \le 5^d$. Set $S_n = \sum_{i=1}^n Y_i$. If $|S_n| > nt$, there exists $u \in S^{d-1}$ with $\langle u, S_n\rangle = |S_n| > nt$. Since $\mathcal{N}$ is a $\frac{1}{2}$-net, there exists $p \in \mathcal{N}$ with $|p - u| \le \frac{1}{2}$, whence

$$\langle p, S_n\rangle = \langle u, S_n\rangle - \langle u - p, S_n\rangle \ge |S_n| - \tfrac{1}{2}|S_n| = \tfrac{1}{2}|S_n| > \tfrac{nt}{2}.$$

Therefore $\{|S_n| > nt\} \subseteq \bigcup_{p \in \mathcal{N}}\big(\{\langle p, S_n\rangle > \frac{nt}{2}\} \cup \{\langle -p, S_n\rangle > \frac{nt}{2}\}\big)$, where the second event accounts for the case $\langle u, S_n\rangle < 0$ by symmetry. A union bound gives $P(|S_n| > nt) \le 2|\mathcal{N}|\,\Psi(\frac{nt}{2}) \le 2 \cdot 5^d\, \Psi(\frac{nt}{2})$. $\square$

REMARK 3.5 (Scalar specialisation). When $d=1$, the sphere $S^0=\{-1,+1\}$ is an exact cover. The union bound over these two points gives $P(|\frac{1}{n}\sum Y_i|>t)\le 2\,\Psi(nt)$ directly, with no approximation loss, since the factor 5^d reduces to 1 and the halving $t\to t/2$ is unnecessary.

4. Concentration via scalar tail bounds. The martingale reduction and covering transfer of Section 3 reduce the problem to obtaining a scalar tail bound Ψ for one-dimensional projections of the martingale differences. Different choices of Ψ yield different concentration inequalities. We formalise this observation as a general principle.

THEOREM 4.1 (General concentration principle). *Let $\{X_i\}_{i=1}^n$ be independent in $L^2(\Omega;\mathbb{R}^d)$ under $\hat{\mathbb{E}}$ with $|X_i|\le M$ a.s. Let (Y_i) be the martingale difference sequence from Lemma 3.3. Suppose there exists a decreasing function $\Psi:(0,\infty)\to[0,1]$ such that, for every $P\in\mathcal{P}$ and every unit vector $p\in S^{d-1}$,*

$$P\left(\sum_{i=1}^n\langle p,Y_i\rangle\ge s\right)\le\Psi(s),\quad s>0.$$

Then for every $t>0$,

$$\hat{V}\left(\rho_\Theta\left(\frac{1}{n}\sum_{i=1}^n X_i\right)>t\right)\le 2\cdot 5^d\,\Psi\big(\tfrac{nt}{2}\big).$$

PROOF. Fix $P\in\mathcal{P}$. By Lemma 3.3(b), $\rho_\Theta(\bar{X}_n)\le|\frac{1}{n}\sum Y_i|$. Lemma 3.4 applied to (Y_i) under P with the given Ψ yields $P(|\frac{1}{n}\sum Y_i|>t)\le 2\cdot 5^d\,\Psi(nt/2)$. Taking the supremum over $P\in\mathcal{P}$ completes the proof. $\square$

The Azuma–Hoeffding and Bernstein inequalities are now immediate corollaries, obtained by plugging in the appropriate scalar tail bound.

COROLLARY 4.2 (Azuma–Hoeffding). *Under the assumptions of Theorem 4.1, for every $t>0$,*

$$\hat{V}\left(\rho_\Theta\left(\frac{1}{n}\sum_{i=1}^n X_i\right)>t\right)\le 2\cdot 5^d\,\exp\left(-\frac{nt^2}{32M^2}\right).$$

PROOF. The scalar increments $D_i=\langle p,Y_i\rangle$ satisfy $|D_i|\le 2M$. The Azuma–Hoeffding inequality [9, Corollary 2.20] gives $\Psi(s)=\exp(-s^2/(8nM^2))$. Theorem 4.1 with $s=nt/2$ yields $2\cdot 5^d\exp(-(nt/2)^2/(8nM^2))=2\cdot 5^d\exp(-nt^2/(32M^2))$. $\square$

COROLLARY 4.3 (Bernstein). *Under the assumptions of Theorem 4.1, for every $t>0$,*

$$\hat{V}\left(\rho_\Theta\left(\frac{1}{n}\sum_{i=1}^n X_i\right)>t\right)\le 2\cdot 5^d\,\exp\left(-\frac{nt^2}{8\bar{\sigma}_n^2+\frac{8Mt}{3}}\right).$$

PROOF. The scalar increments $D_i=\langle p,Y_i\rangle$ satisfy $|D_i|\le 2M$ and, by Lemma 3.3(d),

$$\sum_{i=1}^n E_P[D_i^2\mid\mathcal{F}_{i-1}]\le\sum_{i=1}^n E_P[|Y_i|^2\mid\mathcal{F}_{i-1}]\le n\bar{\sigma}_n^2,\quad P\text{-a.s.}$$

Freedman's inequality [5, Theorem 1.6] with bound $b = 2M$ and predictable quadratic variation $V = n\bar{\sigma}_n^2$ gives $\Psi(s) = \exp\big(-s^2/(2n\bar{\sigma}_n^2 + \frac{4Ms}{3})\big)$. Theorem 4.1 with $s = nt/2$ yields, after simplification,

$$\frac{(nt/2)^2}{2n\bar{\sigma}_n^2 + \frac{2nMt}{3}} = \frac{nt^2}{8\bar{\sigma}_n^2 + \frac{8Mt}{3}}. \qquad \square$$

REMARK 4.4 (Two regimes). The Bernstein bound exhibits two regimes depending on the ratio $t/(\bar{\sigma}_n^2/M)$.

(i) *Sub-Gaussian regime.* When $t \le 3\bar{\sigma}_n^2/M$, the term $8\bar{\sigma}_n^2$ dominates the denominator $8\bar{\sigma}_n^2 + \frac{8Mt}{3}$, so Corollary 4.3 yields the bound $2 \cdot 5^d \exp(-nt^2/(8\bar{\sigma}_n^2))$. This is sharper than Corollary 4.2 precisely when $\bar{\sigma}_n^2 < 4M^2$.

(ii) *Sub-exponential regime.* When $t \ge 3\bar{\sigma}_n^2/M$, the linear term $\frac{8Mt}{3}$ dominates the denominator, and the bound becomes $2 \cdot 5^d \exp(-3nt/(8M))$, decaying exponentially in t rather than in t^2.

COROLLARY 4.5 (Recovery of the moment bound). *Under the assumptions of Theorem 4.1,*

$$\hat{\mathbb{E}}\left[\rho_\Theta^2\left(\frac{1}{n}\sum_{i=1}^n X_i\right)\right] \le \frac{16\bar{\sigma}_n^2}{n}(1 + d\log 5 + \log 2) + \frac{C_0 M^2}{n^2},$$

where C_0 is a universal constant. In particular, this recovers the $O(\bar{\sigma}_n^2/n)$ rate of [8].

PROOF. Let $Z = \rho_\Theta^2(\bar{X}_n)$. Write $\alpha = 2 \cdot 5^d$ for the prefactor in Corollary 4.3. The layer-cake formula gives

$$\hat{\mathbb{E}}[Z] = \int_0^\infty \hat{V}(\rho_\Theta(\bar{X}_n) > \sqrt{s})\, ds \le \int_0^\infty \alpha \exp\left(-\frac{ns}{8\bar{\sigma}_n^2 + \frac{8M\sqrt{s}}{3}}\right) ds.$$

We split at $s_0 = \bar{\sigma}_n^2$. For $s \le s_0$, the denominator satisfies $8\bar{\sigma}_n^2 + \frac{8M\sqrt{s}}{3} \le 16\bar{\sigma}_n^2$ (since $\sqrt{s} \le \bar{\sigma}_n \le 2M$), so the integrand is at most $\alpha\exp(-ns/(16\bar{\sigma}_n^2))$, which integrates to $16\alpha\bar{\sigma}_n^2/n$. For $s > s_0$, the denominator exceeds $\frac{8M\sqrt{s}}{3}$, and this tail contributes $O(M^2/n^2)$. $\square$

5. Dimension-free bound for identically distributed vectors. The 5^d prefactor in Corollaries 4.2 and 4.3 arises from the covering argument and becomes prohibitive in high dimensions. When the X_i are identically distributed, we can bypass this entirely.

The idea is to embed each vector-valued martingale increment Y_i into matrix space via the rank-one map $Y_i \mapsto Y_i e_1^T$, producing a $d \times 1$ matrix martingale. The operator norm of $\sum Y_i e_1^T$ equals the Euclidean norm $|\sum Y_i|$, so controlling the matrix martingale controls the original vector sum. The matrix Freedman inequality [11] then bounds all directional projections simultaneously through a single spectral inequality, replacing the exponential covering prefactor 5^d with the dimensional parameter $d+1$ at no cost to the exponent.

THEOREM 5.1 (Dimension-free concentration). *Let $\{X_i\}_{i=1}^n$ be independent in $L^2(\Omega; \mathbb{R}^d)$ under $\hat{\mathbb{E}}$ with $|X_i| \le M$ a.s. Then for every $t > 0$,*

$$\hat{V}\left(\rho_\Theta\left(\frac{1}{n}\sum_{i=1}^n X_i\right) > t\right) \le (d+1)\exp\left(-\frac{nt^2}{2\bar{\sigma}_n^2 + \frac{4Mt}{3}}\right).$$

PROOF. Fix $P \in \mathcal{P}$ and let $Y_i = X_i - E_P[X_i \mid \mathcal{F}_{i-1}]$ as in Lemma 3.3, so that $|Y_i| \le 2M$ and $E_P[|Y_i|^2 \mid \mathcal{F}_{i-1}] \le \bar{\sigma}_n^2$. Define $Z_i = Y_i e_1^T \in \mathbb{R}^{d\times 1}$, where $e_1 \in \mathbb{R}^1$ is the scalar unit. Since each Z_i has rank one, $\|Z_i\|_{\mathrm{op}} = |Y_i| \le 2M$, and $\sum Z_i = (\sum Y_i) e_1^T$ gives $\|\sum Z_i\|_{\mathrm{op}} = |\sum Y_i|$.

For the predictable variance, note that $Z_i Z_i^T = Y_i Y_i^T$, so for any unit $p \in \mathbb{R}^d$,

$$p^T E_P[Y_i Y_i^T \mid \mathcal{F}_{i-1}]\, p = E_P[\langle p, Y_i\rangle^2 \mid \mathcal{F}_{i-1}] \le E_P[|Y_i|^2 \mid \mathcal{F}_{i-1}] \le \bar{\sigma}_n^2,$$

whence $\|\sum_{i=1}^n E_P[Y_i Y_i^T \mid \mathcal{F}_{i-1}]\|_{\mathrm{op}} \le n\bar{\sigma}_n^2$. Applying [11, Theorem 1.2] to the $d \times 1$ matrix martingale (Z_i) with dimensional parameter $d_1 + d_2 = d + 1$, bound $R = 2M$ and variance $\sigma_*^2 = n\bar{\sigma}_n^2$ gives

$$P\left(\left|\sum_{i=1}^n Y_i\right| \ge s\right) \le (d+1)\,\exp\left(-\frac{s^2}{2n\bar{\sigma}_n^2 + \frac{4Ms}{3}}\right).$$

Setting $s = nt$:

$$P\left(\left|\frac{1}{n}\sum Y_i\right| > t\right) \le (d+1)\,\exp\left(-\frac{nt^2}{2\bar{\sigma}_n^2 + \frac{4Mt}{3}}\right).$$

The conclusion follows from Lemma 3.3(b) and taking the supremum over $P \in \mathcal{P}$. □

REMARK 5.2 (Identically distributed case). When the X_i are identically distributed, $\Theta_i = \Theta_1$ for all i and $\bar{\sigma}_n^2 = \bar{\sigma}_1^2 := \inf_{\theta\in\Theta_1} \hat{\mathbb{E}}[|X_1 - \theta|^2]$, so the bound simplifies to

$$\hat{V}\left(\rho_{\Theta_1}(\bar{X}_n) > t\right) \le (d+1)\,\exp\left(-\frac{nt^2}{2\bar{\sigma}_1^2 + \frac{4Mt}{3}}\right).$$

REMARK 5.3 (Comparison with the covering approach). The prefactor $(d+1)$ is polynomial versus the exponential 5^d in Corollary 4.3. The constants in the exponent are also tighter: $2\bar{\sigma}_n^2 + \frac{4Mt}{3}$ versus $8\bar{\sigma}_n^2 + \frac{8Mt}{3}$. The improvement has two sources. The matrix inequality controls all projections simultaneously without the $t \to t/2$ covering loss, and there is no union bound over the net.

6. Optimality of the sub-Gaussian rate (up to constants). The upper bounds in Sections 4 and 5 all decay as $\exp(-cnt^2)$ for a constant c depending on the boundedness and variance parameters. A natural question is whether this sub-Gaussian rate can be improved. We show that it cannot. By constructing an explicit sublinear expectation space built from shifted uniform distributions, we produce a matching lower bound of the form $\exp(-c'nt^2)$, so the $\exp(-cnt^2)$ rate is optimal up to the value of the constant.

LEMMA 6.1 (A concrete sublinear expectation space). *Let $a > 0$ and $r > 0$. Set $M = a + r$ and $\Omega = [-M, M]^n$, and define the coordinate projections $X_i(\omega) = \omega_i$. For each $\mu = (\mu_1, \dots, \mu_n) \in [-a, a]^n$, let*

$$P_\mu = \bigotimes_{i=1}^n \mathrm{Uniform}[\mu_i - r, \mu_i + r].$$

Define $\mathcal{P} = \overline{\mathrm{conv}}\{P_\mu : \mu \in [-a, a]^n\}$ (closed convex hull in the weak topology) and $\hat{\mathbb{E}}[f] = \sup_{P\in\mathcal{P}} E_P[f]$. Then:

(a) *$\hat{\mathbb{E}}$ is a regular sublinear expectation on $C_{b.\mathrm{Lip}}(\Omega)$.*

(b) $\{X_i\}_{i=1}^n$ *is independent under* $\hat{\mathbb{E}}$ *in the sense of Definition 2.2.*
(c) $\Theta_i = [-a,a]$ *for each* i, *and* $\Theta = [-a,a]$.
(d) $\mathrm{Var}_{P_\mu}(X_i) = r^2/3$ *for every product measure* P_μ *and every* i.

PROOF. (a) Since Ω is compact, Dini's theorem ensures that $f_k \downarrow 0$ pointwise implies $f_k \to 0$ uniformly, giving $\hat{\mathbb{E}}[f_k] \downarrow 0$. Sublinearity and monotonicity are inherited from the supremum over linear expectations.

(b) Under each P_μ the coordinates are independent. Fix $\psi \in C_{b.\mathrm{Lip}}(\mathbb{R}^{i+1})$ and write $h_{\mu_{i+1}}(x) = E_{P_{\mu_{i+1}}}[\psi(x, X_{i+1})]$, where $P_{\mu_{i+1}}$ denotes $\mathrm{Uniform}[\mu_{i+1} - r, \mu_{i+1} + r]$. The distribution of X_{i+1} under P_μ depends only on μ_{i+1}, so $\sup_{\mu_{i+1}} h_{\mu_{i+1}}(x) = \hat{\mathbb{E}}[\psi(x, X_{i+1})]$. The remaining supremum over $(\mu_1, \ldots, \mu_i)$ factors by the product structure, which is the factorisation required by Definition 2.2.

(c) $E_{P_\mu}[X_i] = \mu_i$ ranges over $[-a,a]$ as μ_i varies. The Minkowski average of n copies of $[-a,a]$ scaled by $1/n$ is $[-a,a]$.

(d) $\mathrm{Uniform}[\mu_i - r, \mu_i + r]$ has variance $(2r)^2/12 = r^2/3$. □

LEMMA 6.2 (Rate function bound for the uniform distribution). *Let* $Z \sim \mathrm{Uniform}[-r,r]$. *The large-deviation rate function* $\Lambda^*(x) = \sup_\lambda(\lambda x - \log(\sinh(\lambda r)/(\lambda r)))$ *satisfies*

$$\Lambda^*(x) \le \frac{3x^2}{2r^2} \quad \textit{for } |x| \le r/2. \tag{6.1}$$

PROOF. Write $u = x/r$ with $|u| \le 1/2$. The exact rate function is

$$\Lambda^*(x) = \tfrac{1}{2}\big[(1+u)\log(1+u) + (1-u)\log(1-u)\big] = \sum_{k=1}^\infty \frac{u^{2k}}{2k(2k-1)}.$$

For $|u| \le 1/2$, each term satisfies $u^{2k}/(2k(2k-1)) \le u^2 \cdot 4^{-(k-1)}/(2k(2k-1))$, so

$$\frac{\Lambda^*(x)}{u^2} \le \sum_{k=1}^\infty \frac{1}{2k(2k-1)\cdot 4^{k-1}} = \frac{1}{2} + \frac{1}{48} + \frac{1}{480} + \cdots < \frac{3}{2}.$$

Hence $\Lambda^*(x) \le \frac{3}{2} u^2 = 3x^2/(2r^2)$. □

THEOREM 6.3 (Sharpness of the sub-Gaussian rate). *For any* $a > 0$ *and* $\sigma > 0$, *there exists a regular sublinear expectation space with independent random variables* $\{X_i\}_{i=1}^n \subset L^2(\Omega;\mathbb{R})$ *under* $\hat{\mathbb{E}}$, *with* $|X_i| \le M$ *for some* $M > 0$ *and* $\Theta_i = [-a,a]$ *for all* i, *such that for every* $t \in (0, \sigma/(4\sqrt{n})]$*:*

$$\hat{V}\left(\rho_\Theta\left(\frac{1}{n}\sum_{i=1}^n X_i\right) > t\right) \ge \frac{1}{4}\exp\left(-\frac{2nt^2}{\sigma^2}\right). \tag{6.2}$$

PROOF. Set $r = \sigma\sqrt{3}/2$, so that $r^2 = 3\sigma^2/4$ and $r^2/3 = \sigma^2/4$. Lemma 6.1 with this r and the given a provides a regular sublinear expectation space with $\Theta = [-a,a]$ and $\mathrm{Var}_{P_\mu}(X_i) = \sigma^2/4$.

Choose $\mu_i = a$ for all i. Under $P_{(a,\ldots,a)}$, the centred variables $Z_i = X_i - a$ are i.i.d. $\mathrm{Uniform}[-r,r]$ with mean zero. Since $\rho_{[-a,a]}(x) \ge (x-a)^+$ and $\bar{X}_n - a = \bar{Z}_n$,

$$\hat{V}(\rho_\Theta(\bar{X}_n) > t) \ge P_{(a,\ldots,a)}(\bar{Z}_n > t).$$

By the Cramér–Chernoff lower bound [2, Theorem 3.7.4], for $t \le r/2$,

$$P_{(a,\ldots,a)}(\bar{Z}_n > t) \ge \tfrac{1}{4}\exp(-n\Lambda^*(t)).$$

Lemma 6.2 gives $\Lambda^*(t) \le 3t^2/(2r^2)$. Substituting $r^2 = 3\sigma^2/4$:

$$\frac{3}{2r^2} = \frac{3}{2 \cdot 3\sigma^2/4} = \frac{2}{\sigma^2}, \tag{6.3}$$

so $\Lambda^*(t) \le 2t^2/\sigma^2$, yielding (6.2). The condition $t \le r/2 = \sigma\sqrt{3}/4$ is satisfied since $t \le \sigma/(4\sqrt{n}) \le \sigma/4 < \sigma\sqrt{3}/4$. □

Note that this result establishes the *rate* $\exp(-cnt^2)$ as optimal. It does not claim that the numerical constants in our upper bounds are sharp, and closing the gap between the constants remains open (Section 7).

REMARK 6.4 (Rate versus constants). The upper and lower bounds both decay as $\exp(-cnt^2)$, so the sub-Gaussian exponent is the correct scaling. The numerical constants differ, however. The lower bound has $2/\sigma^2$ in the exponent while Corollary 4.3 gives $1/(8\bar{\sigma}_n^2)$. Two effects account for this. The covering argument introduces a factor of 4 through the $t \to t/2$ halving and the union bound over the net. Additionally, $\bar{\sigma}_n^2$ is a worst-case parameter, whereas the construction uses a specific distribution with variance $\sigma^2/4$. The dimension-free bound (Theorem 5.1) avoids the covering loss entirely and achieves the tighter constant $1/(2\bar{\sigma}_n^2)$, which suggests that much of the gap comes from the covering technique rather than from any fundamental limitation.

REMARK 6.5 (Extension to $d > 1$). The construction generalises to $\mathbb{R}^d$. Take $\Theta_1 \subset \mathbb{R}^d$ convex and compact, let $a_0 \in \partial\Theta_1$ be a boundary point with outward unit normal ν, and replace the one-dimensional uniforms by d-dimensional measures on balls of radius r centred at $\mu \in \Theta_1$. Projecting onto ν recovers a sub-Gaussian lower bound in every dimension.

7. Discussion and open problems. Several directions remain open. Our results assume $|X_i| \le M$. Extending to a sub-Gaussian condition $\hat{\mathbb{E}}[\exp(\lambda\langle p, X_i\rangle)] \le \exp(C\lambda^2)$ requires transferring MGF estimates through the iterated conditioning of Definition 2.2, which amounts to a truncation device that preserves conditional sub-Gaussianity. On the CLT side, Peng's theorem [10] gives convergence of $\sqrt{n}\,\bar{X}_n$ to a G-normal distribution, and a multivariate Berry–Esseen rate (extending the scalar $n^{-1/2}\log n$ rate of [4]) appears within reach via our dimension-free technique combined with Stein-type couplings. Finally, while Theorem 6.3 pins down the rate $\exp(-cnt^2)$, the optimal constant c^* remains undetermined.

REFERENCES

[1] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford Univ. Press, Oxford.

[2] DEMBO, A. and ZEITOUNI, O. (2010). *Large Deviations Techniques and Applications*, 2nd ed. Springer, Berlin.

[3] DENIS, L., HU, M. and PENG, S. (2011). Function spaces and capacity related to a sublinear expectation: Application to G-Brownian motion paths. *Potential Anal.* **34** 139–161.

[4] FANG, X., PENG, S., SHAO, Q.-M. and SONG, Y. (2019). Limit theorems with rate of convergence under sublinear expectations. *Bernoulli* **25** 2564–2596.

[5] FREEDMAN, D. A. (1975). On tail probabilities for martingales. *Ann. Probab.* **3** 100–118.

[6] HU, M., LI, X. and LI, X. (2021). Convergence rate of Peng's law of large numbers under sublinear expectations. *Probab. Uncertain. Quant. Risk* **6** 261–266.

[7] HU, M. and PENG, S. (2009). On representation theorem of G-expectations and paths of G-Brownian motion. *Acta Math. Appl. Sin. Engl. Ser.* **25** 539–546.

[8] LI, X. and HU, M. (2024). Inequalities for independent random vectors under sublinear expectations. Preprint. Available at arXiv:2410.11349.

[9] MCDIARMID, C. (1998). Concentration. In *Probabilistic Methods for Algorithmic Discrete Mathematics* (M. Habib, C. McDiarmid, J. Ramirez-Alfonsin and B. Reed, eds.) 195–248. Springer, Berlin.

[10] PENG, S. (2019). *Nonlinear Expectations and Stochastic Calculus under Uncertainty: With Robust CLT and G-Brownian Motion*. Springer, Berlin.

[11] TROPP, J. A. (2011). Freedman's inequality for matrix martingales. *Electron. Commun. Probab.* **16** 262–270.

[12] VERSHYNIN, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Univ. Press, Cambridge.

[13] ZHANG, L.-X. (2016). Exponential inequalities under the sub-linear expectations with applications to laws of the iterated logarithm. *Sci. China Math.* **59** 2503–2526.